data graphics and interactive information environments

A dissertation submitted to Cardiff University
in candidature for the degree of Doctor of Philosophy

Jignesh Khakhar
The Welsh School of Architecture
Cardiff University
2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Thesis copyright declaration

to be signed by each user of this thesis

author KHAKHAR,J.

title Data graphics and interactive information environments.

year 2008 degree Ph.D. accession no.

I undertake not to publish any portion of or use any information derived from, this thesis without the written consent of the Director. I further undertake to allow no one else to use this thesis while it is issued to me.

date	name and address in capitals	signature

Abstract

The flow and exchange of information is characteristic of the digital age. Information is increasingly consumed and produced. It has to be stored, channelled and processed. It also has to be reproduced in new forms again and again. The exponential growth in the volume of digital data has led to new challenges to visualize this data, such that it makes a significant difference to one's understanding of this data. This gain is known as information, which ultimately transforms into knowledge. Information Visualization attempts to create tools and processes that are an aid to cognition. Although this is a relatively new but established discipline, its roots can be traced to developments in the early 17th century, to what are now termed as Data Graphics. As a precursor to modern, dynamic, computer-based visualizations, Data Graphics underpin the science of visualization.

This thesis looks at the design principles that govern the construction of historic and contemporary data displays. Although the medium on which Data Graphics have historically been constructed and presented has changed from paper to the computer, design principles have remained the same regardless of the medium or the source of content. The thesis then applies these principles to construct two applications based on large, complex and multivariate data-sets. The first is a proposal for a three-dimensional radar display to visualize Air Traffic Control data. The second application is a visual navigation tool to search within a hypertext document. Both case studies apply design principles inherent in data graphics and utilise human perceptual and cognitive abilities to extract information inherent in data. Finally, both applications are tested by conducting user studies. The contributions of the thesis lie in the construction of the two novel information visualization applications stated above, and by demonstrating that data graphic design principles trascend the medium in which they are produced and presented and can greatly enhance the beauty, efficacy and effectiveness of data displays.

for,
my mother *Manjula*
&
father *Mahendra*

Acknowledgements

Professor Richard Weston *for being an exceptional mentor and friend*
Dr. Andrew Marsh *for unparalleled energy, enthusiasm and above all, vision*
Joop van Duin *without whom OPUS would not exist*
Professor Phil Jones *for the opportunity to work at the WSA*
Donald Alexander *for guidance on OPUS*
Sylvia Harris *for support on academic resources, and with whom I have shared some of my deepest thoughts*
Dave Bull *for unquestioned support on resources for this thesis*
Jacob Hotz *for being a wonderful colleague and friend, and most of all, for his wild humour*
Dominic Sleeman *for support on resources*
Paola Sassi *for her guidance and support during my stay on the Gallery, and for wonderful discussions on many aspects of life*

Alison Lewis, Mark Green, Devang Lakhani and Simon Taylor *for their guidance and support at NATS*

Juliet Odgers, Dr. Sophia Psarra, Dr. Adam Sharr, Dr. Wayne Forster and Dr. Ian Knight *whom, in various capacities, have indirectly influenced this thesis*

Simon Lannon, Dylan Dixon and Janice Coyle *for technical support on the thesis*

Annie Pitchford, Katrina Lewis, Susan Bowden, Jan O'Brien, Mary Brice, Susan Morgans and Frances Simpson *for making sure my time at the WSA was as smooth a ride as possible*

Edward Wainwright, Samuel Austin and Vijay Patel *for scholarly discussions on the thesis*

Dr. Chris Watts and Melanie Richardson *for being wonderful teachers*

Yuan Lin and Spyros Stravoravdis *for discussions over coffee*

Amit Gurav and Zainah Zain *for memorable discussions into the long hours of the night, and without whom life in Cardiff would be dull*

Daniel & Aleksandra Sudhershan and Sudhir Rao *for their undying support and above all, for their friendship*

Sumit and Chirag *for their generosity and respect, and with whom I have shared countless memories over the last eight years*

Mahesh fua and Geeta faiba *for treating me as a son and without whom I would not have had the opportunity to study in Britain*

List of Tables

Chapter 1

Chapter 2

Chapter 4

All tables in this chapter are by the author.

List of Figures

Chapter 1

p 12 facing — Chernoff faces
Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 13 — Galileo's sunspot diagram
Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

p 14 top — Galileo's sunspot diagrams of October 1611
Tufte, Edward R., Dynamic Graphs for Data Analysis, *Statistical Science*, Vol 2. No 4, November 1987, pp 389 - 392

p 14 bottom, right — Copernican vs Ptolemaic planetary systems
Drake, Stillman, *Galileo Studies: Personality, Tradition and Revolution*, Ann Arbor: The University of Michigan Press, 1970

p 15 — E. W. Maunder's sunspot diagram
Maunder, E. Walter, Note on the Distribution of Sun-spots in Heliographic Latitude, 1874 - 1902, *Monthly Notices of the Royal Astronomical Society*, Vol 64, June 1904, pp 747 - 761

p 16 — Images from Charles Howard Hinton's 'The Recognition of the Fourth Dimension' in 1902
Hinton, Charles Howard, (Editor) Rudolph v.B. Rucker, *Speculations on the Fourth Dimension: Selected Writings of Charles H. Hinton*, New York: Dover Publications, 1980

p 17 top, right — unfolding of a hypercube in four dimensions
Kaku, Michio, *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps and the Tenth Dimension*, Oxford: Oxford University Press, 1994

p 17 bottom — stereoview of tesseract
http://en.wikipedia.org/wiki/Image:3D_stereographic_projection_tesseract.PNG, Accessed <June 04, 2007>

p 18 — Hawk-Eye images of a cricket ball in flight
http://www.hawkeyeinnovations.co.uk, Accessed <March 08, 2005>

p 19 top — Feynman diagrams to represent an electron, positron and photon
www2.slac.stanford.edu/vvc/theory/feynman.html, Accessed <January 10, 2006>

p 19 bottom — Feynman diagrams to represent electromagnetic interactions
www2.slac.stanford.edu/vvc/theory/feynman.html, Accessed <January 10, 2006>

p 20 — two Feynman diagrams depicted electromagnetic interactions in space-time
Feynman, Richard P., *QED: The Strange Theory of Light and Matter*, Princeton, New Jersey: Princeton University Press, 1985

p 21 facing — stills from 4D scans of a prenate yawning inside a mother's womb
'Life Before Birth', Channel 4, broadcast <April 07, 2005 at 21:00 hrs>

p 21 middle — compiled actogram of a one hour observation of a fetus at 17 weeks
de Vries, J.I.P., G.H.A. Visser and H.F.R. Prechtl, The emergence of fetal behaviour: II: Quantitative aspects, *Early Human Development*, Vol 12, 1985, pp 99 - 120

p 21 bottom — Electrocardiogram of a patient with an irregular heartbeat
Meek, Steve and Francis Morris, ABC of clinical electrocardiography: Introduction: 1: Leads, rate, rhythm, and cardiac axis, *British Medical Journal*, Volume 324, 16 February 2002, pp 415 - 418

p 23 pair of images depicting preattentive processing in human vision based on difference in hue
http://www4.ncsu.edu/~healey/PP/index.html, Accessed <January 26, 2007>

p 23 pair of images depicting preattentive processing in human vision based on difference in form
http://www4.ncsu.edu/~healey/PP/index.html, Accessed <January 26, 2007>

p 24 pair of images depicting preattentive processing in human vision based on hue boundary
http://www4.ncsu.edu/~healey/PP/index.html, Accessed <January 26, 2007>

p 24 pair of images depicting a conjunction search for hue and form, which is not preattentive
http://www4.ncsu.edu/~healey/PP/index.html, Accessed <January 26, 2007>

p 25 stills from a supercomputer animation of a severe storm; top image is the original; bottom image is the redesign by Tufte *et al*
Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 25 facing, top William Playfair's graph of imports and exports of England to and from the East Indies between the years 1700 and 1780
Cleveland, William S., *The Elements of Graphing Data*, Monterey, California: Wadsworth, 1985

p 25 facing, bottom William S. Cleveland's graph showing the 'difference of curves' based on William Playfair's original graph above
Cleveland, William S., *The Elements of Graphing Data*, Monterey, California: Wadsworth, 1985

p 26 E. J. Marey's graphic depicting the catastrophic march of Napoleon's army through Russia in 1812.
Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

Chapter 2

---

p 28 evolution of computer processing speed in instructions per second and primary and secondary memory size in bytes from 1947 to the present
Denning, Peter J. and Robert M. Metcalfe, *Beyond Calculation: The Next Fifty Years of Computing*, New York: Copernicus, Springer-Verlag, 1997

p 30 cows and not cows
http://thecoloringspot.com, Accessed <August 22, 2006>

p 36 middle, right Tangram
Gardner, Martin, Mathematical Games, On the fanciful history and the creative challenges of the puzzle game of tangrams, *Scientific American*, 1974, pp 98 - 103B

p 36 bottom Tangram illustrations
Gardner, Martin, Mathematical Games, On the fanciful history and the creative challenges of the puzzle game of tangrams, *Scientific American*, 1974, pp 98 - 103B

p 37 top, right pair of Tangram illustrations
Gardner, Martin, Mathematical Games, On the fanciful history and the creative challenges of the puzzle game of tangrams, *Scientific American*, 1974, pp 98 - 103B

p 37 Morse code
http://www.babbage.demon.co.uk/morseabc.html, Accessed <November 04, 2006>

p 37 early 19th century tangram cards
facing Gardner, Martin, Mathematical Games, On the fanciful history and the creative challenges of the puzzle game of tangrams, *Scientific American*, 1974, pp 98 - 103B

p 40 the Koch snowflake
http://en.wikipedia.org/wiki/Koch_snowflake.png, Accessed <March 28, 2007>

Epilogue to Part 1

---

p 42 Chernoff faces
right Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 42 Galileo's sunspot diagram
left Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

p 42 advance of the Gecko
bottom Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 43 E. W. Maunder's sunspot diagram
top Maunder, E. Walter, Note on the Distribution of Sun-spots in Heliographic Latitude, 1874 - 1902, *Monthly Notices of the Royal Astronomical Society*, Vol 64, June 1904, pp 747 - 761

p 43 Butterfly sunspot diagram
bottom Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

p 44 Illustrations from Christopher Scheiner on the observation of sunspots
top Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

p 44 tracks depicting the movement of a horse at different paces
right Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 45 Price Waterhouse Coopers fax delivery data
Wainer, Howard, *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton: Princeton University Press, 2005

p 47 Copernican vs Ptolemaic planetary systems
top, right Drake, Stillman, *Galileo Studies: Personality, Tradition and Revolution*, Ann Arbor: The University of Michigan Press, 1970

p 47 rubber O-rings
right, centre Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 47 bottom — hand-written chart by engineers at Morton Thiokol
Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 48 — chart depicting 'History of O-ring damage in field joints'
Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 48 bottom, right — legend in chart above
Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 49 — scatterplot showing O-ring damage in relation to temperature
Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

p 50 — the 'dot map' by Dr. John Snow depicting the cholera outbreak in Central London in 1854
Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 50 facing — redesign of the 'dot map' by Dr. John Snow
Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 51 — two-dimensional relational graphic entitled 'Revenue Growth Forecast'
Tufte, Edward R., *Beautiful Evidence*, Connecticut: Graphics Press, 2006

Chapter 3

---

p 60 facing — graphic representation of radar controller interface
NATS document, Section 2: Command Entry and Display Terminology, Vol 1, Issue 11

p 61 — representation of radar image as seen on a radar display by an air traffic controller
NATS document, Section 2: Command Entry and Display Symbology, Vol 1, Issue 11

p 63 — puffing up one-dimensional data into three
Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

p 65 facing — radar screen at the Swanwick Air Traffic Control Centre on June 03, 2004
The Flight Collection

p 67 top, right — symbols used to depict data variables
author

p 67 bottom — SID's with opaque and transparent skin
author

p 68 — twelve colour-set by Colin Ware, against white and black backgrounds
Ware, Colin, *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000

p 69 facing — illustrations from an investigation by Linda Reynolds into the application of colour in radar displays at NATS
Wildbur, Peter, and Michael Burke, *Information Graphics: Innovative Solutions in Contemporary Design*, London: Thames and Hudson, 1998

p 71 — plan and perspective view of the proposed three-dimensional radar display
author

p 71 facing — orthographic and camera view of the proposed three-dimensional radar display
author

p 72 — detection time as a function of distance from fixation point
Peterson, Harold E. and Doris J. Dugas, The Relative Importance of Contrast and Motion in Visual Detection, *Human Factors*, Vol 14, No 3, 1972, pp 207 - 216

p 72 facing — stills from an animation simulating a Short-Term Conflict Alert (STCA) scenario
author

Chapter 4

All illustrations in this Chapter are by the author with the exception of:

p 77 — search results from the search engine Google
http://www.google.co.uk

p 77 facing, top — search results from the search engine Live
http://www.live.com

p 77 facing, bottom — search results from the search engine Ask
http://uk.ask.com

p 78 — search results from the search engine Exalead
http://www.exalead.co.uk/search

p 78 facing — search results from the search engine Snap
http://www.snap.com

p 79 — search results from the search engine Clusty
http://www.clusty.com

p 79 facing, top — search results from the search engine Shakespeare Searched
http://shakespeare.clusty.com/

p 79 facing, bottom — search results from the search engine Vivisimo
http://vivisimo.com/

p 80 search results from the search engine Grokker
http://www.grokker.com/

p 81 search results from the search engine Mooter
top http://www.mooter.com/

p 81 search results from the search engine Quintura
bottom http://www.quintura.com/

p 82 TheBrain navigation technique implemented in the Mappa Mundi Magazine website
http://mappa.mundi.net/

p 82 search results from TheBrain search engine
facing, http://www.webbrain.com
top

p 82 search results from the search engine Kartoo
facing, http://www.kartoo.com
bottom

p 84 clustering representation in the search engine Clusty
top http://www.clusty.com
right

p 85 text-cloud technique in search engine Quintura
top http://www.quintura.com/

p 85 hyperlinks in the search engine Mooter
bottom http://www.mooter.com

p 85 display of air pollution levels over southern California for four different times of the day
facing Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

p 95 two images of the Equiangular Spiral
http://en.wikipedia.org/wiki/Image:Fibonacci_spiral.svg, Accessed <May 11, 2005>

Chapter 5

All illustrations in this Chapter are by the author with the exception of:

p 119 representation of radar image as seen on a radar display by an air traffic controller
NATS document, Section 2: Command Entry and Display Symbology, Vol 1, Issue 11

Contents

0.0 Introduction

0.0 Introduction

0.1 The Age of Information

The latter half of the 20th century is often referred to as the Information Age. The use of the term 'information' to coin phrases such as 'information society', 'information economy', 'information revolution' and so on seem to imply a profound sense of change in the way we live, work and interact with our environment - both natural and artificial; where the latter refers to electronic environments. The ability of the computer to act as an agent of change was perhaps foreseen by Marshall McLuhan in 1964, when he proclaimed "the medium is the message."[1] If the message was change, then the carrier of this message was the computer. Since its commercial introduction in 1977 by Apple, and then by IBM in 1981, the personal computer or PC has struggled to keep up with the growth of digital information.The flow and exchange of information is characteristic of the digital age. Information is increasingly consumed and produced. It has to be stored, channelled and processed. It also has to be reproduced in new forms again and again. According to researchers at the University of California at Berkeley, the world generated five exabytes (5×10^{18} bytes) of data in 2002, double the amount in 1999. Put in another context, according to Eric Schmidt, CEO of Google, "absorbing five exabytes of data on TV would require sitting in front of a screen for 40, 700 years."[2] As semiconductor and bandwidth capacities try to keep pace with the growth of digital information, a primary area of concern is the representation and comprehension of information.

1 McLuhan, Marshall, *Understanding Media: The Extensions of Man*, , Cambridge, Massachusetts: MIT Press, 1999, p 7

2 Eric Schmidt, 'Don't bet against the internet', *The Economist: The World in 2007*, p 130

The representation and comprehension of data is at the heart of producing exemplary data graphics, and has been practised for over 300 years. In the past, the medium of presentation and exploration was print, today it is overwhelmingly digital. Although the medium has changed, principles of data design are remarkably similar and are reviewed in this thesis. The design principles and techniques that are inherent in data graphics take advantage of the extraordinary skill of human beings to perceive the information inherent in our environments through their senses; and apply this capacity of the eye-brain cortex to our understanding of data. Although "human perceptual skills are quite remarkable, [they are] largely underutilised in current information and computing systems."[3] Thus from both a historic perspective and a human perception and cognition point of view, Data Graphics underpin the maturing field of visualization. As a discipline, visualization grew in response to requirements from the scientific community, particularly medicine and biochemistry, in the 1980s. The field grew rapidly and

3 http://www.cs.brown.edu/people/ifc/hci/shneiderman.html <accessed 29/06/2004>

extended itself to include other than scientific data, to produce the field of Information Visualization. Several techniques and methods, discussed in the next chapter, were introduced to make sense of phenomena inherent in data. The field matured rapidly, as Stephen Eick suggests, from a "craft practised by skilled craftsmen using heuristic methods"[4] to a point where established tools and techniques have become widely available. Clearly, there are exciting opportunities ahead in this field owing to the phenomenal volume of multivariate data generated on a year-on-year basis.

4 Eick, Stephen G., 'Information Visualization at 10', *IEEE Computer Graphics and Applications*, Vol 25, No 1, January/February 2005, pp 12 - 14

Visualization can be approached and practised as art, science or technology, and as a language. Given the inherent cross-disciplinary nature of the discipline, there is a gap in the knowledge of 'experts' involved in data graphics or visualization. Those in the field of data graphics are generally statistically illiterate and fail to understand the significance of what they are representing. According to the American Statistical Association, "formal training in graphic presentation has had a marked decline at all levels of education over the last few decades."[5] In the visualization community, lack of knowledge of theory of graphics in general and data graphics in particular is common. In 'On the Death of Visualization', Bill Lorensen, one of the founders of the field, argues that "Visualization, the Community, lacks application domain knowledge."[6] Particularly disconcerting is the fact that there is almost no evidence of any involvement of those intimately concerned with the 'making of space' *ie* architects. A review of major publications in the field[7] reveals little evidence of architects as active members of this community. There is, however, some evidence of scholars with backgrounds in design (Bertin, Tufte, Wijk, for example). Given the nature of the discipline of architecture, the architect is in a unique position to combine these disciplines, however far apart they may seem in practice, they all merge in the dimensions of space, either on paper or on computer screen. This is a discipline where subtle differences in understanding data or its representation can result in mis-information or 'negative knowledge'.

5 Wainer, Howard, 'Making Newspaper Graphs Fit to Print', in Kolers, Paul A., Merald E. Wrolstad and Herman Bouma (Editors), *Processing of Visible Language: Volume 2*, New York; London: Plenum Press, 1980, p 139

6 Lorensen, Bill, 'On The Death of Visualization', *Position Papers NIH/NSF Proceedings, Fall 2004 Workshop Visualization Challenges*, 2004

7 refer to the 'Journals' section in the Bibliography

0.2 Background to Study

An interest in this subject area arose out of a design project undertaken at final year undergraduate level, which explored the navigation of data in three spatial dimensions.[8] An early research proposal outlined an interest in exploring this project in depth, by investigating the three-dimensional spatial navigation of data, particularly involving and influenced by the human senses. However, as the thesis took shape, the direction of study leaned towards data graphics for two primary reasons: a) as a starting

8 A final year undergraduate design project: 'A British Pavilion for a Future Exposition', Year 3, The Leicester School of Architecture, 2001

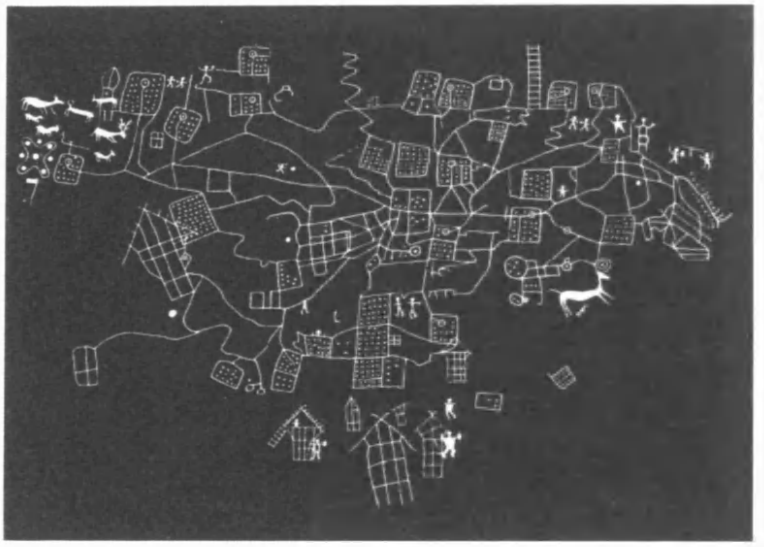
point, investigating theory of data graphics provided a basis to study and apply the learnings to the science of visualization and b) an understanding of data or what was being represented is crucial to producing exemplary data graphics. This line of inquiry subsequently led to theoretical investigations trying to define the nature of the term 'information', and its distinction from data and knowledge. These theoretical studies also feed into practice. The method of investigation took an exploratory route of applying design principles to the navigation of data in two large and complex multivariate datasets: air traffic control and a visual search tool for a hypertext document.

0.3 Thesis Structure

This thesis is structured in two parts: Part 1 discusses *Theory* and Part 2 applies this knowledge to construct *Applications*. In Part 1, Chapter 1 is an account of data graphics and the design principles employed to construct superb visualizations. Chapter 2 attempts to define information and expose its close association with probability, entropy and form. In Part 2, two practical applications are described: Chapter 3 proposes a three-dimensional radar display for future air traffic control; a project that was envisaged in collaboration with National Air Traffic Services Ltd. Chapter 4 discusses the design, development and testing of a navigation tool based on the principles of information design. Known as an Observer Participant User System or OPUS, the software application aims to support searching within the web pages of an online portal related to environmental design. This case study integrates theoretical considerations from the disciplines of data graphics, user interface design and human-computer interaction, to support an information seeking task. Since the mode of output for both case studies is digital, a CD accompanies this thesis, where the reader can view videos that demonstrate both visualizations in action.

Part 1: Theory

1.0 Envisioning Information

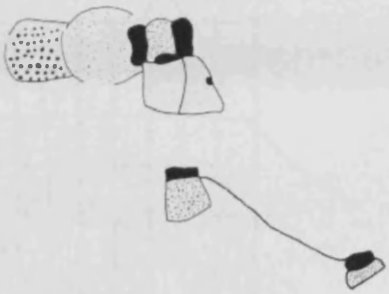


Bedolina petroglyph, Valcamonica, northern Italy

1.0 "Envisioning Information"[9]

9 after Edward R. Tufte, *Envisioning Information*, 1983

Our inquiry into 'reasoning with information' begins with an examination of data graphics, which find their earliest origins in statistical graphics. These in turn have geographical maps as their predecessors. Petroglyphs are the earliest known maps that represent landscape. The petroglyphic map from Val Fontanalba at Mount Bego in the Ligurian Alps (right), and the Bedolina petroglyph at Valcamonica in northern Italy (left) are the earliest known examples of remarkable representations of landscape. Maps quantify distances between places in the physical world. As navigational aids, they orient us - a primary concern of cartography. The journey from the mapping of physical space to the mapping of quantitative variables in virtual or paper-space is a long one and is explored in this Chapter. Historic and contemporary examples of data design weave narratives of space and time to reveal design principles that are consistent regardless of their source and content.



petroglyphic map from Val Fontanalba at Mount Bego in the Ligurian Alps

1.1 Origins

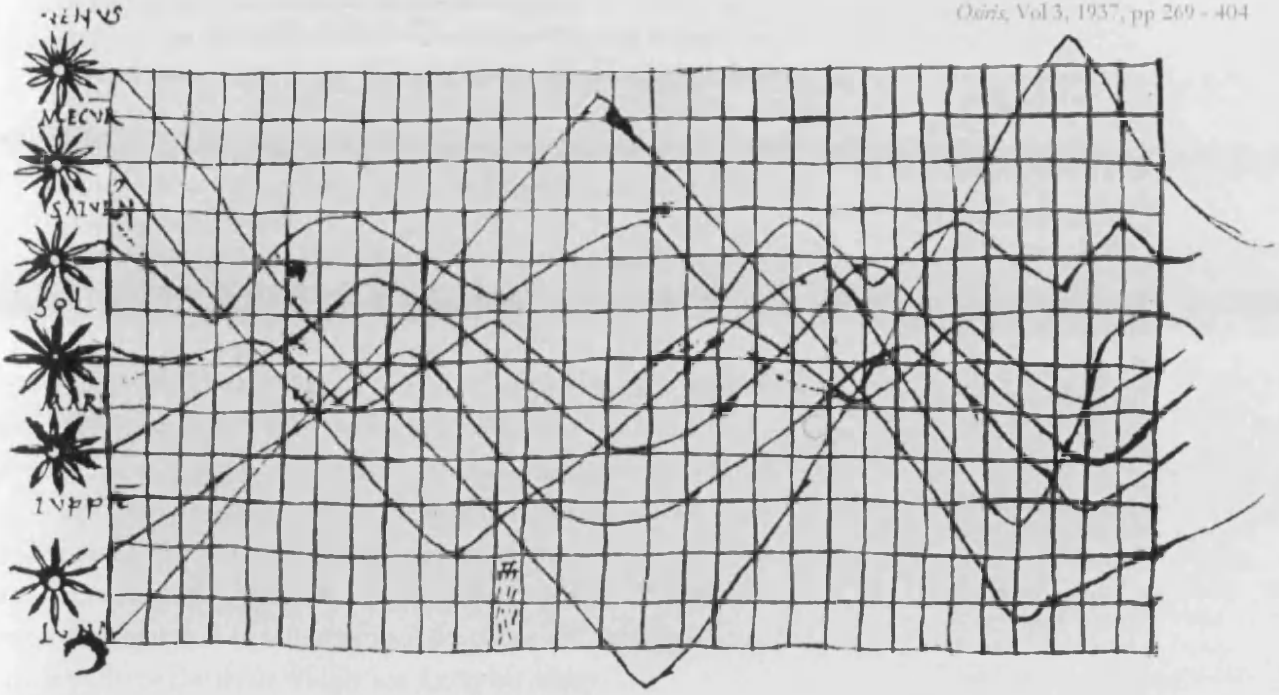
Visualization is an established but relatively new discipline. In October 1986, the National Science Foundation (NSF) in the United States sponsored a meeting "to provide input on establishing and ordering priorities for acquiring graphics and image processing hardware and software at research institutions doing advanced scientific computing."[10] Since computer graphics and image processing are within *computer science* and the application of computers to the discipline sciences is called *computational science*, the panel members defined "applying graphics and image techniques to computational sciences [as] *Visualization in Scientific Computing (VISC).*"[11] This was followed by the first Workshop on VISC in February 1987. This initiative was a catalyst for the discipline of visualization. Initially, as the acronym suggests, this was heavily involved with imaging produced from scientific data, but has now been extended to encompass all possible types of data. The first Institute of Electrical and Electronics Engineers (IEEE) conference on Visualization in 1990 brought together people from the diverse fields of statistics, data analysis, data graphics, data mining, pattern recognition, machine learning and artificial intelligence, and laid a formal direction for further developments in this field. Amongst these interdependent areas of study, of particular interest to this thesis, is the discipline of data graphics. Although the origins of data graphics are generally traced to the 18th century, a few notable and isolated earlier pieces of work deserve mention here.

10 McCormick, Bruce H., Thomas A. DeFanti and Maxine D. Browne, *Visualization in Scientific Computing*, New York: ACM Siggraph, 1987

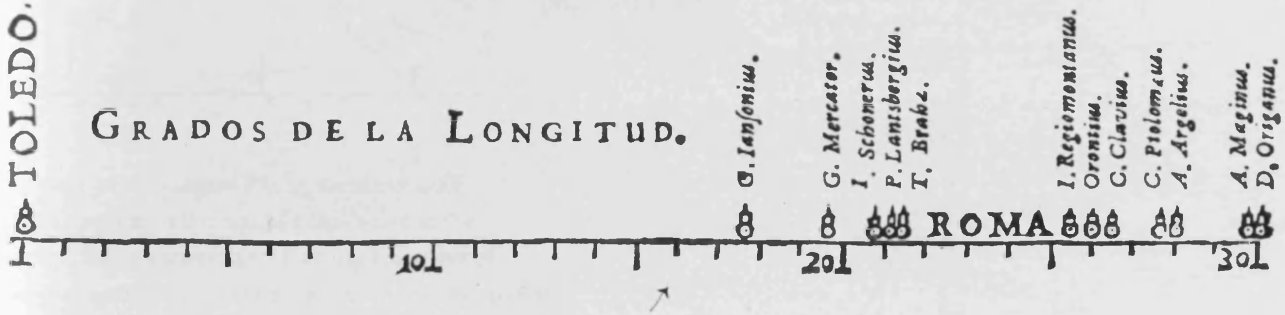
11 ibid, p IX, emphasis in original

This anonymous 11th century time series of planetary orbits is the earliest graphic of its kind and was meant to represent a plot of the inclinations of the planetary orbits as a function of time.[12]

12 Funkhouser, H. Gray, 'Historical Development of the Graphical Representation of Statistical Data', *Osiris*, Vol 3, 1937, pp 269 - 404



The Flemish astronomer Michael Florent van Langren drew this remarkable one-dimensional graphic representing the distance between Toledo and Rome in 1644. Note that the distance between Rome and Toledo is measured in degrees longitude. Twelve diverse but incorrect



estimates are shown. The correct distance is 16°30'.

The two graphics above are exceptional for their time owing to the fact that they do not make any attempt to identify or represent some sort of physical or geographical map. It is surprising how difficult and long the journey of breaking free from analogies to the physical world took. In the words of the eminent scholar on statistical evidence and information

design, Edward R. Tufte:

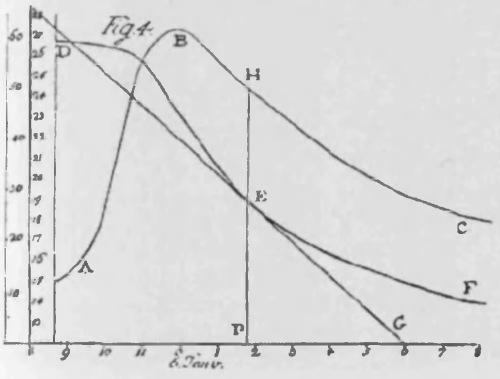
> Despite their quantifying scales and grids, maps represent miniature pictorial representations of the physical world. To depict relations between *any* measured quantities, however, requires replacing the map's natural spatial scales with abstract scales of measurement not based on geographic analogy. To go from maps of existing scenery to graphs of newly measured and collated data was an enormous conceptual step. Embodied in the very first maps were all the ideas necessary for making statistical graphics - quantified measures of locations of nouns in two-dimensional space - and yet it took 5,000 years to change the name of the coordinates from *west-east* and *north-south* to empirically measured variables X and Y.[13]

This great intellectual step forward began to manifest itself in the 18th century, when such examples of empirical measurement and analysis found their way into scientific writings. In 1765, for example, Johann Heinrich Lambert wrote: "We have in general two variable quantities, *x*, *y*, which will be collated with one another by observation, so that we can determine for each value of *x*, which may be considered as an abscissa, the corresponding ordinate *y*."[14] This writing by Lambert, who was exceptional in his time for using sophisticated graphs to describe scientific experiments (as in this graph below, depicting the rate of evaporation of water as a function of temperature) describes the 'relational graphic' which is perhaps the most widely used graphic today.[15]

13 Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, pp 14-15, emphasis in original

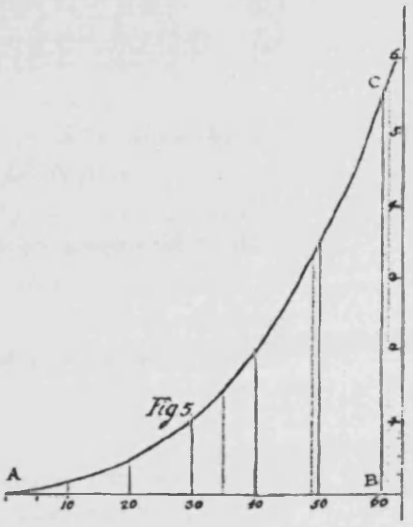
14 Lambert, Johann Heinrich, *Beytrage zum Gebrauche der Mathematik und deren Anwendung* (Berlin, 1765) as quoted in Tilling, Laura, 'Early Experimental Graphs', *British Journal for the History of Science*, Vol 8, No 30, 1975, pp 193-211, emphasis in original

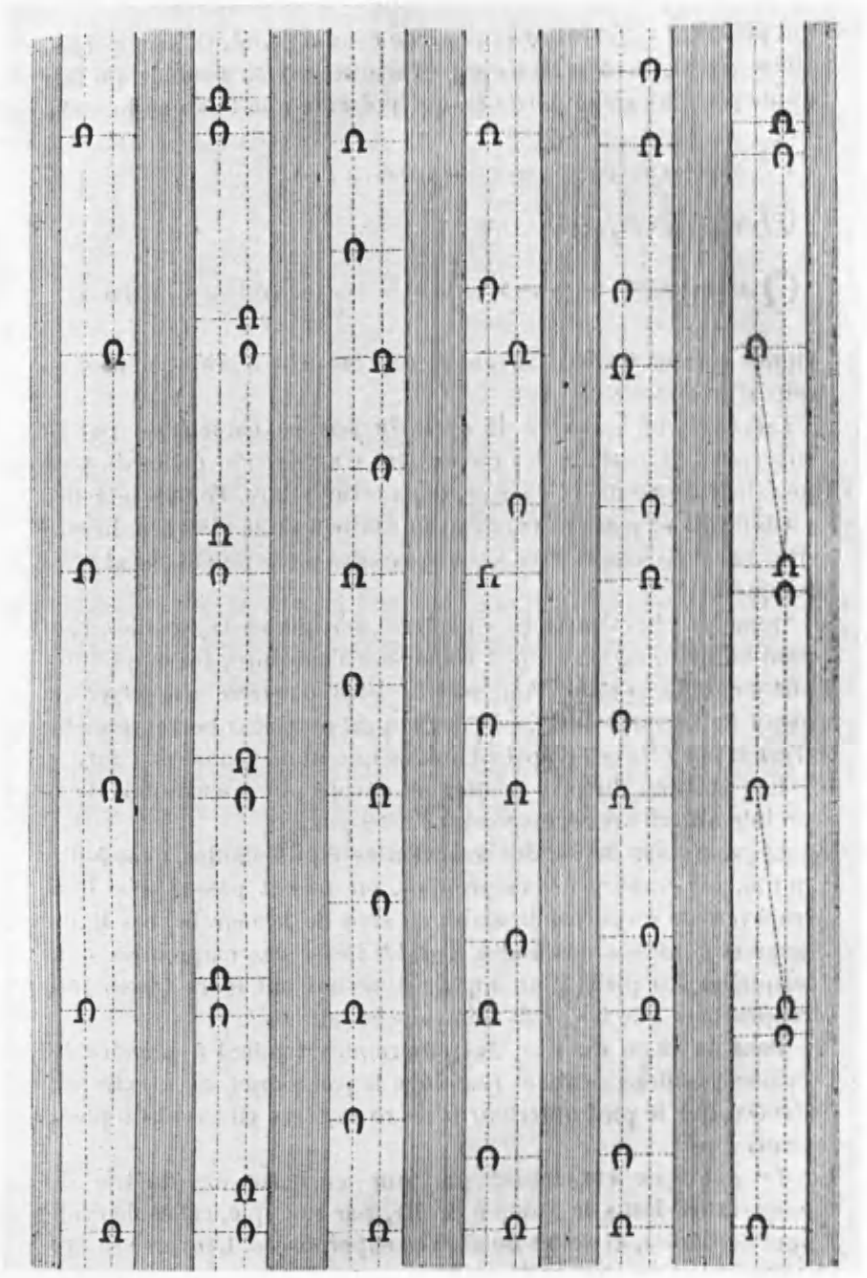
15 A relational graphic is one where any variable quantity can be plotted against another variable quantity, the most common form being 2d graphs plotting variable *x* against variable *y*, on the X and Y axes respectively.



The decrease in the height of water in a capillary tube (curve DEF) in response to the rising temperature (curve ABC) is the subject of the graph to the left.

The slope of the tangent EG to the curve DEF gives a measure of the rate of evaporation at the corresponding temperature. Plotting a number of tangents against the temperature gives the graph to the right.



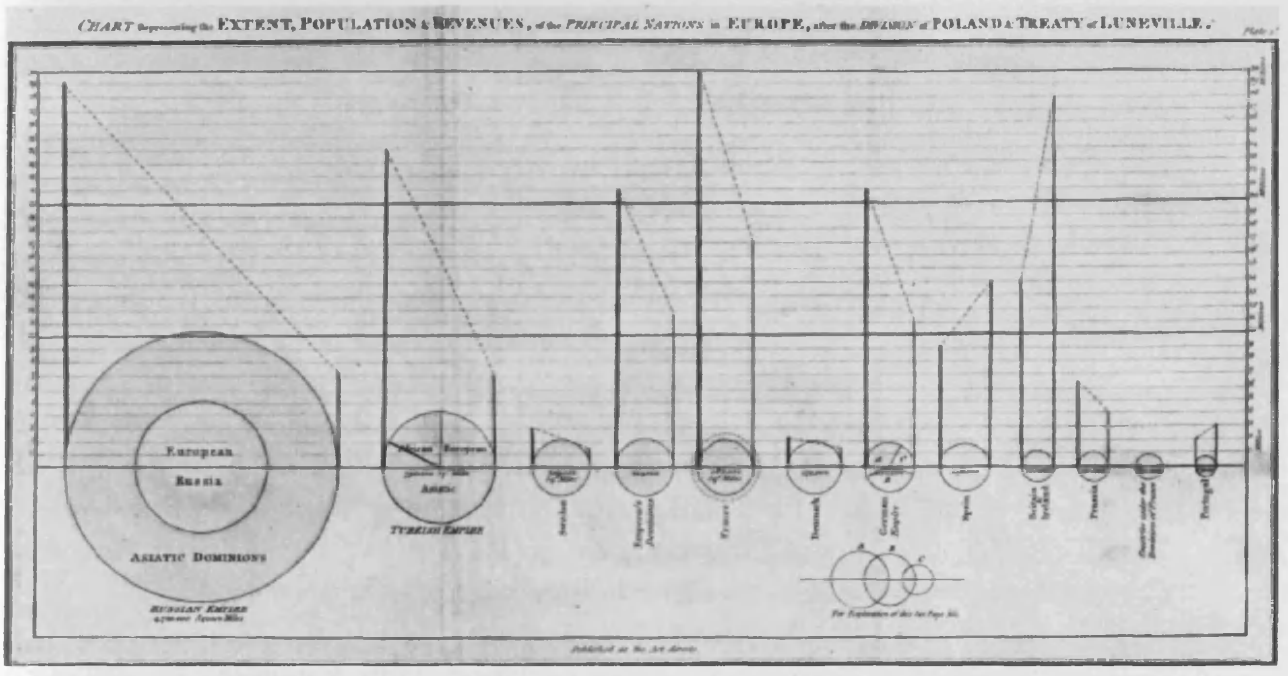


ordinary walk, walk, (long stride) quick walk, amble, jug-trot, gallop

E. J. Marey, *Movement*, London, 1895

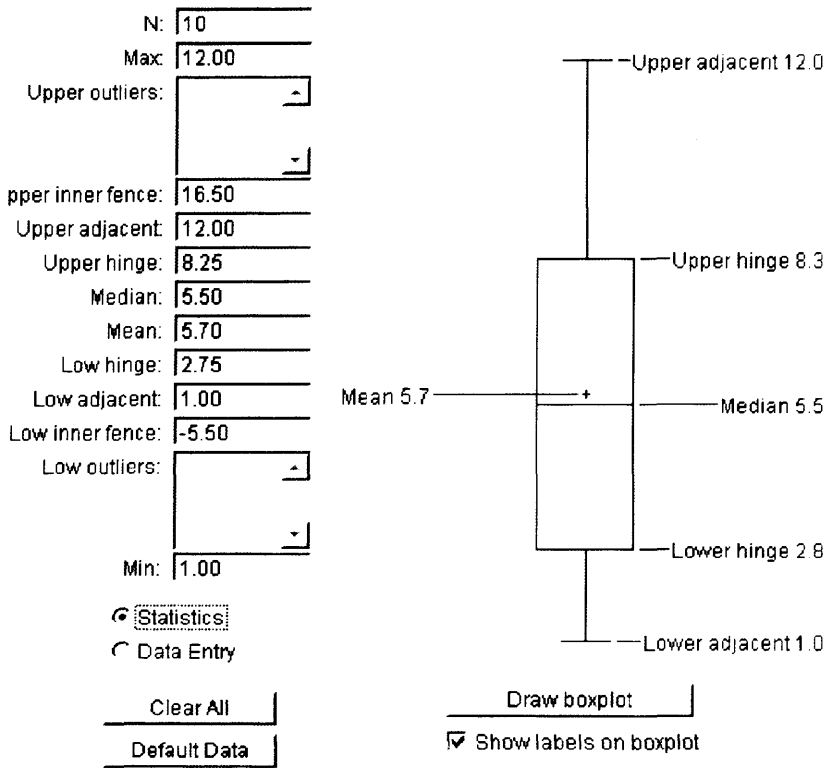
The origin of data graphics can be traced to developments, sometimes happening simultaneously, in the eighteenth century, by several pioneers: Johann Heinrich Lambert (1728 - 1777), a Swiss-German scientist and mathematician; Ettiene-Jules Marey, a French scientist and chronophotographer (1830 - 1904); William Playfair (1759 - 1823), an English political economist; and Charles Joseph Minard (1781 - 1870), a French engineer. Amongst them, William Playfair was the pioneer in using economic data for statistical analysis, inventing most of the graphical methods of statistical analysis still in use today: the bar chart, the histogram, the surface chart and the pie chart. Playfair published *The Commercial and Political Atlas* in 1786, which was succeeded by *The Statistical Breviary* in 1801. In both works, graphics were ingeniously used as "instruments for reasoning about quantitative information,"[16] aptly demonstrated below:

16 Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983, p 9



This graphic depicts multivariate data using simple geometric elements, and with clarity. The size of the circle represents the area of each country. The line on the left of each circle represents the population of the nation, and the line on the right represents revenues collected in taxes. The dotted sloping line joining these two tangential lines does not reveal much through the magnitude of its slope, as it depends on the size of the circle and the distance between the two vertical lines. However, the direction of its slope reveals an immediate fact: states that are either under-taxed or heavily-taxed. Hence, Great Britain and Ireland, represented by the fourth circle from the right, are overburdened with excessive taxation.

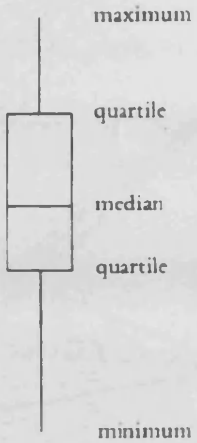
Although the interest in statistical graphics continued through the 19th century after Playfair, the next major step was taken with the advent of the computer and its use as an analytical tool. In 1977 the eminent statistician John W. Tukey's seminal work *Exploratory Data Analysis*



box-and-whisker plot for the data-set {1, 2, 3, 4, 5, 6, 7, 8, 9, 12}

introduced various novel ways of gaining rapid insight into a set of numbers, for example stem-and-leaf plots, box-and-whisker plots, and suspended rootograms. Below is a brief description of Tukey's 'five number summaries' or box-and-whisker plots:

A Box- and-Whisker plot is a graphic way of gaining rapid insight into a set of 5 numbers:
1. the highest or maximum value in the data-set
2. the 75th percentile or upper quartile is the median of the upper half values; also known as the upper hinge.
3. the median, which is not the average of all values in the data-set, but a measure of central tendency
4. the 25th percentile or lower quartile is the median of the lower half values; also known as the lower hinge.
5. the lowest or minimum value in the data-set

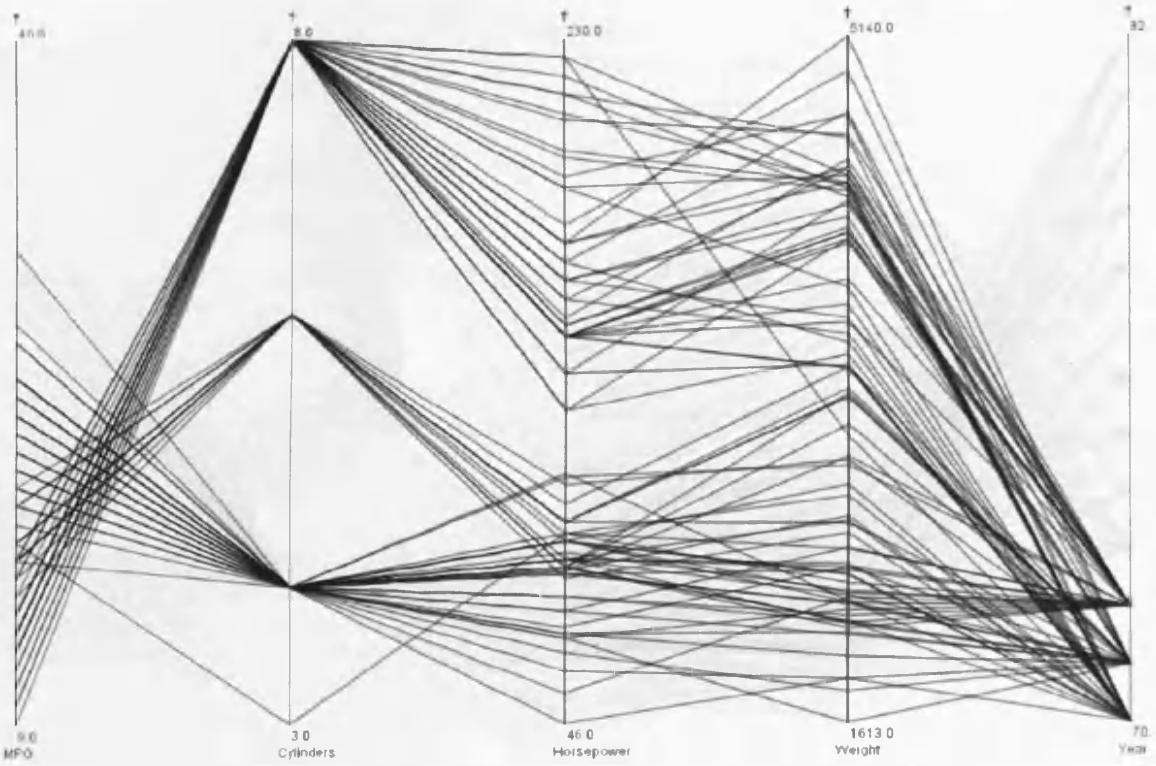


The lines joining the 'quartiles' to the minimum and maximum values are called whiskers. To the left is a box plot of a very simple data-set: {1, 2, 3, 4, 5, 6, 7, 8, 9, 12}. A number of variations of the box plot exist of which the one depicted is the simplest.

Progress in statistical graphics continued in incremental phases throughout the 18th, 19th and early 20th centuries. The advent of the computer created the problems it is now helping to solve. "[Digital] data grew unmanageable for the human generated table and for tabulation manipulation and analysis."[17] Thus graphs, charts and tables gave way to contemporary visualization tools and techniques to analyse and model data-sets much larger in size and with many more variables. For example, the Parallel Coordinates method of visualizing large data-sets, introduced by Alfred Inselberg, provides a method of modelling relations. These 'two-way' displays of 'n-way' data plot each variable within a data-set on a vertical axis. Thus, the search for relations(hips) amongst variables becomes a two-dimensional pattern recognition problem as described overleaf.

17 Fayyad, Usama, Georges G. Grinstein and Andreas Wierse (Editors), *Information Visualization in Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann, Academic Press, 2002, p 3

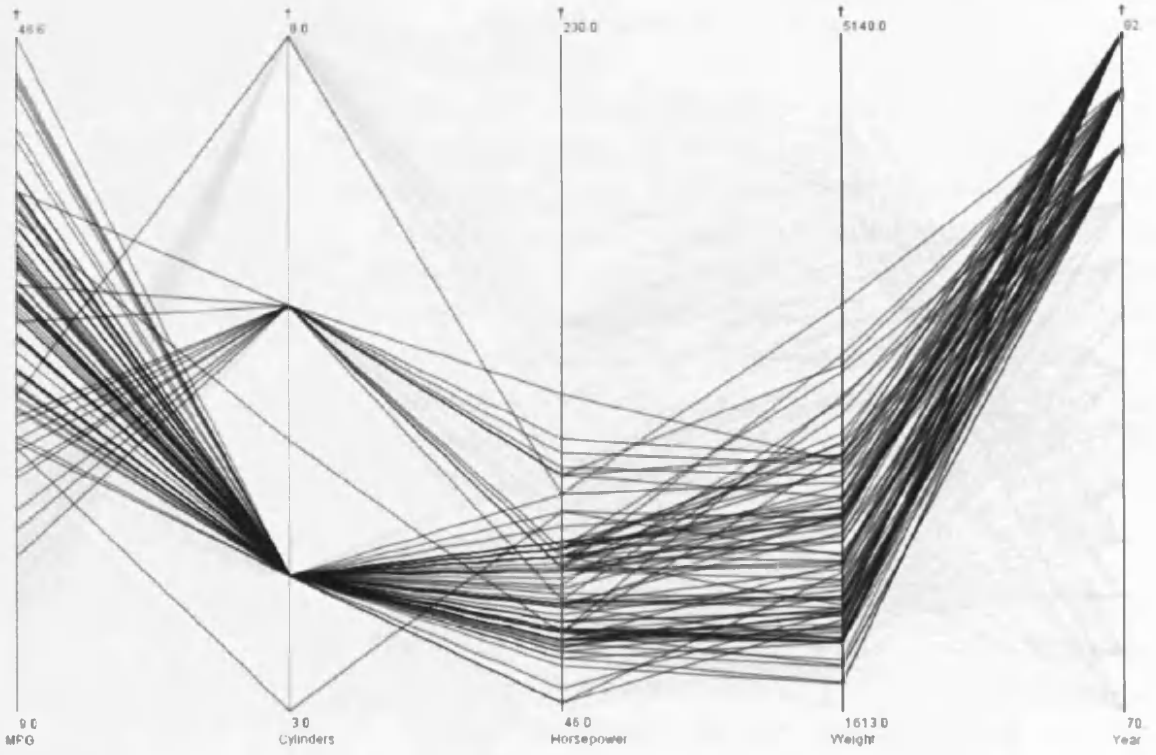
In spite of nearly 300 years of inquiry in the field, literature theorizing data graphics appears only in the latter half of the 20th century. In 1967, French cartographer Jacques Bertin published a formal theory governing the design of two-dimensional graphics in *The Semiology of Graphics*. John Wilder Tukey's *Exploratory Data Analysis* in 1977 re-established the integrity of statistical analysis using data graphic techniques. In 1985, William S. Cleveland provided further insight into the structure of data using graphical methods and principles, in *The Elements of Graphing Data*. In *The Visual Display of Quantitative Information* in 1983, Edward R. Tufte demonstrated the importance of design principles in enhancing the dimensionality and density of information in the design



parallel co-ordinates for:

1970 - 1972

1980 - 1982



The selective omission and comparison of information starts to reveal interesting trends or patterns in the data. For example, cars in the 1980s were more fuel efficient (mpg) than those in the 1970s; there is only one car with an 8 cylinder engine in 1981; cars were more evenly spread out, in terms of horsepower, in the 1970s, whereas most of the cars between 1980 to 1982 had 4 cylinder engines.

and display of large and complex sets of data. This was followed by his *Envisioning Information* (1990), *Visual Explanations* (1997) and most recently, *Beautiful Evidence* (2006). In contemporary terms, the work of Bertin, Tukey and Tufte has been highly influential in the Information Visualization community. Numerous tools and techniques that have emerged over the last decade, of which Parallel Co-ordinates (left) is an example, are a result of several areas of study mutually influencing each other.

1.2 Definitions

Visualization has been described as "constructing a mental image in the mind [or] a graphical representation of data or concepts."[18]

Grinstein and Ward (2002) describe visualization as:

> "the graphical (as opposed to textual or verbal) communication of information."[19]

Schneiderman, Mackinlay and Card (1999) define visualization as:

> "The use of computer-supported, interactive, visual representations of data to amplify cognition."[20]

Information Visualization is defined by the same authors as:

> "The use of computer-supported, interactive, visual representations of *abstract* data to amplify cognition."[21]

The difference in the last two definitions is the addition of the term 'abstract' to 'data'; 'abstract data' being defined as that which is inherently non-spatial, for example, stock-market or financial data.

18 Ware, C., *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000, p 1

19 Fayyad U., G.G. Grinstein and A. Wierse (Editors), *Information Visualization in Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufman, 2002, p 23

20 Schneiderman B., S.K. Card and J.D. Mackinlay (Editors), *Readings in Information Visualization: Using Vision To Think*, San Francisco: Morgan Kaufmann, 1999, p 6

21 ibid, p 6, emphasis in original

1.3 Purpose

Visualization is an aid to cognition. Intricately linked to perception, it can harness human abilities to extract patterns inherent in data. The mathematician and computer pioneer Richard Hamming has remarked, "The purpose of computation is insight, not numbers."[22] In view of cognition, "the purpose of visualization is insight, not pictures, [and] the main goals of this insight are *discovery, decision making, and explanation.*"[23]

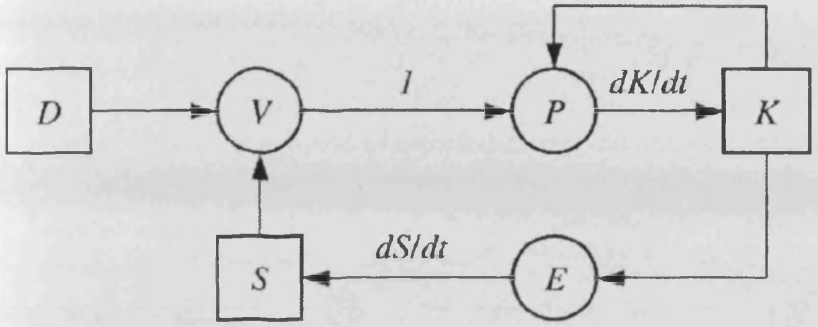
The purpose of visualization is to peel back the layer of the appearance of objects to reveal their underlying nature.

22 Schneiderman B., S.K. Card and J.D. Mackinlay (Editors), *Readings in Information Visualization: Using Vision To Think*, San Francisco: Morgan Kaufmann, 1999, p 6

23 ibid, p 6, emphasis in original

1.4 Model

The figure below describes a generic model of visualization as proposed by Jarke J. van Wijk, Professor of Visualization at Eindhoven University. In this diagram, squares denote containers and circles denote processes that transform inputs into outputs.



generic model parameters:

D = data;
V = visualization;
I = time varying image;
P = particular properties of perceptual and cognitive abilities of user;
K = knowledge gained, where
K_0 = initial knowledge and
K(t) = current knowledge;
S = specification of visualization;
E(k) = intereactive visualization

The central process in the model is visualization *V*. Data *D* is transformed according to a specification *S* into a time varying image *I(t)*. The image *I* is perceived by a user, with an increase in knowledge *K* as a result. The amount of knowledge gained depends on the image *I*, the current knowledge *K* of the user, and the particular properties of the perceptual and cognitive abilities *P* of the user. The current knowledge *K(t)* is the sum of the initial knowledge K_0 and all knowledge gained from the images so far. An important aspect is interactive visualization, *E(K)*. Starting from an initial specification S_0, the user may continuously change the specification of the visualization, based on his current knowledge, in order to explore the data further.[24]

The model is about mapping of data (tables) to a visual structure[25], as described by McKinlay, Card and Schneiderman (1999). Thus mathematical relations inherent in data, for example, are given a visual 'form' or structure based on graphical properties effectively processed by the eye-brain cortex. The visual structure also affects the cost structure associated with a task. The cost (structure) of a task is a way of measuring the value of a visualization - its efficiency and effectiveness. A simple example would be the amount of time it takes to solve mathematical or physics problems using diagrammatic versus non-diagrammatic representations. Larkin and Simon (1987) demonstrate that diagrams are superior to verbal descriptions for problem solving inferences owing to their capacity to group information and reduce the amount of time to search for elements needed to make an inference.[26]

24 Wijk, Jarke J. van, 'Views on Visualization', *IEEE Transactions on Visualization and Computer Graphics*, Vol 12, No 4, July/August 2006, pp 421- 432

25 visual structure is a term used to describe the visual representation of relationships in a data-set.

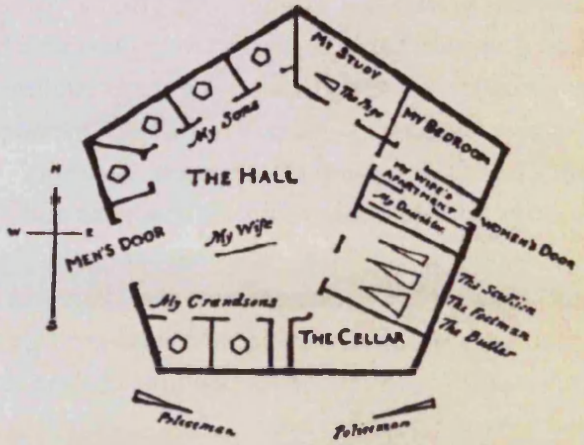
26 Larkin, Jill H. and Herbert A. Simon, 'Why a Diagram is (Sometimes) Worth Ten Thousand Words', *Cognitive Science*, Vol 11, 1987, pp 65 - 99

"O day and night, but this is wondrous strange"

A ROMANCE OF MANY DIMENSIONS

By A Square

(Edwin A. Abbott)



"And therefore as a stranger give it welcome."

BASIL BLACKWELL . OXFORD

1.5 Illustrations

The illustrations that follow are narratives through 'space-time' of six selected works, each spanning a different author, date, genre and purpose. All share a common intent: "to *reveal* data."[27] Edwin A. Abbot's Flatland uses the fourth dimension as a satirical tool to challenge the authority of Victorian England; Charles Hinton's compelling step by step depiction of a cube in four dimensions; Galileo's application of visualization to document a process and unravel the mysteries and workings of celestial bodies; Richard Feynman's elegant diagrams depicting complex ideas open up an entirely new world of interaction and inhabitation; Hawk-Eye's modelling of the trajectory of a ball offers subtle provocation to the hitherto unchallenged authority of a cricket umpire; the dynamic high-resolution imaging of the development of a human embryo, and its effect on parental bonding - all of these collectively illustrate the power of visualization. It is also worth noting that Galileo's sunspots, Hawk-Eye, and imaging of 'Life before Birth' are scientific visualizations based on physical data - the sun, a ball in motion and the human body. The information is inherently spatial. Abbot and Hinton's visualizations are abstract - constructs of the mind, whereas Feynman diagrams inhabit both worlds.

27 Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983, p 13, emphasis in original

Three of these illustrations - Flatland, Galileo's sunspots and Feynman diagrams - are also discussed by Edward R. Tufte in his work *Envisioning Information*. Tufte is also attributed with first discussing Galileo's process of discovery of sunspots and his subsequent deductions, in the context of visualizing information.

1.5.1 Flatland

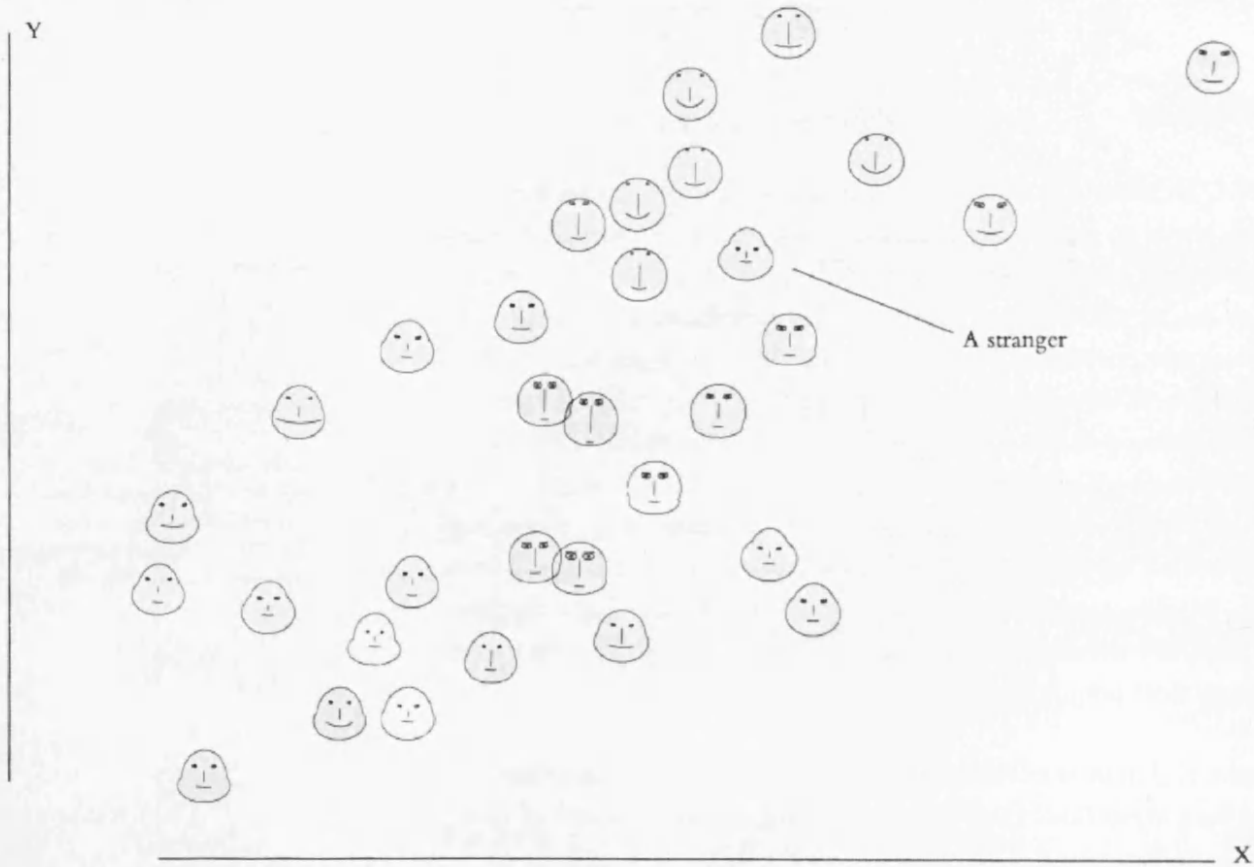
In 1884, Dr. Edwin A. Abbot, headmaster of the City of London School, wrote an immensely successful scientific novel entitled *Flatland: A Romance Of Many Dimensions, By A. Square*. Despite its humorous title, as Ian Stewart suggests in *Flatterland, Like Flatland Only More So*, "the scientific purpose of Flatland was serious and substantial."[28] Abbot used the fourth dimension as a tool for exposing the "subtle bigotry and suffocating prejudice prevalent in Victorian England."[29] The novel pictures beings whose experience is confined to a two-dimensional plane, and who have no ability to experience the third dimension by moving off the surface on which they live. In Flatland, "everyone is a geometric object. Women, occupying the lowest rank in the social hierarchy, are mere lines, the nobility are polygons, while the High Priests are circles. The more sides people have, the higher their social rank."[30]

28 Stewart, Ian, *Flatterland: Like Flatland Only More So*, London: Macmillan, 2001, p vii

29 Kaku, Michio, *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps and the Tenth Dimension*, Oxford: Oxford University Press, 1994, p 56

30 ibid, p 56

The central character of the story is a gentleman, Mr. A. Square, who is visited by a mysterious Lord Sphere, from an unknown world of three dimensions, Spaceland. Lord Sphere uses the method of analogy to convey his form *ie* dimension to Mr. Square. Since Mr. Square has no conception of a third dimension, Lord Sphere moves up and down the

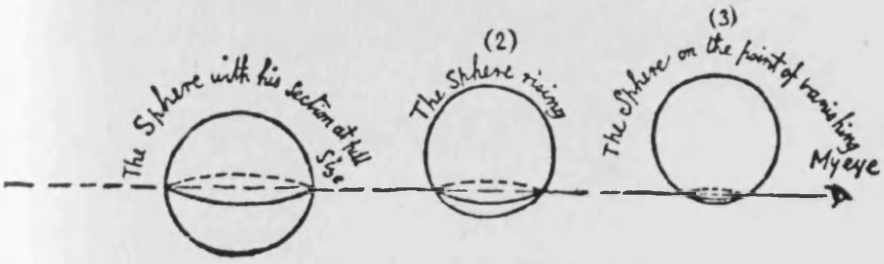


Chernoff faces

'plane' or flatland on which Mr. Square resides. Mr. Square visualizes the intersections of the Sphere as the motion or growth of form in his world of a circle which appears and gradually disappears. The following passage from the novel gives an interesting insight into the scene:

> Stranger. (To Himself.) How shall I convince him? Surely a plain statement of facts followed by ocular demonstration ought to suffice.--Now, Sir; listen to me.
>
> You are living on a Plane. (...)I am not a plane Figure, but a Solid. You call me a Circle; but in reality I am not a Circle, but an infinite number of Circles, of size varying from a Point to a Circle of thirteen inches in diameter, one placed on the top of the other. When I cut through your plane as I am now doing, I make in your plane a section which you, very rightly, call a Circle. For even a Sphere--which is my proper name in my own country--if he manifest himself at all to an inhabitant of Flatland--must needs manifest himself as a Circle.
>
> But now prepare to receive proof positive of the truth of my assertions. You cannot indeed see more than one of my sections, or Circles, at a time; for you have no power to raise your eye out of the plane of Flatland; but you can at least see that, as I rise in Space, so my sections become smaller. See now, I will rise; and the effect upon your eye will be that my Circle will become smaller and smaller till it dwindles to a point and finally vanishes.[31]

31 Abbott, Edwin A., *Flatland: A Romance of Many Dimensions by A Square*, Oxford: Basil Blackwell, 1978, pp 71-72



Lord Sphere demonstrates his form to Mr. Square

The quest for higher dimensions has perplexed scholars and artists for centuries. Charles Hinton's tesseract (Section 2.3) is an example of an escape from the third dimension into the unknown fourth. However, how does one represent 'n' dimensions? A delightful attempt was by Herman Chernoff, who, in 1973, devised a method to represent 'k' dimensions (where $k \leq 18$). Known as Chernoff faces, "multivariate data is represented by a cartoon of a face whose features, such as length of nose and curvature of mouth, correspond to components of the point."[32] Ribarsky *et al* (1994) have produced a tool called Glyphymaker to build customized representations of multivariate data, and state that "glyphs owe their effectiveness to the human eye-brain system's ability to discern finely resolved spatial relationships and differences in shape."[33] Unfamiliar faces lurking in the neighbourhood are thus strangers!

32 Chernoff, Herman, 'The Use of Faces to Represent Points in K-Dimensional Space Graphically', *Journal of the American Statistical Association*, Vol 68, No 342, June 1973, pp 361-368

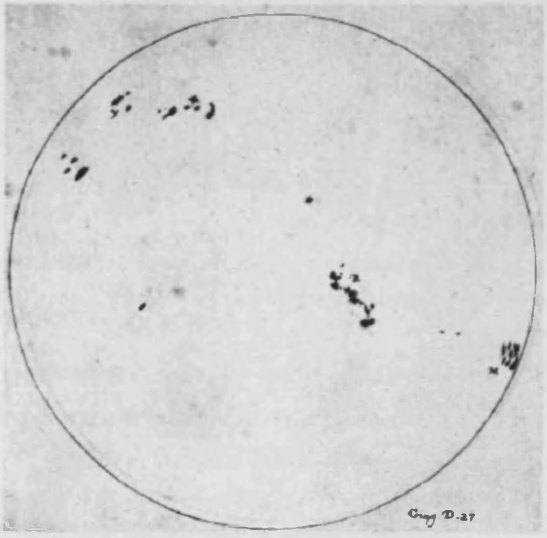
33 Ribarsky, William *et al*, 'Glyphmaker: Creating Customized Visualizations of Complex Data', *Computer*, July 1994, pp 57-64

1.5.2 Galileo's Sunspots

Galileo Galilei and others are said to have observed sunspots in detail between 1610 and 1612 by means of the telescope. What is particularly fascinating is Galileo's process of visualizing sunspots, as much as the deductions from these recorded observations. Writing from Florence in 1612, Galileo explains:

> The method is this: Direct the telescope upon the sun as if you were going to observe that body. Having focused and steadied it, expose a flat white sheet of paper about a foot from the concave lens; upon this will fall a circular image of the sun's disk, with all spots that are on it arranged with exactly the same symmetry as in the sun. The more the paper is moved away from the tube, the larger this image will become, and the better the spots will be depicted. Thus they will all be seen without damage to the eye, even if the smallest of them-which, when observed through telescope, can scarcely be perceived, and only with fatigue and injury to the eyes.[34]

34 Drake, S., *Discoveries and Opinions of Galileo*, New York: DoubleDay, 1957, p 115



Galileo's sunspot diagram

> In order to picture them accurately, I first describe on the paper a circle of the size that best suits me, and then by moving the paper towards or away from the tube I find the exact place where the image of the sun is enlarged to the measure of the circle I have drawn. This also serves me as a norm and rule for getting the plane of the paper right, so that it will not be tilted to the luminous cone of sunlight that emerges from the telescope. For if the paper is oblique, the section will be oval and not circular, and therefore will not perfectly fit the circumference drawn on the paper. By tilting the paper the proper position is easily found, and then with a pen one may mark out spots in their right sizes, shapes, and positions. But one must work dextrously, following the movement of the sun and frequently moving the telescope, which must be kept directly on the sun.[35]

35 ibid, pp 115, 116

Information consists of small multiples and variables, which is exemplified in Galileo's sunspot diagrams. The multiples being sunspots, and variables

- the shape and position of the spots. The position of sunspots indicate movement as a function of time, and dispel myths about two facts, simultaneously linking physics, astronomy and mathematics:
a) "The sunspots must be located either on the surface of the sun or at a negligible distance from it"[36] *ie* the sunspots must be contiguous to the surface of the sun.
b) The cyclical changes in the paths of sunspots are attributed to the fact that a rotating sun is observed from a rotating and orbiting earth.

36 Drake, S., *Galileo Studies: Personality, Tradition and Revolution*, Ann Arbor: The University of Michigan Press, 1970, p 115

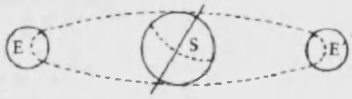
Illustrations from Galileo's *Three Letters on Sunspots* (1613), reporting the sunspots of October 1611. Small multiples record the movement of sunpots on the surface of the sun.

As Galileo explains in his *Letters on Sunspots* in 1612:

> To begin with, the spots at their first appearance and final disappearance near the edges of the sun generally seem to have very little breadth, but to have the same length that they show in the central parts of the sun's disk. Those who understand what is meant by foreshortening on a spherical surface will see this to be a manifest argument that the sun is a globe, that the spots are close to its surface, and that as they are carried on that surface toward the centre they will always grow in breadth while preserving the same length. (...)this maximum thinning, it is clear, takes place at the point of greatest foreshortening....[37]

37 ibid, p 107

As mentioned earlier, visualization leads to discovery, decision-making and explanation. Galileo's observations and subsequent deductions made him a supporter of the Copernican planetary system very early on, for if the Ptolemaic system were true, the sunspots would not vary gradually throughout the year, but would twist about the same set of paths everyday. This is not what is observed. Finally, a modern interpretation of Galileo's sunspot diagrams: a fine macro overview and micro detail of the measurements of sunspots from 1877 to 1902, by E. W. Maunder. Shown here is one-dimensional data - the latitudes covered by a spot on the surface of the sun, each vertical stroke representing one observation. A



a. Copernican

b. Ptolemaic

pattern emerges, akin to a butterfly's wingspan, showing distributions on both hemisphere's of the sun, and most importantly: a visual measure of *variation*, which is at the heart of quantitative reasoning.

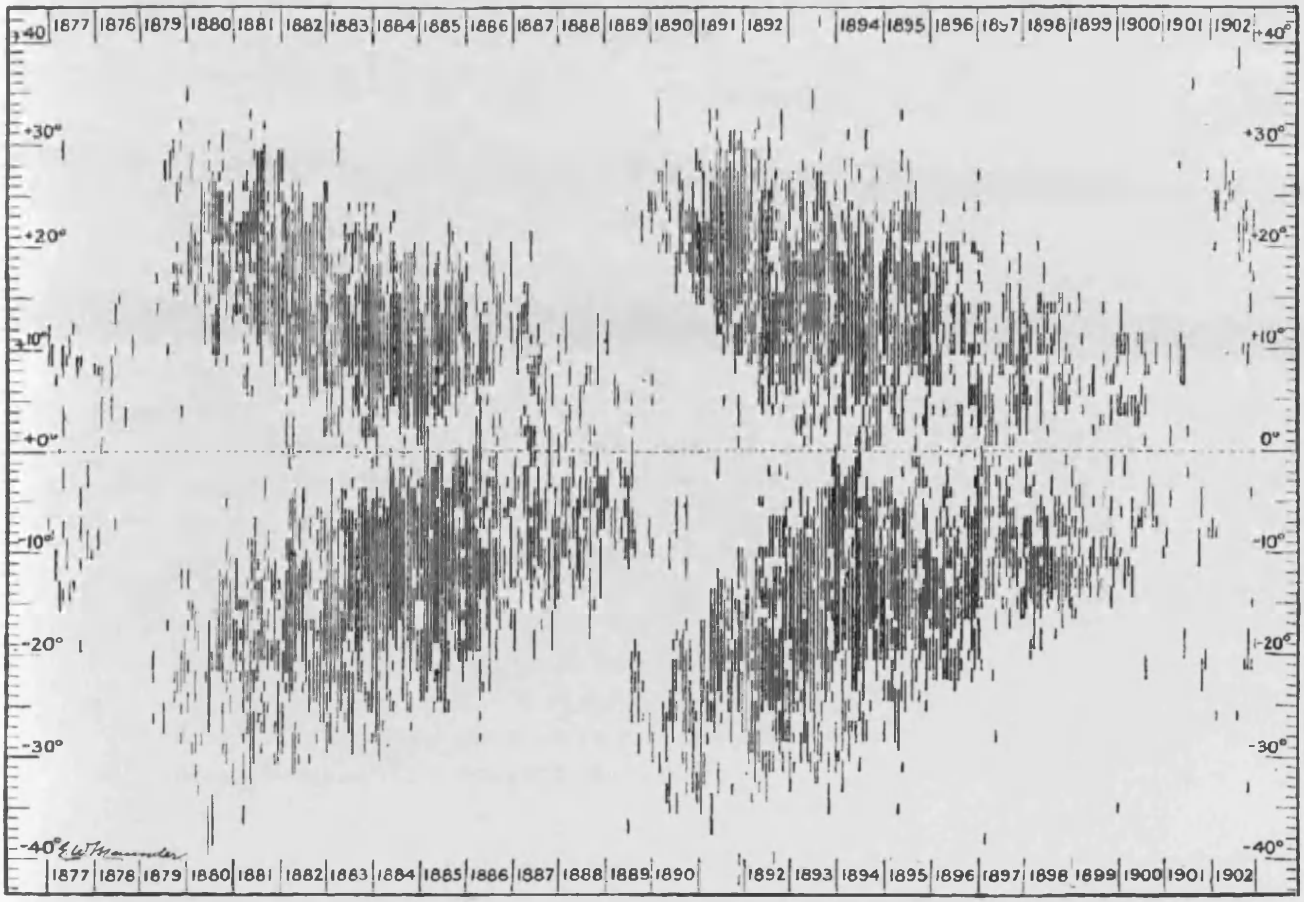


FIG. 8.—DISTRIBUTION OF SPOT-CENTRES IN LATITUDE, ROTATION BY ROTATION, 1877–1902.

E.W. Maunder's sunspot diagram

1.5.3 Hinton's Cubes

Charles Howard Hinton, an English professional mathematician, is widely remembered for a lifetime of work dedicated to the inquiry of higher dimensional space. "The bulk of his writings are aimed at developing in the reader the power to think about 4-D space."[38] Hinton is believed to have coined the word 'tesseract' for his unravelled hypercubes in four dimensions. Hinton treats dimensions as 'variables' and our experience of space as the 'variation of these variables', for "is that which we call space simply the organised knowledge of the relations of these variables?"[39]

38 Hinton, Charles H., (Editor) Rudolph v.B. Rucker, *Speculations on the Fourth Dimension: Selected Writings of Charles H. Hinton*, New York: Dover, 1980, p V

39 ibid, p 72

The following extract from *The Recognition of the Fourth Dimension* in 1902 induces the reader to visualize a tesseract in the fourth dimension:

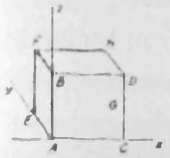


figure 1.5.3a

Refer for a moment to figure 1.5.3a. The point A, moving to the right, traces out the line AC. The line AC, moving away in a new direction, traces out the square ACEG at the base of the cube. The square AEGC, moving in a new direction, will trace out the cube ACEGBDHF. The vertical direction of this last motion is not identical with any motion possible in the plane of the base of the cube. It is an entirely new direction, at right angles to every line that can be drawn in the base. To trace out a tesseract the cube must move in a new direction - a direction at right angles to any and every line that can be drawn in the space of the cube.

The cubic sections of the tesseract are related to the cube we see, as the square sections of the cube are related to the square of its base which a plane being sees.

Hence in four dimensions a body rotates about a plane. There is no such thing as rotation round an axis.

Figure 1.5.3b1 represents a cube in our space, the three axes x, y, z, denoting its three dimensions. Let w represent the fourth dimension.

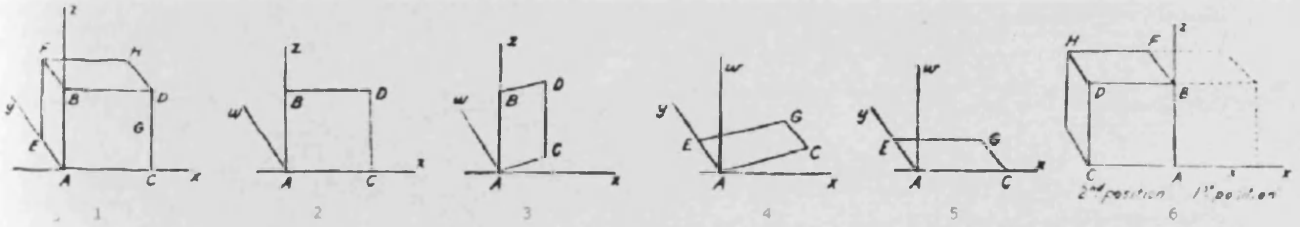


figure 1.5.3b

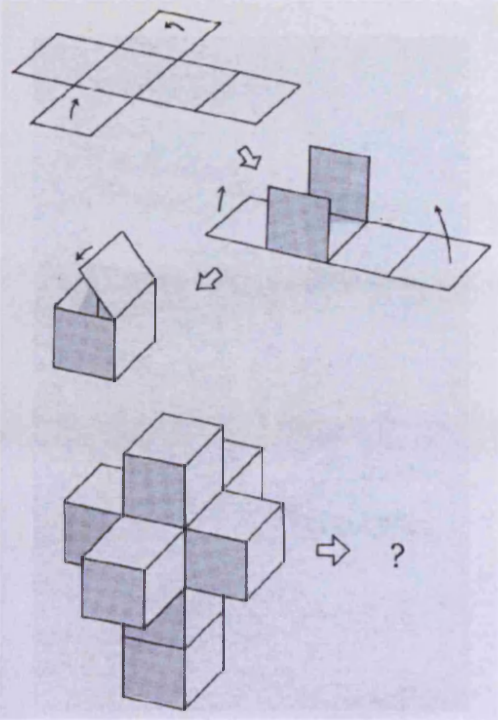
Thus, if the cube be turned by an x to w turning, both the edge AB and the edge AC remain stationary; hence the whole face ABEF in the yz plane remains fixed. The turning has taken place about the face ABEF.

Suppose this turning continue till AC runs to the left from A. The cube will occupy the position shown in figure 1.5.3b6. This is the looking-glass image of the cube in figure 1.5.3a. By no rotation in three-dimensional space can the cube be brought from the position in figure 1.5.3b1 to that shown in figure 1.5.3b6.

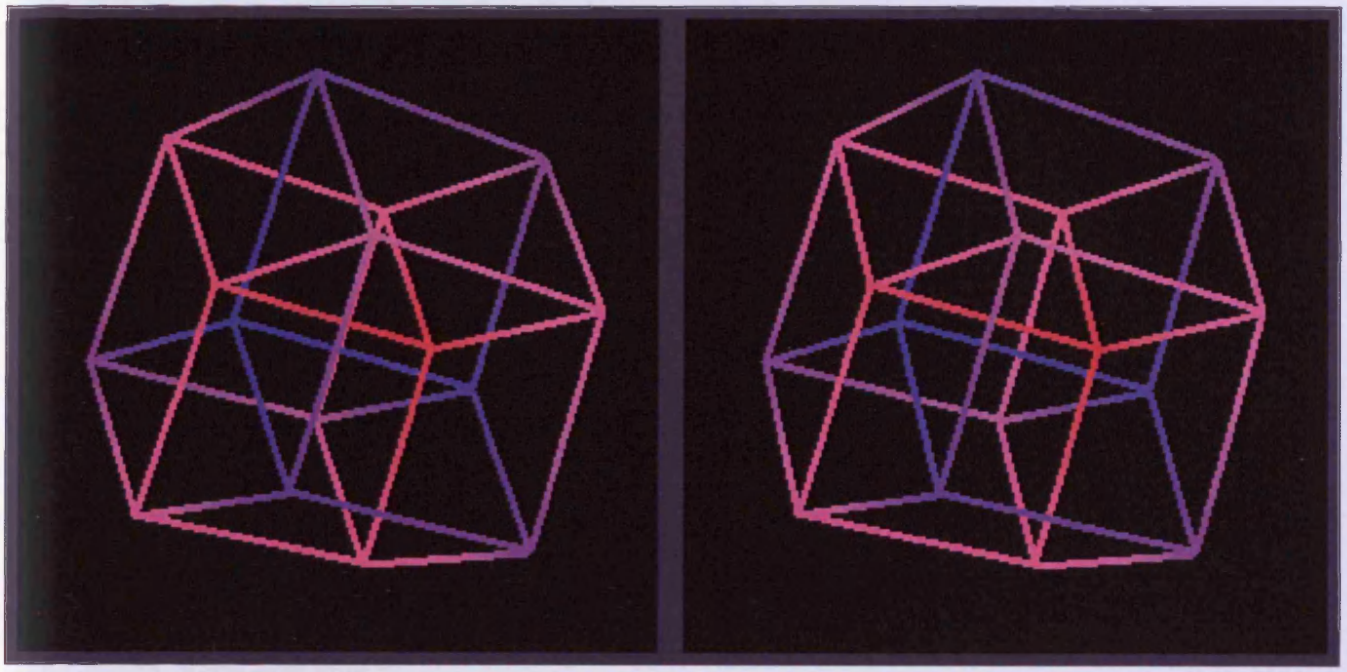
One of the conditions, then, of our inquiry...is that we form the conception of rotation about a plane. The production of a body in a state in which it presents the appearance of a looking-glass image of its former state is the criterion for a four-dimensional rotation.[40]

40 Hinton, Charles H., (Editor) Rudolph v.B. Rucker, *Speculations on the Fourth Dimension: Selected Writings of Charles H. Hinton*, New York: Dover, 1980, pp 149, 150

In a three-dimensional world, the two-dimensional faces of the cross shown in the figure can be folded to form a cube. In a four-dimensional world, a hypercube can be *unfolded* to form the cross consisting of six cubes *ie* a tesseract. It is impossible for a being in this world to visualize a hypercube, but a being in the fourth dimension could fold these cubes into one hypercube. Also, in three-dimensions, the shadows of a cube would be two-dimensional. In four dimensions, the shadow of a hypercube would be "a cube within a cube."[41] It is also worth noting that the number of vertices double with every dimension: a point has one, a line or segment has two, a square has four, a cube has eight, and a tesseract has sixteen. In general, and 'n' dimensional cube has 2^n vertices. The stereoview below is an attempt to depict the tesseract beyond the flatland of this page.



41 Kaku, Michio, *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps and the Tenth Dimension*, Oxford: Oxford University Press, 1994, p 73



stereoview of a hypercube

1.5.4 Hawk-Eye

In 2001, analysis in cricket entered a new dimension with the introduction of a technology named Hawk-Eye. Hawk-Eye is a simple visualization technique that allows the possibility of looking at the ball from practically any viewpoint in three dimensions. It uses "sophisticated image processing techniques to process the output of a series of cameras positioned around the sports ground."[42] Traditional slow motion techniques lack the three-dimensional interactivity which Hawk-Eye can offer.

42 Hawk-Eye Innovations, 'Hawk-Eye', *http://www.hawkeyeinnovations.co.uk*, Accessed <May 28, 2004>

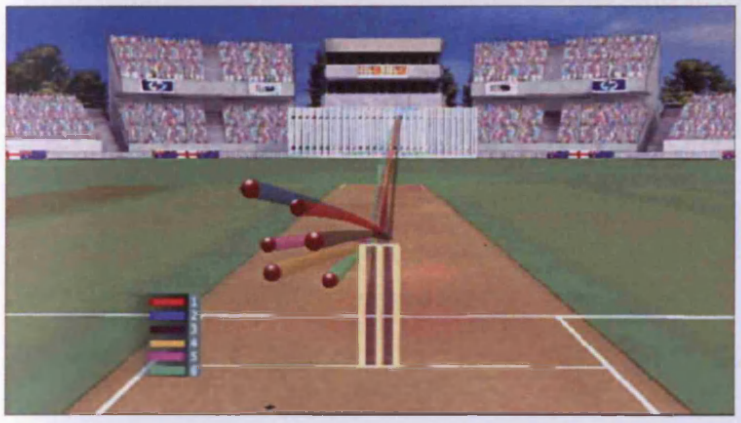


Image 1.5.4a: Hawk-Eye used for bowling analysis

"Hawk-Eye aims to aid the umpire by giving inputs on the physics of the ball by tracking and predicting its motion."[43] Information such as the exact pitch and bounce of each ball of the bowler can be visually analysed. Hawk-Eye has led to instant replays of significant events on the field. Controversial decisions such as leg before wicket (lbw), are now subject to cross examination by this visualization technique. The inclusion of Hawk-Eye as an official tool to aid the umpire is still a matter of serious debate. However, the technology has had, according to Hawk-Eye Innovations, "important ramifications for umpiring and refereeing."[44]

43 PCQuest, 'How Hawk Eye Works', *http://www.pcquest.com/content/cricket/103030404.asp*, Accessed <May 28, 2004>

44 Hawk-Eye Innovations, 'Hawk-Eye', *http://www.hawkeyeinnovations.co.uk*, Accessed <May 28, 2004>

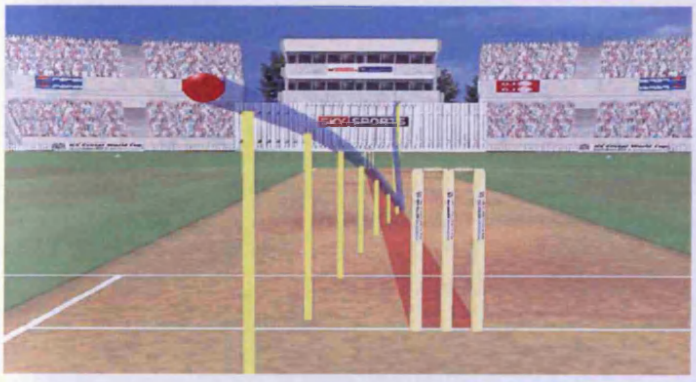
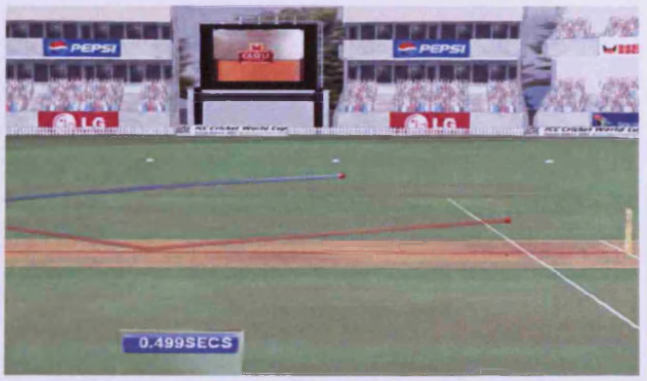


Image 1.5.4b

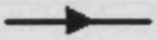


1.5.4c

Images 1.5.4a, b & c allow local comparisons of the cricket ball in flight. Unlike a conventional time series based on economic data, where the same variable is compared against itself over time, these images compare the positions of the different balls within the time-span of a bowler's over, which is six deliveries, hence six recorded measurements and hence six comparisons.

1.5.5 Feynman Diagrams

Feynman Diagrams, as the name suggests, are diagrams invented by the late physicist Richard Feynman for visualizing the interactions of particles in Quantum Electrodynamics (QED). These extraordinarily elegant diagrams indicate a process, which begins on the left and ends on the right. Each line in a Feynman diagram indicates a particle; the simplest particles in QED theory being:

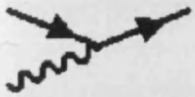
an electron, represented by 

a positron, represented by 

a photon, represented by 

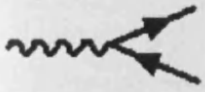
Thus, an electromagnetic interaction is represented by a vertex:


an electron emits a photon


an electron absorbs a photon


a positron emits a photon


a positron absorbs a photon


a photon produces an electron and a positron

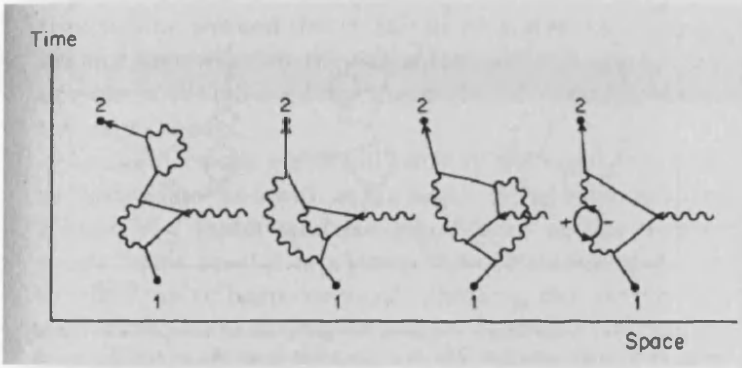

an electron and positron meet and annihilate

In *Beautiful Evidence*, Edward Tufte describes the essence of Feynman diagrams as follows: "These diagrams portray the interactions of photons, electrons, positrons, their colleagues and anti-colleagues by means of visual reasoning, logical enumeration, and mathematical operations."[45] Although the formal derivation of these diagrams by means of path integrals[46] has been subject to question, their numerical agreement under laboratory conditions is a solid fact. As Martinus Veltman states in *Diagrammatica: The Path to Feynman Diagrams*, "Physics is a quantitative science, and such agreement defines its validity."[47]

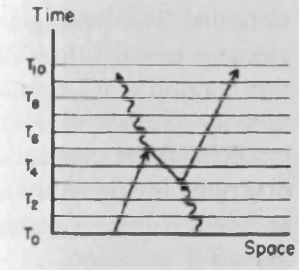
45 Tufte, Edward R., 'Beautiful Evidence', *www.edwardtufte.com*, Accessed <January 10, 2006>

46 Path Integrals are mathematical tools for solving problems in classical systems which are subject to random influences from their surrounding medium, for example, Brownian motion.

47 Veltman M., *Diagrammatica: The Path to Feynman Rules*, Cambridge: Cambridge University Press, 1994, p XI



These diagrams are 'pictures of verbs' and they represent basic quantum electrodynamic actions. The stage upon which these actions takes place is space-time. For example, in figure below, an electron and photon travel towards each other from T0. At T3, the photon disintegrates to produce an electron and a positron. At T5, the electron collides with the original electron to produce a photon. The electron produced in the earlier disintegration carries on forward in time. These results have been tested scientifically in a laboratory. Feynman diagrams are an exercise in data reduction. They reduce a complex process into an elementary diagram in the most economical way without compromising meaning.



1.5.6 "Life before Birth"[48]: 4D scans of the human embryo

48 after Channel 4's documentary *Life Before Birth*, broadcast <April 07, 2005 at 21:00 hrs>

Sonography or Ultrasound is an imaging technology extensively utilised in medical practice for obstetric imaging. Recent technological advances in this area have enabled high resolution images of a prenate inside a mother's womb to be visualized dynamically *ie* the images are updated to capture the movements of the foetus. The three dimensions of space and fourth dimension of time enable us to visualize 'life before birth' as never before. The emotional, psychological and physiological impact of such imaging technologies is profound. Research has shown that such

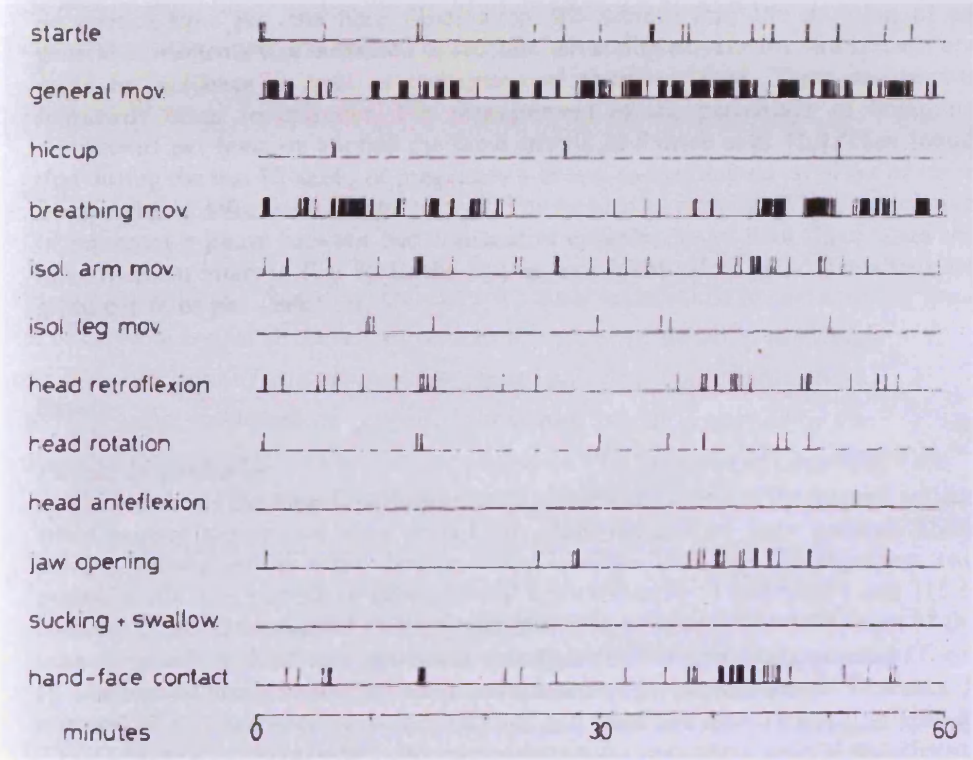
Still images of 4d scans of a prenate inside the mother's womb.

visualization techniques utilised during pregnancy strengthen the bond between mother and child.[49] "In addition to yawning, sucking and swallowing, which have previously been described by 2D imaging, it is now feasible by 4D ultrasound to study a full range of facial expressions including smiling, crying and eyelid movements."[50] According to Professor Stuart Campbell, a pioneer in the field of obstetric imaging, the "biggest gains in terms of 4D scanning are in two relatively unexplored areas of parental behaviour and fetal behaviour."[51]

49 Campbell S. *et al*, 'Ultrasound Scanning in pregnancy: the short-term psychological effects of early realtime scans', *J Psychosom Obstet Journal*, 1-2, pp 57-61

50 Campbell S., '4D, or not 4D: that is the question', *Ultrasound Obstet Gynecol*, Vol 19, 2002, pp 1-4

51 ibid, pp 1-4



Compiled actogram of a one hour observation of a fetus at 17 weeks. The rate of occurrence of various categories of fetal movements during the first half of pregnancy is an indication of developmental trends in fetal motility.[52]

52 de Vries, J.I.P., G.H.A. Visser and H.F.R. Prechtl, 'The emergence of fetal behaviour: II: Quantitative aspects', *Early Human Development*, Vol 12, 1985, pp 99-120

The electrocardiogram or ECG is another fine example of visualizing the rate at which the heart beats, in this case of a patient with an irregular heartbeat of 126 beats/minute:

Medical Imaging acquires vast quantities of raw data from scanning technologies such as X-rays, CAT scans, MRI scans and Sonography. Recalling the 1986 NSF Report 'Visualization in Scientific Computing', which anticipated the emergence of a new discipline based on the

need to see and understand the information hidden in the data acquired by imaging technologies such as those mentioned above, the discipline expanded rapidly and was no longer restricted to developing tools and techniques to visualize scientific data. Financial data, calculations from energy simulations, retail analysis - these are some examples of vast quantities of abstract or non-spatial data that required visual representation that allowed the user to interact with the data. Visualization is the tool which allows a user to ultimately derive knowledge from data. In aiming to do so, it relies on human visual perception capabilities to extract patterns in data. This subtle and complex relationship is discussed below.

1.6 "Graphical Perception"[53]

53 after William S. Cleveland, *The Elements of Graphing Data*, 1985

The illustrations in preceding sections use subtle techniques to enhance the effectiveness and efficiency of data graphics, most notably: *multiples and variables, context and comparison, cause and effect.* Underlying these fundamental techniques are powerful principles at play governed by the science of human perception and cognition. In *The Elements of Graphing Data*, William S. Cleveland explains,

> When a graph is constructed, quantitative and categorical information is *encoded*, chiefly through position, size, symbols, and colour. When a person looks at a graph, the information is visually *decoded* by the person's visual system. A graphical method is successful only if the decoding process is effective. No matter how clever and how technologically impressive the encoding, it is a failure if the decoding process is a failure. Informed decisions about how to encode data can be achieved only through an understanding of the visual decoding process, which is called *graphical perception*.[54]

54 Cleveland, William S., *The Elements of Graphing Data*, Monterey, California: Wadsworth, 1985, emphasis in original

A thorough review of how humans process data or information is not the subject of this study. However, it seems pertinent to discuss an area of particular importance: Preattentive Processing.

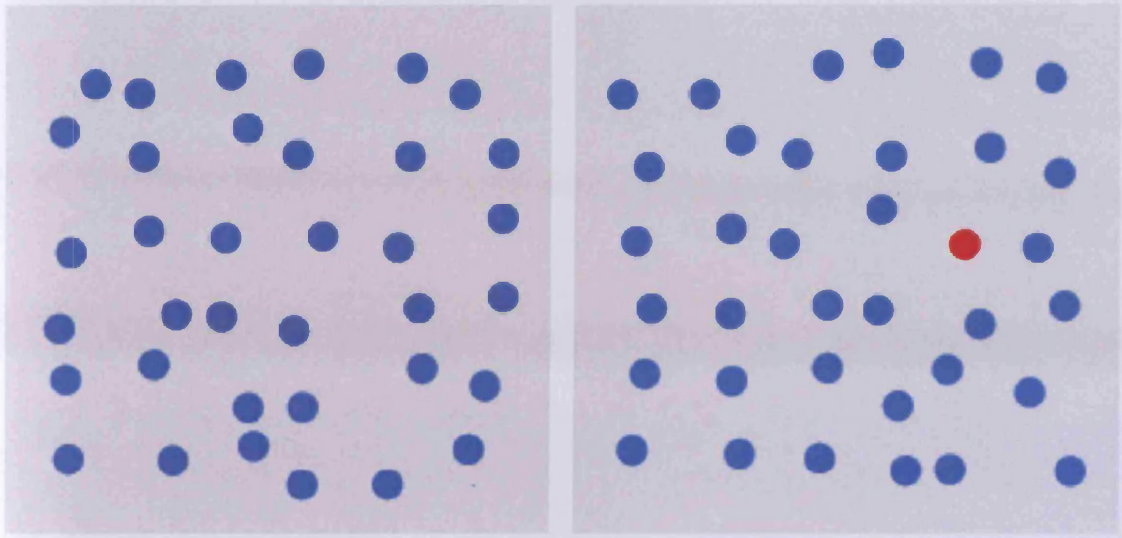
1.6.1 Preattentive Processing

Preattentive processing refers to an area of human perception and cognition that is related to vision research. Healy *et al* (1996, 1997) describe preattentive processing as,

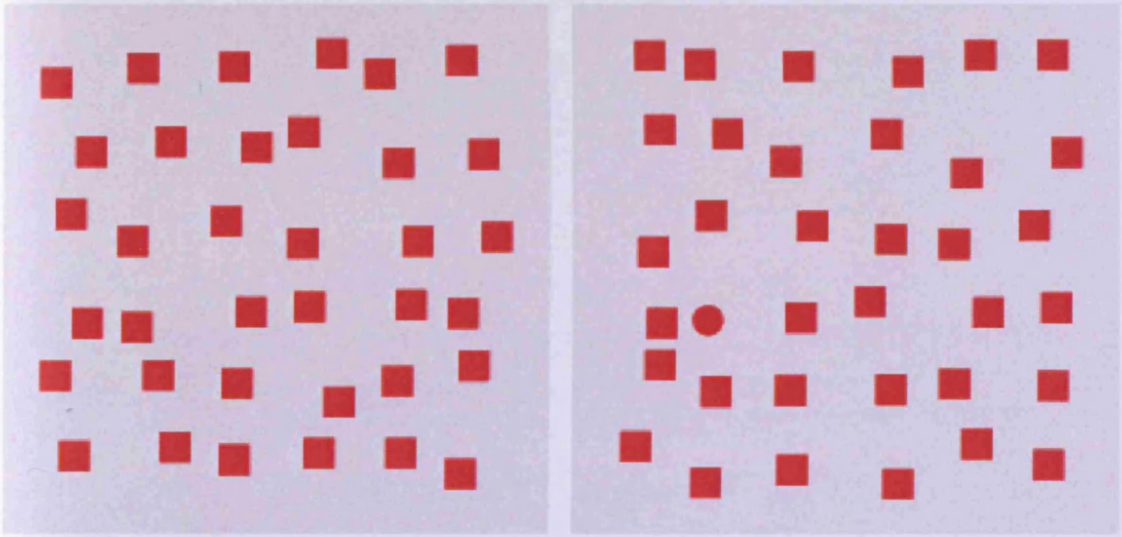
> ...cognitive operations that can be performed prior to focussing attention on any particular region of an image. Typically, tasks that can be performed on large multielement displays in 200 milliseconds or less are considered preattentive. This is because eye movements take at least 200 milliseconds to initiate. Any perception that is possible within this time frame involves only the information available in a single glimpse.[55]

55 Healey, Christopher G., Kellog S. Booth and James T. Enns, 'High-Speed Visual Estimation Using Preattentive Processing', *ACM Transactions on Computer-Human Interaction*, Vol 3, No 2, June 1996, pp 107-135

In simple terms, it is the amount of information that can be detected (gained) instantly without focussed attention *ie* in a single glimpse. In the example below, a red circle is detected preattentively in a group of blue circles,



above: same form, different hue
below: same hue, different form



Using preattentive visual features as part of a display or design allows for more intuitive interaction with the visualization. Triesman (1996) states two problems that one is confronted with: "One is to define which features or properties are the basic elements or visual primitives in this language of early vision. The second concerns how they are put together again into the correct combinations to form the coherent world that we perceive."[56] In response to the first question, visual features that have been found to be preattentive are listed to the left. It has also been established that certain visual features may cause interference, if used in

56 Triesman, Anne, 'Preattentive Processing in Vision', *Computer Vision, Graphics and Image Processing*, Vol 31, 1985, pp 156-177

Preattentive Features by Category

Form

- Line thickness
- Line length
- Line width
- Line collinearity
- Size
- Curvature
- Spatial grouping
- Added marks
- Numerosity
- Terminators
- Intersection
- Closure

Colour

- Hue
- Intensity

Motion

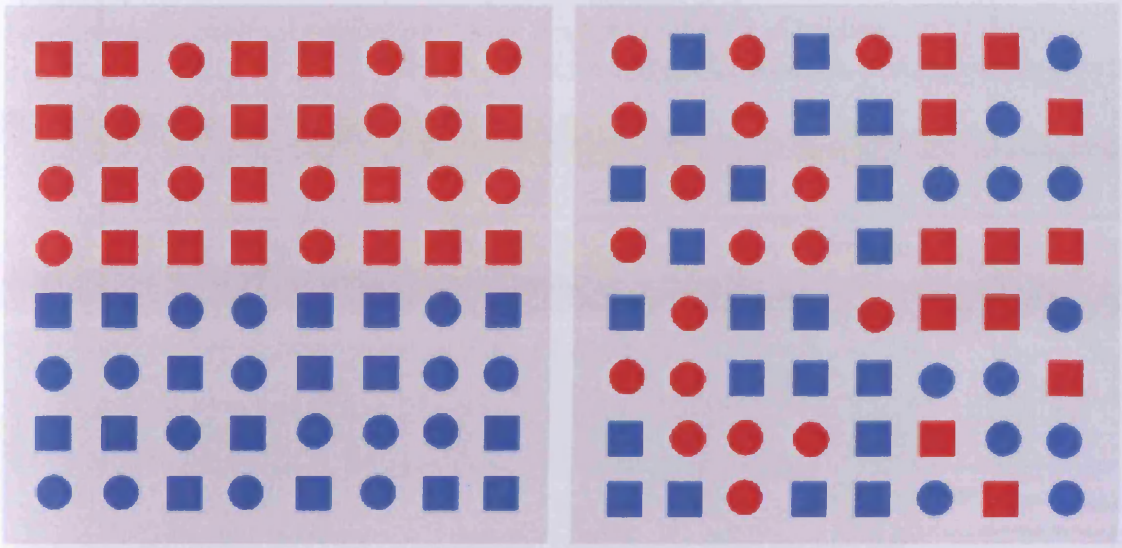
- Flicker
- Direction of motion

Spatial Position

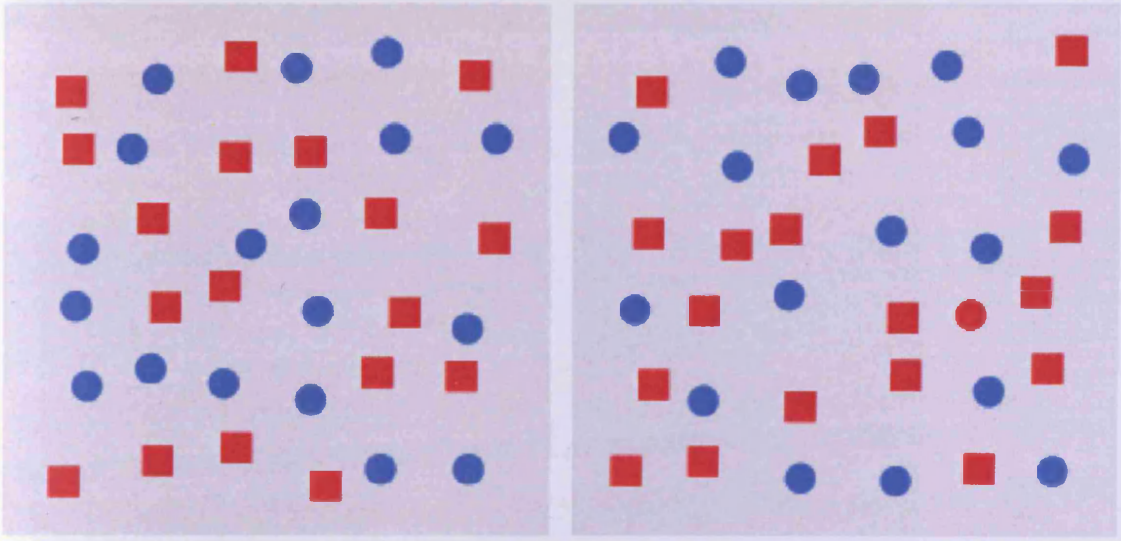
- 2D position
- Binocular Luster
- Stereoscopic depth
- 3D depth cues
- Lighting direction

adapted and modified from Healy *et al* (1996) and Ware (2002)

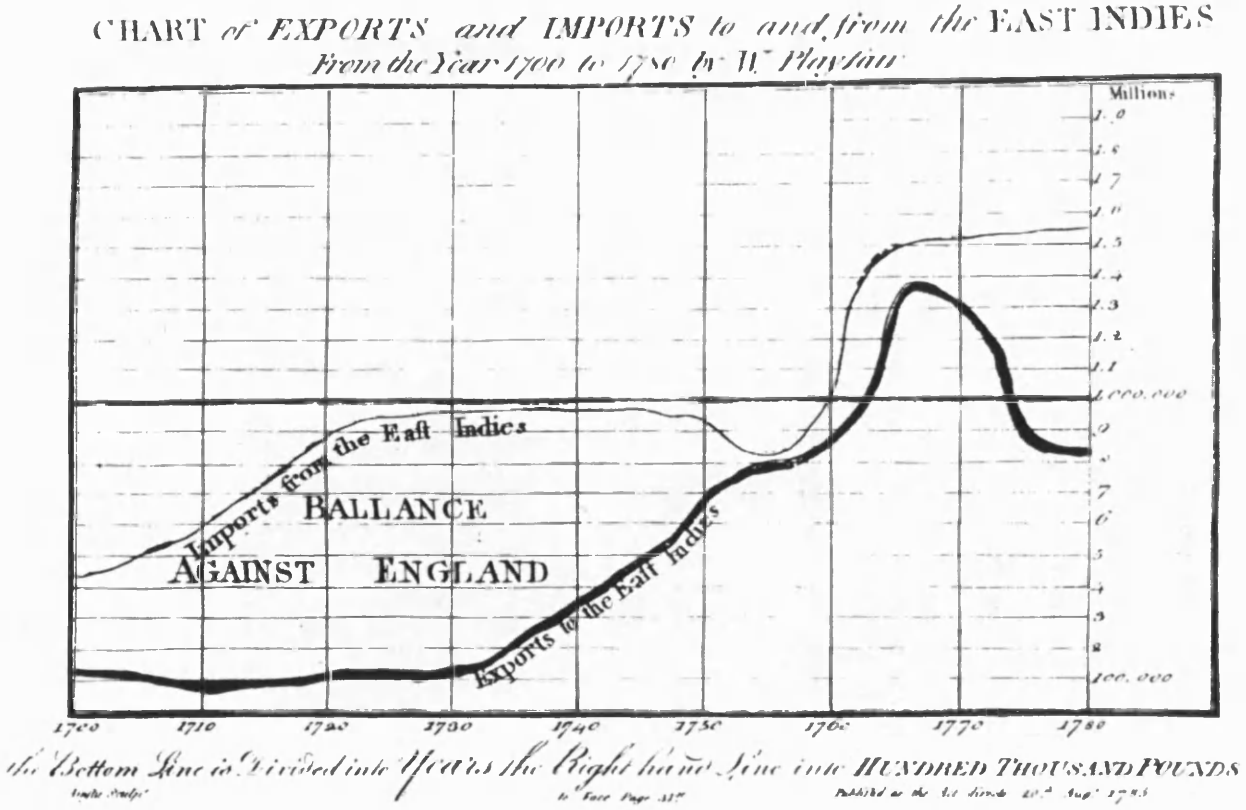
the same display (Callaghan, 1989). For example, variation of intensity interferes with hue segregation, but variation of hue does not interfere with intensity segregation. Hence, a hue boundary can be identified preattentively, regardless of whether form varies, as demonstrated below:

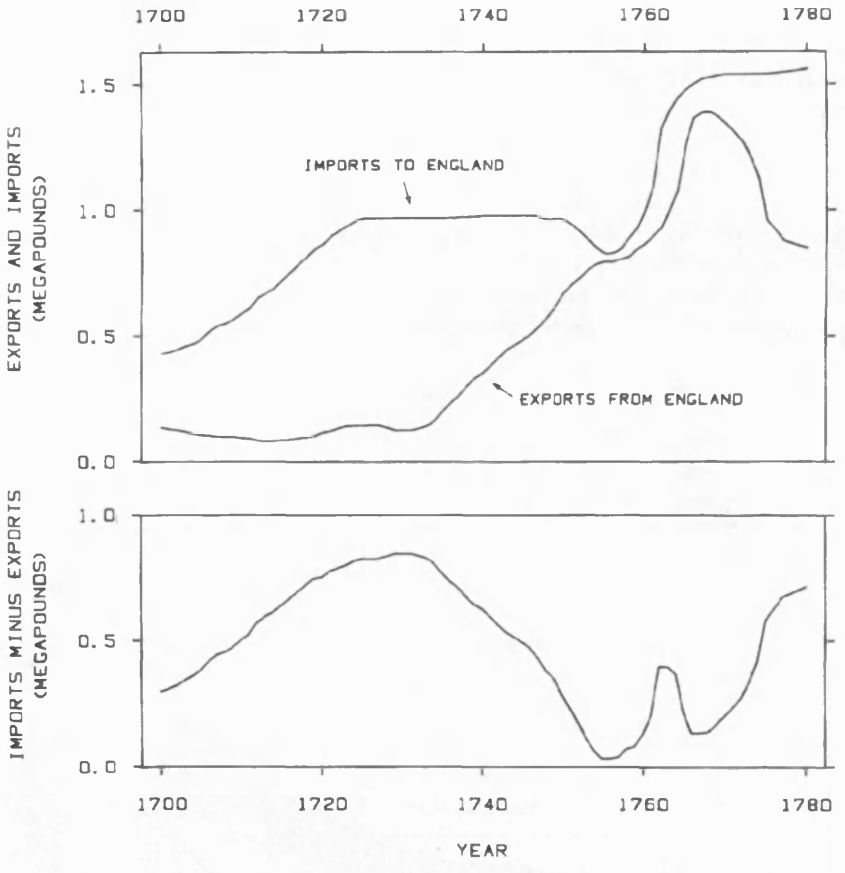


Multivariate data displays are likely to employ more than one variable to encode data values, and preattentive features are likely to be combined. For example, in a visualization representing the quantity and quality of iron ore in a survey, the quality of ore can be represented by colour, and the quantity by height. Searching for the best quality ore available in significant amount would involve looking for a serial or conjunction search for both colour (hue) and height, which is *not* preattentive.



searching for a red target in a sea of 'blue circle and red square distractors' is not preattentive, and requires a time consuming serial search.





The chart above shows the imports and exports of England to and from the East Indies between the years 1700 and 1780.

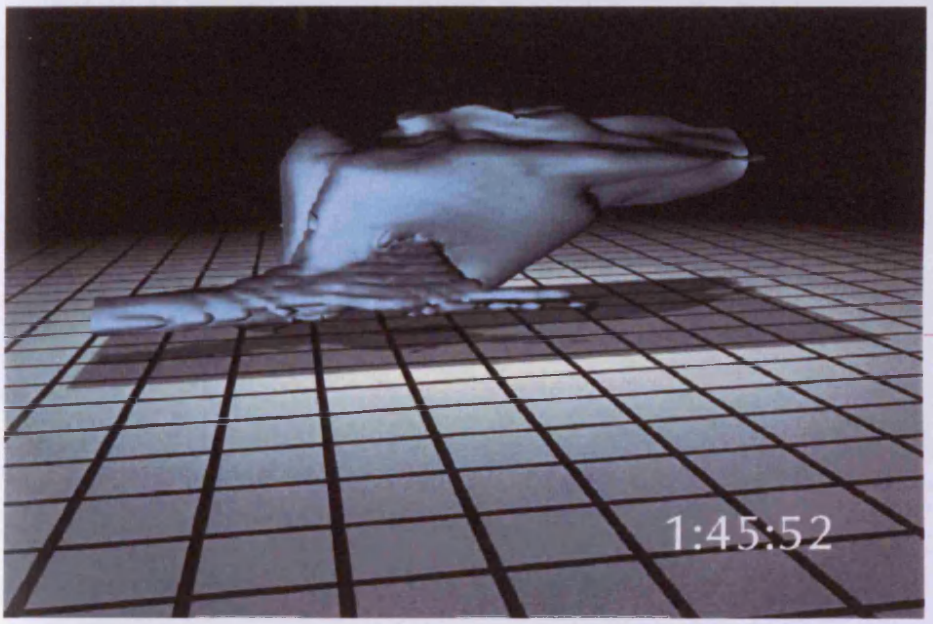
An important deduction from this time series, apart from the increase or decrease in imports or exports over time, is the net sum of money *ie* balance in favour of or against England. To decode this information, one is required to make a judgement of distance along a vertical scale. A casual observation may deduce that the difference between the imports and exports from 1760 onwards is not much, and does not change by much thereafter. This is not the case.

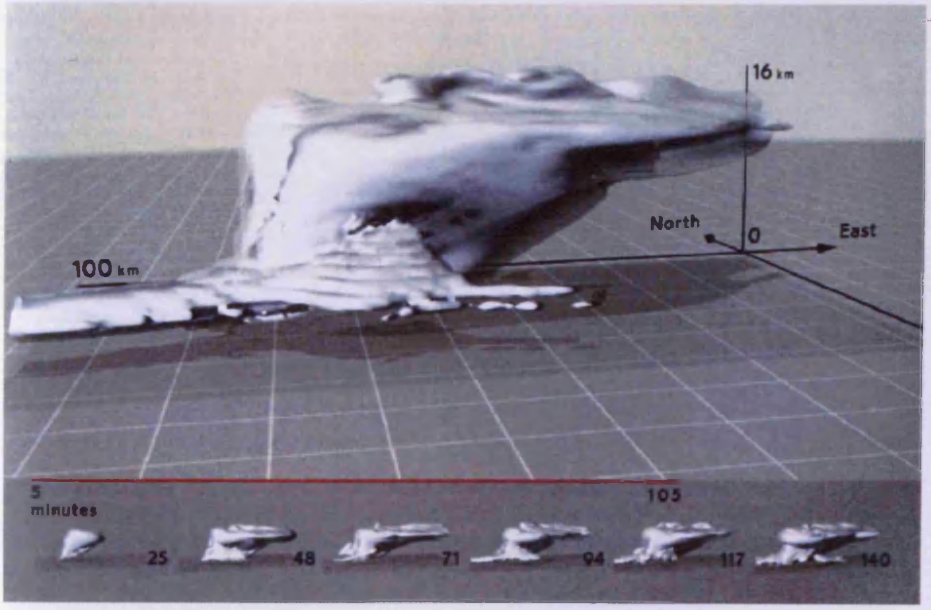
A graph showing the *difference of curves*, shows accurately how the curve suddenly rises after 1760, plunges downwards and regains strength by 1780, with a balance *not* in favour of England.

Arbitrary mapping of preattentive features to data dimensions may not build a successful visualization tool. However, applying principles of graphical perception can greatly improve multivariate data displays. In response to the second question posed by Triesman, the following two examples by William Cleveland and Edward Tufte illustrate the importance of graphical perception. In the example to the left, Cleveland demonstrates the error that results from the weakness in our visual apparatus to judge distances on a vertical scale. In the example below, Tufte *et al* redesign the graphical display of footage on the simulation of a severe thunderstorm. Fundamentals of "scale, orientation and labels - for centuries routine in maps and statistical graphics - are missing"[57] in the image below, and equally enhancing the latter.

57 Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, p 20

58 ibid, p 20





This still from a supercomputer animation of a severe storm lacks visual quality by failing to answer the following basic questions: "How big is that cloud? What direction is it moving? What are the dimensions of the grid?"[58] The redesigned example answers the above questions by:
a) introducing a three-dimensional axes that denotes the direction in which the storm is moving.
b) the intensity and contrast of the grid is softened to distinguish rather than to dominate. An overlapping of text on the axes gives the grid a scale and restores quantitative order.
c) small still images at the bottom of the grid below the red line provide a context with which to compare the moving animation above.

Carte Figurative des pertes successives en hommes de l'Armée Française dans la

Dressée par M. Minard, Inspecteur Général des Ponts et Chaus

Les nombres d'hommes présents sont représentés par les largeurs des zônes colorées à raison d'un millimètre p
des zônes. Le rouge désigne les hommes qui entrent en Russie, le noir ceux qui en sortent. ___ Les renseig
dans les ouvrages de M.M. Thiers, de Ségur, de Fezensac, de Chambray et le journal inédit de Jo
Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du M
à Mobilow et ont rejoint vers Orscha et Witebsk, avaient toujours m

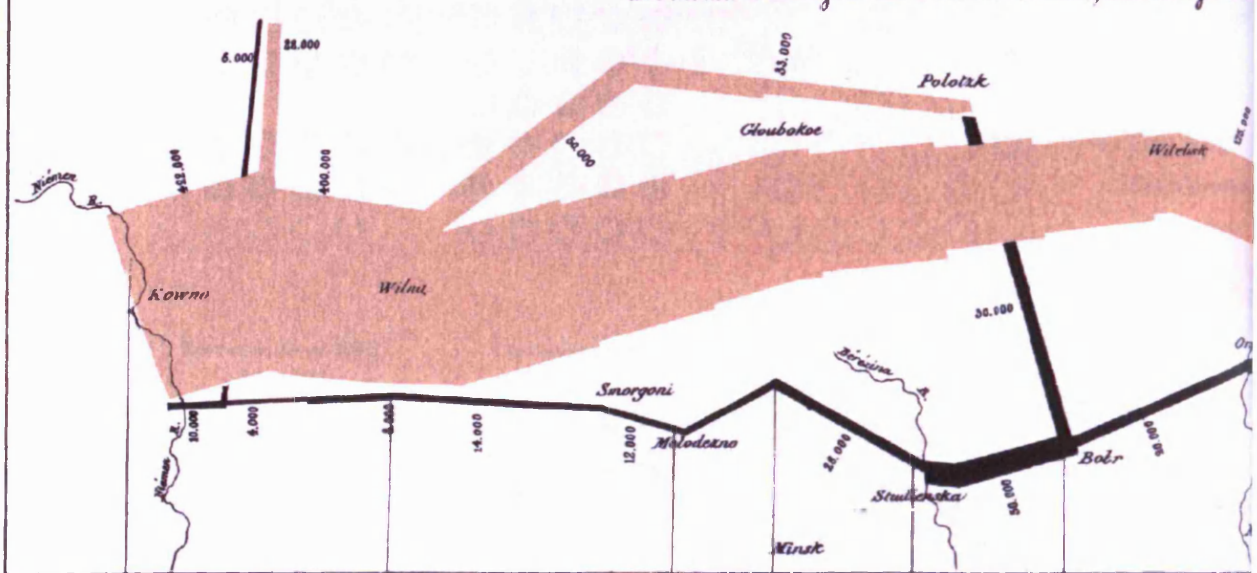


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de R

Les Cosaques passent au galop le Niémen gelé.

— 26.° le 7 X.bre

— 30.° le 6 X.bre

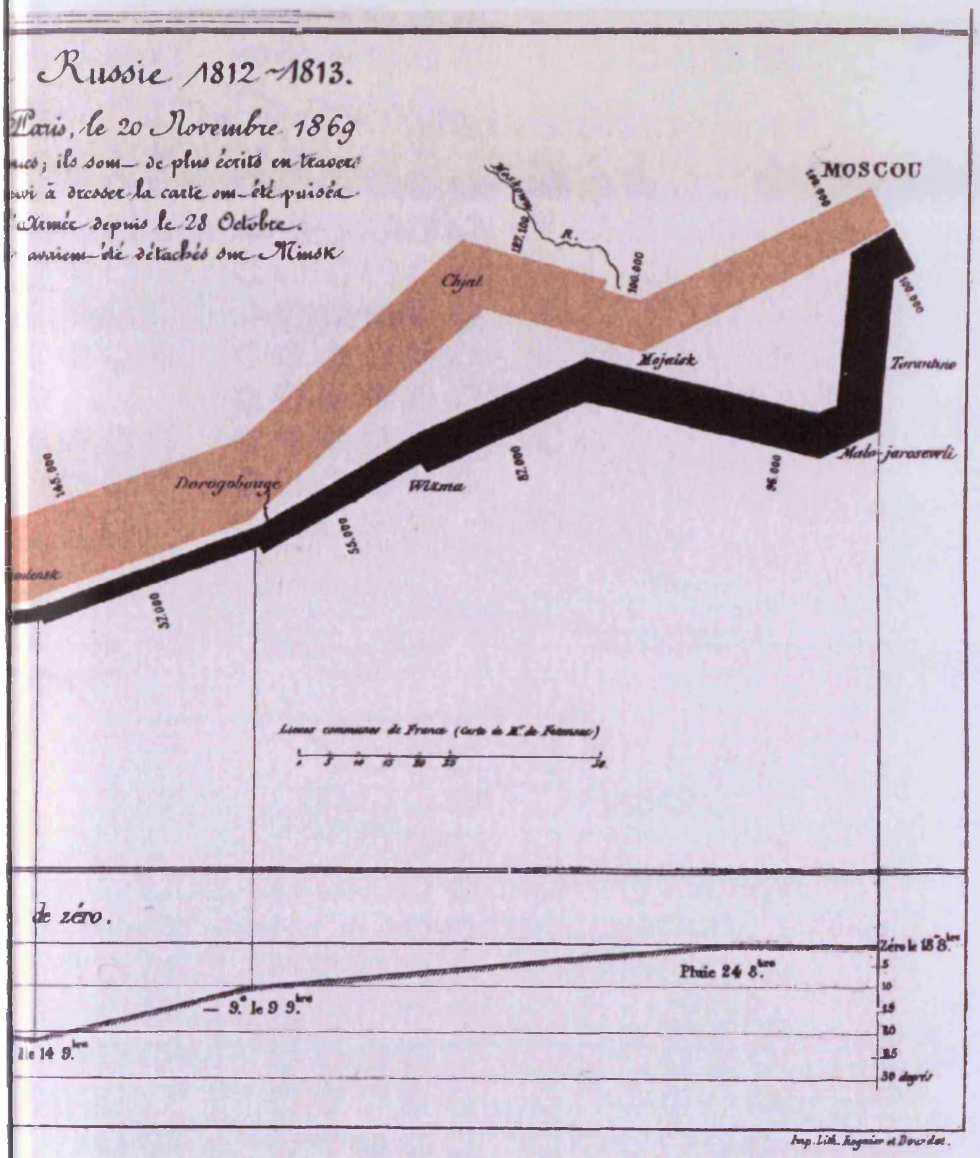
— 24. le 1er X.bre

— 20.° le 28 9.bre

— 11.°

Imp. par Regnier, R. Pas. St Marie St Gmain à Paris.

Finally, a fine example of how to visually construct a graphic that leaves behind a permanent imprint on one's mind - a superb visualization of data in several dimensions which forces the viewer to visualize the catastrophic march of Napoleon's army through Russia in the bitterly cold winter of 1812. Described by E. J. Marey as "seeming to defy the pen of the historian by its brutal eloquence"[59], it may well be "the best statistical graphic ever drawn."[60]



59 E.J. Marey, *La Methode Graphique*, Paris: G. Masson, 1885, as quoted by Edward R. Tufte in *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983, p 40

60 ibid, p 40

All the preceding examples on data graphics in diverse fields (literature, astronomy, mathematics, sports, physics and medicine) are related to graphical perception. They are all examples of exercises in communication, the ultimate aim of which, as stated before, is *to reveal data.*

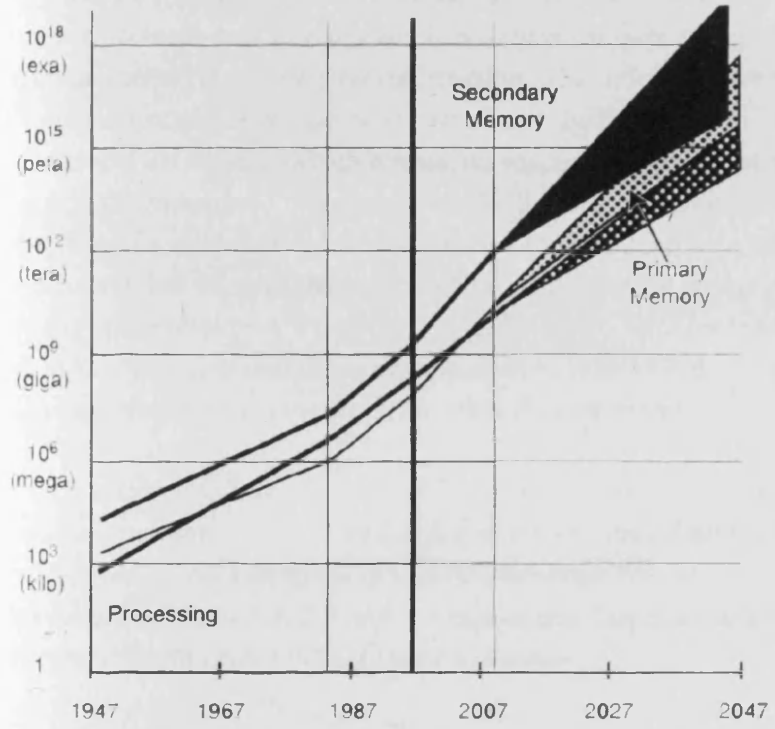
2.0 The Nature of Information

2.0 "The Nature of Information"[62]

62 after Paul Young, *The Nature of Information*, 1987

In 1997, the Nobel Laureate physicist Murray Gell-Mann summed up the future of computing over the next 50 years in an essay entitled 'Pulling Diamonds From The Clay'. The challenge we face, according to Gell-Mann, is finding 'meaning in a flood of data, and wisdom in an ocean of information.'[63] Gordon Moore, founder and chairman of Intel, first observed and posited a phenomenon now known as Moore's Law, in which every three years semiconductor capacities increase fourfold. The exponential growth of processing speeds, storage capacities, and network bandwidths at the rate of 1.60 every year, in accordance with Moore's Law, represents the ever increasing challenge to meet the growth of digital information.

63 Denning, Peter J. (Editor), *Talking Back To The Machine: Computers and Human Aspiration*, New York: Copernicus, Springer-Verlag, 1999, p 155



Evolution of computer processing speed in instructions per second and primary and secondary memory size in bytes from 1947 to the present, with a surprise-free projection to 2047. Each division represents three orders of magnitude and occurs in roughly fifteen-year steps.

Much hype surrounds the word information. In economic circles, there is talk of information replacing capital as the basis for local and global economies. Eaton and Bawden (1991) cautiously define information as a *resource*.[64] This is indeed reflected in the stock market trends of companies such as Google Inc. Information is finding its way into our everyday lives through pervasive computing. In *City of Bits*, William J. Mitchell comments, "More and more of the instruments of human

64 Eaton, J.J. and D. Bawden, 'What Kind of Resource is Information?', *International Journal of Information Mangement*, Vol 11, 1991, pp 156-165

interaction, and of production and consumption [are] being miniaturized, dematerialized, and cut loose from fixed locations."[65] The mobile or cellular phone is a classic example. It can now mutate into a traditional phone, a personal organiser, a camera, a game pad, a word processor, a radio, an mp3 player, and much more. The range of content sent and received across this device is immense, both in quantitative and visceral terms. Weather reports, travel updates, match scores, stock markets, news headlines: all can come in text, image or video format; access to information, anywhere/everywhere; anytime/all the time.

65 Mitchell, William J., *City Of Bits: Space, Place, and the Infobahn*, Cambridge, Massachusetts: MIT Press, 1995, p 4

The verbal penetration of the terms data, information and knowledge belies the real meanings associated with them. Specifically, the synonymous use of data and information in everyday language does not recognise the fundamental technical distinctions between these words. Data is information in its raw state. It is only when data makes some significant sense, that it *becomes* information. The colloquial definitions of information are too vague to be part of this study. They are like "operational definitions, which require no real understanding of what it is that is being measured."[66] For example, the journey from the operational definition of temperature around 1600, as a number read on a scale, to its technical definition in the middle of the 19th century, as the average speed of molecules, took a quarter of a millennium. A robust, succinct, and precise technical definition of information, akin to that of energy, is still a matter of serious pursuit in the scientific community.

66 Baeyer, Hans Christian von, *Information, The New Language of Science*, London: Phoenix, Orion Books, 2003, p 19

The sections that follow attempt to both define information and expose its nature. Sections 2.1, 2.2 and 2.6 define information through the terms *difference*, *freedom of choice* and *the communication of relationships* respectively. Sections 2.3, 2.4 and 2.5 expose the close association of information with probability, entropy and form.

2.1 Difference

In 1969, the anthropologist-philosopher Gregory Bateson defined information as "a difference which makes a difference."[67] The robustness of this simple definition comes from its obvious implication of differentiating between that which makes sense and that which does not. However, what is a difference? In *Form, Substance and Difference*, Bateson addresses it as: "A difference is a very peculiar and obscure concept. It is certainly not an event or a thing. A difference, then, is an abstract matter."[68] At this point, it seems necessary to illustrate what a difference actually is. For this purpose, I shall turn to the study of language, a reference which will recur later.

67 Bateson, G., *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution and Epistomology*, St Albans: Paladin, 1973, p 428

68 ibid, pp 426, 427

Sanskrit is the oldest and most refined of the Indo-European languages. The Indian Logician Dignaga developed a theory of meaning based on the concept of *Apoha*[69] around approximately the 5th and 6th centuries AD. Apoha can be translated as differentiation. "In Dignaga's theory, the meaning of a word consists of its repudiation of what it is not. Indeed it is only perceived to exist to the extent that it is cognized or differentiated from other things. 'Not-A' implicitly says 'A' is."[70] In *The word and the World, India's Contribution to the Study of Language*, Bimal Krishna Matilal explains,

> ...a class-name such as 'cow' can be applied to a particular, not by virtue of its possession of some abstract property such as cow universal, but by a process of *exclusion* by which all contrary particulars, horses, etc., are excluded. For Dinnaga, inference and word or language (sabda) generate knowledge in a similar fashion. Both essentially depend upon a process of *exclusion*. Just as the presence of smoke generates the knowledge of fire in a particular spot,...,the word 'cow' generates the knowledge of a cow by excluding similar not-cow individuals, and this is also aided by our learning to use the word 'cow' in the presence of a cow on some occasion.[71]

69 Coward, H.G. and K.K. Raja (Editors), *Encyclopedia of Indian Philosophies: The Philosophy of the Grammarians*, Princeton: Princeton University Press, 1990, pp 6, 27

70 Holtzman, Steven R., *Digital Mantras: The Languages of Abstract and Virtual Worlds*, Cambridge, Massachusetts: MIT Press, 1994, p 37

71 Matilal, Bimal M., *The word and the World: India's Contribution to the Study of Language*, New Delhi: Oxford University Press, 2001, p 38, emphasis in original

cow not-cow not-cow not-cow

The French linguist Ferdinand de Saussure, published his *Memoir on the Primitive Vowel System in Indo-European Languages* in 1878. This work is widely regarded as one of the most important in comparative philology, and Saussure is acknowledged as the founder of modern linguistics and semiology. Saussure's view of language in terms of system, structure and relationships echoes Dignaga's theory of exclusion. In Saussure's *Course in General Linguistics*, he remarks:

> We discover not *ideas* given in advance but *values* emanating from the system. When we say that these values correspond to concepts it is understood that these concepts are purely differential, not positively defined by their content but negatively defined by their relations with other terms of the system. Their most precise characteristic is that they are what the others are not.[72]

In Saussure's view, it is the distinctions which are important. "So long as there are differences of some kind - any kind will do. Identity is wholly a function of differences within a system."[73] We can hence define a difference as a *diacritical relationship*.

72 Culler, J., *Saussure*, London: Fontana, 1985, p 26, emphasis in original

73 ibid, p 28

To give another concrete example, studies[74] on colour vocabulary have shown that there are differences in colour nomenclature around the world. Not only can the name or term used to designate a colour vary, but also the 'semantic form' (hue) that a particular colour term designates in a particular language, and its place in that colour system. In English, for example, light blue and dark blue are shades of the same colour blue. In Russian, both are treated as separate primary colours. Studies have also indicated that people have difficulties in distinguishing the boundaries of colours, *ie* where a particular colour starts or ends in colour space.[75] It is by knowing the distinctions between different colours that one can easily identify which colour term a hue (or form) may belong to. Context and comparison play a crucial role in this process of differentiation. To quote Jonathan Culler in *Saussure*; a colour, for example brown, "is not an independent concept defined by some essential properties but one term in a system of colour terms, defined by its relations with the other terms which delimit it."[76]

74 Berlin, Brent and Paul Kay, *Basic Color Terms: Their Universality and Evolution*, London: University of California Press, 1969

75 Colour Space is a mathematical model describing colour representation, usually using three or four colour components. RGB (red, green, blue) and CMYK (cyan, magenta, yellow, black) are commonly used colour spaces.

76 Culler, J., *Saussure*, London: Fontana, 1985, p 25

2.2 Freedom of Choice

In 1948, Claude E. Shannon published a classic paper, 'A Mathematical Theory of Communication' in the *Bell System Technical Journal*, which laid the foundation for Information Theory. It was later popularized and made more accessible by his colleague Warren Weaver. In *The Mathematical Theory of Communication* by Shannon and Weaver, Warren Weaver defines information as, "..this word in information theory relates not so much to what you *do* say, as to what you *could* say. That is, information is a measure of one's freedom of choice when one selects a message."[77] Hence, in theory, the more the number of choices, the more the degree or measure of freedom (assuming the selection of each choice is independent of previous or preceding choices). A pertinent question to ask is: 'what governs our choices: are they independent or dependent on previous choices'. A scenario where choices are dependent on previous events is known as a Markoff process. Language is a good example of how grammar determines our choice of words to form sentences.

77 Shannon, Claude E. and Warren Weaver, *The Mathematical Theory of Communication*, Urbana: The University of Illinois Press, 1962, p 100

For freedom of choice to exist, it is important that a 'set of messages' exists - an ensemble. However, measure of freedom of choice is not an objective measure of information. For this, Shannon invented a (conceptual) tool, in which, "...the amount of information is defined, ..., to be measured by the logarithm of the number of available choices."[78] Consider the following example: A person has to choose from a set of eight alternatives. One could say that the uncertainty in reducing the number of alternatives from eight to one is a measure of the amount of

78 ibid, p 100

information associated with this event - 1/8 (since uncertainty is reduced by a factor of 8) or 7 (since seven alternatives have been eliminated). However, from a communication theory perspective, the binary code is the most economical in storing, processing and transmitting information. This implies that the message be translated into the binary code of the computer. In doing so, the problem above is translated into a series of binary choices: Given a set of eight alternatives, one divides them into two groups of four choices each. With the flip of a coin, one of these groups is selected and the other is eliminated. The selected group of four alternatives is further divided into two groups of two choices each, and the coin is flipped again. This process is repeated until one choice remains. It takes three flips of a coin or three binary choices to reduce the number of alternatives from eight to one. Hence, according to Shannon's definition, three bits is the measure of information associated with the above event as $\log_2 8 = 3$.

However freedom of choice is not an isolated or insulated event. Context influences freedom of choice, for example, choice under certainty (one knows the possible outcome) or choice under uncertainty (the outcome is unknown). In the example above, let the set of eight alternatives be denoted by *S*, and let choosing an element *x* from this set *S* be denoted by *x/S*. Suppose the feasibility of the set *S* shrinks to *T*. The chosen element may of course, nominally remain the same *ie x*. However, *x/S* is not equal to *x/T*. Consider the case of a person choosing to fast. "Choosing to skip meals can be described as 'fasting' only if the person really does have the alternative of eating more. Indeed, if the opportunity of normal eating diminishes or disappears, the person may no longer be *able* to fast. The loss of the opportunity to eat freely is a substantive loss *even* for the person who chooses to fast. *Doing x* and *choosing to do x* are, in general, not equivalent."[79]

Shannon's operational definition of counting bits does not include the influence of context on a message; it ignores the information content of a message - its meaning. The number of bits required to choose one alternative from a set of eight is three regardless of which message is selected at the end. The association between information and meaning is explored in the last three sections of this Chapter. Shannon's definition does however, quantify the reduction in uncertainty of a system. This measure of uncertainty and its effect on decision-making is explored in the next section.

79 Sen, Amartya, 'Freedom of Choice: Concept and Content', *European Economic Review*, Vol 32, No 2 - 3, March 1988, pp 269 - 294, emphasis in original

2.3 Probability

The origins of data graphics, as seen in Chapter 1, can be traced to statistics, with some of the earliest examples of quantitative reasoning by William Playfair. But what are statistics and their relation to information? In the opening paragraph of their paper, 'Representations of Knowledge in Complex Systems', Grenander and Miller (1994) state, "The object of statistics is information. The objective of statistics is the understanding of information contained in that data."[80] Statistical data graphics such as Tukey's box plots have since long been aids to statistics in achieving this objective. More than often, the primary aim of statistical studies is confirmation - to instill certainty into the phenomenon represented by the numbers. The discussion here, however, will focus on the study of the role of chance in statistics *ie* uncertainty or probability.

80 Grenander, Ulf and Michael I. Miller, 'Representations of Knowledge in Complex Systems', *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol 56, No 4, 1994, pp 594 - 603

Probability can be defined as the frequency of occurrences. Natural or spoken languages such as English can be and have been analysed statistically. For example, according to the Oxford English Corpus (OEC), the 100 most common lemmas[81] in the English language are,

1-10	the	be	to	of	and	a	in	that	have	I
11-20	it	for	not	on	with	he	as	you	do	at
21-30	this	but	his	by	from	they	we	say	her	she
31-40	or	an	will	my	one	all	would	there	their	what
41-50	so	up	out	if	about	who	get	which	go	me
51-60	when	make	can	like	time	no	just	him	know	take
61-70	people	into	year	your	good	some	could	them	see	other
71-80	than	then	now	look	only	come	its	over	think	also
81-90	back	after	use	two	how	our	work	first	well	way
91-100	even	new	want	because	any	these	give	day	most	us

Hence, 'the' is the most frequently occurring lemma in the English language. Probability, defined in terms of frequency of occurrences, is an objective tool for assessment, to do with number or counting. However, there is also a subjective approach to probability. The English Minister and able mathematician Reverend Thomas Bayes established a subjectivist approach to statistical analysis. His 'An Essay towards solving a Problem in the Doctrine of Chances' was published posthumously in 1764 by his close friend and aid Richard Price. Bayes's theorem has "since become a focal point of debates on the foundations of inference"[82] and has divided the statistical community into Bayesians and Frequentists. The significance of Bayes theorem has recently come to the fore, with computational analysis rendering Bayesian methods possible, in very

81 A lemma is a base word, for example: *climbs, climbing* and *climbed* are examples of the lemma *climb*. Ten lemmas (the, be to, of, and, a, in, that, have, I) account for 25% of the 1 billion words used in the Oxford English Corpus.

82 Stigler, Stephen M., 'Thomas Bayes's Bayesian Inference', *Journal of the Royal Statistical Society, Series A (General)*, Vol 145, No 2, 1982, pp 250 - 258

complex statistical problems such as the location of geographical clusters of cases of a disease.[83]

83 Bland, J.M. and D.G. Altman, 'Statistics notes, Bayesians and frequentists', *British Medical Journal*, Vol 317, 1998, p 317, <downloaded from bmj.com on June 6, 2006>

One of the most controversial probability problems, a subject of much academic and public debate, is known as the Monty Hall problem, named after Monty Hall, the host of a television game show 'Let's Make A Deal' in the United States. On September 9, 1990, Marilyn vos Savant, a writer for *Parade* magazine, posed a mathematical puzzle as follows:

> Suppose you're on a game show, and you're given the choice of three doors. Behind one door is a car, behind the others, goats. You pick a door, say #1, and the host, who knows what's behind the doors, opens another door, say #3, which has a goat. He says to you, "Do you want to pick door #2?" Is it to your advantage to switch your choice of doors?
>
> Yes; you should switch. The first door has a 1/3 chance of winning, but the second door has a 2/3 chance.[84]

84 Savant, Marilyn vos, 'Letters to the Editor', *The American Statistician*, November 1991, Vol 45, No 4, p 347

The solution to the Monty Hall problem led to a serious academic and public debate on probability. It appeared in journals and even found its way to the front page of the New York Times. The 'correct' answer to this puzzle is of less concern to this study. However, it does unfold a subtle principle that binds probability and information. Let us revert to Bayes theorem. Thomas Bayes' 'An Essay towards solving a Problem in the Doctrine of Chances' remains a piece of work difficult to read and interpret. As Hans Christian von Baeyer unfolds in *Information, The New Language of Science*,

> Bayes' theorem answers the following question: Suppose you know, or assume to know, the probability that a certain conclusion follows from an initial hypothesis. Suppose that a new bit of *information* is obtained, and added to the hypothesis. How do you then compute the updated probability that the conclusion is true, based on the combination of the old hypothesis with the new information?[85]

85 Baeyer, Hans Christian von, *Information: The New Language of Science*, London: Phoenix, Orion Books, 2003, p 76, emphasis in original

Thus, Bayes's postulate does not ask 'how likely is an event?', but rather 'how has the likelihood of an event changed by the information gained?' It is the addition of information to the participant's knowledge in the game show 'Lets Make A Deal' that (apparently) increases his or her chance of winning the game. Bayes' theorem involves computing a *prior probability*, by allowing *a subjective input in a statistical problem.*

The subtle inter-relationship of information and uncertainty is the clue to understanding (and solving) the Monty Hall problem. In these statistical terms the two words information and uncertainty find themselves to be partners. To quote Shannon and Weaver,

> The concept of information developed in this theory at first seems disappointing and bizarre - disappointing because it has nothing to do with meaning, and bizarre because it deals with not a single message but

> rather with the statistical character of a whole ensemble of messages, bizarre also because in these statistical terms the two words *information* and *uncertainty* find themselves to be partners.[86]

Probability and the 'lack of information' are related to another entity which establishes the Second Law of Thermodynamics: Entropy.

86 Shannon, Claude E. and Warren Weaver, *The Mathematical Theory of Communication*, Urbana; London: The University of Illinois Press, 1949, p 116, emphasis in original

2.4 Entropy

As seen in the previous section, the statistical study of language, in this case English, reveals not only the frequency of occurrence or statistical count of a lemma or word (which determines its order or rank in the corpus), but to some extent, the structure and characteristics of the language. The syntax or grammar of a language provides a framework for determining subsequent letters in a word or words in a sentence *ie* the *order* in which symbols occur. Each 'freedom of choice' in choosing a symbol is thus governed by previous choices. In mathematics, this event is known as a Markoff process. Thus, the set of rules affects the freedom of choice one has in selecting from a set of symbols, for example, letters of the English alphabet. This in turn affects the information associated with such an event. *The Mathematical Theory of Communication* by Shannon and Weaver states,

> The quantity which uniquely meets the natural requirements that one sets up for "information" turns out to be exactly that which is known in thermodynamics as *entropy*. That information be measured by entropy after all, natural when we remember that information, in communication theory, is associated with the amount of freedom of choice we have in constructing messages. Thus for a communication source one can say, just as he would say it of a thermodynamic ensemble, 'This situation is highly organised, it is not characterised by a large degree of randomness or of choice - that is to say, the information (or the entropy) is low.'[87]

87 ibid, p 103, emphasis in original

Terms that are of particular importance, extracted from the above paragraph are: order, organisation, and shuffled-ness or randomness. Although entropy is a deeply engaging subject, we refer here to the "bearing of entropy on the concept of order and disorder"[88] in the statistical approach to thermodynamics. In *Entropy and Art*, Rudolph Arnheim describes entropy as "the quantitative measure of the degree of disorder in a system."[89] He further states that "Disorder is not the absence of all order but rather the clash of uncoordinated orders."[90] In a physical sense, order and disorder have to do with the structure of a system - the way its smallest units are arranged. It is this arrangement that constitutes entropy. In the purely statistical approach, "the term order can be used to describe a sequence or arrangement of items unlikely to come about by mere chance."[91]

88 Schrodinger, Erwin, *What Is Life?, with Mind and Matter and Autobiographical sketches*, Cambridge: Cambridge University Press, 1967, p 72

89 Arnheim, Rudolph, *Entropy And Art: An Essay On Disorder And Order*, Berkeley: University of California Press, 1971, p 8

90 ibid, p 13

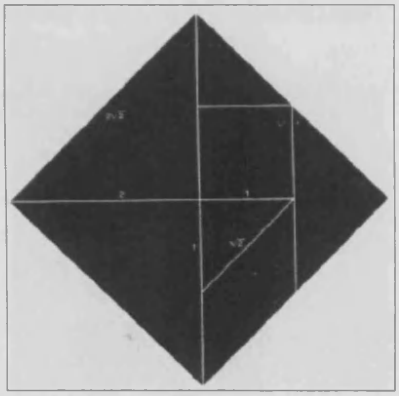
91 ibid, pp 15, 16

For example, let there be two transparent vessels, each filled with a different kind of gas, both at a constant temperature and pressure. One of these two gases is coloured. The vessels are connected by a tube with a valve, which when opened allows the flow of gases in both directions. On opening the valve, both gases intermingle and occupy each other's space. It is highly unlikely that we will now see a mixture of gas in each vessel, half of which is pale (uncoloured) and another half coloured. What is observed is a new pale-coloured mixture occupying both vessels. It is the arrangement of molecules in the new mixture that eventually produces a pale-coloured gas. It is this re-arrangement of molecules in the system that constitutes entropy. The Austrian physicist Ludwig Boltzmann is associated with relating probability and entropy in 1871. In Boltzmann's words,

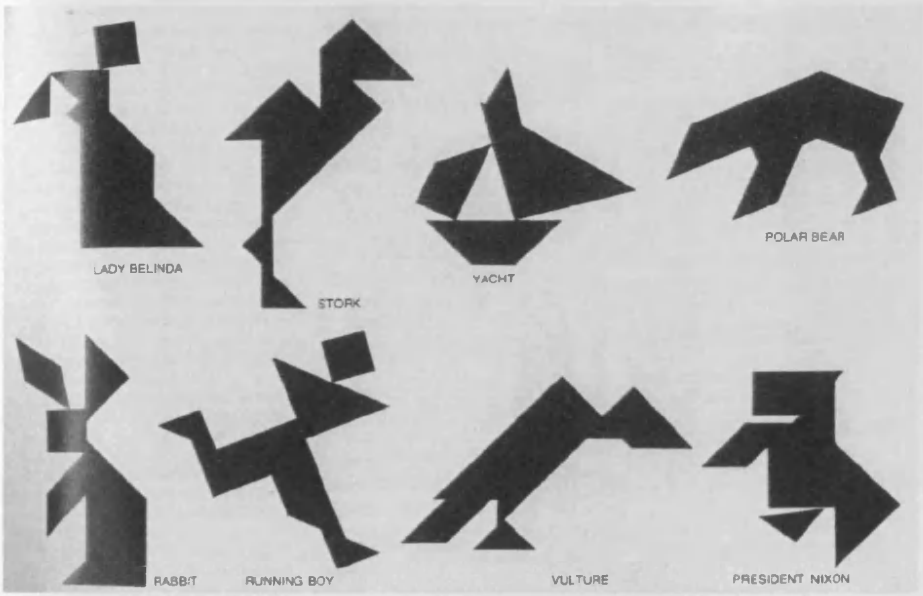
> It is likewise highly improbable that when two different gases are present in a space, only molecules of the one are found in one part of the space and in the remainder only molecules of the other. The mixing of two gases therefore also represents a transformation from an improbable to a more probable state.[92]

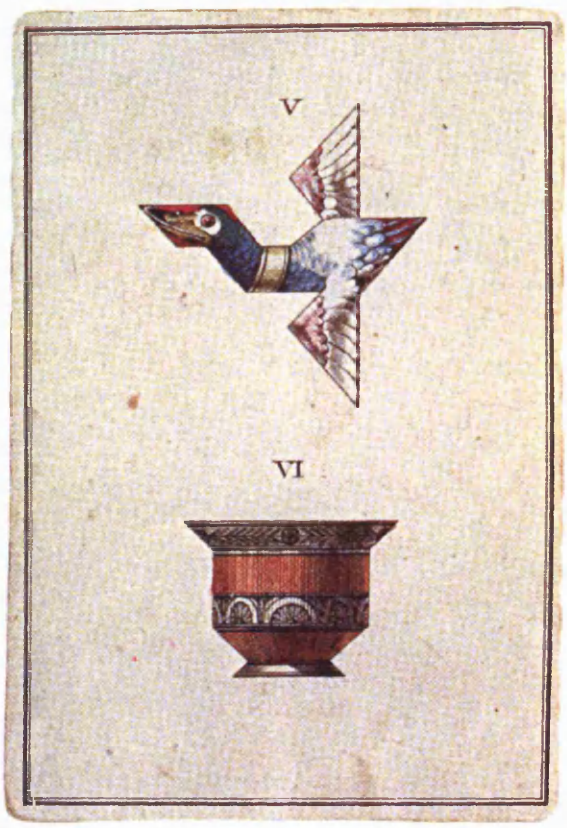
92 Daub, Edward, E., 'Probability and Thermodynamics: The Reduction of the Second Law', *Isis*, Vol 60, No 3, Autumn 1969, pp 318-330

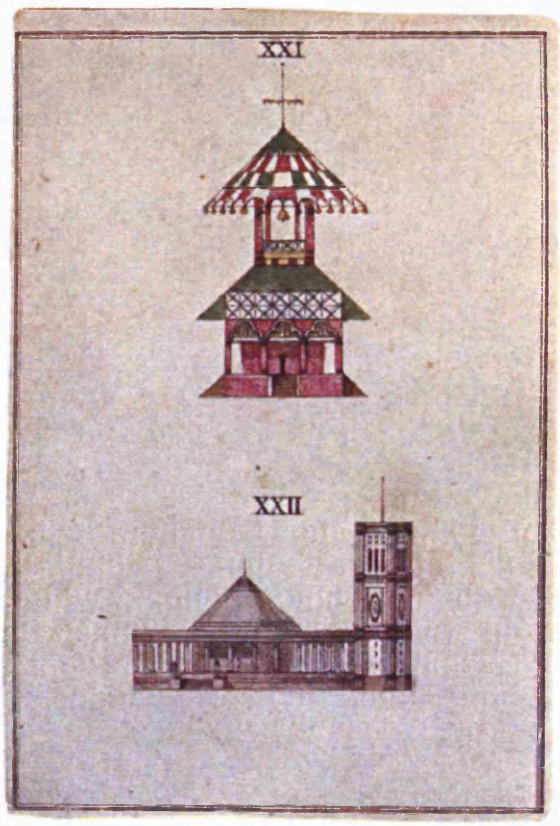
To take another analogical example of arrangement of units of a system, a tangram. This is an ancient Chinese dissection puzzle which involves rearranging seven pieces or 'tans' (right) from a square to form different configurations. As Martin Gardner dissects in the *Scientific American*: "The tans are obtained by slicing a square to produce two large triangles, a middle-size triangle, two small triangles, a square and a rhomboid. Note that all corners are multiples of 45 degrees. If a side of the square tan is taken as unity, a side of any tan has one of four lengths: 1, 2, $\sqrt{2}$ and $2\sqrt{2}$."[93] Tangrams rely on the user's geometric intuition and artistic ability to create artistic and amusing pictures as below.



93 Gardner, Martin, Mathematical Games, 'On the fanciful history and the creative challenges of the puzzle game of tangrams', *Scientific American*, 1974, pp 98 - 103B

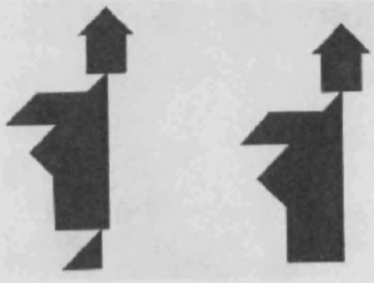






early 19th century tangram cards

Each of the pictures to the right is made from exactly seven tans. However, the 'form' of each tangram is different, and the 'message' they embody or communicate is not the same. Consider the following example to the right where tangrams depicted are *nearly* similar, however, each one is constructed from exactly seven tans. At another level, 'the number of ways of (re)arranging a system', has to do with the relationships amongst parts of a system, and ultimately the meaning they communicate. The Morse Code is another fine example of a very economical use of two symbols - a dot and a dash - to produce all the letters of the English alphabet, which can then be used for telegraphic messaging. An SOS signal would thus be,



H	E	L	P
....	.	.-..	.--.

A .-	N -.	0 -----
B -...	O ---	1 .----
C -.-.	P .--.	2 ..---
D -..	Q --.-	3 ...--
E .	R .-.	4 ....-
F ..-.	S ...	5 .....
G --.	T -	6 -....
H	U ..-	7 --...
I ..	V ...-	8 ---..
J .---	W .--	9 ----.
K -.-	X -..-	Fullstop .-.-.-
L .-..	Y -.--	Comma --..--
M --	Z --..	Query ..--..

2.5 Form

The word 'information' derives from the Latin *informare* (in + formare), "meaning to give form, shape or character to, therefore to be the formative principle of, therefore to imbue with some specific character or quality."[94] Prior to its entry in Science, information meant knowledge. After the 1940s, with developments in information and communication technologies, especially with the work of Shannon and Weiner, information became associated with the terms 'signal' and 'message', and very quickly became a widely used term in scientific literature. However, its amorphous quality transformed it into a device to describe a variety of phenomena across several domains - to mean anything anyone wanted it to mean. It became and remains ambiguous. One question to ask is that 'isn't this obvious?', for just as the physical shape of the amoeba changes in response to external forces, 'something' becomes information to someone, at a specific point in time, within a specific context. It is in associating information with the mind or mental processes, that it is confused with meaning. As Fred I. Dretske explains in *Knowledge and the Flow of Information*,

> To speak of information as *out there*, independent of its actual or potential use by some interpreter, and antedating the historical importance of all intelligent life is bad metaphysics. Once [this] distinction [between information and meaning] is clearly understood, one is free to think about information (though not meaning) as an objective commodity, something whose generation, transmission and reception do not require or in any way presuppose interpretive processes.[95]

But it is precisely this ambiguity, or to be more specific, contradiction, that we see in historical debates on form, in which Plato and Aristotle are key figures. Plato held the view that forms exist independently of their physical manifestation. The idea or *eidos* was the essence, and all physical forms were a likeness of this essence. Plato's pupil Aristotle, held an opposite view, where he rejected the idea that a form can exist independent of its physical manifestation. As he states in *Metaphysics*, "each thing itself and its essence are one and the same."[96] A conundrum in the classical views on form that carries over into the discussion on information is "the belief that 'form' exists to transmit meaning."[97] It is important to note that meaning does not inhere in form. While information processes are form-dependent, all form processes do not necessarily convey information. For example, the information carried by several physical, chemical and biological systems (DNA molecule for example) are embodied in their form. However, possessing a certain shape, pattern, structure or form may not necessarily convey any useful information.

94 Young, Paul, *The Nature of Information*, New York: Praeger, 1987, p 6

95 Dretske, Fred I., *Knowledge and the Flow of Information*, Oxford: Basil Blackwell, 1981, p vii, emphasis in original

96 Forty, Adrian, *Words and Buildings: A Vocabulary of Modern Architecture*, London, Thames & Hudson, 2000, p 151

97 ibid, p 149

The Scottish polymath D'Arcy Wentworth Thompson is widely remembered for his epic work *On Growth and Form*, first published in 1917, which controversially proposed a mathematical framework for form processes in biological systems. In this work, D'Arcy Thompson states, "the form of an object is a 'diagram of forces'."[98] According to Thompson, we can see this mathematical conception of force at work in the living cell of an amoeba or spirogyra, or even in the bee's honeycomb cell, where hexagonal arrangement results in the most efficient packing and maximum storage of honey. According to D'Arcy Thompson,

> The mathematical definition of 'form' has a quality of precision which...is expressed in few words or in still briefer symbols, and these words or symbols are so pregnant with meaning that thought itself is economized.[99]

A parallel view, in terms of visualization, has been expressed by Edward Tufte, when he says that "graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space"[100] - in other terms, an 'economy of perception' results. On a similar note, in the context of communications theory, one has only to look to the telegraph where a single 'dot' is assigned to represent the most frequently occurring vowel 'e' in the English alphabet. Studies in form, whether biological, chemical, physical or mathematical reveal the underlying structure of an organism, or the underlying mechanism of a process, and expose its workings. More fundamentally, form expresses the relationships that constitute the functioning of a system.

98 Thompson, D'Arcy Wentworth, (Ed.) John Tyler Bonner, *On Growth and Form*, Cambridge: Cambridge University Press, 1992, p 11

99 ibid, p 269

100 Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, p51

2.6 Relationships

"Shape, structure, configuration, pattern, arrangement, order, organisation [and] relations"[101] - all these are synonyms used by the writer Paul Young as he tries to pin down the nature of information. The last - relations or relationships - is of particular importance, as it infiltrates or defines all the preceding terms. Relationships are what establish context, and it is context that is so crucial for information to exist - for it to *become*. As Paul Young explains in *The Nature of Information* (in reference to language, as seen in Section 2.1),

> All languages are form dependent. In spoken language, arbitrarily selected symbols are manipulated as units that can be interconnected or arranged only in specific relationships according to specific rules. It is the form (relations), whether semantic, syntactic, experiential, or contextual, of the elements of the language, and not the matter of which they are constructed, from which the mind generates meaning;..[102]

In Section 2.1, we looked at information as 'any difference which makes a difference', and defined a difference as a 'diacritical relationship'. In *Mind*

101 Young, Paul, *The Nature of Information*, New York: Praeger, 1987, p 15

102 ibid, p 43

and Nature, Gregory Bateson reinforces this perspective: "To produce news of difference *ie information*, there must be two entities (real or imagined) such that the difference between them can be immanent in their mutual relationship;..."[103] It is the relationship amongst entities that establishes context. Each entity is, in turn, defined by other entities. A graphical demonstration of this complex web of (inter)relationships, in this case those between human beings that form a community, is given by the Italian novelist Italo Calvino in *Invisible Cities*:

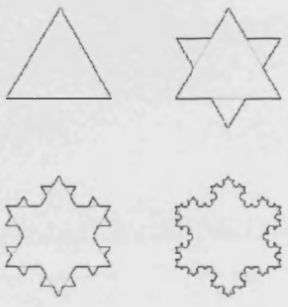
> In Ersilia, to establish the relationships that sustain the city's life, the inhabitants stretch strings from the corners of the houses, white or black or grey or black-and-white according to whether they mark a relationship of blood, of trade, authority, agency. When the strings become so numerous that you can no longer pass among them, the inhabitants leave: the houses are dismantled; only the strings and their supports remain.
>
> From a mountainside, camping with their house-hold goods, Ersilia's refugees look at the labyrinth of taut strings and poles that rise in the plain. That is the city of Ersilia still, and they are nothing.
>
> They rebuild Ersilia elsewhere. They weave a similar pattern of strings which they would like to be more complex and at the same time more regular than the other. Then they abandon it and take themselves and their houses still farther away.
>
> Thus, when travelling in the territory of Ersilia, you come upon the ruins of the abandoned cities, without the walls which do not last, without the bones of the dead which the wind rolls away: *spiderwebs of intricate relationships seeking a form*.[104]

The nomads of Ersilia leave behind an everchanging imprint of their relationships with each other, everytime they move. We may recall D'Arcy Thompson's description of form as a 'diagram of forces'. Relationships are like these forces, where the nature and magnitude of the force changes over time.

What is the nature of these relationships, one may ask, and how do we visualize them? Relationships can be spatial, temporal, logical or causal. Consider the space-time relationship between the Moon and the Earth's oceans. The invisible pull generated by the Moon (and Sun's) gravitational force causes a cyclical rise and fall in the earth's ocean surface, *ie* tides. A simple but powerful spatial relationship exists in the points that form a circle, each point being equidistant from a centre or loci?. The essence of the circle is in the relationships of its parts. Similarly, the Koch snowflake (right) follows a simple rule of division and multiplicity, and forms a shape of infinite perimeter in a finite area - an incredible condensation of information. We have already seen how the roots of the concept of information take us to its close association with the concept of 'form' dating back to classical antiquity. Since form is also about relationships; since relationships are what govern the semantic component of whatever it is we call information; since in the engineering problem of

103 Bateson, G., *Mind and Nature: A necessary unity*, Glasgow: Fontana, 1980, p 78

104 Calvino, Italo, *Invisible Cities*, translated from The Italian by William Weaver, London: Vintage, 1997, p 76, emphasis deliberate



The Koch snowflake

transmission and fidelity over a channel or pipe, these relationships have to be preserved while being transmitted over various media; and since the desired action (if required) on behalf of the receiver, upon receiving the message is an important phase in the process of communication, information can be defined as the *"communication of relationships."*[105]

105 Baeyer, Hans Christian von, *Information: The New Language of Science*, London: Phoenix, Orion Books, 2003, p 25, emphasis in original

By now information must be emerging as a plastic entity, which finds its utility in several domains, from mathematics, to communication theory, to biology and physics. The almost simultaneous publication in 1948 of Norbert Weiner's *Cybernetics*, and Claude E. Shannon's 'The Mathematical Theory of Communication', both measuring 'the amount of information in a message'; and the information-energy-entropy connection arrived at independently though the study of heat, and the equation of entropy as 'missing information' by Ludwig Boltzmann - both of these developments are crucial factors that reinforced the view that information was a commodity that could be generated, transmitted, calculated, stored, channelled and distributed. But therein lies a paradox, for unlike energy, which "as a robust scientific concept we can describe in precise mathematical terms, and as a commodity we can measure, market, regulate and tax,"[106] a similar situation does not exist for information. There is no equation for information. I = ?. Information resides partly in the mind and is very person- and context-specific. Hence, its quantitative and qualitative attributes are much more difficult to assess.

106 ibid, p 11

Understanding the nature of what is being represented, whether numbers, pictures or words, is a deeply engaging and demanding task. While the literature on data graphics or information design discusses design principles and techniques, it does not seem to expose the nature of information which has been discussed in this Chapter. The second part of this thesis describes and evaluates two new tools for envisioning information, whose design aims to put into practise the principles explored in Part 1.

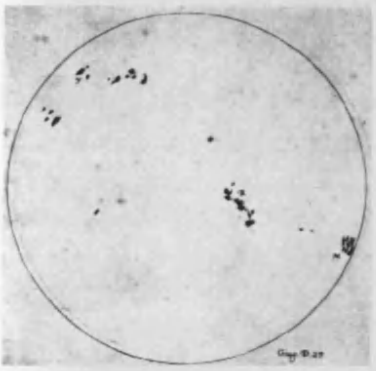
Epilogue to Part 1

Epilogue to Part 1

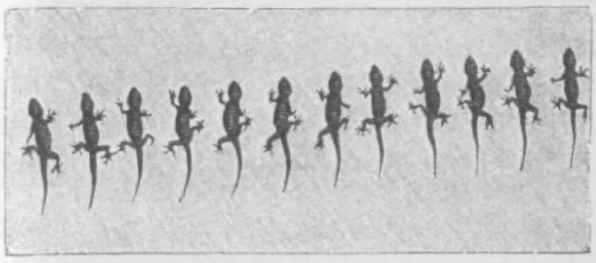
The reader may wish to revisit the data graphics by five pioneers in Chapter 1: E. J. Marey (p 7), William Playfair (p 7), John Wilder Tukey (p 8), Edward R. Tufte (p 25) and Charles Joseph Minard (p 26). These examples, and indeed all the illustrations in Chapter 1, sum up principles of data design that emerge again and again in historic and contemporary examples of data graphics. The primary objective of all these illustrations, as explained before, is 'to reveal data'. All the examples in Chapter 1 are based on what William S. Cleveland termed Graphical Perception - utilising human perceptual and cognitive abilities to comprehend information that is not obvious, ultimately leading to the extraction of knowledge from data. Regardless of the medium on which they are created or presented, *three* simple and broad design principles underpin the success of data graphics. These are outlined below.

1. ***Multiples and Variables***

A multiple is a plural of a singular unit, where a unit is any (variable) entity. It is the variation in features or characteristics of this unit, which generally manifest themselves externally, that differentiates a specific unit from its neighbouring units. In the examples discussed, Chernoff faces (right) are examples of multiples of single unit - a face - with variable features of mouth, nose and eyes. Similarly the multiple dots or marks on the circular disk (left) representing a spherical sun vary in movement (path), size and shape, which, as described by Edward R. Tufte, led to Galileo's support for the Copernican planetary system by deducing that such an observation of sunspots is only possible if a rotating sun is observed from a rotating and orbiting earth.



In the image below, the small units of the gecko are repeated with varying positions of its legs and tail, leading to the perception that it is advancing upwards.



In E. W. Maunder's sunspot diagram of data recorded from 1877 to 1902, small multiples in the form of vertical strokes vary in length and position to give this butterfly diagram.

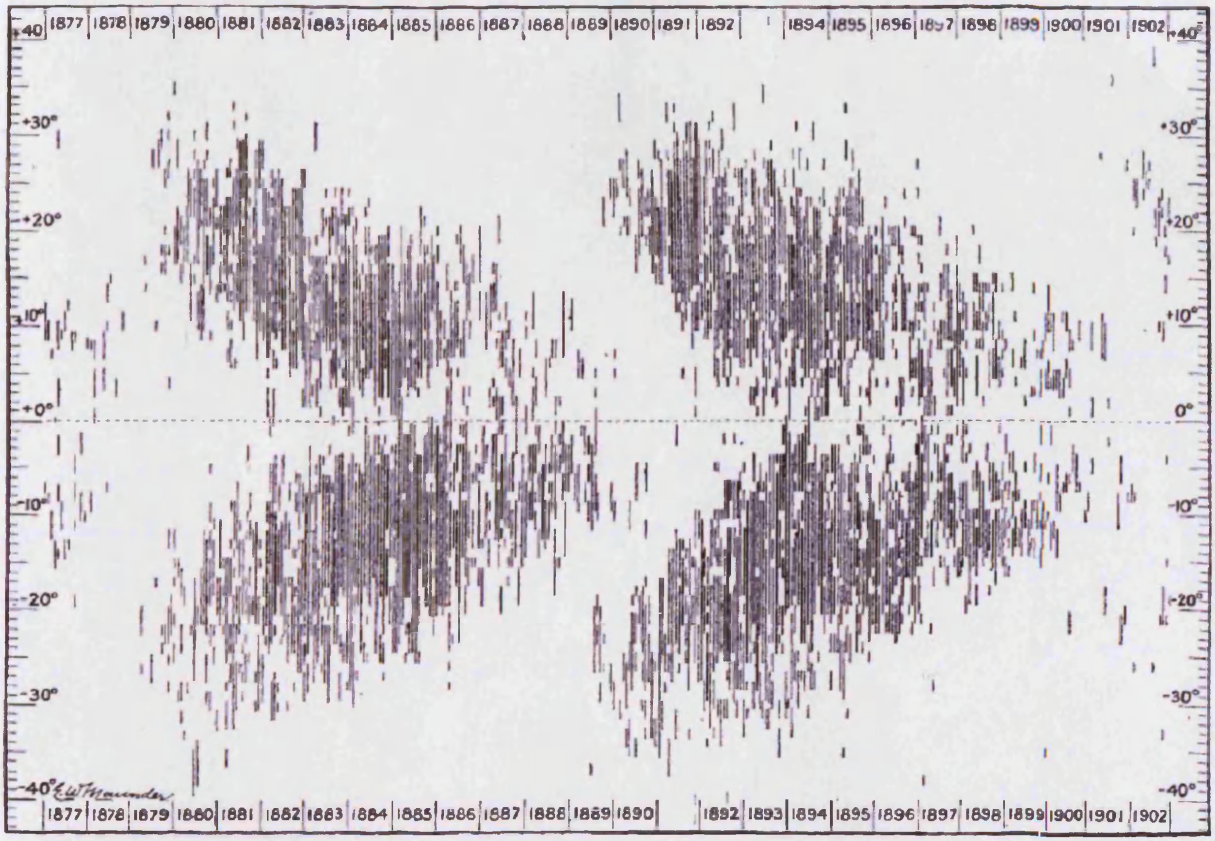
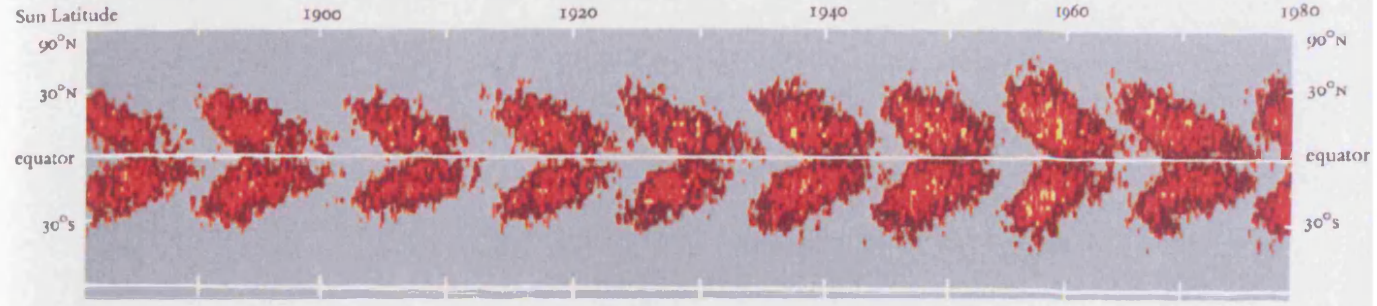


FIG. 8.—DISTRIBUTION OF SPOT-CENTRES IN LATITUDE, ROTATION BY ROTATION, 1877–1902.

A modern contextual comparison of sunspots from 1880 to 1980 gives this macro overview below where the multiples resemble a butterfly wingspan, and the overall reading reveals "between- and within-cycle variation."[107]

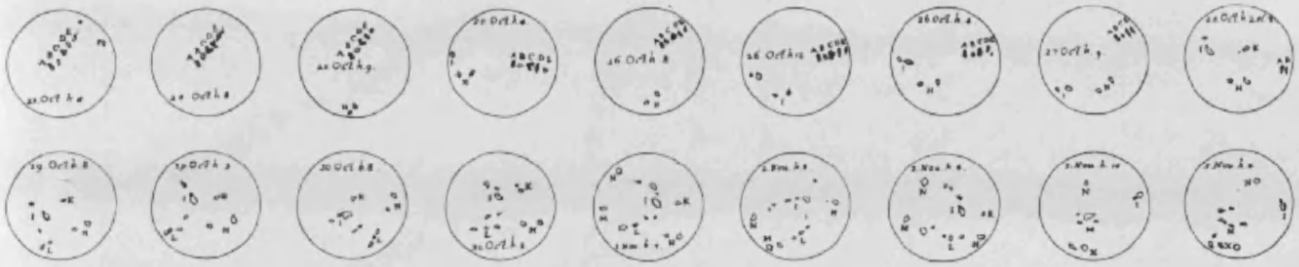
107 Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990, p 23



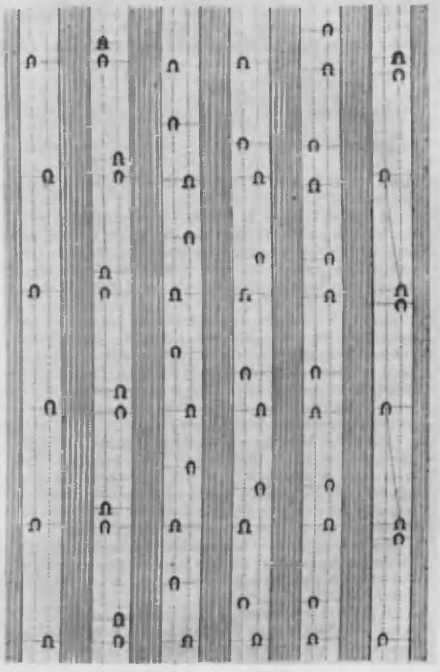
The brilliant micro recording (top) and macro comparison (below) lead to the second important principle in data design: Context and Comparison.

2. *Context and Comparison*

'Always place numbers in their context' - a fundamental rule of statistical analysis. In other words, data, be it in any form, must never be quoted out of context. This principle was seen in Galileo's sunspot diagrams, where the sunspots were depicted on a circular disk representing the sun. These macro multiples were then placed in a linear sequence leading to a comparison of size, shape and position over time.



The gecko in the picture overleaf is perceived to advance only because its varying multiples are placed next to each other. In the following construct (right), E. J. Marey places small multiples ∩Ω∩ of a basic unit that vary in shape. These varying multiples, of studies of the movements of a horse at different paces, are then placed longitudinally next to each other, thus forcing the viewer to make comparisons within an eyespan.

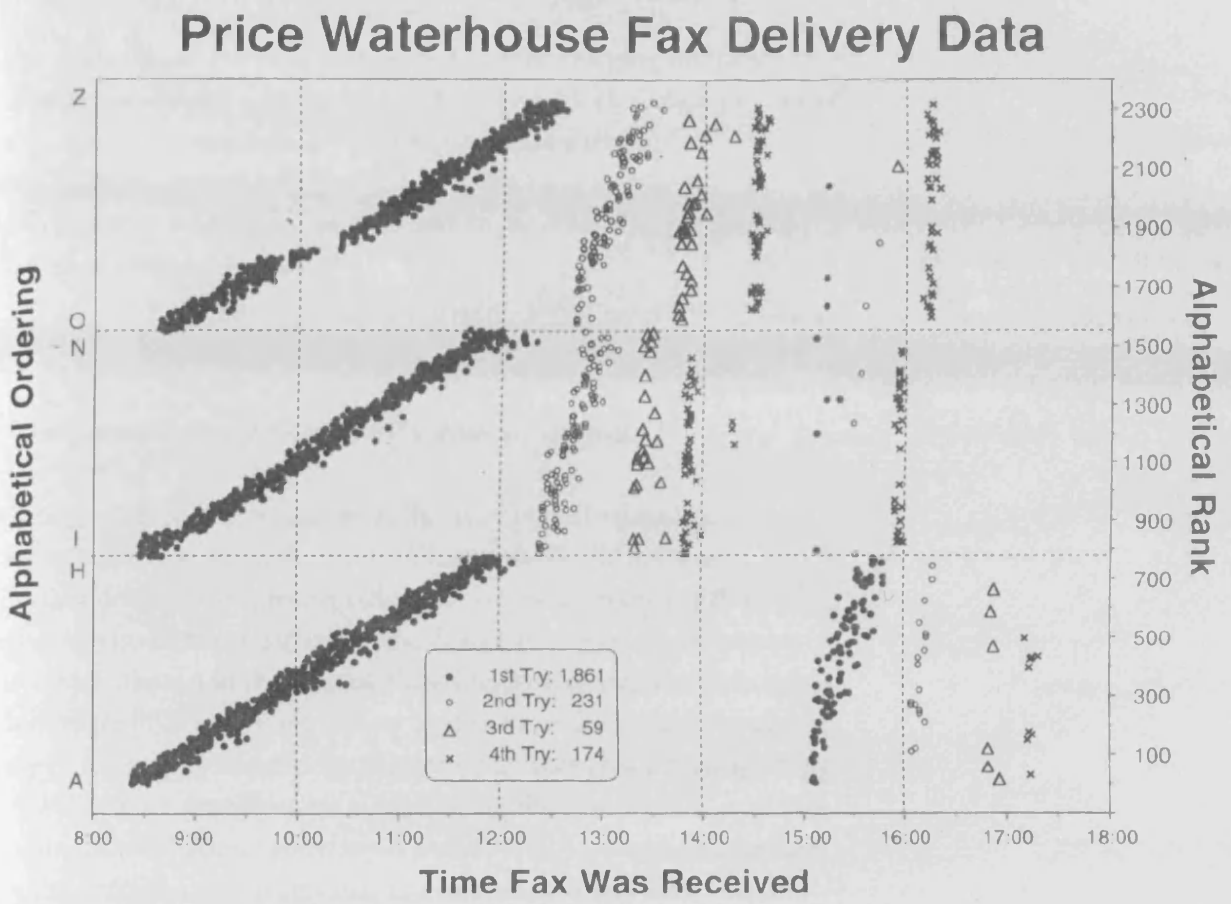


Another fine example using the principle of multiples and variables, as well as placing data in their context and allowing comparisons is a fax data sheet compiled by Alan J. Davis, director of tax communications at Price Waterhouse in Toronto. Every year Canada's federal government publishes its annual budget. Of special concern to legal and accounting firms are tax changes announced in this budget. Price Waterhouse Coopers, an accounting firm, has a tradition of distributing bulletins outlining these changes to their clients. "In 2002, Price Waterhouse engaged a new fax service provider, which was given an electronic list of more than 2, 300 fax recipients, in an order that reflected their time zones. Generally, those farther east were at the top of the list. Faxing began the following morning."[108] A few days later, the fax company provided a report. The data consisted of the following variables: "the client or fax recipient, the time at which the final attempt to fax was made, and whether it was successful or not; a code indicating the number of attempts that had been made (one, two, three, four) and either the word completed or a short description of the reason for failure. The data was not in any immediately obvious order."[109]

108 Wainer, Howard, *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*. Princeton: Princeton University Press, 2005, p 84

109 ibid, p 84

After a number of attempts to comprehend the data provided and gain an overview of the process that had taken place through a series of graphs and charts, Alan J. Davis, the director of tax communications at Price Waterhouse in Toronto prepared the following data graphic below.



This fax delivery data graphic tells the following story:

a) The faxes were sorted alphabetically in three groups: A to H, I to N and O to Z.

b) Faxing began at 08:30am and was nearly completed by 12:00pm, except for group O to Z which finished at 12:30pm.

c) At noon, the operators took a break. Groups I to N and O to Z resumed faxing between 12:30 and 13:00pm.

d) Groups I to N and O to Z had a number of unsuccessful faxes on first attempt, compared to group A to H. They persistently tried to fax the data, and at 16:00pm or just after, terminated faxing.

e) Group A to H had less unsuccessful first attempts compared to the other two groups and the operator also seemed to have forgotten to fax some data through, which was resumed at 15:00pm. Faxing for unsuccessful attempts began just after 16:00pm and finished after 17:00pm.

A final comment by the author of the graph sums up its utility: "The graph proved useful in discussions with the company about the timing of the faxes."[110] This graphic is a very simple example of the "remarkable power that well-designed graphical displays can wield in aiding our understanding of large data sets."[111]

110 Wainer, Howard, *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton: Princeton University Press, 2005, p 84

111 ibid, p 84

In this example, *all* the data is presented as small varying multiples • ◦ △ × so that the viewer can *compare* and *deduce*. Two subtle design principles emerge under the umbrella of Context and Comparison:

2a. A constancy in design allows the user to focus on changes in data rather than changes in design.[112]

112 Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990, p 29

and

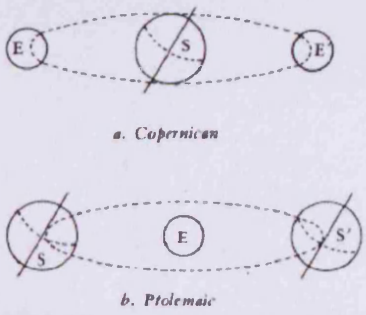
2b. Comparisons should be enforced within an eyespan.[113]

113 ibid, p 33

Both these principles are evident in the examples discussed in this section, and the majority of examples in the Chapter 1. While both remain important design principles regardless of the medium on which the data graphic or visualization is created, the design principle *2a* advocating 'a constancy in design so that the user can focus on changes in data rather than changes in design' is even more pertinent in the digital domain. It is easy to fall prey to technological gimmickry that arises from applying 'digital effects' to visualizations created using the computer. Constantly changing colour backgrounds, rapid transitions or zooming in and out are some examples of features that can be unknowingly built into the application and be a hindrance rather than an aid to cognition.

3. *Cause and Effect*

Analytical thinking requires us to reason about causality *ie* what cause leads to a specific effect. Visual representations of analytical data thus need to show causality. The cause and effect principle is once again seen in action in Galileo's sunspot diagrams. By elegant visual reasoning, he demolished the association of data with official Church doctrine and deduced that sunspots are contiguous to the surface of the sun. Moreover the path of these bodies implied that sun was seen from a rotating and orbiting earth, rather than the other way around (right).



a. Copernican

b. Ptolemaic

Perhaps one of the most disastrous consequences of failing to represent and consequently analyse and think about causal behaviour is evident in the decision to launch the Space Shuttle Challenger. The Challenger was launched on January 28, 1986 and 73 seconds after take off it blew up leading to the death of seven astronauts. The primary reason attributed to this catastrophic event was the failure to analyse and represent convincingly, through visual means, the damage to O-rings[114] in relation to temperature. The lack of pre-launch analysis and effective representation by the rocket maker Thiokol's engineers could have prevented the disaster. Both representation and subsequent analysis could have been achieved through visual means where dry tables should have been represented by graphs or scatterplots to reveal an appropriate correlation that would convince NASA officials that the space shuttle should not be launched.

114 Rubber O-rings are nearly 11.6 metres in diameter and 6.4mm thick, used to seal the joints between two solid-fuel booster rockets.

The following displays from three different sources show the same data in a completely different light, highlighting the importance of data displays. The first is a hand-written chart by engineers at the rocket maker Morton Thiokol. Missing here are the 22 actual previous launches of the Challenger. Only two[115] out of twenty four previous launches are depicted

115 In the chart below, Developmental Motors DM-4 and DM-2, and Qualifying Motors QM-3 and QM-4 are test rockets, *not* actual launches. Only Solid Rocket Motors SRM-15 and SRM-22 are shuttle pre-launch data. SRM-25 is forecast data.

BLOW BY HISTORY
SRM-15 WORST BLOW-BY
- 2 CASE JOINTS (80°), (110°) ARC
- MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY
- 2 CASE JOINTS (30-40°)

SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

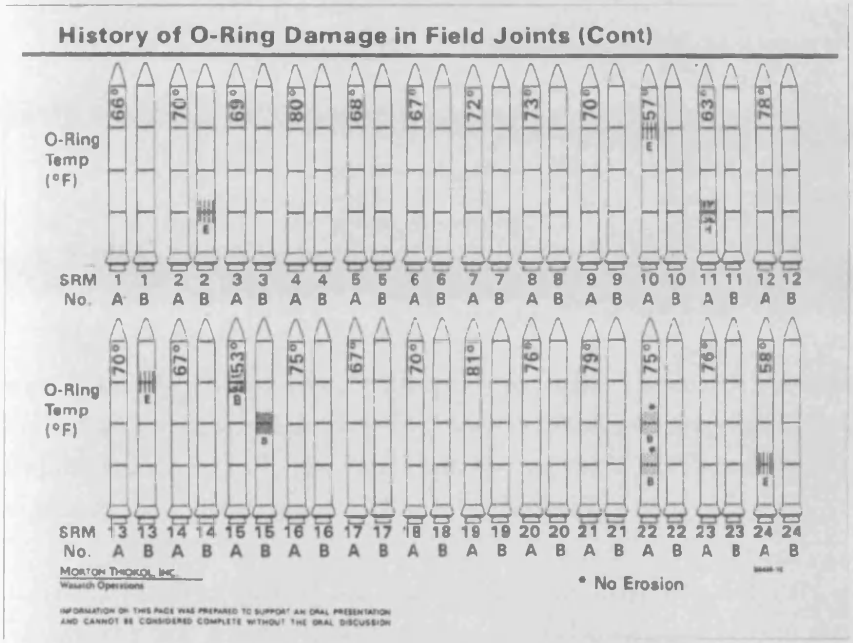
HISTORY OF O-RING TEMPERATURES
(DEGREES - F)

MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

here. The database is incomplete, out of context and hence inaccurate and

unreliable. The chart fails to show and ask the question: is their a correlation between temperature and O-ring damage?

The second chart was prepared for the Presidential Commission investigation after the Challenger accident, and is perhaps the most significant. Unlike the previous graphic, this chart shows *all* the data *ie* all twenty four (24) previous shuttle launches. However, it still



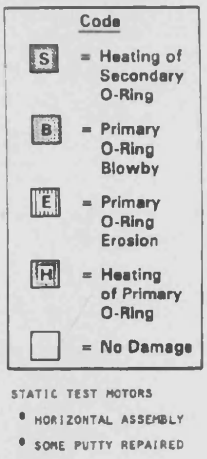
fails to show the relation between temperature and damage to O-rings; it still fails to show cause and effect. Poor data design in this graphic is attributed to:

a) The data is depicted in the form of multiples of small rockets which create visual clutter rather than focus on the data.

b) Rotated text indicating temperature interferes with the outlines of the rocket and is difficult to read.

c) Five different types of cross hatching are used to depict the severity of the damage to the O-rings indicated by the following legend (right). The code further adds to visual noise and hinders the understanding of data.

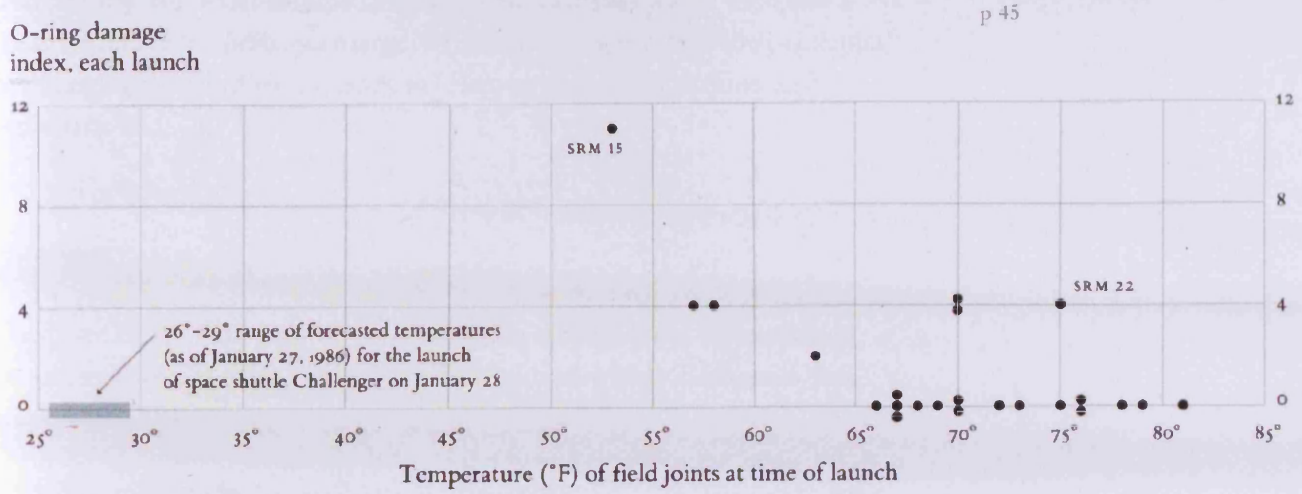
d) The linear arrangement of rockets indicating previous launches is based on the date of their launch - earliest first SRM No. 1 A 1 B to 24 A 24 B - the launch immediately prior to the Challenger. The ordering of data is perhaps the fatal flaw[116] in the data graphic. It fails to ask the question: is their a relation between decreasing temperature and O-ring damage based on all available pre-launch data?



116 Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, p 48

This highly relevant question is displayed and asked in the following data graphic - a scatterplot[117] indicating all twenty four (24) launches prior to the Challenger. Here, the data is arranged not in chronological order of

117 redesigned by Edward R. Tufte in: Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, p 45



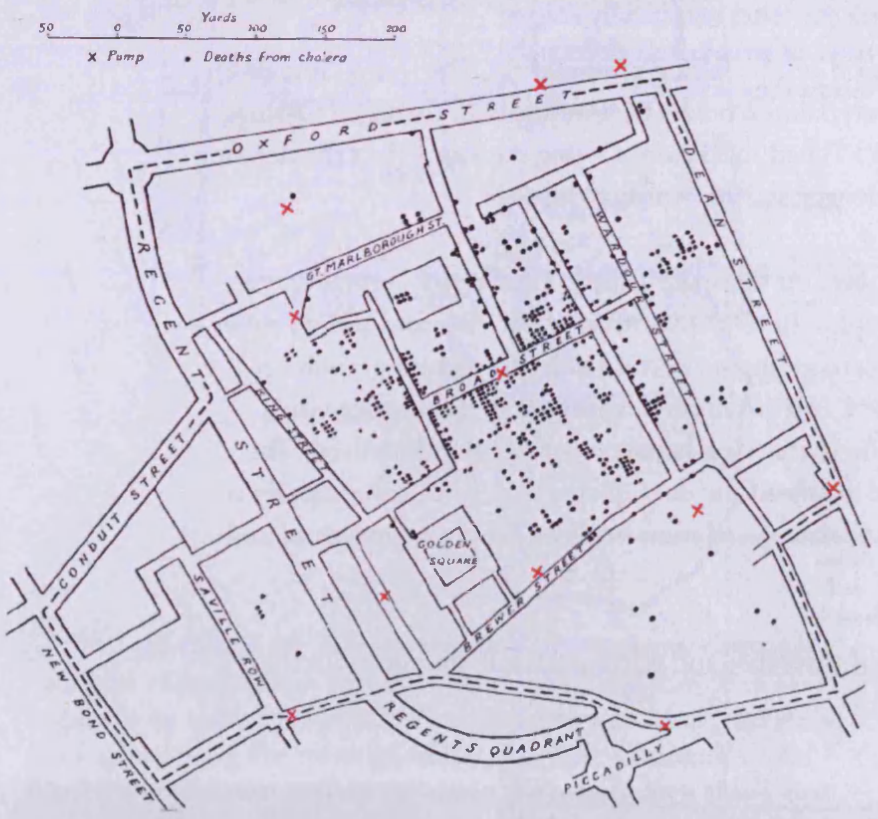
launch date, but based on the falling temperature. The matrix reveals an immediate fact: Every launch below 66°F resulted in damaged O-rings and the launch at 53°F had maximum O-ring damage. Thus the decision to launch the Challenger at 29°F was unthinkable.

The three data graphics on the space shuttle Challenger launch vary as follows: the first graphic (p 47) hides most of the data and fails to show it in context thereby revealing no correlation. The second graphic (p 48) shows all the data and attempts to compare it, but creates visual clutter and reveals no correlation between the variables involved. The third graphic above shows all the data, places it in context, compares it, and reveals an appropriate correlation between falling temperature and damage to O-rings.

The digital variant of the cause and effect principle manifests itself as follows:

3a. It is important to show cause and effect within an appropriate time span

This principle is particularly important in the digital domain where users may interact with the data through an interface. The lack of response from the system within an appropriate time frame can lead to a disinterested individual or audience, System response time with appropriate feedback is an important usability criteria in the design of user interfaces.


50 0 50 100 150 200 Yards
X Pump • Deaths from cholera
OXFORD STREET
REGENT STREET
GT. MARLBOROUGH ST.
BROAD STREET
WARWICK STREET
DEAN STREET
CONDUIT STREET
LITTLE WINDMILL STREET
GOLDEN SQUARE
BRENT'S STREET
SAVILLE ROW
NEW BOND STREET
REGENTS QUADRANT
PICCADILLY

It is clear from these examples how the display of data can provide crucial insights into the understanding of phenomenon inherent in data. To do this, one needs to understand the data, the variables involved and most importantly, the relationships between these variables. Only then can a meaningful representation emerge, which is consistent with the principles of data graphics, and which leads to 'discovery, decision making and explanation'.

Summary

The three broad data graphic principles that emerge from these studies are summarized in this simple, but effective, and widely celebrated data graphic of 1854:

The redesigned graphic on the opposite page shows the 11 water pumps indicated in red crosses, in relation to the deaths during that period.

This 'dot map' was constructed in 1854, by Dr. John Snow, during the worst Cholera outbreak in London in its history. Each dot · represents a single death, while community water pumps are indicated by crosses (x). Unable to establish a direct link by examining the water from a pump[118] situated on the junction with Broad Street and Cambridge Street, Dr. Snow obtained a list of 83 deaths that occurred as a result of Cholera. This tabular representation of data was then represented visually by constructing the dot map above, where *multiples* of dots arranged in two dimensions give the density of distribution which *varies* in relation to the

118 The pump in question is marked by a cross situated to the right of 'D' in Broad Street on the map.

water pumps (Principle 1). The data graphic places the deaths in their *context* spatially and *compares* them (Principle 2) to the proximity of the 11 water pumps located in the area, revealing a strong association between the Broad Street pump and the high number of casualties in the vicinity. A *cause* and *effect* (Principle 3) relationship becomes apparent as most of the deaths occur near the Broad Street pump. Upon investigation, Snow found that the deceased drank water from this pump. In September 1854, acting upon Dr. John Snow's advice, the authorities ordered that the handle of the pump be removed, after which the plague was stayed. Dr. Snow is celebrated for establishing the mode of transmission of cholera - water, and consequently, its cure.

Producing an exemplary data graphic requires of course, an understanding of what it is that is being represented. In many cases, it is numbers or quantities that require representation, and if those adept graphically may not be so numerically, then data distortion is common. Worse, is if one is inept at both, producing the graphically preposterous such as the example below.

A four-fold increase in area of the athlete in this data-thin graphic fails to answer what variables are being compared, in which context and why, for whom and when; prompting a question such as "Comrade, why are we having this meeting?"

By nature, well designed data graphics are high density, multi-dimensional pieces of work that repay careful attention. Human beings rely on their tactile, sensory and spatial abilities to gather and process information from the environment both subconsciously and deliberately. If visualization is a tool for envisioning information, then constructing a visualization is a craft. As seen, the principles behind exemplary pieces of information design are simple and consistent, but often ignored. If living in the information age implies using electronic equipment for information seeking then these established principles of data design cannot be ignored.

Research Questions and Methodology

Research Questions and Methodology

This thesis attempts to answer the following questions:

a) What are the design principles employed in Data Graphics?
and
b) How can they be applied to construct effective information visualization(s)?

The literature review in the thesis focusses on the study of Data Graphics as they are a precursor to modern dynamic computer-based visualizations. Although the medium on which Data Graphics have historically been constructed and presented has changed from paper to the computer, design principles, as seen, have remained the same, regardless of the medium or the source of content. Having uncovered the design principles, two case studies will be explored, where the application of these design principles can be tested. Although the choice of the case studies may seem widely divergent or disparate, they have been chosen as apart from the author's interests in these two areas, they represent data-sets which have different spatial bearings - Air Traffic Control data is inherently three-dimensional and Hypertext does not instrinsically have any spatial mapping. Also, the data-sets are multivariate in nature, and are large and complex to deal with. They also represent challenging real world problems. With the growth in aviation worldwide, visualizing air traffic control data for the efficient management of air traffic is a critical issue for passenger safety, national security and the economy. Similarly, with the exponential growth in digital data, constructing tools to assist users to navigate through the internet or web documents remains a challenging task.

Case study 1 is an investigation into the display of two-dimensional radar images for air traffic control. National Air Traffic Services Ltd. (NATS) - the government body involved in managing United Kingdom airspace will be approached in order to inquire about the current status of radar displays, technologies in use, and future directions in air traffic control displays. A proposal to construct a three-dimensional radar display will be proposed, which, if accepted by the Organisation, will enable operational data to model and simulate controlled airspace to be acquired by the author. It is necessary to prepare a dynamic simulation of the proposed three-dimensional radar display by means of a computer model. The software application 3d Studio Max will be used for this purpose as the author is familiar with the modelling tool. It is envisaged that meetings will be held between the author and ATC planners at NATS where the visualization will be demonstrated, and subsequent refinements will be based on their guidance and feedback. A final evaluation of the

visualization is to be conducted in collaboration with domain experts at NATS.

Case study 2 is a software application that is designed to facilitate the navigation of data in a hypertext document. The author will have to learn a suitable programming language[119] to implement the visualization. Three distinct stages are manifest in this application: design, development and evaluation. The design of the navigation tool is grounded in the theoretical underpinnings of Part 1 of the thesis *ie* design principles in Data Graphics. It may also utilise principles from domains such as software user interface design and human-computer interaction. Development of the software will be through a process of iterative design, where the prototype will be refined through informal user studies at intermediate stages. The final stage involves a formal validation of the navigation tool, where statistically relevant data can be collected and analysed to test the functionality of the proposed application. This will be conducted through user studies described in detail in Chapter 4, Section 4.6.

119 described in 'Prologue to Part 2'

It can be deduced from the previous two paragraphs that the exploratory nature of this thesis combines both quantitative and qualitative methods. Both case studies will go through an intensive, open-ended iterative process, simultaneously involving data collection, data analysis and representation, and data validation described below.

Data Collection: In case study 1, a series of consultative meetings with domain experts in air traffic control form the basis of data collection. The author intends to audio record the interviews with air traffic controller's with permission from the participants. This data will also form part of the data analysis and validation. In case study 2, a questionnaire will be formulated which will be part of the user study conducted on the navigation tool, and will form the primary source of data collection. The actions of the user performing search tasks on computer will be recorded using a screen capture software for later analysis.

Data Analysis: In case study 1, the three-dimensional simulation of the proposed radar interface is to be validated and refined in the meetings held with ATC planners. The interview audio recordings will be transcribed for later analysis. In case study 2, data from the questionnaires and video recordings will be coded and represented as a data matrix.[120] The video recordings of users performing search tasks need to be statistically analysed in terms of parameters essential to assessing the effectiveness of the software application. Qualitative feedback on the user interface will also be coded and displayed in the data matrix. The

120 pages 110 to 113

data matrix will present an overview of the results which is useful for comparisons and to identify patterns, trends and paradoxes. The analysis of videos will provide direct insight into users search behaviour.

Data Validation: Domain experts are the primary source of validation for the proposed three-dimensional radar display. The audio transcription[121] will serve as an audit of the feedback provided at a live demonstration to an ATC planner at NATS. In the second case study, user studies will confirm the search abilities of the tool if the majority of users are able to answer majority of the task questions correctly within an appropriate time frame.

121 Appendix A

The case study method where two practical projects will be implemented is well suited to this thesis as:
a) *Two* case studies are explored as it is essential to test the application of design principles of Data Graphics to visualizations based on *two different types of data-sets* - spatial (ATC data) and non-spatial (hypertext).
b) Both case studies will go through an iterative process where the design, product or process is refined through intermediate stages based on user feedback and expert guidance. This combined method of data generation and data analysis is well suited to an exploratory method with case studies which is grounded in a body of knowledge *ie* Data Graphics.
c) The formative method of evaluation for the radar display will guide its development in a cyclical process whereas the predominantly summative method of evaluation for the navigation tool will test its functionality after the prototype has been prepared.

Part 2: Applications

Prologue to Part 2

The second part of the thesis consists of two case studies. Although they address widely divergent tasks - an air traffic control display and a software designed to help with navigating a hypertext document - the principles underpinning the designs are similar and grounded in the ideas already explored.

Case study 1: Air Traffic Control
An interest in this case study was partly incidental. Given the events of September 2001, the political and economic climate surrounding aviation meant that it was subject to constant media coverage. During this time, the Swanwick Air Traffic Control Centre had also been in the news as it was finally operational, albeit with severe delays and major software breakdowns in its early years. Further investigation into air traffic control prompted an interest in how data is displayed to air traffic controllers in conventional radar displays. It was thought that National Air Traffic Services Ltd (NATS) would be the best source of operational information. On contacting NATS, the author was directed to Alison Lewis, at their Long Term Research and Innovation Unit. The organisation was very cooperative with the author's proposal to construct a three-dimensional visualization of air traffic data. Between March 2004 and September 2004, three meetings were held with authorities at NATS, who provided limited but relevant information on air traffic control and radar displays, which enabled the proposal in Chapter 3. Coincidentally, NATS was just beginning work on a similar proposal entitled the 'air tunnel approach' and was keen to see results from an alternative viewpoint.

Case study 2: A Navigation Tool for Information Seeking
This case study is the result of an interest in the way we search for information online, the subsequent representation of search results or data, and how we interact with the data during the search process. It thus looks at the process of search, display and interaction in a navigation tool designed to seek information within a specific data set. The software application required programming or scripting a prototype to develop and test ideas as they evolved. Initial 'field work' in the programming languages C++ and OpenGL proved too demanding a task. However, an encounter with ActionScript, the native scripting language of Macromedia Flash MX, was found to be appropriate both in terms of its capability to produce the intended application, and, more crucially, the author's ability to understand and develop the application using this programming language. Three distinct phases are evident in this case study: design (applying theory from Part 1), development (implementing the ideas by scripting using ActionScript) and finally, evaluating the software (by conducting user studies). Although the case study is cross disciplinary across several domains, a focus on subject matter of this thesis has been retained.

Both case studies discussed in this thesis are examples of Information Visualization. The visualizations differ in the type of data they represent: air traffic control data is inherently spatial, whereas the web pages on environmental design, which form the content of the navigation tool have no spatial bearing - they are abstract.

3.0 Visualizing SID's

3.0 Visualizing SID's

Since the Wright brothers undertook their historic flight in 1903, aviation has become "a gigantic enterprise undertaken, above all, in the name of economy and commerce."[122] Once so simple to comprehend, airways are now "a complex, crafted network invisible to the spectator on the ground."[123] The successful navigation of these mechanical birds depends on a visualization system that translates geographic co-ordinates into Cartesian co-ordinates. It is this exclusive zone - its management using information visualization - that is the subject of this case study.

122 Pascoe, D., *Airspaces*, London: Reaktion, 2001, p 27

123 ibid, p 9

This case study has been made possible through collaboration with National Air Traffic Services (NATS) Ltd., which is the official body responsible for managing United Kingdom airspace.

3.1 Air Traffic Control (ATC)

3.1.1 Historical Background

During the beginnings of commercial aviation, air crew largely relied on direct observation, their own deductions and simple navigational instruments. With growth in air traffic, legislation to regulate airspace was introduced after the First World War. In 1919, the International Commission for Air Navigation (ICAN) was formed, after which the exchange of weather reports between all major airports began. In 1922, flights between London and Paris were instructed to remain west and east of specific landmarks when flying inbound and outbound of London respectively. This was a precursor of the airways system. A network of radio beacons was created, which transmitted signals that could be sensed by aircraft flying along that route, leading to the creation of airways. In 1933, specific procedures to maintain safe separation between aircraft were adopted. The airways system enabled safe lateral, longitudal and time separation between aircraft. Aircraft flying in the same direction or in opposite directions could be on separate airways, or even on the same airway or air corridor, as long as the safe separation between aircraft involving the three spatial dimensions and time were ensured.

The Second World War led to technological advances in the identification of aircraft, navigation and target finding. In 1935, Robert Watson Watt submitted a memorandum 'The Detection and Location of Aircraft by Radio Methods' to the British Government, explaining how radio waves could be used to detect aircraft, subsequently marking the birth of radar.

In 1938, the division of airspace into two categories, aerodrome and area, was proposed. In 1939, 'Air Traffic Control' (ATC) was adopted as an official term by the Air Ministry in Great Britain. The Chicago convention in 1944 formed the basis for present day ATC throughout the world. In 1951, the UK adopted the airways system. Five new airways were created: Amber One (Daventry-Dunsfold-Dieppe-Paris), Amber Two (Daventry-Brookmans Park-Maidstone-Paris), Blue One (Worburn-Watford-Crowborough), Red One (Dunsfold-Maidstone-North Foreland-Amsterdam), and Red Two (Woodley-Epsom-Ashford). In 1962, National Air Traffic Control Services (later known as National Air Traffic Services Ltd., or NATS) was formed. "The safe, orderly and expeditious flow of air traffic became established as the objective of air traffic control."[124]

124 Hopkin, V.D., *Human Factors in Air Traffic Control*, London: Taylor & Francis, 1995, p 4

3.1.2 "The Unalloyed Blue"[125]

125 Pascoe, D., *Airspaces*, London: Reaktion, 2001, p 43

This section describes the conceptual organisation of United Kingdom airspace.

UK airspace is divided into zones, with each zone being further divided into sectors. A sector is a region of airspace defined by geographical and height boundaries which is controlled by an Air Traffic Controller (ATCO), or a team of ATCO's.

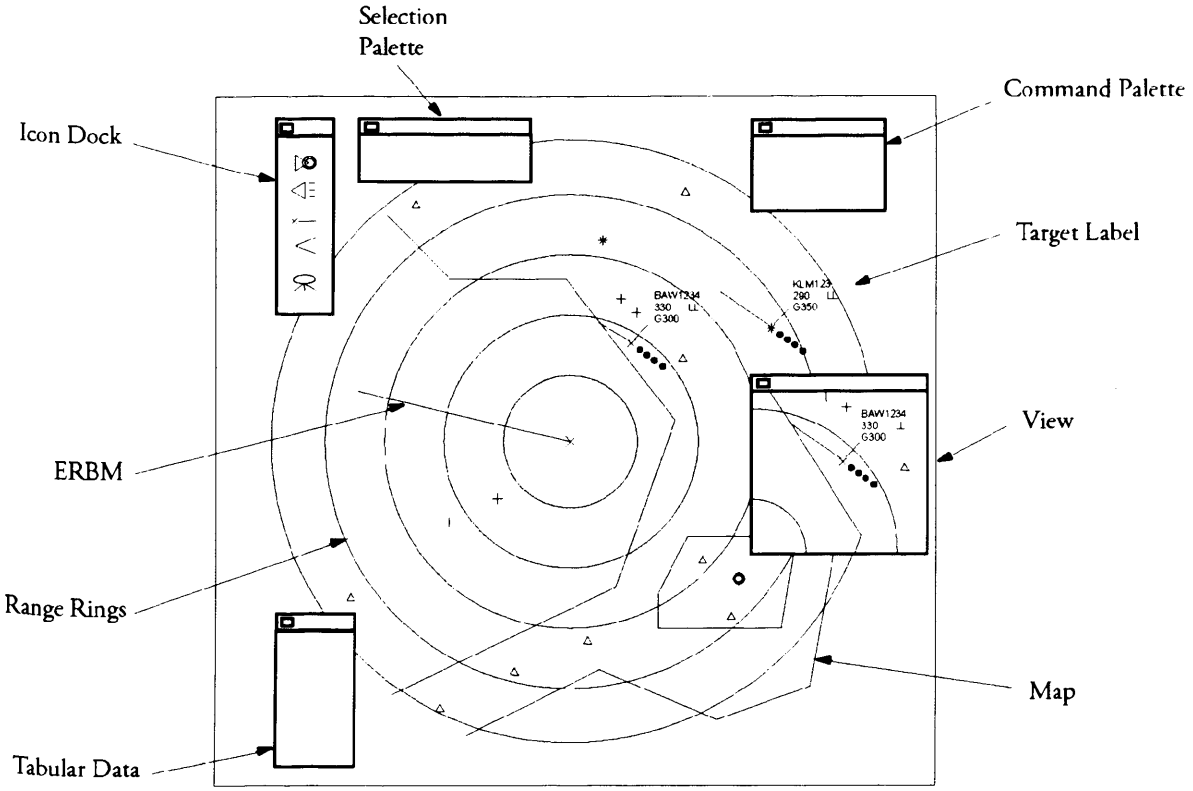
A series of airways (analogous to motorways) exist in the skies, which guide the movement of aeroplanes departing or approaching an airport. The safe lateral, longitudal and time separation between aircraft are ensured at all times. "The actual magnitudes of the separations required for safety are not universal, but depend on the quality of the navigational information."[126]

126 Hopkin, V.D., *Human Factors in Air Traffic Control*, London: Taylor & Francis, 1995, p 4

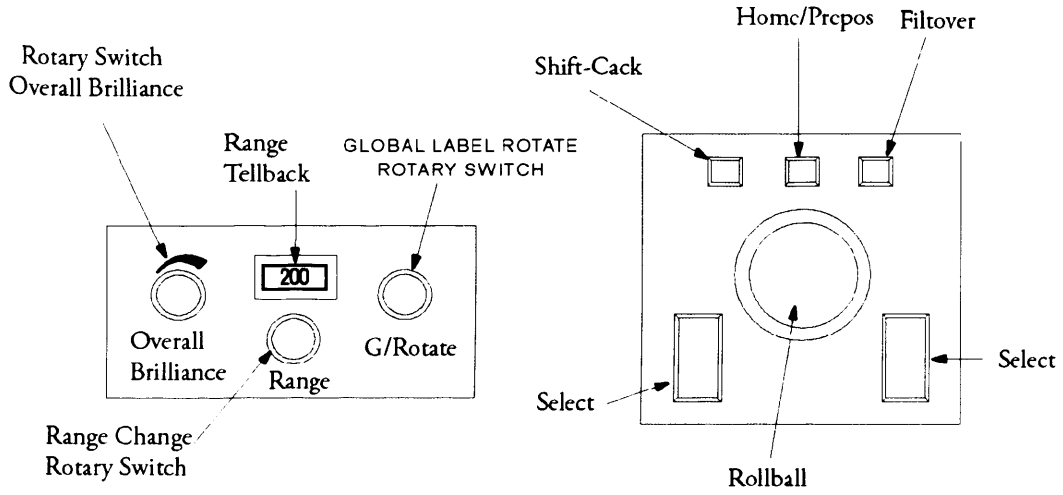
Departure routes are known as Standard Instrument Departures (SID's), and Arrival Routes are known as STandard Arrival Routes (STAR's).

An aircraft flying through a specific airspace has to follow the ATC guidelines of that region or country, and is obliged to carry specified navigation and communications equipment.

Each airport resides within a Terminal Radar Approach and Control Region (TRACON). A Terminal Control Centre on the ground tracks and controls the position of all aircraft entering or leaving the airspace it is responsible for. Air Traffic Controllers track the position of aircraft and are responsible for communication with the pilot.



Representative Display
(Not to Scale)



Rotary Switch Panel

Rolling Ball Assembly

Figure 3.2.2 Radar Controller Interface

A virtual map representing 'way points' or altitude constraints exists (on ground) which informs the pilot whether the aeroplane should be *at, above* or *below* the height represented by the altitude constraint.

3.2 Radar

3.2.1 Definition

Radar or RAdio Detecting And Ranging, is "a method of detecting distant objects and determining their position, velocity or other characteristics by analysis of very high frequency radio waves reflected from their surfaces. The equipment used in such detection is known as RADAR."[127] Radar enables the air traffic controller to visualize the relative positions of all objects in airspace.

127 www.answers.com, 'Radar', *www.answers.com/radar*, Accessed <April 15, 2004>

3.2.2 The Radar Interface

The radar interface is a Graphical User Interface (GUI) system. As shown in figure 3.2.2 (left), it comprises of three main elements:
1) A Radar Display: a high resolution, colour, raster display.
2) A Rollball Assembly: a rolling ball with additional keys.
3) A Rotary Switch Panel Assembly: three rotary switches, for label rotation, brilliance and range setting, together with a numeric range tellback.

The rollball assembly and the rotary switch panel assembly constitute what the air traffic controllers refer to as 'Off Screen Man Machine Interface'. Together, they provide the physical method of access and control - pointing and selecting. Figure 3.2.2 is a graphical representation of the radar interface.

3.2.3 The Radar Image

The radar image is a crucial element in the radar interface. It is a visual representation in *real-time* of the data accumulated by the radar device, which enables an air traffic controller to manage traffic on the ground and in the air.

A series of dots, squares or asterisks represent the geographic position of the aircraft. Each reference or blip is accompanied by a tag. This tag carries two vital pieces of information: the aircraft identification number

(similar to a flight number) and the altitude of the aircraft.
Concentric circles or range rings on screen define the area of view *ie* the radius of airspace corresponding to its coverage on ground that the air traffic controller is looking at. The range setting button within the rotary switch panel assembly allows the air traffic controller to zoom in and zoom out, thus enabling him or her to concentrate on a specific area or sector of responsibility. Also, in the case of the United Kingdom, a coastal outline gives the air traffic controller an indication of the aircraft's position with respect to the land mass.

A series of menus and palettes can be hidden or invoked using the rollball assembly allow further interaction with data on the radar screen.
As seen in the image below, a trail of circles, gradually diminishing in size, forms a tail behind the asterisk representing an aircraft. The subtle deviation of this tail from linearity gives the controller an indication of its intended direction.

A representation of a radar image as seen on a radar display by an air traffic controller.

3.3 Related Work

A number of studies on air traffic displays have been conducted. Studies in this area have focussed on: the cognitive demands of air traffic control (Wickens *et al,* 1998; Morineau *et al*, 2003; Wickens *et al*, 2005), enthnographic studies on air traffic controllers (Bentley *et al*, 1992; Mackay *et al,* 1998) and the role of flight strips in air traffic control (Mackay 1999), amongst others. In the context of this case study, studies on the design of three-dimensional displays are of primary concern. A number of formal studies in this direction have been conducted by Wickens *et al*, most significantly: 'Two- and Three-Dimensional Displays for Aviation: A Theoretical and Empirical Comparison' (Wickens and Haskell, 1993); 'Terrain Representation for Air Traffic Control: A Comparison of Perspective with Plan View Displays' (Wickens and May, 1994); 'Weather Displays for Air Traffic Control: The Effect of 3D Perspective' (Wickens *et al*, 1995); 'Cognitive Factors in Aviation Display Design' (Wickens *et al*, 1998).

3.3.1 Advantages and disadvantages of three-dimensional displays

advantages:
Wickens *et al* reveal 'mixed evidence' regarding the utility of perspective displays versus planar displays. An equivalent performance is found in basic component tasks, whereas the planar display supports more complex tasks. The planar display supports more 'accurate' navigation, whereas perspective displays support more 'efficient' navigation. Wickens *et al* report that the spatial component representing altitude is a modest benefit for the perspective display, whereas Brown and Slater (1997) state that this more 'natural' display of data leads to a potential reduction of mental workload. Formal studies by two former controllers' (Strutt, 1991 and Burnett, 1991), suggest the development or formation of a mental model based on the alphanumeric information present in plan view, and suggest that it is the spatial component in perspective display that can potentially reduce the air traffic controller's mental workload. Wickens *et al* support the use of the perspective display as a good training tool to understand three-dimensional characteristics of airspace, therefore facilitating the development of this mental model.

disadvantages:
Brown and Slater (1997) state "the difficulty in interpreting aircraft position in perspective displays due to inadequate depth cues"[128] as a possible reason why 3d displays have not yet been used in operational ATC displays to date. A key concern in air traffic control is the

128 Brown, Mark A. and Mel Slater, Some Experiences with Three-Dimensional Display Design: An Air Traffic Control Visualization, *IEEE International Workshop on Robot and Human Communication*, 1997, pp 296 - 301

judgement of distances and angles. Inadequate depth cues can lead to objects much further in the distance being displayed close to each other, resulting in clutter. Occlusion or the masking of background items by foreground items is also more evident in perspective displays. Wickens *et al* suggest that the difficulty or "negative transfer experienced in using the perspective format due to difference in spatial meaning"[129], which is related to the controller's current mental model, as a drawback for 3d displays. However, all studies comparing 2d and 3d displays show that air traffic controllers are biased towards the planar display, while pilots remain biased towards the perspective display. Evidence continues to suggest the viability, if not superiority, of perspective displays. Published literature does not appear to contain examples of other forms of three-dimensional projections, for instance, isometric or axonometric projections.

129 Wickens, Christopher D. and Patricia May, 'Terrain Representation for Air Traffic Control: A Comparison of Perspective with Plan View Displays', Prepared for Federal Aviation Administration, Systems Operations & Engineering Branch, Washington DC, September 1994

3.4 The Issue

Studies on air traffic control that are generic (giving an overview of the field) and specific (specialist studies, on colour or the effect of perspective, for example) are common. There is little or no evidence, however, of studies on ATC display design from a data graphics viewpoint. This section discusses the issues concerning two-dimensional radar displays from this perspective, and proposes a graphic display consistent with the principles of information design.

(a) *data integrity: the number of data variables is not equal to the number of design variables*

The radar image is two-dimensional. Air Traffic Controllers construct a three-dimensional image in their mind by looking at (a) the proximity of blips on screen, (b) their altitudes and (c) their velocity. In terms of visualization, it lacks the interactivity that a three-dimensional representation can offer. The ability to look at the interaction of aeroplanes with each other in the air, from any direction or vantage point, seems natural for a body of data that is inherently spatial. Examples abound of graphics that distort the data by exaggerating the number of real data dimensions compared to those depicted (see image 3.4a to the right); but the radar display is a *rare* example where "the number of information-carrying (variable) dimensions depicted [does] not exceed the number of dimensions in the data."[130] "Graphical [integrity] begins with telling the truth about the data."[131]

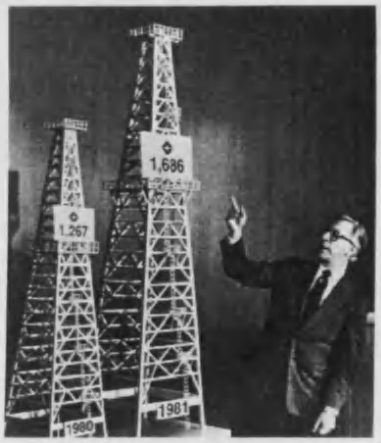


image 3.4a: a perfect example of puffing up one-dimensional data, in this case quantities, into three.

130 Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983, p 77

131 ibid, p 53

(b) *data overload: cognitive demands and the impact of visualization*
Air traffic controllers are subject to intense cognitive demands. According to the International Labour Organisation (ILO),

> Surveys show that the main sources of stress reported by air traffic controllers are related both to the operative aspects of their job and to organizational structures. In the former case, the most important factors are peaks of traffic load, time pressure, resolving conflicts in the application of rules, and the limitations and reliability of equipment. The factors relating to organizational structure mainly concern shift schedules (and particularly night work), role conflicts, unfavourable working conditions and the lack of control over work.
>
> Analysis has emphasized the complexity of the work of air traffic controllers. For example, the cognitive/sensory capacities required for high performance at radar workstations include spatial scanning, movement detection, image and pattern recognition, prioritizing, visual and verbal filtering, coding and decoding, inductive and deductive reasoning, short- and long-term memory, and mathematical and probabalistic reasoning. Air traffic controllers are also among the groups of workers who are most exposed to critical accidents which cause unusually strong emotional reactions, such as air accidents with loss of life or serious injury, near collisions and loss of control due to overload.[132]

The importance of 'graphical perception' has already been seen in Chapter 1, where pre-attentive processing was discussed. A visualization is said "to be more effective if it is faster to interpret, can convey more distinctions, or leads to fewer errors than some other mapping."[133] According to Edward Tufte, excellence in visual design is achieved through graphics that correspond with the mental tasks they are meant to support. "If the thinking task is to understand causality, the task calls for a design principle: 'Show causality.' If a thinking task is to answer a question and compare it with alternatives, the design principle is 'Show comparisons.'"[134] A simple example is the design and display of symbols that exemplify what they represent, for example, a commercial aircraft, a fighter jet and a helicopter, in place of an asterisk or a dot (which relies on the accompanying data tag) to determine the type of aircraft. The importance of cognitive-design principles in enhancing the design and display of data cannot be understated. A good visualization should reduce cognitive overload, not be the equivalent of 'chartjunk.'[135]

(c) *data density: Continuous growth in the volume of data*[136]
In light of paragraph (b), two recent events are worth noting. In November 2004, the Guardian reported that "a moment of forgetfulness by an air traffic controller caused a near disaster at Manchester airport when a MyTravel holiday jet had to abort its takeoff at 100mph to avoid another aircraft. The Air Accident Investigation Board (AAIB) said the incident was down to human error."[137] On June 3, 2004, BBC News reported that "thousands of passengers are facing delays after an air

132 International Labour Organization, 'SafeWork, Stress Prevention In Air Traffic Control', *http://www.ilo.org/public/english/protection/safework/stress/trafctr.htm*, Accessed <November 25, 2004>

133 Schneiderman B., S.K. Card and J.D. Mackinlay, *Readings in Information Visualization: Using Vision To Think*, San Francisco: Morgan Kaufmann, 1999, p 23

134 Zachry, Mark and Charlotte Thralls, 'Cross-Disciplinary Exchanges: An Interview with Edward R. Tufte', *Technical Communications Quarterly*, Vol 13, No 4, 2004, pp 447 - 462

135 'Chartjunk' is a term coined by Edward R. Tufte to describe the false ornamentation of data displays.

136 According to the Department for Transport, the number of air passengers in 2003 was nearly 200 million. This figure is set to rise between 400 and 600 million by 2030.

137 http://www.guardian.co.uk/uk_news/story/0,,1348184,00.html, Accessed <November 22, 2004>

Image 3.4c: A computer failure at the Swanwick Air Traffic Control Centre on June 03, 2004 suspended air traffic control services for a few hours resulting in 'a traffic jam in the skies', the density of which is captured in the image above.

traffic control computer failure caused flights to be suspended across the UK."[138] Image 3.4c (left) shows the dense radar screen at the Swanwick Centre on that day. Carefully designed and displayed multivariate data sets of high density can be very effective in using the eye's ability to read substantive quantities of information in limited space. Increasing the size of radar displays is not an easy solution, as the optimum size[139] is greatly influenced by viewing distances and angles, and the corresponding legibility of information on screen. Here, the principle of selective omission of information applies, where the ATCO can select what data to view for a particular task and timescale. Although the addition of a third dimension for viewing is to be treated with caution, in terms of data density, this escape from flatland does provide much needed extra depth for data display. Also, the shift from tactical air traffic control (resolving problems that have arisen) to strategic air traffic control (preventing problems from arising), where air traffic control is likely to be organised into flows, rather than dealing with flights on a singular basis, is the potential biggest advantage of a 3d display.[140]

138 http://news.bbc.co.uk/1/hi/uk/3772022.stm, Accessed <June 03, 2004>

139 The optimum size of a radar display with a viewing distance of 500mm is about 350mm.

140 Hopkin, V.D., *Human Factors in Air Traffic Control*, London: Taylor & Francis, 1995, p 5

(d) *Lack of appropriate visual structure*

One of the characteristics of air traffic displays is that they are not self-evident *ie* they contain no information about themselves and make no sense to the non-expert. The information displayed has to be learned and this knowledge has to be applied in different scenarios. Most of the information contained in the displays is quantitative in nature (speed, height, acceleration, etc). This quantitative information has to be decoded to form an appropriate mental model of airspace.[141] Ultimately, this mental model and the visual representation must match. Current radar displays are not instructionally supportive and an appropriate or better visual display can enhance the formation or development of the air traffic controller's mental model and also be a valuable learning tool.

141 See 'Related Work', Section 3.3.1

3.5 Visualizing Standard Instrument Departures (SID's)

The sections that follow describe the transformation of the two-dimensional radar image into a three-dimensional radar display. The representation is also known as the 'air tunnel approach'.

3.5.1 Parameters / Constraints

The sector of airspace to be visualized is above Heathrow airport. The radius of airspace under consideration is 25 nautical miles (nm), where 1nm = 1.852km.

Air traffic regulations are a set of rules that define the airspace and the way that 'objects' are supposed to navigate through this space. This set of rules is binding for the users' of the airspace (aircraft), the air traffic controllers, and designers of a visualization system.

An aircraft has a specific route map for its journey. Within controlled airspace, it follows a strict flight path. ATC guidelines in the UK require a minimum lateral separation between two aircraft of 3nm, on both sides, and a minimum 1000 feet vertical separation, both up and down.

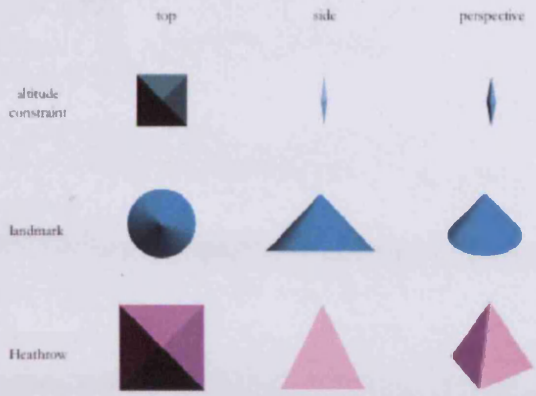
It is these constraints that are the very basis for this particular visualization. The representation that emerges has the following basic components: tunnels, symbols, text and colour.

3.5.2 Tunnels

As mentioned earlier, the trail of circles on the radar image gives the air traffic controller an indication of the direction of the aircraft. A more accurate representation is to plot the flightpath of an aircraft. The minimum lateral and vertical separation distances imply, in theory, that an aircraft travels in an insulated volume of airspace. Translating these distances into a shape, an elliptical form evolves. Extruding this elliptical shape along a flight path, a loft or tunnel is formed, which represents the three-dimensional elliptical volume of airspace through which the aircraft is travelling.

3.5.3 Symbols

The table to the right lists the various symbols used to represent data in the proposed three-dimensional model. The term symbol(s) is used here to "represent data through visual properties such as colour, shape, size and orientation."[142] (Wittenbrink *et al*, 1996) A symbol is thus "an object with parametric geometry and appearance that can be linked to data quantities."[143] (Walsum, *et al*, 1996) The terms icon and glyph have also been used in visualization. However, the term symbol is used here, as both an icon and a glyph are symbols of some kind, whose purpose is to signify. Semiotic aspects of these symbols are discussed in Section 3.5.5. These geometric objects have attributes which can be classified into three groups[144]:

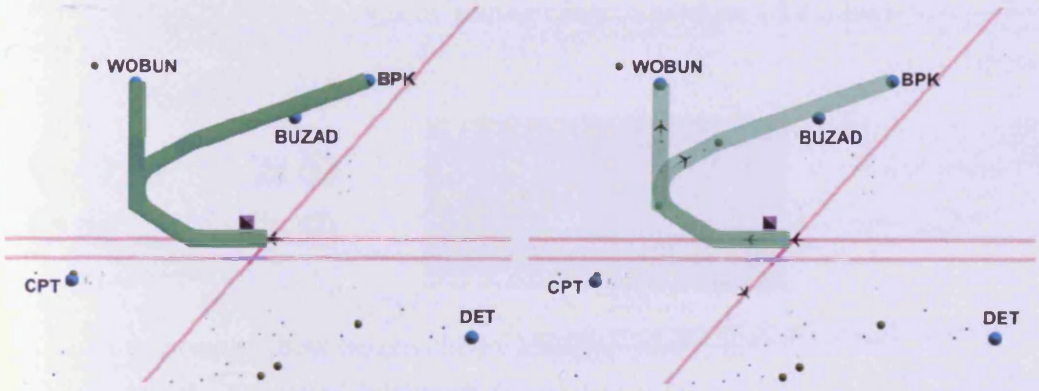


142 Wittenbrink, C.M., A.T. Pang and S.K. Lodha, 'Glyphs for Visualizing Uncertainity in Vector Fields', *IEEE Transactions on Visualization and Computer Graphics*, Vol 2, No 3, September 1996, pp 266-279

143 Walsum, Theo van *et al*, 'Feature Extraction and Iconic Visualization', *IEEE Transactions on Visualization and Computer Graphics*, Vol 2, No 2, June 1996, pp 111-119

144 ibid, pp 111 - 119

a) *spatial parameters* such as position and orientation: The position of static objects (landmarks, altitude constraints, runways and SID's) are governed by their geographic positions. The objects are aligned with the Z axis.
b) *geometric parameters* which control the shape of the object. The symbols chosen to represent Heathrow and surrounding landmarks are simple geometric objects - a pyramid and a cone respectively. Both can be communicated with verbal ease. The altitude constraint however, is represented by an elongated octahedron, whose size differs in proportion to the altitude it represents.
c) *descriptive parameters* such as colour, texture, transparency or sound. Two descriptive parameters are utilised to increase comprehension - colour and transparency. The colour assigned to each geometric primitive is discussed in the next Section. The transparency or opacity of the tunnels is modulated to allow symbols representing aircraft to emerge through the skin of the tunnels,



SID's with opaque (left) and transparent skin (right)

The symbols used to represent the flightpaths, runways and aircraft are

self-evident *ie* their virtual avatars bear considerable (if not absolute) similarity with their real physical manifestations. However, two variables that required to represent data - landmarks and altitude constraints - are not self-evident. The meaning assigned to them has to be learned. Although the shapes are highly distinguishable, and two out of three can be communicated verbally with ease (cone and pyramid), an economic use of such symbols is recommended in ATC displays.

3.5.4 Colour

The application of colour to the three-dimensional ATC model developed, is primarily for segregating objects. "Colour is often extremely effective as a nominal code"[145] *ie* a label. The discussion of the application of colour in visualization from a psychology of perception perspective would unnecessarily broaden the scope of this thesis. However, there are a few studies and established principles in colour theory that deserve mention.

One important question is "how can we choose effective colours that provide good differentiation between data elements during the visualization task."[146] In 1969, the anthropologists Brent Berlin and Paul Kay showed in a study on colour vocabulary, of more than 98 languages worldwide, that the naming of primary colours is remarkably consistent across diverse cultures. In *Basic Colour Terms, Their Universality and Evolution*, Berlin and Kay state, "...a total universal inventory of exactly eleven basic colour categories exists from which the eleven or fewer basic colour terms of any given language are always drawn. The eleven basic colour categories are white, black, red, green, yellow, blue, brown, purple, pink, orange and grey."[147] Hence, colours such as red, green, yellow and blue are "far more valuable in coding data than others."[148] Ware (2000) extends this set of eleven colours by adding cyan, to produce a 12 colour-set as follows:

A related question to 'how we can choose effective colours in visualization', is "how many colours we can display at once, while still allowing for rapid and accurate target identification?"[149] Previous work by Healey (1996) shows that "seven isoluminant colours is the maximum we can display at one time, while still allowing for rapid and accurate

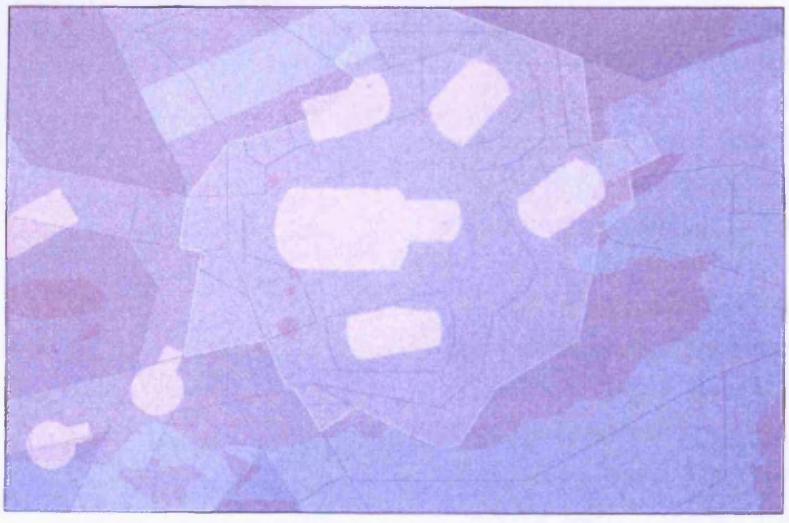
145 Ware, C., *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000, p 133

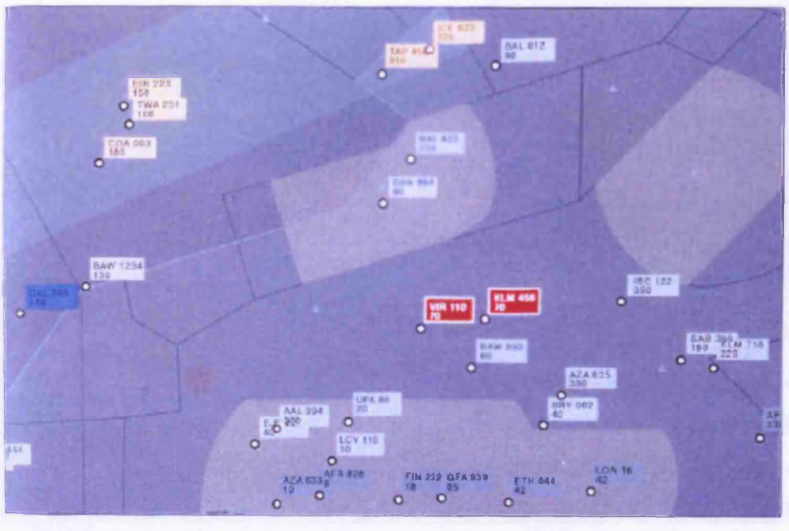
146 Healey, Christopher G., Choosing Effective Colours for Data Visualization, *Proceedings, Seventh IEEE Visualization 1996 (VIS '96)*, 1996, pp 263 - 270

147 Berlin, Brent and Paul Kay, *Basic color terms: their universality and evolution*, London: University of California Press, 1969

148 Ware C., *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000, p 119

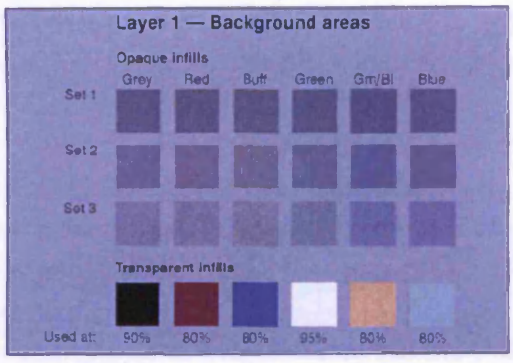
149 Healey, Christopher G., Choosing Effective Colours for Data Visualization, *Proceedings, Seventh IEEE Visualization 1996 (VIS '96)*, 1996, pp 263 - 270

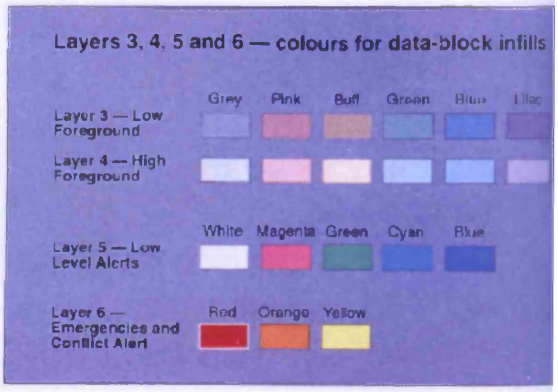




An example of the application of colour to air traffic control displays is given by Linda Reynolds in *Information Graphics: Innovative Solutions for Contemporary Design.*

As a colour specialist with the Colour Group setup under the NATS Research and Development Directorate, the task required developing a draft colour standard for use by display designers. The two images to the left (centre and lower) represent prototypes of full-colour radar displays in static form. The displays use colours of medium luminance and avoid highly saturated colours to avoid visual fatigue. Transparency is introduced to distinguish between background and foreground layers. The results have been used to fine tune specifications for civil and military displays at NATS.





identification of any one of the colours."[150] As mentioned in Chapter 1, colour is preattentive, and can be effectively applied to visualization tools where "high-speed target detection, boundary identification and region detection are important."[151] The use of green in monochromatic radar displays owes to the fact that it is the most sensitive part of the visible spectrum. Colours such as white and yellow suffer from excessive brightness contrast, whereas red and blue suffer from inadequate contrast. Green maintains good brightness contrast throughout most of its saturation range.

150 Healey, Christopher G., Choosing Effective Colours for Data Visualization, *Proceedings, Seventh IEEE Visualization 1996 (VIS '96)*, 1996, pp 263 - 27

151 Healey, Christopher G., Kellog S. Booth and James T. Enns, 'High-Speed Visual Estimation Using Preattentive Processing', *ACM Transactions on Computer-Human Interaction*, Vol 3, No 2, June 1996, pp 107-135

The choice of colour for different objects also depends on the background colour in the model. The colour of the background in our representation is light grey. The Swiss cartographer Eduard Imhof's classic work *Cartographic Relief Presentation* describes design principles for the Swiss maps. The third rule in this work states:

> Large area background or base-colours should do their work most quietly, allowing the smaller brighter areas to stand out most vividly, if the former are muted, greyish or neutral. For this reason, grey is regarded in painting to be one of the prettiest, most important and most versatile of colours. Strongly muted colours, mixed with gray, provide the best background for the coloured theme. This philosophy applies equally to map design.[152]

152 Imhof, Eduard, (Editor) H.J. Steward, *Cartographic Relief Presentation*, Berlin: Walter de Gruyter & Co., 1982

Ultimately, the study of colour and its application depends on context. For a visualization system required to handle a critical task as the one under discussion, the application of colour and its effectiveness have to be thoroughly tested for application in industry. This is beyond the scope of this thesis. In this case study, colour has been discussed from scientific, cultural and aesthetic viewpoints - the synthesis of which is necessary for any visualization.

3.5.5 "What means this shape of hovering shade?"[153]

153 Pascoe, D., *Airspaces*, London: Reaktion, 2001, p 76

Semiotics is the study of symbols and how they convey meaning. A 3d model consisting of tunnels in airspace means little to an air traffic controller. It needs to be supplemented by a library of symbols as seen earlier, each of which imparts some meaning to the representation. Let us follow a hypothetical flight to any destination departing from Heathrow. The runways, the city of London, surrounding cities, towns or landmarks of concern to air traffic control; altitude constraints and the relevant sectors of UK airspace - all of these are represented and acquire meaning through some form of geometry and convey a message to the air traffic controller. The Highway Code is an example of a set of symbols that we "consult in order to locate, identify and understand."[154] In this context, it would not be surprising if an international set of symbols for air traffic

154 Owen, William, (Ed.) Roger Fawcett-Tang, *Mapping: An illustrated guide to graphic navigational symbols*, Mies, Switzerland: RotoVision, 2002

control is introduced.

Visualization can be approached as an art, science or language. However, the argument for visualization as a language has strong philosophical proponents. Colin Ware, an expert in the psychology of perception explains in *Information Visualization, Perception for Design*,

> Visualization is about diagrams and how they can convey meaning. Diagrams are generally held to made up of symbols, and symbols are based on social interaction. The meaning of a symbol is normally understood to be created by convention, established in the course of person-to-person communication. Diagrams are arbitrary and are effective in much the same way as the written words on this page are effective - we must learn the conventions of the language and the better we learn them, the clearer the language will be. Thus, one diagram may ultimately be as good as another; it is just a matter of learning the code, and the laws of perception are largely irrelevant.[155]

155 Ware C., *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000, p 5

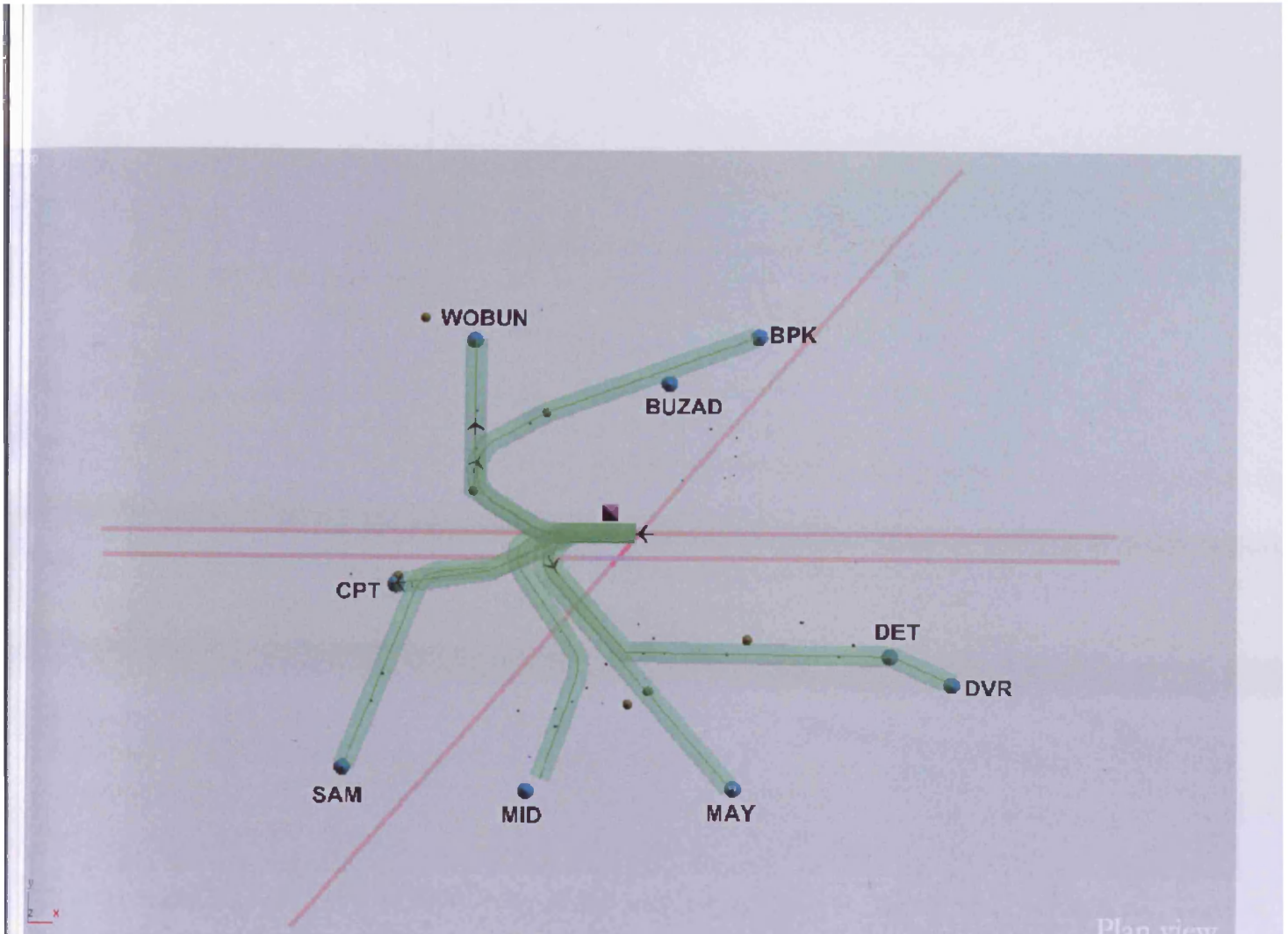
The detailed analysis of the semantics of symbols and their effectiveness from the science of human perception and cognition is beyond the remit of this thesis. The previous two sections have outlined the perceptual basis for the symbols chosen, and decisions governing surface properties such as colour and transparency.

3.5.6 Note to images

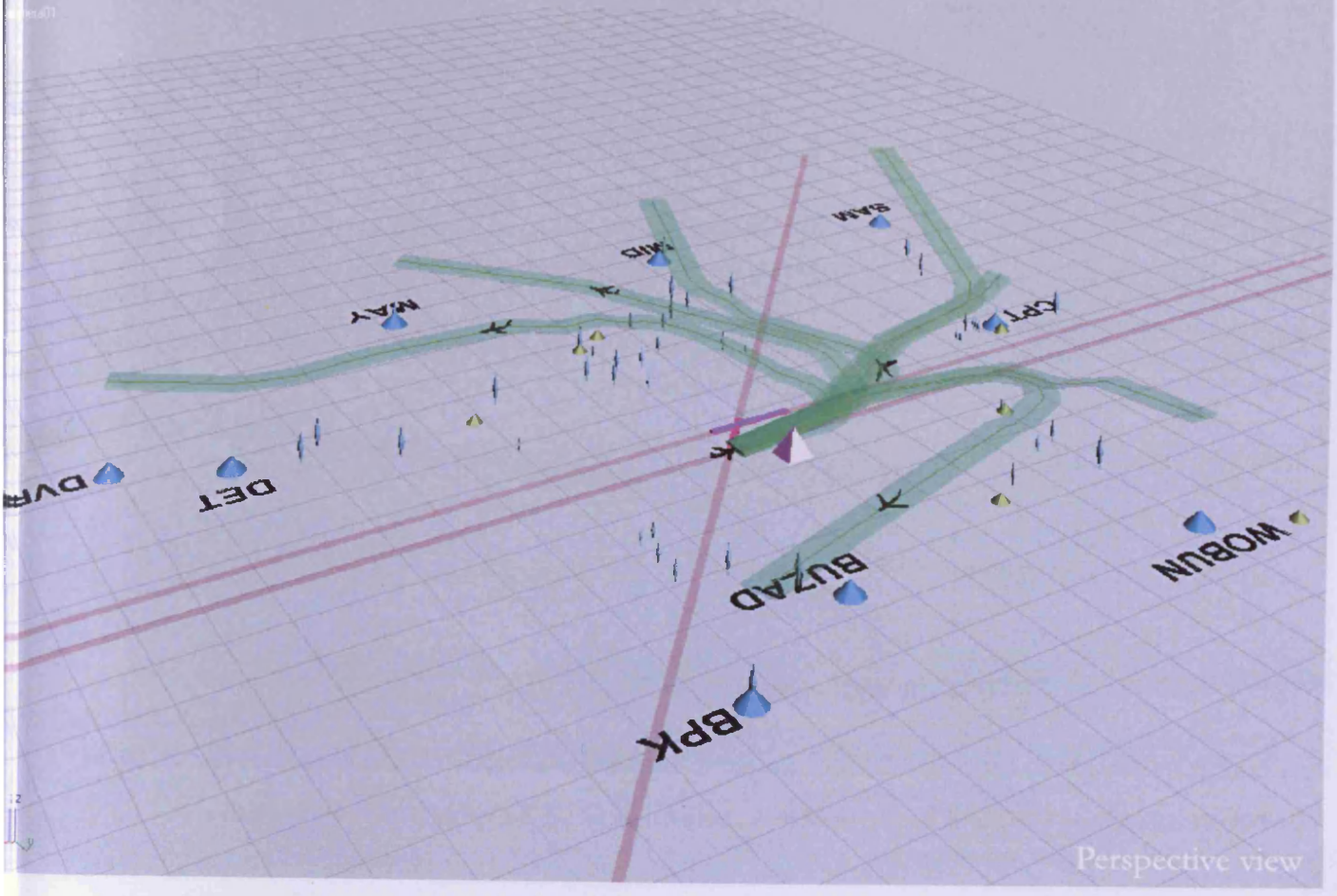
The images overleaf represent still frames from a simulation of Standard Instrument Departures or SID's from Heathrow airport. The region of airspace modelled extends to 25 nautical miles from the airport and includes seven SID's via Compton, Southampton, Midhurst, Mayfield, Dover, Detfield, Buzad, Brookmans Park and Wobun. The animation is modelled using the software application 3d Studio Max (version 6) and represents an egocentric *ie* through-the-window or monitor-based view. The animations have been recorded as video files and can be viewed on the CD accompanying the thesis. The Technology and Programmes Division at NATS anticipates that the 'air tunnel approach' depicted here will be operational by 2015. As further development of the project was inhibited by an intellectual property dispute between the University and NATS (Section 3.7), data related to STandard Arrival Routes or STAR's was not available to the author, and have thus not been simulated.


Orthographic

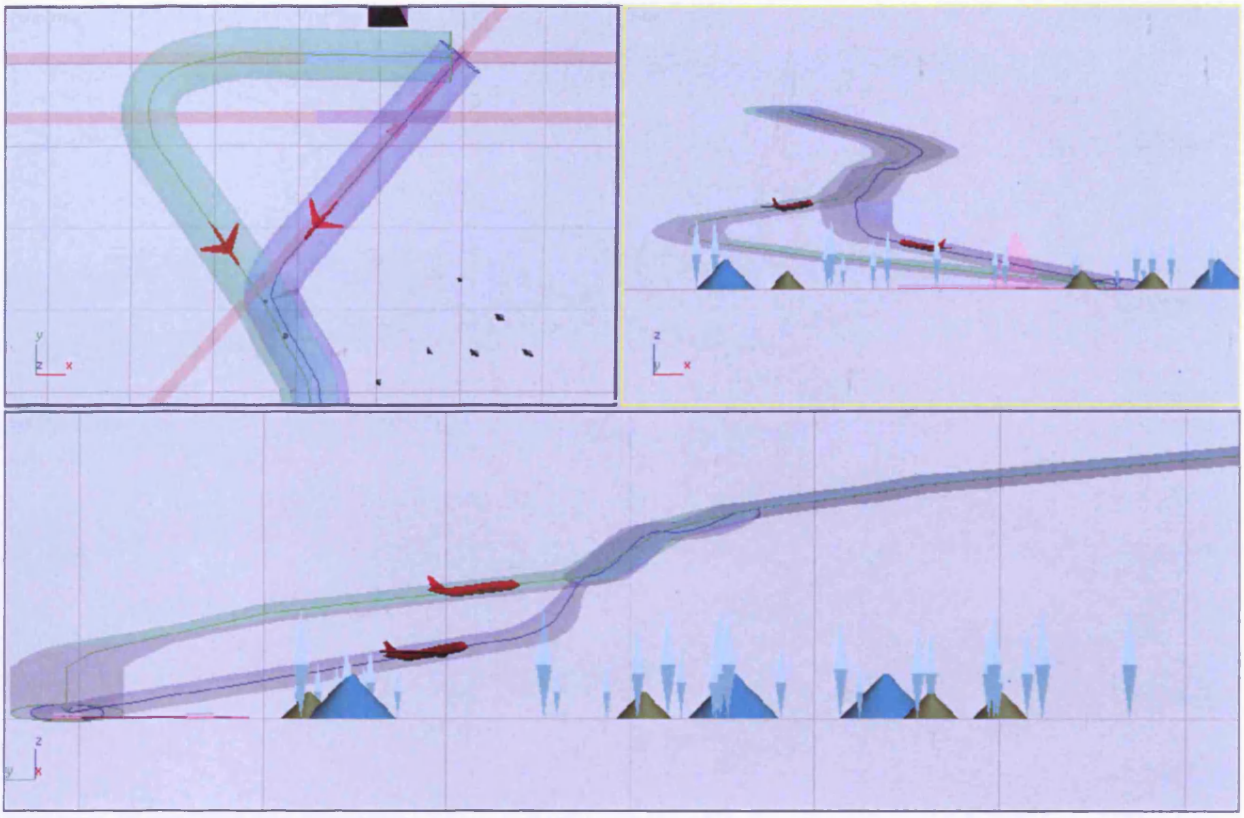

Camera



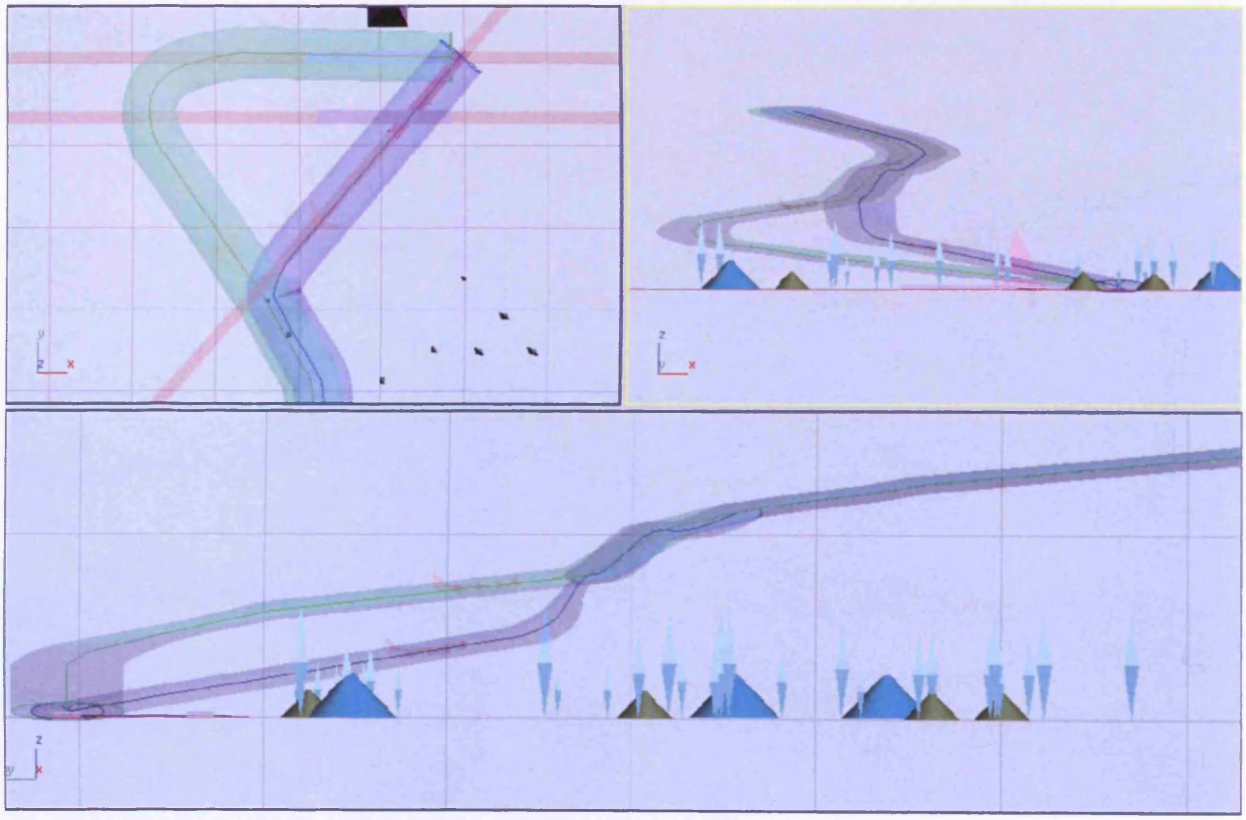
Plan view



Perspective view



above: t = 0 and below: t = 1.5 seconds

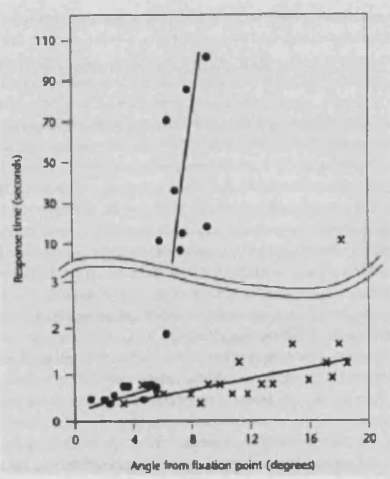


Short-term Conflict Alert: The hue of the aircraft symbols changes as the system detects the possibility of a conflict. The symbols start flashing at a 1.5 second interval.

3.6 Short-Term Conflict Alert (STCA)

A Short Term Conflict Alert refers to the possible mid-air collision of two or more aircraft. In such critical cases, rapid response is required from an air traffic controller to avert a disaster. For that to occur, the visualization should be able to alert the user of such a possibility, effectively and efficiently. In a real-time monitoring scenario, the 'expected cost' of missing such an event is too high. As discussed earlier, all objects in airspace are bound to the minimum lateral and vertical separation distances, that result in an elliptical cocoon of airspace that insulates an aircraft. However, there can be a situation where the radii - horizontal or vertical - of two aircraft can come dangerously close to touching or even overlapping each other. Aircraft which violate each other's minimum separation distances - lateral and vertical, longitudinal or time - are said to be in conflict.

The solution employed in this study is to use a 'flashing target' *ie* a moving target that starts to flash at an appropriate time by detecting the possibility of such a scenario. A moving-flashing target has two significant benefits. Peterson and Dugas (1972)[156] show that moving objects have a wider 'useful field of view' (UFOV)[157], compared to static targets. Since motion does play a role in attracting a user's attention outside the UFOV, the persistence of a flashing target can continue to remind the user that the associated task needs urgent attention. Goldstein and Lamb (1967) show how a flashing target can be "effective as an attention-getting device and as a message source."[158] Moreover, four different flash rates have been found to be easily discernible by the visual system. The more critical the signal or message becomes, the higher its flash rate. Change in the hue of the aircraft symbol is another mechanism of seeking visual attention. The hue of the aircraft changes from black (normal) to red (critical). However, research has shown that peripheral vision is colour-blind, which reduces the effectiveness of colour as a user interrupt compared to a flashing symbol.



Detection time as a function of distance from fixation point. There is a sharp increase in the detection times for a static target as its position goes beyond 60 from the fixation point.

156 Peterson, Harold E. and Doris J. Dugas, 'The Relative Importance of Contrast and Motion in Visual Detection', *Human Factors*, Vol 14, No 3, 1972, pp 207-216

157 UFOV or 'useful field of view' defines the size of the region on screen from which one can rapidly take in information.

158 Goldstein, Donald A. and Jerry C. Lamb, 'Visual Coding Using Flashing Lights', *Human Factors*, Vol 9, No 5, 1967, pp 405-408

3.7 Evaluation and Limitations

In September 2004, the project was presented to senior figures at NATS' London Terminal Control Centre (LTCC). At this point, NATS was willing to fund the further development of the three-dimensional airspace model. The proposal was entitled '3D Representation of ATC Data' and involved production of the current London Terminal Control airspace model, and modelling of future airspace designs. The model would replicate the SIDs and STARs to all of the London airports - Stansted, London City, Heathrow, Gatwick and Luton - and adjacent airports such as Southampton, Farnborough, Birmingham, Coventry and East Midlands. NATS emphasised their interest in a design guide for future use by the Organization. However, the project was embroiled in Intellectual Property issues, and an agreement could not be reached between the University and NATS. At this point, it became clear that further collaboration was not possible, preventing both a formal evaluation of the display and further development of the case study.

User studies have hence not been part of this study. However, input has been received from domain experts at NATS with whom the validity of the proposal has been discussed. They confirmed its graphical superiority and an internal email confirms the benefits of the visualization produced:

> Can provide more graphical detail on tracks, aircraft, buildings, terrain, etc.
> Faster processing and allows for a multiple of camera views.
> Cylindrical tracks whilst Luciad has lines only.
> A more detailed simulation of an aerodrome, London TMA, etc.
> More emergent and realistic aircraft behaviour.[159]

In terms of interaction, one of the most significant benefits of the tool is the ability to rotate in three-dimensions in real-time.[160] The perspective view depicted on page 60 was considered 'ideal' for purposes of viewing the interactions between SID's.[161]

159 Email from Devang Lakhani (Project Manager) to the author on September 28, 2004. Refer to Appendix A.

160 Refer to the video 'Four Views.exe' in the 'Chapter 3_Case Study 1' folder on the CD accompanying this thesis.

161 Refer to Appendix A for a transcript of discussions between the author and Mark Green (ATC Planner) at NATS. Refer to the video 'Camera View.exe' in the 'Chapter 3_Case Study 1' folder on the CD accompanying this thesis.

3.8 Summary

Air Traffic Control is a complex collaborative activity which is conducted in real-time. New technologies are introduced with extreme caution. The design of the ATC display integrates established principles in ATC display design with principles of visual data design. It is not an attempt to provide *the* solution to issues inherent in air traffic control displays, but shows that principles of information design, if applied in careful measure, can greatly enhance the efficacy, effectiveness and beauty of data displays.

4.0 Visualizing Probable Space

4.0 Visualizing Probable Space

This case study looks at the process of *search, display and interaction* in a software tool designed to help with the navigation of data in a hypertext document. It is important to note that the goal or aim of the project is *not* to create a search engine or search tool per se, but to create an interactive learning tool.[162] The focus is on context and its representation and not on semantics. This may seem absurd in light of the first half of the preceding sentence (as meaning is derived from context) but will become clear in further sections. The case study applies the theoretical aspects of information explored in Chapter 2 to the design and display of the spatial navigation tool.

162 see 'Section 4.8: Discussion'

4.1 "Electric Rain"[163]

Googol is a number represented by 10^{100}. In 1998, Messrs Larry Page and Sergy Brin applied for the registration of their company by this name, but misspelt it as Google. This name has since become synonymous with one particular web-related activity: search.[164] The number of web pages indexed by the search engine Google at the time of writing is nearly ten billion or 10^{10}. While it is true that not all information or knowledge is available 'online', it is a fact that an increasing volume of data is published electronically, the categories of which range from education, business and media to the highly personal internet diaries or 'blogs' (web logs). The steady proliferation of the internet worldwide and increasing bandwidth capacities are factors driving the generation and dissemination of digital data in numbers in 'powers of 10'.

163 Baeyer, Hans Christian von, *Information: The New Language of Science*, London: Phoenix, Orion Books, 2003, p 3

164 According to the Oxford English Dictionary, 'to Google' or 'Googling' refers to using the search engine Google to find information on the Internet.

Although the foundations of the internet can be traced to developments at the Advanced Research Projects Agency (ARPA) in the United States in the 1960s, its commercial application was conceived at the European Organization for Nuclear Research (CERN) by Tim Berners Lee, who described a set of access protocols and display standards for an internet application called World Wide Web (WWW). These standards were later incorporated in an application named Mosaic by Netscape Communications Corporation. In the early 1990s, Microsoft Corporation launched its widely used browser, Internet Explorer. Today, 'browsing the Net' has become alike to exploring locations in an unchartered territory. More specifically, searching for information on the Net has become equivalent to digging for precious minerals in deep mines. Unsurprisingly, the terms 'data mining' and 'knowledge discovery' are closely integrated with the science of visualization.

4.2 Visual Representation of Search Queries

The term 'query' is used here to represent what one is looking for on the internet using a search engine such as Google, Yahoo or MSN. Before delving into the visual representation of search results, a brief overview of how search engines work is necessary. A generic description is presented here.[165]

165 'How do Internet search engines work?', Scientific American, January 2003, p82

Publicly available search engines such as Google, Yahoo, MSN, Microsoft Live and Altavista use three most common methods to search for user queries: Preprocessing data, 'smart' representation and prioritizing the results. A software named a 'crawler' crawls through the web looking for web pages. These pages are then parsed[166] to extract keywords which are stored along with the corresponding links to web pages in an index file. User queries are matched against this index file. To update the database, the crawler performs its routine periodically. Smart representation refers to selecting an index structure that minimizes search time. For example, is it more efficient to search through a list sequentially, or to use a tree structure? It is established that the latter is a more efficient way to organise data. The next step involves page ranking strategies. Here statistics and probability take precedence by employing an algorithm that gleans for the number of times a specific term occurs in a document. Known as 'term-frequency-inverse-document-frequency', relative weights are attached to a term signifying its importance in a document. Search results may also be refined further using techniques such as link analysis. In this strategy, the importance of a link *ie* a web page or a website depends on whether it is an authority (a number of pages point to it) or a hub (it points to a number of other pages).

166 In computer literature, 'to parse' means to describe the syntactic role of a word in a sentence or a phrase.

An information quest usually begins with an information problem that is conceptualized as a statement or question by the user. This is then transformed into a search expression *ie* a series of keywords, or a phrase, which can be submitted to the system. The user may also select from an existing directory of categories. In response to the search expression, the system returns a list of documents for the user to scan. The user may then examine the documents in more detail, or may revise and re-submit the query to the system in search for more or better results.[167] This time-honoured pattern of behaviour has changed very little in the last 40 years, since Douglas Engelbart's NLS (oN Line System) in 1968 - a hypertext system inspired by Vannevar Bush's visionary 'Memex' system, first describing hypertext in the 1945 article 'As We May Think'.

167 Toms, Elaine G., 'Supporting the Sub-Tasks used in Search', *http://xtasy.slis.indiana.edu/jcdlui/papers/toms.doc*, Accessed <April 27, 2004>

4.2.1 Data display techniques in Search Engines

This section reviews how search engines, from the most popular at this time, to the relatively unknown, present information to the user. A cross-section of displays shows a variety of methods utilised, from plain text, to methods combining pictures and words. Twenty search engines are reviewed, based on their market capitalization[168] and more importantly, data display technique. Search engines employing a common data display technique are grouped into a similar category:

a) Text (String of characters)
Google, Yahoo, Live, MSN, Excite, SearchMash, Ask, Lycos, AlltheWeb and Guruji

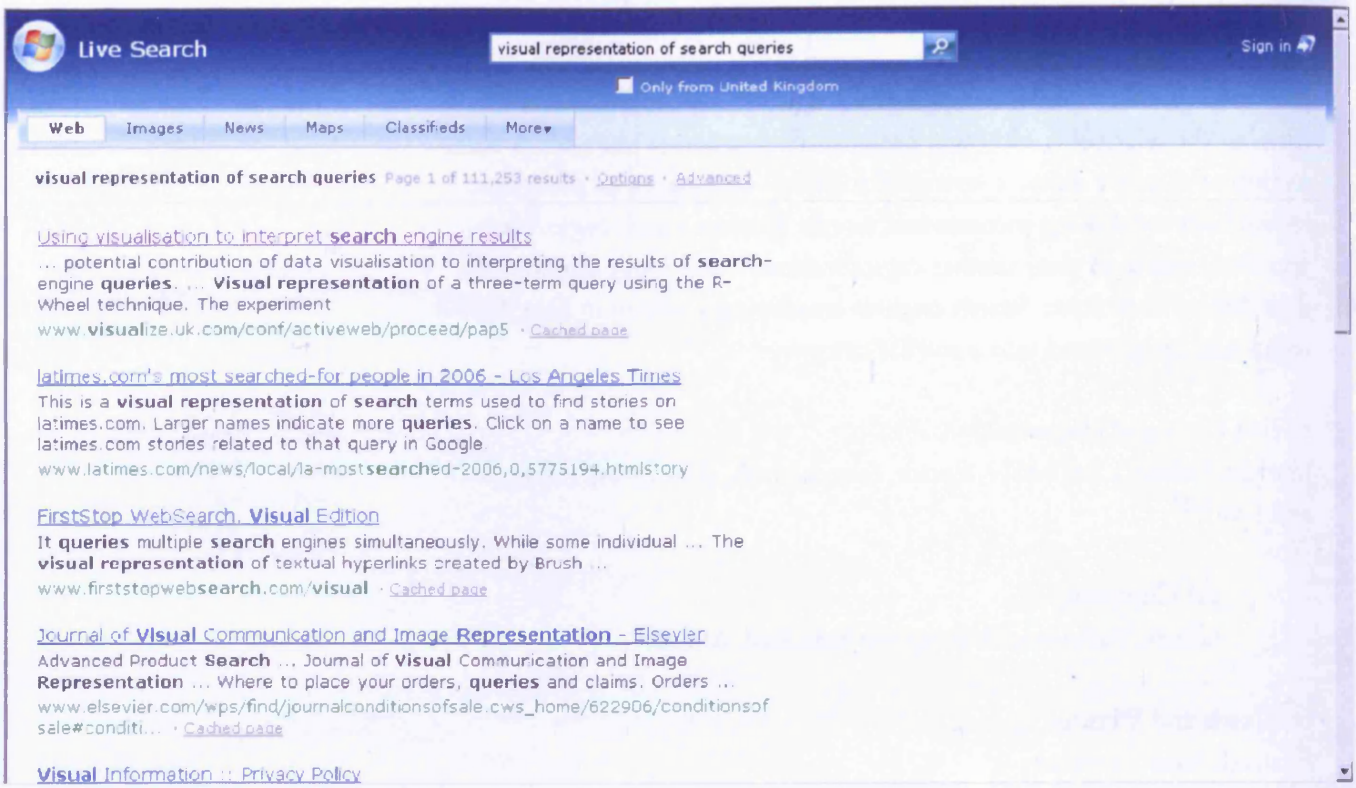
aa) Clustering
Clusty, Vivisimo, Shakespeare Searched, Mooter

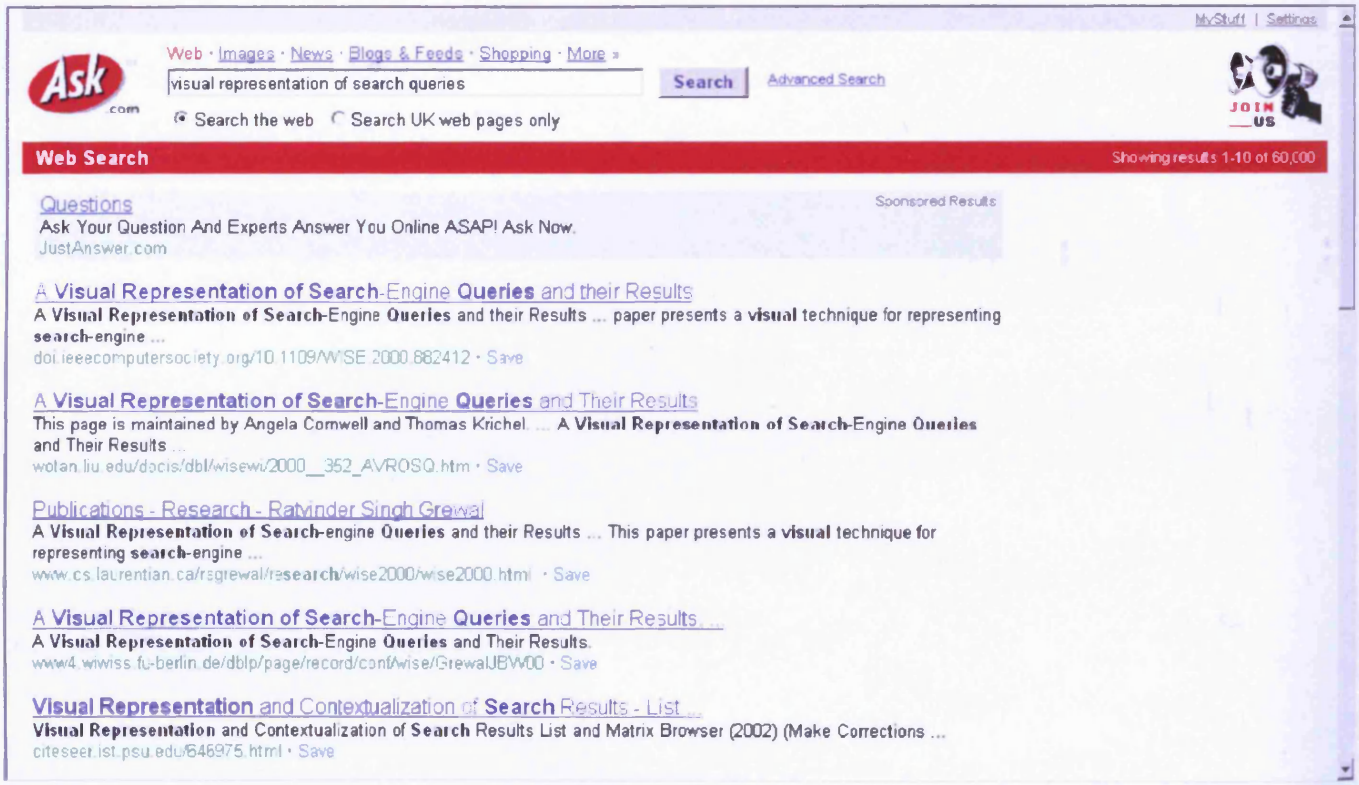
b) Words and Pictures
Exalead, Snap, Grokker

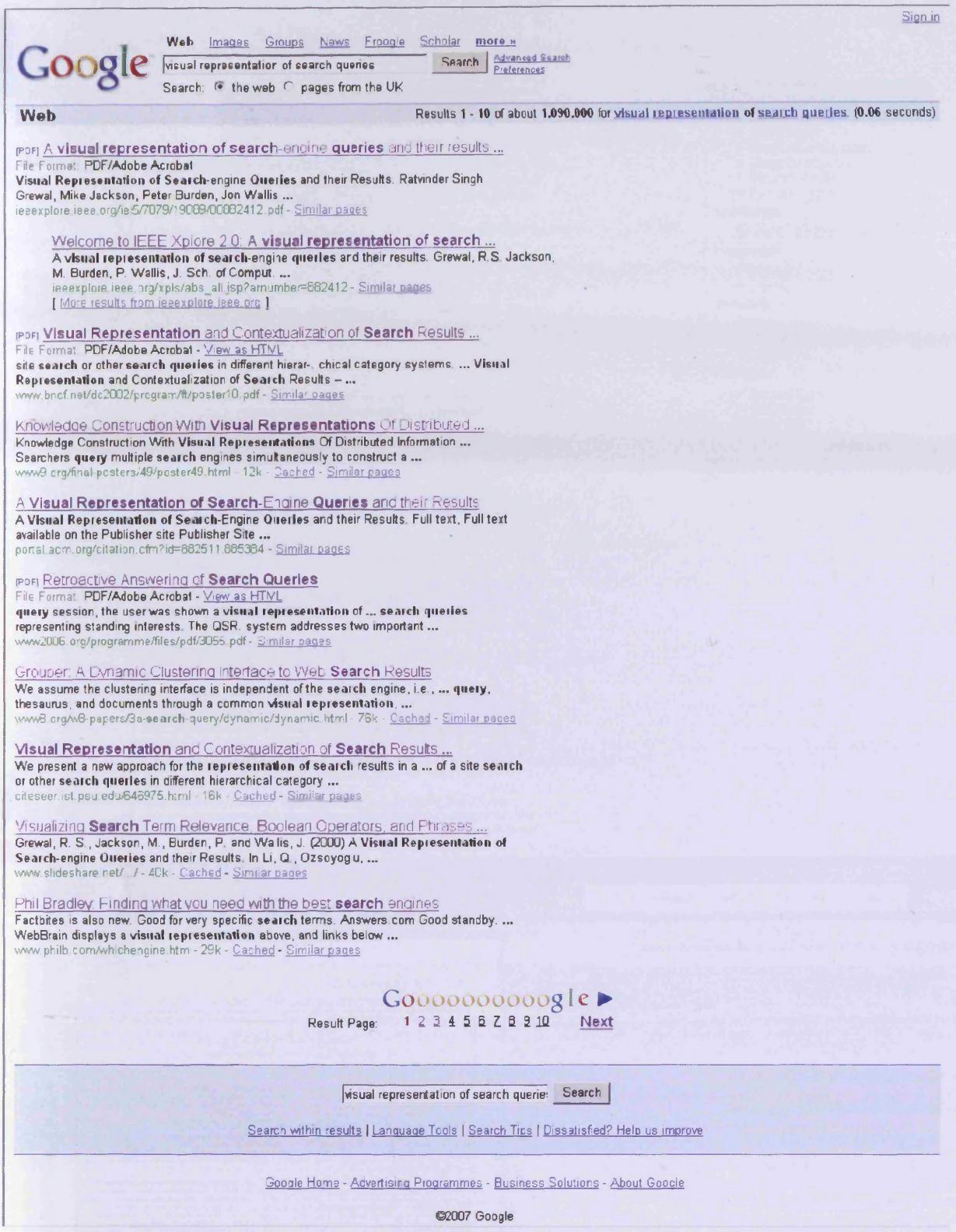
bb) Topological / Predominantly graphic
Kartoo, Brain, Quintura

168 According to the Nielsen//Net Ratings for July 2006, the market share of search engines in the USA was:

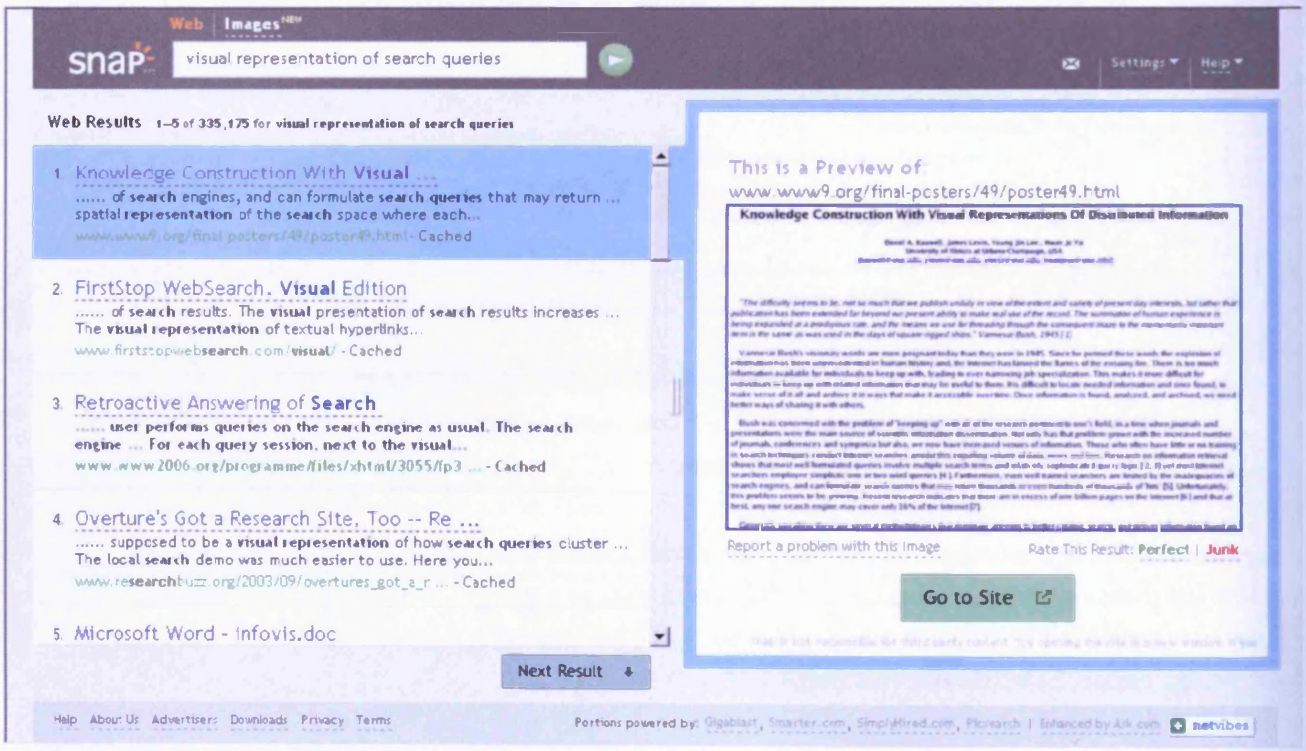
Google	49.2%
Yahoo	23.8%
MSN	9.6%
AOL	6.3%
Ask	2.6%
Others	8.5%


Live Search
visual representation of search queries
Sign in
Only from United Kingdom
Web
Images
News
Maps
Classifieds
More
visual representation of search queries Page 1 of 111,253 results · Options · Advanced
Using visualisation to interpret search engine results
... potential contribution of data visualisation to interpreting the results of search-engine queries. ... Visual representation of a three-term query using the R-Wheel technique. The experiment
www.visualize.uk.com/conf/activeweb/proceed/pap5 · Cached page
latimes.com's most searched-for people in 2006 - Los Angeles Times
This is a visual representation of search terms used to find stories on latimes.com. Larger names indicate more queries. Click on a name to see latimes.com stories related to that query in Google.
www.latimes.com/news/local/la-mostsearched-2006,0,5775194.htmlstory
FirstStop WebSearch, Visual Edition
It queries multiple search engines simultaneously. While some individual ... The visual representation of textual hyperlinks created by Brush ...
www.firststopwebsearch.com/visual · Cached page
Journal of Visual Communication and Image Representation - Elsevier
Advanced Product Search ... Journal of Visual Communication and Image Representation ... Where to place your orders, queries and claims. Orders ...
www.elsevier.com/wps/find/journalconditionsofsale.cws_home/622906/conditionsofsale#conditi... · Cached page
Visual Information :: Privacy Policy

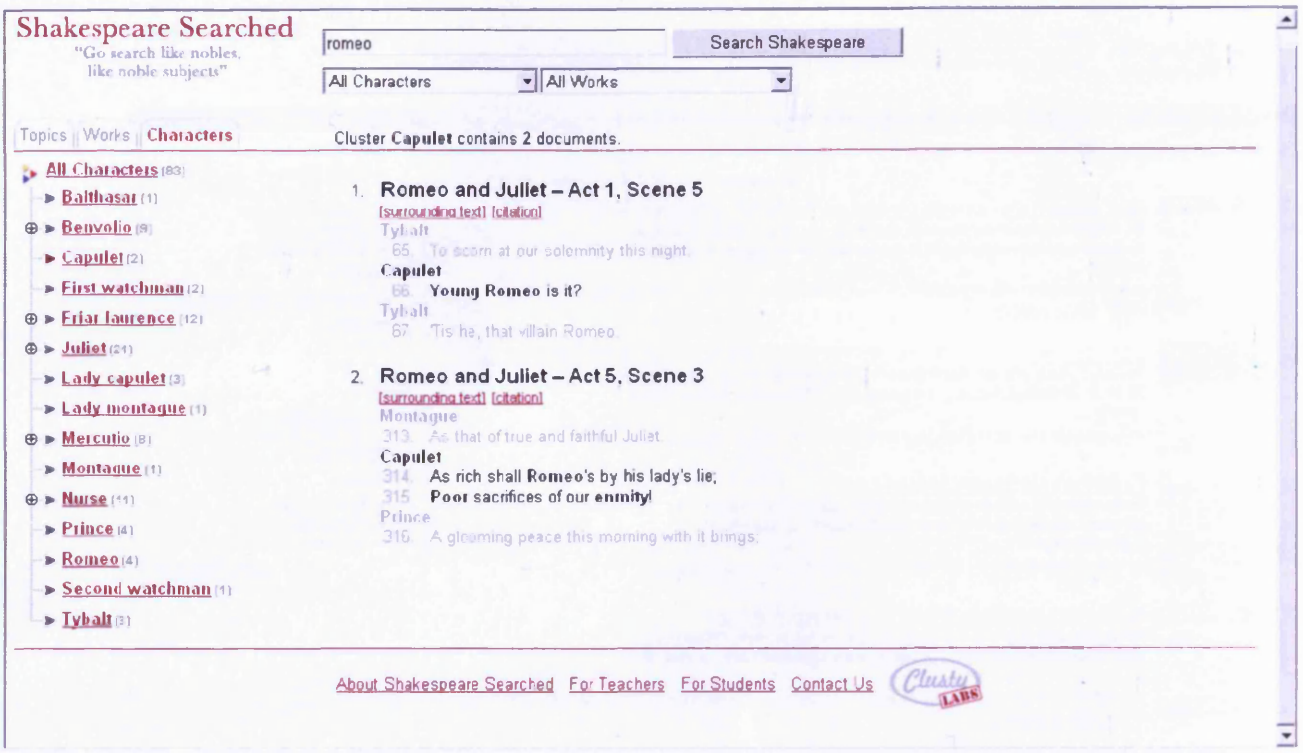

MyStuff | Settings
Web · Images · News · Blogs & Feeds · Shopping · More »
visual representation of search queries
Search
Advanced Search
Search the web
Search UK web pages only
JOIN US
Web Search
Showing results 1-10 of 60,000
Questions
Sponsored Results
Ask Your Question And Experts Answer You Online ASAP! Ask Now.
JustAnswer.com
A Visual Representation of Search-Engine Queries and their Results
A Visual Representation of Search-Engine Queries and their Results ... paper presents a visual technique for representing search-engine ...
doi.ieeecomputersociety.org/10.1109/WISE.2000.882412 · Save
A Visual Representation of Search-Engine Queries and Their Results
This page is maintained by Angela Cornwell and Thomas Krichel. ... A Visual Representation of Search-Engine Queries and Their Results ...
wotan.liu.edu/docis/dbl/wisewi/2000__352_AVROSQ.htm · Save
Publications - Research - Ratvinder Singh Grewal
A Visual Representation of Search-engine Queries and their Results ... This paper presents a visual technique for representing search-engine ...
www.cs.laurentian.ca/rsgrewal/research/wise2000/wise2000.html · Save
A Visual Representation of Search-Engine Queries and Their Results. ...
A Visual Representation of Search-Engine Queries and Their Results.
www4.wwiss.fu-berlin.de/dblp/page/record/conf/wise/GrewalJBW00 · Save
Visual Representation and Contextualization of Search Results - List ...
Visual Representation and Contextualization of Search Results List and Matrix Browser (2002) (Make Corrections ...
citeseer.ist.psu.edu/646975.html · Save

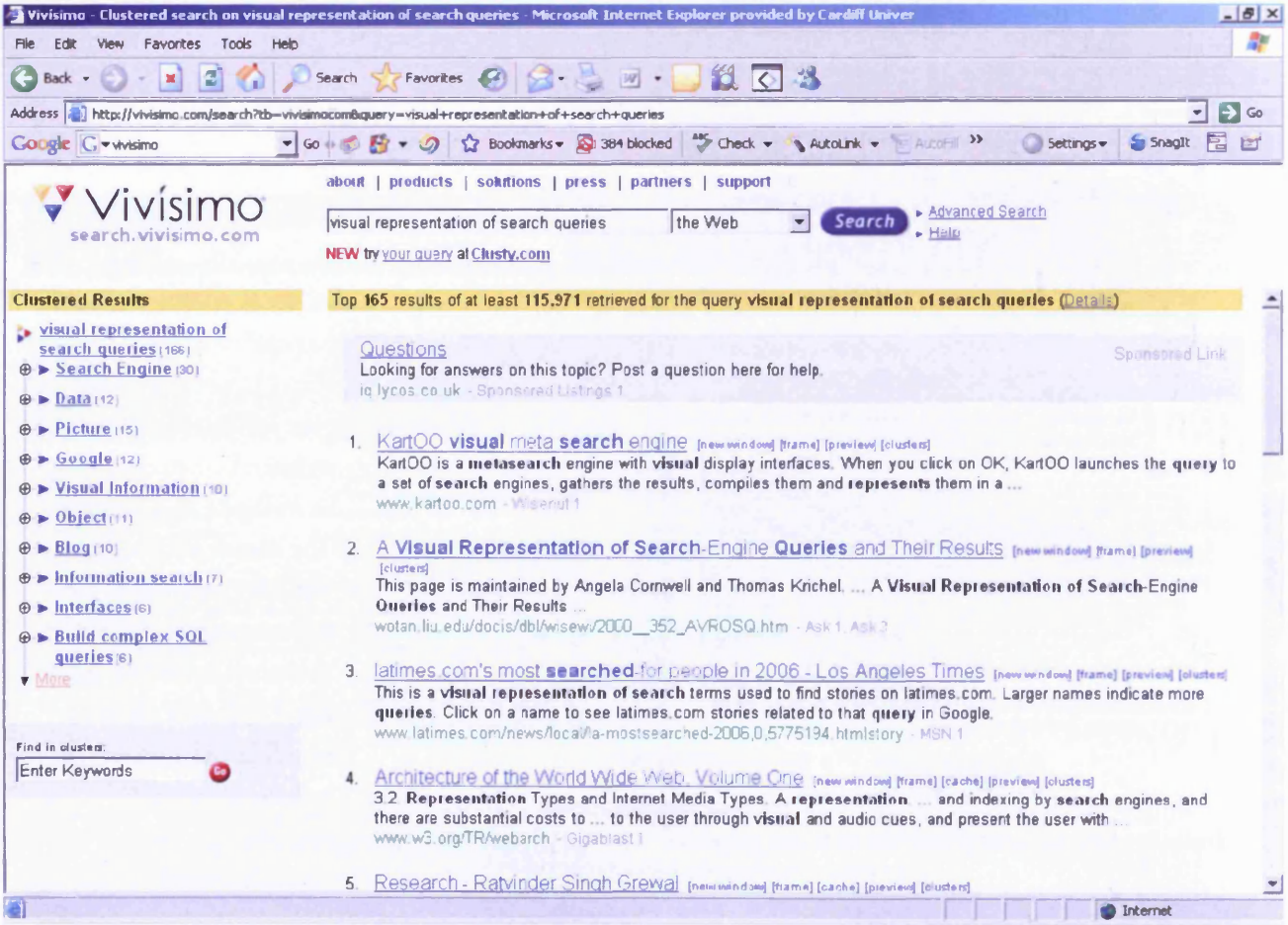


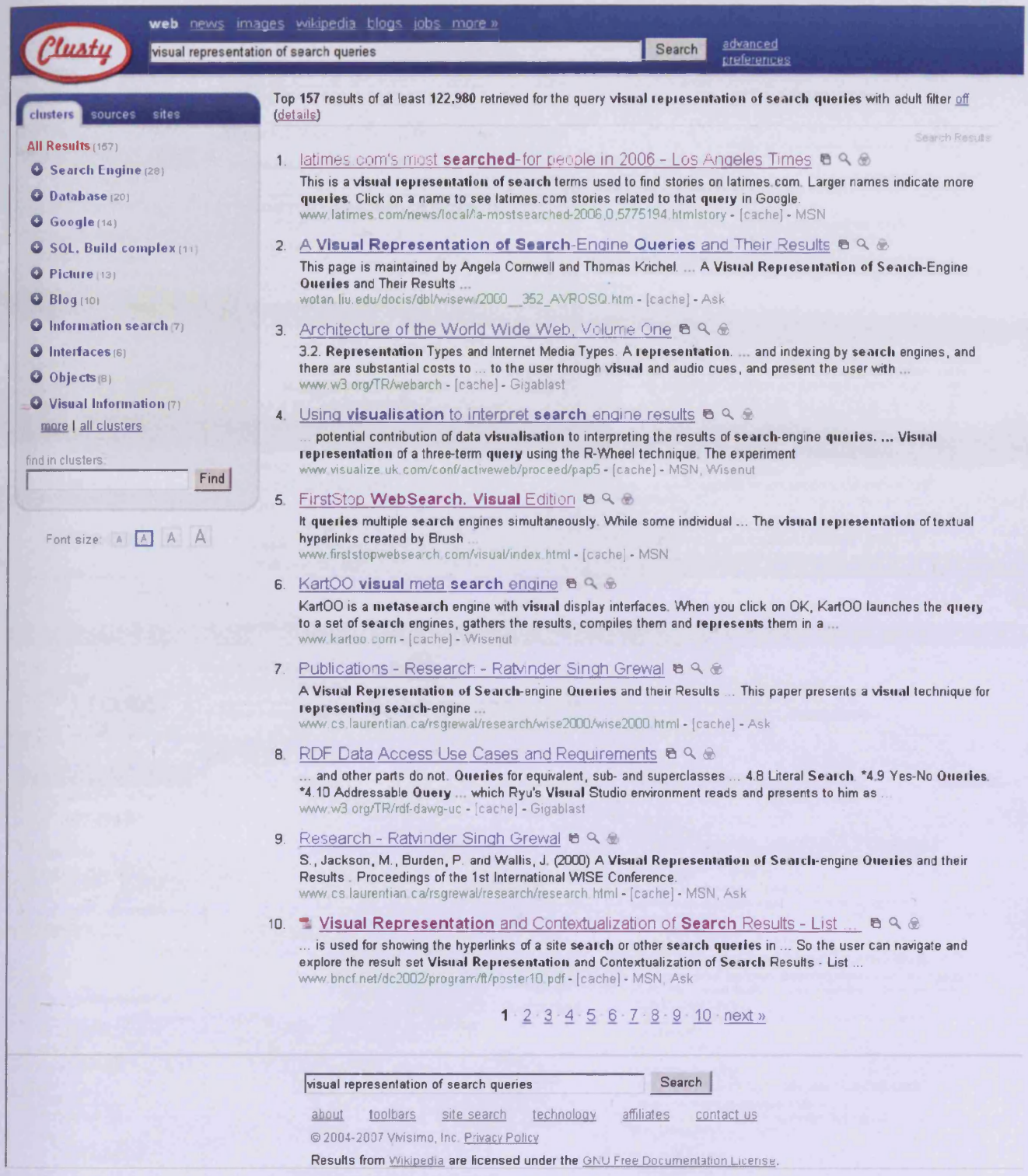
The popular search engine Google (above) displays search results in a linear text-list manner. Similar techniques of displaying data as a *string of characters* is employed by other search engines such as Live (top left), Yahoo!, MSN, Excite, SearchMash, Ask (left), Lycos, AlltheWeb and Guruji.


Web
Images NEW
snap
visual representation of search queries
Settings
Help
Web Results 1–5 of 335,175 for visual representation of search queries
1. Knowledge Construction With Visual ...
..... of search engines, and can formulate search queries that may return ... spatial representation of the search space where each...
www.www9.org/final-posters/49/poster49.html - Cached
2. FirstStop WebSearch. Visual Edition
..... of search results. The visual presentation of search results increases ... The visual representation of textual hyperlinks...
www.firststopwebsearch.com/visual/ - Cached
3. Retroactive Answering of Search
..... user performs queries on the search engine as usual. The search engine ... For each query session, next to the visual...
www.www2006.org/programme/files/xhtml/3055/fp3 ... - Cached
4. Overture's Got a Research Site, Too -- Re ...
..... supposed to be a visual representation of how search queries cluster ... The local search demo was much easier to use. Here you...
www.researchbuzz.org/2003/09/overtures_got_a_r ... - Cached
5. Microsoft Word - infovis.doc
Next Result
This is a Preview of
www.www9.org/final-posters/49/poster49.html
Knowledge Construction With Visual Representations Of Distributed Information
Report a problem with this image
Rate This Result: Perfect | Junk
Go to Site
Help About Us Advertisers Downloads Privacy Terms
Portions powered by: Gigablast, Smarter.com, SimplyHired.com, Picsearch | Enhanced by Ask.com
netvibes

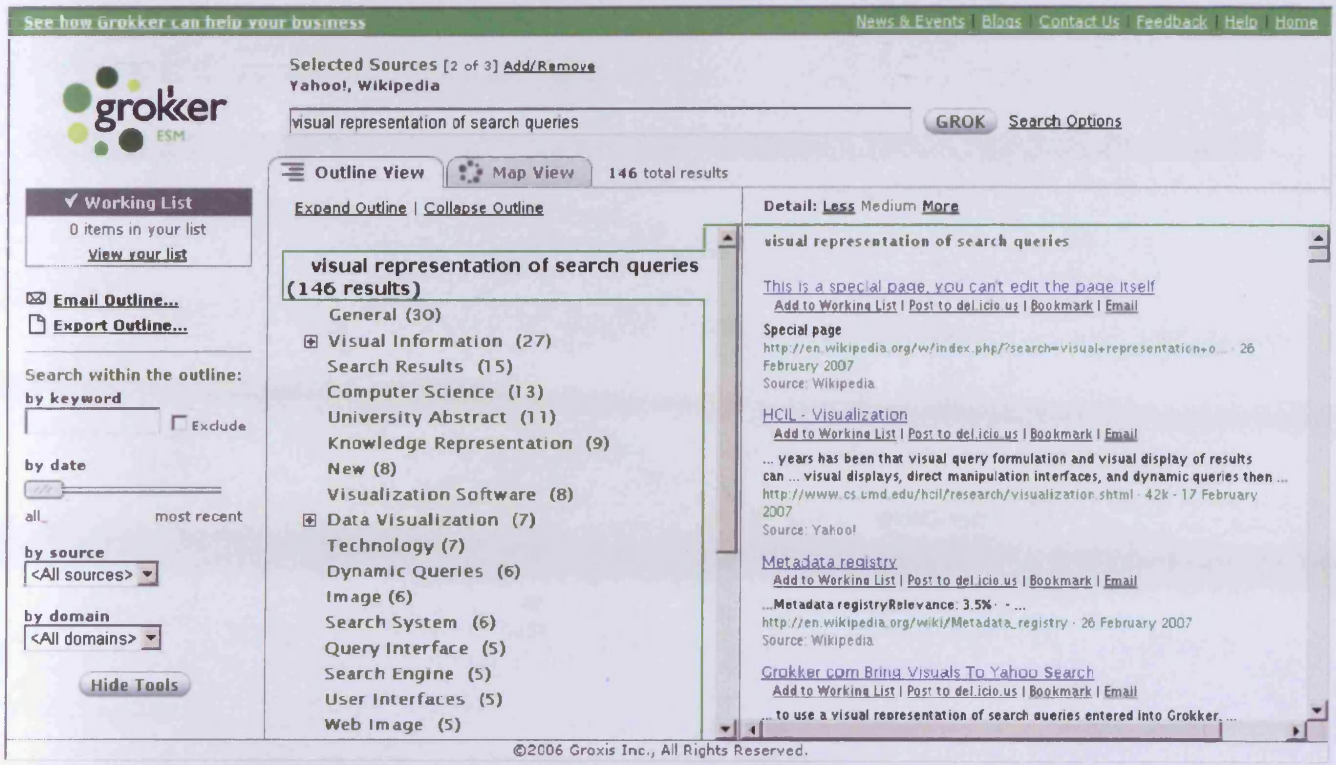
Exalead displays thumbnails or snapshots of the homepage of each website next to its text-based hyperlink. The user can preview the website before choosing to navigate to a website. The search engine Snap (far left) uses the same technique but places the thumbnails on the right-hand side and allows a much larger preview of websites.

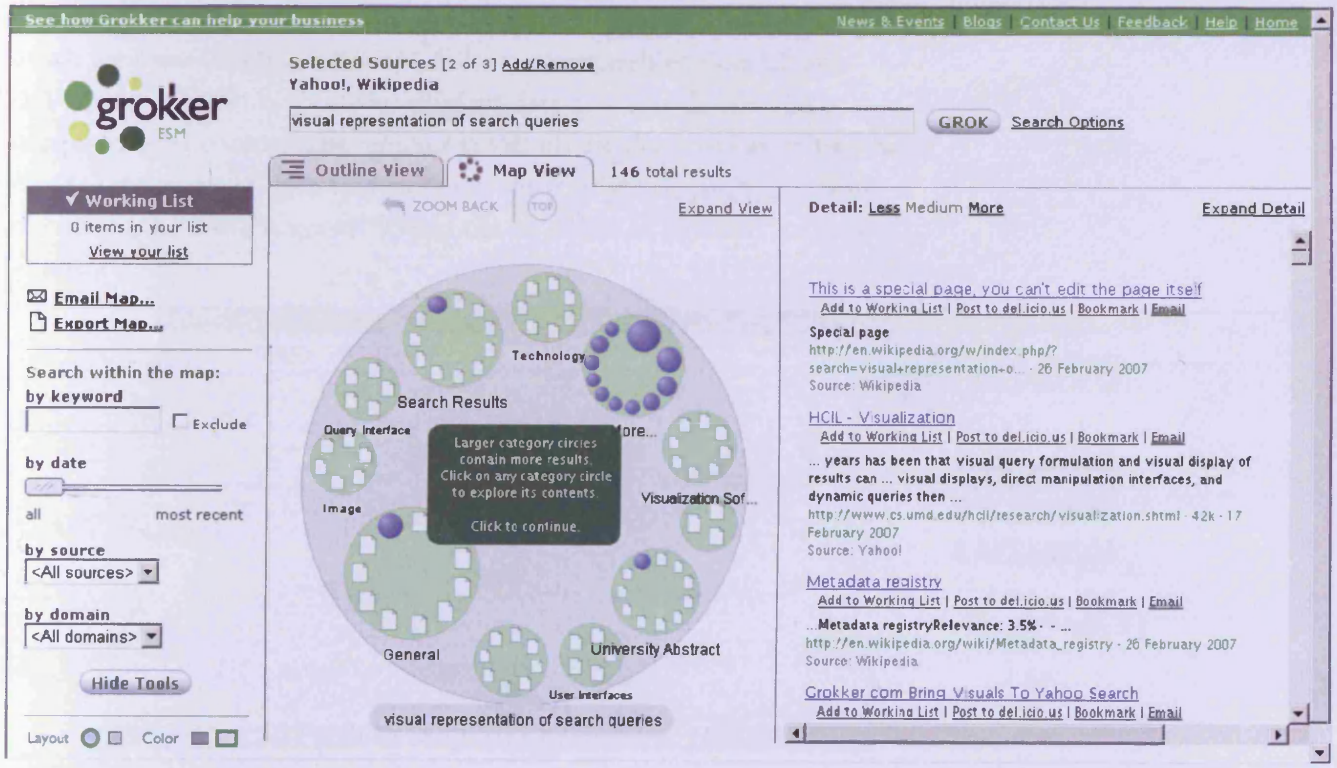

Shakespeare Searched
"Go search like nobles, like noble subjects"
romeo
Search Shakespeare
All Characters
All Works
Topics | Works | Characters
Cluster Capulet contains 2 documents.
All Characters (83)
Balthasar (1)
Benvolio (9)
Capulet (2)
First watchman (2)
Friar laurence (12)
Juliet (21)
Lady capulet (3)
Lady montague (1)
Mercutio (8)
Montague (1)
Nurse (11)
Prince (4)
Romeo (4)
Second watchman (1)
Tybalt (3)
1. Romeo and Juliet – Act 1, Scene 5
[surrounding text] [citation]
Tybalt
65. To scorn at our solemnity this night.
Capulet
66. Young Romeo is it?
Tybalt
67. 'Tis he, that villain Romeo.
2. Romeo and Juliet – Act 5, Scene 3
[surrounding text] [citation]
Montague
313. As that of true and faithful Juliet.
Capulet
314. As rich shall Romeo's by his lady's lie;
315. Poor sacrifices of our enmity!
Prince
316. A glooming peace this morning with it brings;
About Shakespeare Searched
For Teachers
For Students
Contact Us
Clusty LABS


Vivisimo - Clustered search on visual representation of search queries - Microsoft Internet Explorer provided by Cardiff Univer
File Edit View Favorites Tools Help
Address http://vivisimo.com/search?tb=vivisimocom&query=visual+representation+of+search+queries
Vivisimo
search.vivisimo.com
about | products | solutions | press | partners | support
visual representation of search queries
the Web
Search
Advanced Search
Help
NEW try your query at Clusty.com
Clustered Results
Top 165 results of at least 115,971 retrieved for the query visual representation of search queries (Details)
visual representation of search queries (165)
Search Engine (30)
Data (12)
Picture (15)
Google (12)
Visual Information (10)
Object (11)
Blog (10)
Information search (7)
Interfaces (6)
Build complex SQL queries (6)
More
Find in clusters:
Enter Keywords
Questions
Looking for answers on this topic? Post a question here for help.
iq.lycos.co.uk - Sponsored Listings 1
Sponsored Link
1. KartOO visual meta search engine
KartOO is a metasearch engine with visual display interfaces. When you click on OK, KartOO launches the query to a set of search engines, gathers the results, compiles them and represents them in a ...
www.kartoo.com - Wisenut 1
2. A Visual Representation of Search-Engine Queries and Their Results
This page is maintained by Angela Cornwell and Thomas Krichel. ... A Visual Representation of Search-Engine Queries and Their Results ...
wotan.liu.edu/docis/dbl/wisew/2000__352_AVROSQ.htm - Ask 1, Ask 2
3. latimes.com's most searched-for people in 2006 - Los Angeles Times
This is a visual representation of search terms used to find stories on latimes.com. Larger names indicate more queries. Click on a name to see latimes.com stories related to that query in Google.
www.latimes.com/news/local/la-mostsearched-2006,0,5775194.htmlstory - MSN 1
4. Architecture of the World Wide Web, Volume One
3.2 Representation Types and Internet Media Types. A representation ... and indexing by search engines, and there are substantial costs to ... to the user through visual and audio cues, and present the user with ...
www.w3.org/TR/webarch - Gigablast 1
5. Research - Ratvinder Singh Grewal
Internet

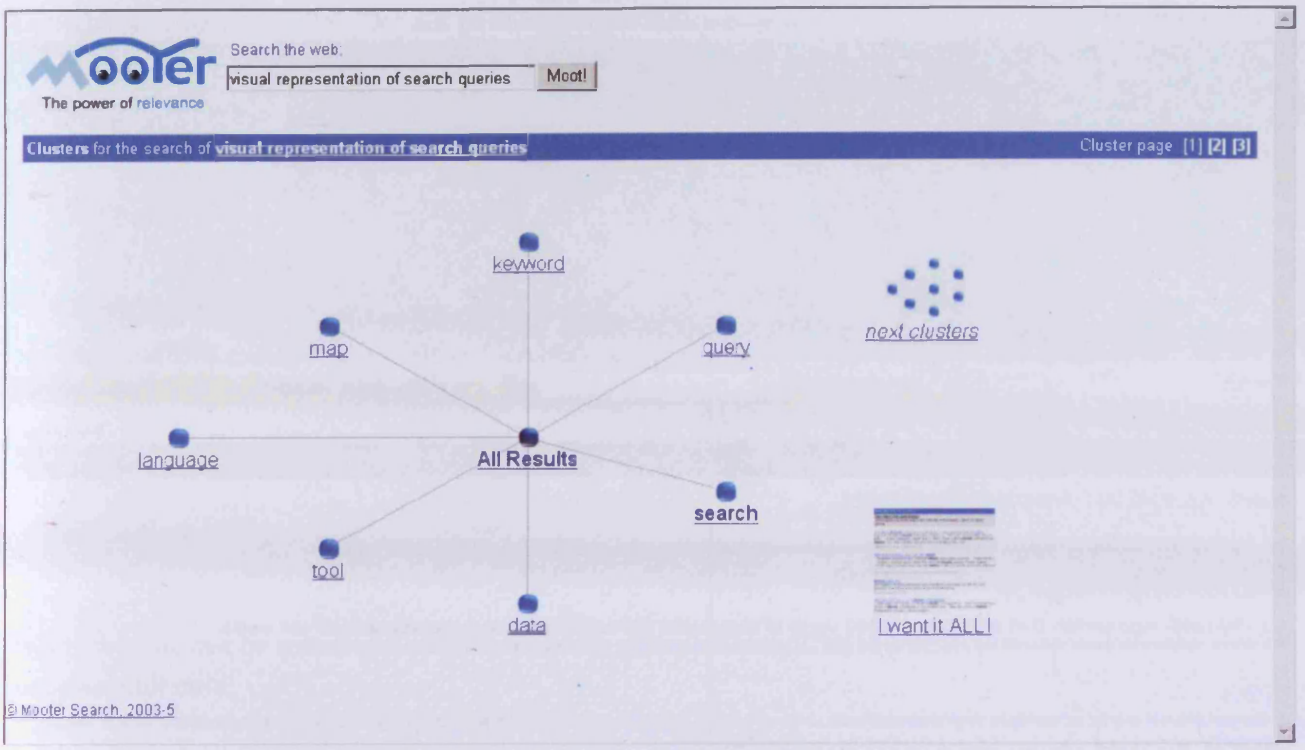


Clusty, as its name suggests, groups relevant links into 'clusters' enabling users to navigate results by subject area. The search engine Vivisimo (left) utilises a similar technique. The Shakespeare Search(ed) engine (top-left) is specific to the works of the playwright Shakespeare, and displays results in their context.

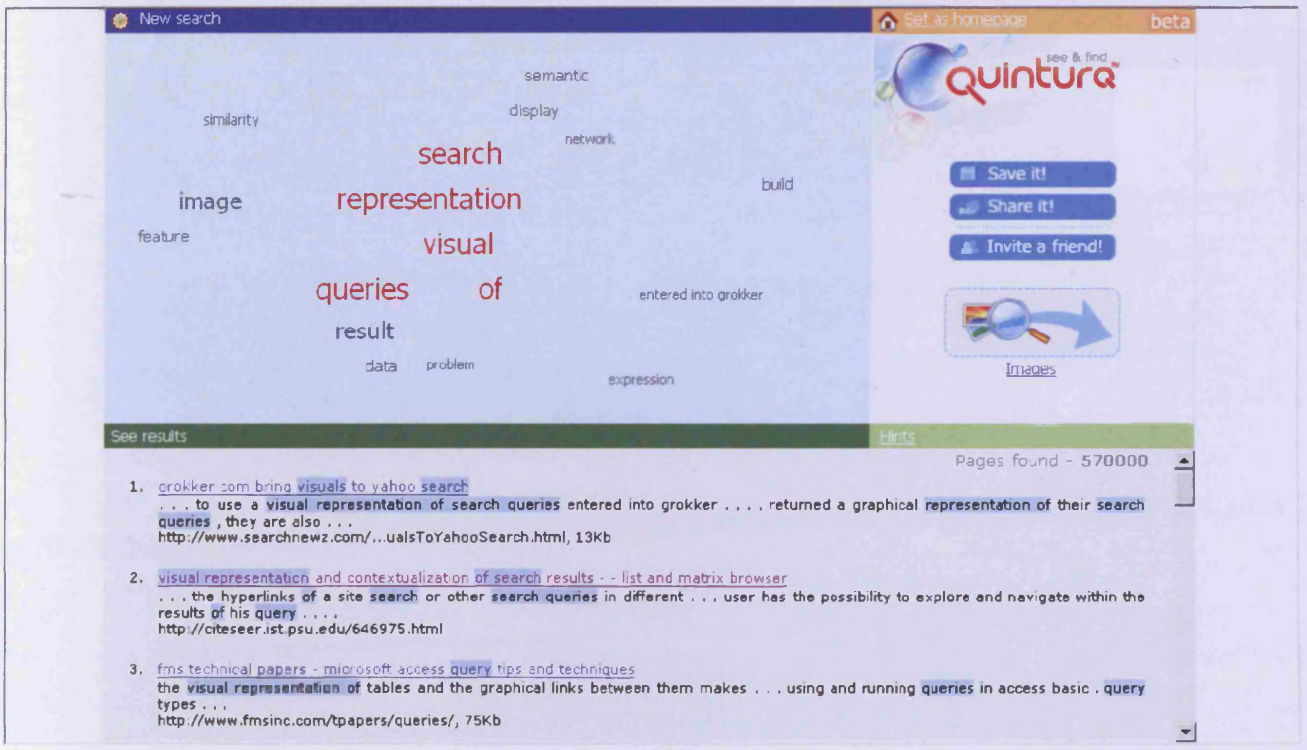


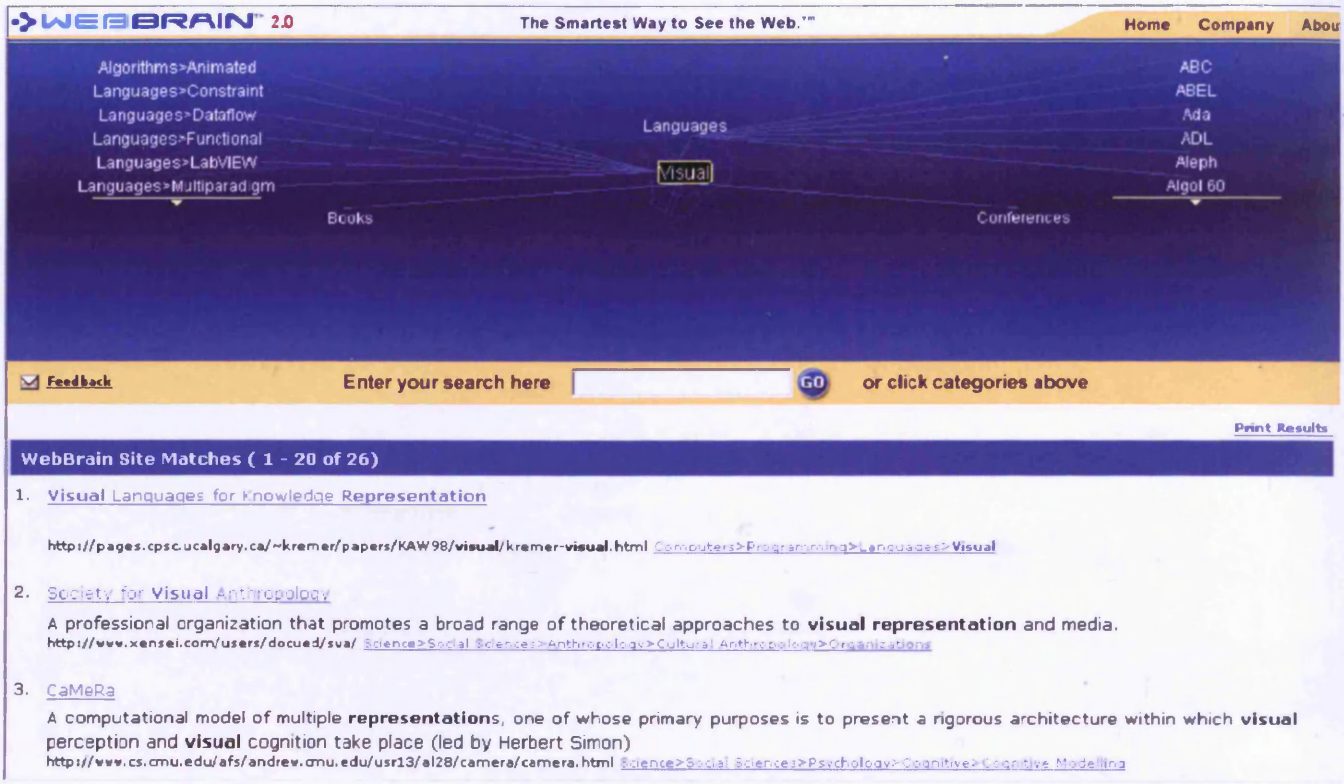


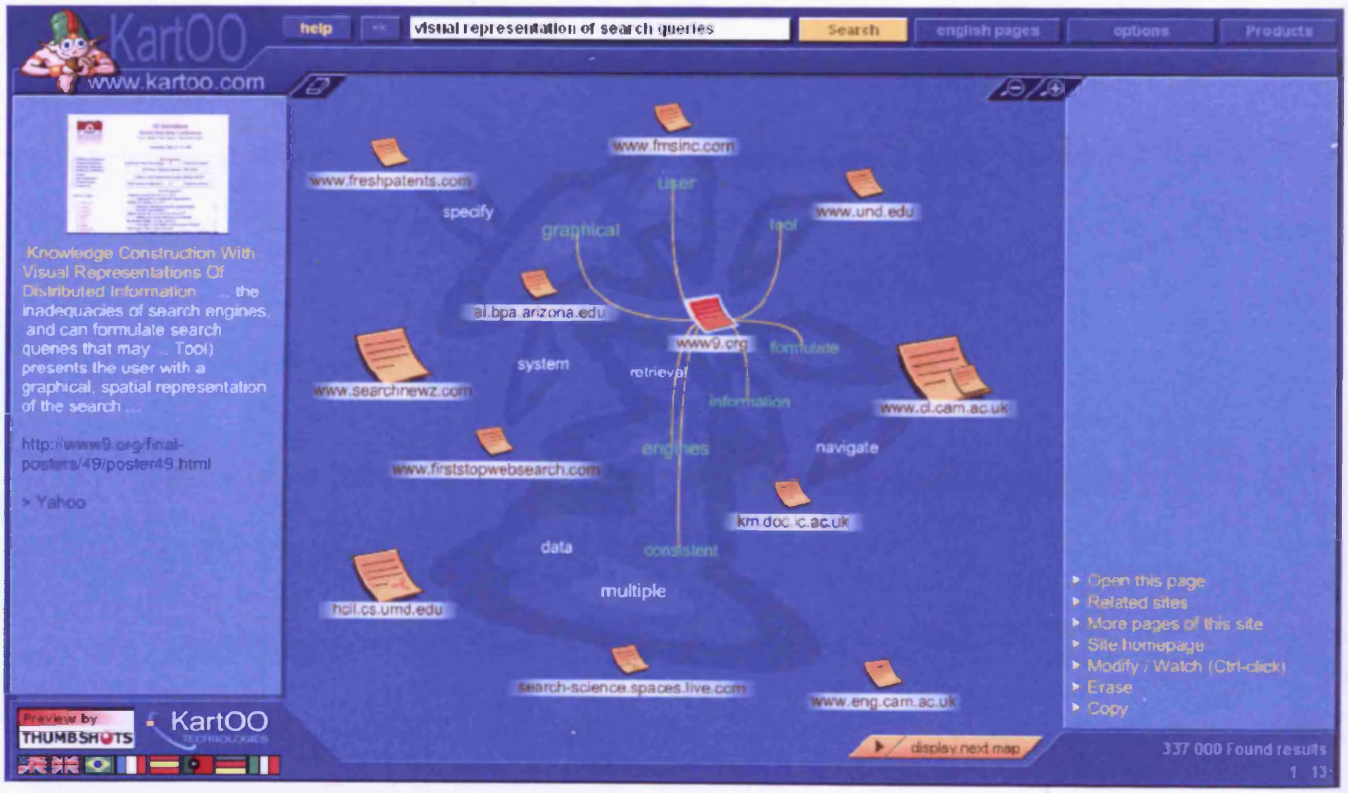
Grokker uses the text-based clustering technique seen earlier, and presents it in an 'outline view' (top). An alternate 'map view' (bottom) presents information in circular baskets, where each basket represents a cluster, whose results are displayed as lists on the right-hand side.



Mooter (above) is a cluster-based search engine that presents data as a 'node diagram'. Each keyword in the cluster map groups relevant links which are presented in a manner similar to the search engines Clusty or Vivisimo. Quintura (bottom) displays data as a 'text cloud' where semantically or contextually relevant keywords are displayed as part of the cloud. Text sizes and distances represent the strength of association with the current keyword selected. Words can be added or removed to update results.

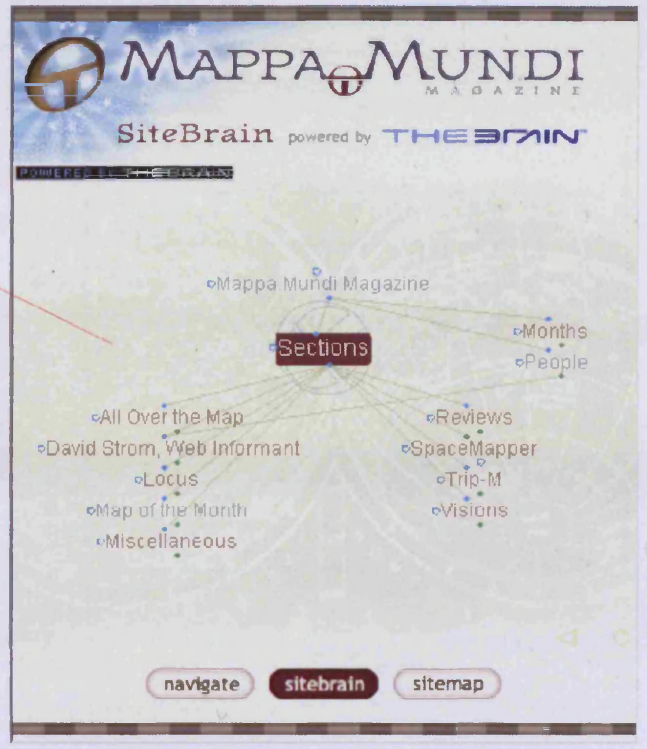



WEBBRAIN 2.0
The Smartest Way to See the Web.™
Home
Company
Abou
Algorithms>Animated
Languages>Constraint
Languages>Dataflow
Languages>Functional
Languages>LabVIEW
Languages>Multiparadigm
Languages
Visual
Books
Conferences
ABC
ABEL
Ada
ADL
Aleph
Algol 60
Feedback
Enter your search here
GO
or click categories above
Print Results
WebBrain Site Matches (1 - 20 of 26)
1. Visual Languages for Knowledge Representation
http://pages.cpsc.ucalgary.ca/~kremer/papers/KAW98/visual/kremer-visual.html Computers>Programming>Languages>Visual
2. Society for Visual Anthropology
A professional organization that promotes a broad range of theoretical approaches to visual representation and media.
http://www.xensei.com/users/docued/sva/ Science>Social Sciences>Anthropology>Cultural Anthropology>Organizations
3. CaMeRa
A computational model of multiple representations, one of whose primary purposes is to present a rigorous architecture within which visual perception and visual cognition take place (led by Herbert Simon)
http://www.cs.cmu.edu/afs/andrew.cmu.edu/usr13/al28/camera/camera.html Science>Social Sciences>Psychology>Cognitive>Cognitive Modelling


KartOO
www.kartoo.com
help
visual representation of search queries
Search
english pages
options
Products
Knowledge Construction With Visual Representations Of Distributed Information ... the inadequacies of search engines, and can formulate search queries that may ... Tool) presents the user with a graphical, spatial representation of the search ...
http://www9.org/final-posters/49/poster49.html
> Yahoo
www.frnsinc.com
www.freshpatents.com
user
specify
graphical
tool
www.und.edu
al.bpa.arizona.edu
www9.org
formulate
system
retrieval
information
www.searchnewz.com
www.cl.cam.ac.uk
www.firststopwebsearch.com
engines
navigate
km.doc.ic.ac.uk
data
consistent
multiple
hcil.cs.umd.edu
search-science.spaces.live.com
www.eng.cam.ac.uk
Open this page
Related sites
More pages of this site
Site homepage
Modify / Watch (Ctrl-click)
Erase
Copy
Preview by THUMBSHOTS
KartOO
display next map
337 000 Found results
1 13

The Brain search engine (left) refines its clustering technique by introducing hierarchy into its clusters. There are four categories:

a) Parent (top)
b) Sibling (right)
c) Jump (left)
d) Child (bottom)



This technique can be tested and seen in action at mappamundi.com.

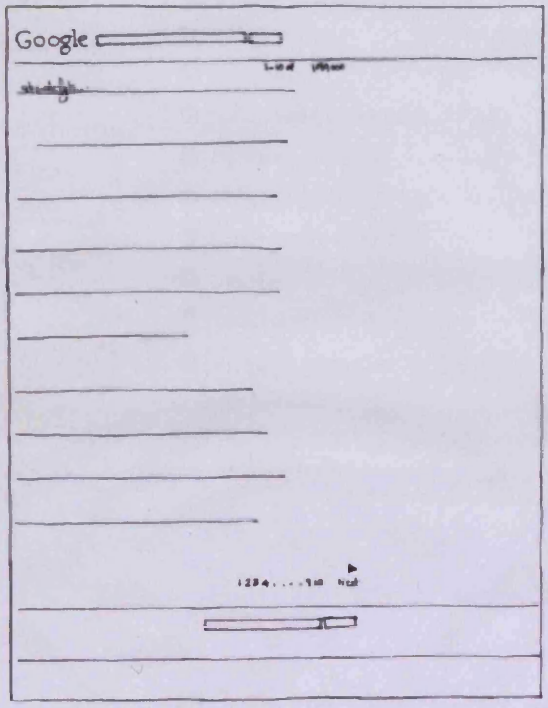
The search engine Kartoo displays hyperlinks to web pages in the form of a text-map that graphically mediates between Mooter and Quintura. Hyperlinks to relevant results are displayed as icons. Lines are drawn dynamically between links to show if web pages are associated with each other; an analogy similar to the clustering technique seen earlier. A thumbnail preview of the website is also shown on the top left-hand corner of the display.

4.2.2 Summary of data display techniques

The 'string of characters' approach suffers from two major setbacks or flaws: a) the linearity of presentation inherent in lists is not perceptually strong and does not relate to the user's model of the web[169], and b) with so many results being displayed, and 10 or 20 links displayed per web page (depending on the user's default preferences), few users choose to navigate to the next 'result page',

Goooooooooogle ▶
Result Page: 1 2 3 4 5 6 7 8 9 10 Next

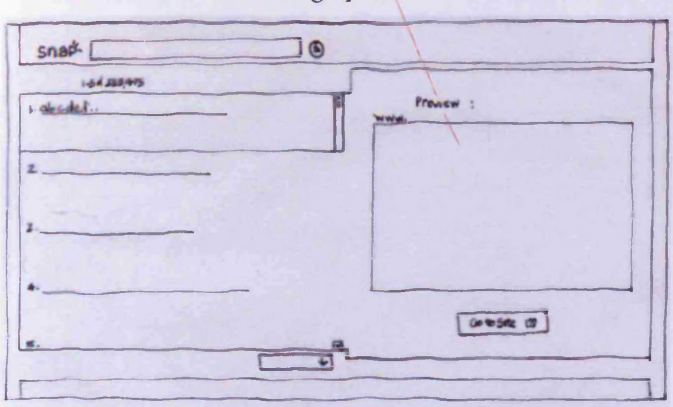
This defeats the purpose of a search tool, if most results are hidden from view due to restrictions in human computer interfaces or user interface design. Also, the implicit assumption that a user may want a complete set of relevant documents is not well suited for the web. "It is more important to guide the user to a small number of high quality documents than to achieve completeness."[170]



169 see 'Section 4.3: Related Work'

170 Nielsen, J., 'User Interface Directions for the Web', *Communications of the ACM*, Vol 42, No 1, January 1999, pp 65 - 72

Some tend to show previews to assist or even lure the user to jump to a website. In this case, a large preview is better, as it is more legible.

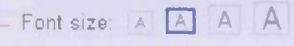


Options View: to see results as text only, text with thumbnails or thumbnails with even longer lists, are always welcome, but seldom noticed if too small to perceive. Issues of scale loom large on the limited screen space of a desktop computer. The lack of effective overviews inherent in lists does not support complex search tasks, where the user may require "gaining overviews of and exploring large sets of search results, identifying unusual documents, and understanding their context."[171]

171 Kules, Bill and Ben Schneiderman, Categorized graphical overviews for web search results: An exploratory study using U. S. government agencies as a meaningful and stable structure, *Proceedings of the Third Annual Workshop on HCI Research in MIS*, December 10 -11 2004, pp 1 - 5

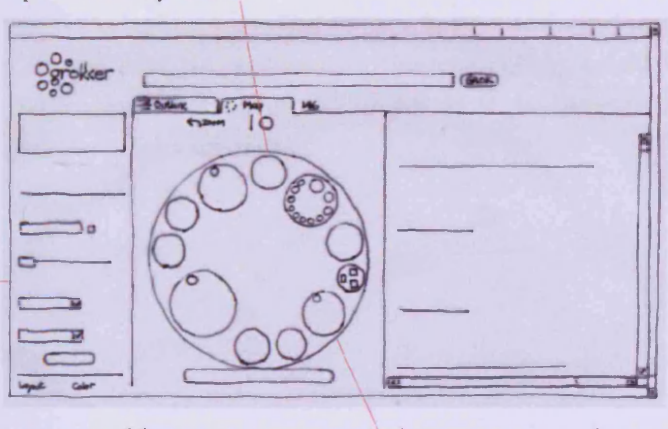
There are some who tend to group their results by similarity of subject area that a website or body of data may be referring to. Clustering, as this technique is known, can be quite useful when generic searches are performed, as they force the user to ask the question: "am I looking for 'x' in context of 'y'?." The clustering approach aims to bring context into the picture. Users can select from 'n' number of choices from a particular category. Recent work[172] on the presentation of text-based lists has shown that the clustering technique improves the speed of retrieval from text lists, and can support exploration and navigation in large sets of data.

Tools to search within search results and the ability to reduce or enlarge font sizes on screen are always welcome.

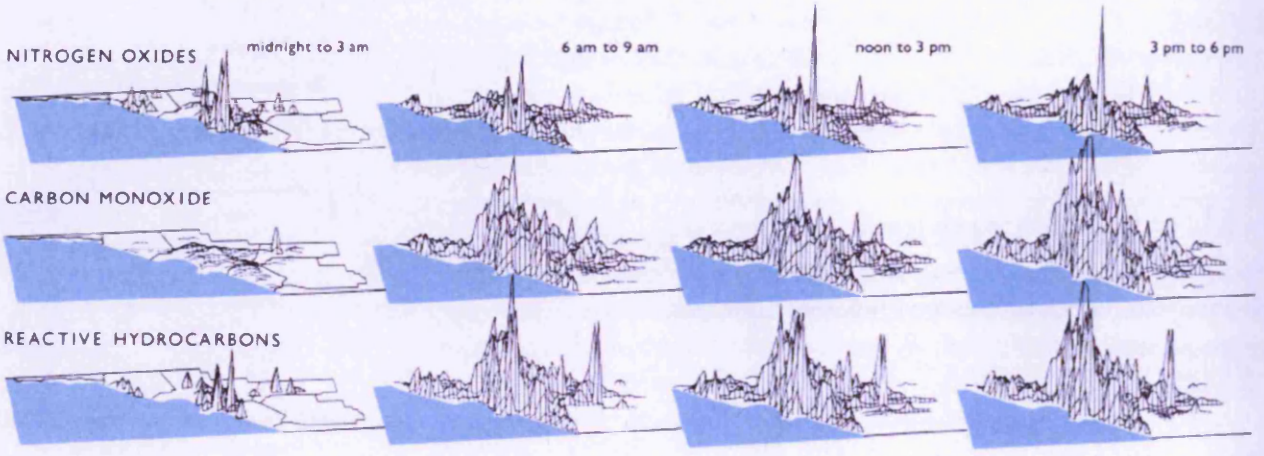


172 Recent work by Dumais, Cutrell and Chen (2001) has shown through comparative studies of seven *Category* versus *List* interfaces, that the former are always more effective than the latter.

Those who are tired of searching, can resort to 'grokking'. Two alternatives are presented here: an 'outline view', which is very similar to the clustering approach; and a 'map view' (below), where each cluster is represented by a circle.



'n' number of clusters = 'n' number of circles; all displayed within a 'global cluster' which is represented by a giant parent circle representing the search query. Options to email or export a map, narrow search results by date, source or domain, and even to change the colour or layout of the map are well intended. However, the constant zooming in and out ZOOM BACK leads to a loss in retention of the visual structure of the search results returned by the system in comparison to the simple clusters above.



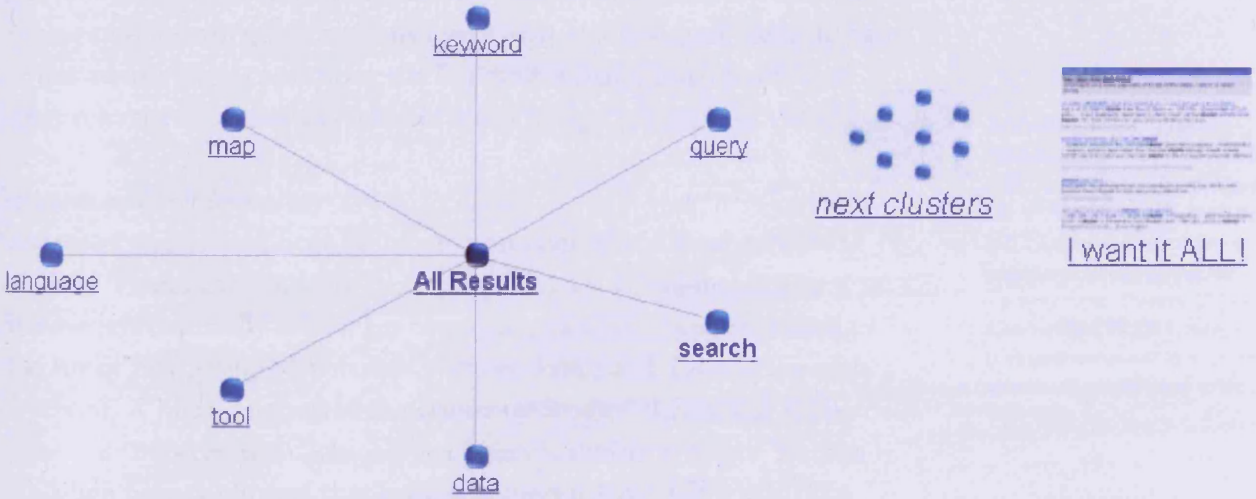
The display above shows the air pollution levels over southern California for four different times of the day. Once the viewer understands the pattern of multiples and variables in 'one slice of data, they have familiar access to data in all other slices. This constancy in design allows viewers to focus on changes in information rather than changes in graphical composition.' (Tufte, 1990)

One search engine represents the search query and other terms of relevance to that specific query, in a 'text cloud' form. The distances of these words from the original query, and their individual sizes represent the strength of association to the search query. Quintura uses the mouse rollover technique where the user can hover the mouse over a word to include it in their search and update results simultaneously.

An even more dynamically changing representation is TheBrain search engine. Here, the user is left to figure out the hierarchy intended by the creators!. The constant shifting of words based on the user's current selection, and the lack of a back button to convey a previous state leads to confusion. However, it does show associations dynamically, which is a powerful principle if employed appropriately. Grokker, Quintura and TheBrain search engines seek to provide some form of overview, but overlook a simple rule: a constancy in design allows the user to focus on changes in data (see left), and not the other way around. The Zoomable User Interface (ZUI) of Grokker, and the constant shifting of data in Quintura and TheBrain lead to a common problem with hypertext: disorientation.[173]

173 "Along with the power to organize information much more complexly comes the problem of having to know (1) where you are in the network and (2) how to get to some other place that you know (or think) exists in the network. Hypertext offers more degrees of freedom, more dimensions to move, and hence, a greater potential to become lost or disoriented." - Conklin, Jeff, 'Hypertext: An Introduction and Survey', *Computer*, September 1987, pp 17 - 41

And finally, the cumbersome 'death by a thousand clicks' approach: The text list is replaced by clusters which are rearranged into a 'node diagram'. Click on 'next clusters' and get the next cluster, or click on 'I want it ALL' and see the search results. Why not show the results apropos? Novelty here outweighs substance.



4.3 Related Work

The design of the navigation tool integrates established principles from several domains, most notably, human-computer interaction, user interface design, human factors and information seeking. Below is a review of the principle areas of concern to this case study, from these overlapping spheres of knowledge, which are summarized under the headings 'successive searching, hypertext and user behaviour, cognitive aspects, information seeking and direct manipulation'. Each of these sub-sections discusses literature that is pertinent to the design and display of data in the navigation tool.

successive searching

Three studies by Jansen and Spink are worth noting: 'Searching the Web: a survey of EXCITE users' (Spink, Bateman and Jansen, 1999), 'Information Seeking and Mediated Searching Study: Part 3: Successive Searching' (Spink *et al*, 2002), and the more recent 'How are we searching the World Wide Web? A comparison of nine search engine transaction logs' (Jansen and Spink, 2006). The former is a focussed study on users of the search engine Excite. The study found that "[web] users tend to employ simple search strategies, and often conduct more than one search [*ie*] successive searches over time to find information related to a particular topic."[174] This is in stark contrast to the fact that current Information Retrieval (IR) systems (library portals, for example) and search engines do not, by and large, support successive searches. Current search tools are built on the "single search paradigm, *ie* they are designed and operate on the assumption that every search is *unrelated* to a user's previous or future search."[175] These earlier findings were later supported by their work in 2002 on successive searching. The latter and more recent study by Jansen and Spink is a comparative study of five U.S. and European search engines - Excite, Fireball, AlltheWeb, BWIE and AltaVista - over a seven year period. Their findings confirm that the number of one term queries remains very high and few users navigate past the first results page; confirming the hypothesis that a small number of highly relevant documents is critical from a navigation point of view.[176]

174 Spink, Amanda, Judy Bateman and Bernard J. Jansen, 'Searching the Web: a survey of EXCITE users', *Internet Research: Electronic Networking Applications and Policy*, Vol 9, No 2, 1999, pp 117 - 128, text in brackets...

175 ibid, p 118, emphasis deliberate

176 According to the Graphic, Visualization and Usability Centre (GVU) at Georgia Tech University, the 8th WWW User Survey in 1997 found that the three most important issues facing the internet were: Privacy (30.49%), Censorship (24.18%) and Navigation (16.65%). Concerns about navigation tend to increase with age.

hypertext and user behaviour

Numerous studies have been conducted on Hypertext. Those reviewed here are: 'Hypertext: An Introduction and Survey' (Conklin, 1987), 'The matters that really matter for hypertext usability' (Nielsen, 1989), 'The Art of Navigating Hypertext' (Nielsen, 1990) and 'Interacting with Hypertext: A Meta-Analysis of Experimental Studies' (Chen and Rada, 1996) and 'Between the clicks: Skilled Users Scanning of Pages. Studies by Nielsen have confirmed that hypertext users behave differently than

when reading printed texts. Experienced hypertext users tend to leave their book habits behind. Users tend to scan text, rather than read it sequentially as in a book or newspaper. Nygren's work suggests that spatial grouping is a cue used by skilled searchers and "highlighting, informative patterns and spatial constant positions are design principles"[177] that can be employed in the design of web pages, to optimize average search time, rather than adopt a layout that optimizes legibility for the occasional user.

177 Nygren, Else, 'Between the clicks: Skilled Users Scanning of Pages', *Proceedings of Designing for the Web: Empirical studies*, October 1996

cognitive aspects

A comprehensive three-part study of information seeking and retrieving has been conducted by Saracevic and Kantor (1988). The study uses a framework of five variable classes: users, questions, searchers, searches and items retrieved. Three practical implications of their findings are as follows: a) "The context of a question is confirmed to be important. This suggests that it is important for searchers (or *interfaces*) to explore the background of a question and get as much information as possible about the problem at hand and the intent in use of information. b) Skills in word association and a preference for abstract thinking appeared to be important abilities in searchers with high search performance. This suggests that cultivating semantic association, be it in the language in general or in a subject in particular, seems to be a profitable enterprise for searchers. c) Cycles in searching showed a significant impact on outcome. This suggests that it may be quite important to view and review intermediate results as the search progresses and adjust the search accordingly."[178]

178 Saracevic, Tefko and Paul Kantor, 'A Study of Information Seeking and Retrieving: III: Searchers, Searches, and Overlap', *Journal of the American Society for Information Science*, Vol 39, No 3, 1988, pp 197 - 216, emphasis deliberate

information seeking

The beginning of this chapter states that the explicit goal or aim of this project is not to create a search engine or search tool per se, but to create an interactive learning tool. This view is consistent with a framework where information seeking is a fundamental human process closely related to learning and problem solving.[179] Marchionini (1995) defines information seeking as "a process in which humans purposefully engage in order to change their state of knowledge. The term *information seeking* is preferred to *information retrieval* because it is more human oriented and open ended. Retrieval implies that the object must have been 'known' at some point; most often, those people who 'knew' it organised it for later 'knowing' by themselves or someone else. Seeking connotes the process of acquiring knowledge; it is more problem oriented as the solution may or may not be found. For example, seeking spiritual enlightenment makes sense, but retrieving enlightenment does not."[180] Information seeking is fundamentally both human-oriented and an interactive process.

179 Marchionini, Gary, *Information Seeking in Electronic Environments*, Cambridge: Cambridge University Press, 1995, pp 5, 6

180 ibid, p 6, emphasis in original

direct manipulation
Direct Manipulation is a concept introduced by Schneiderman (1983). It literally refers to the direct manipulation of objects of interest. One example of direct manipulation is driving an automobile or car, where the scene is directly visible through the (front) window(s), and the consequence of actions such as braking, accelerating or steering is common knowledge. "To turn left, the driver simply rotates the steering wheel to the left. The response is immediate and the scene changes, providing feedback to refine the turn. Imagine trying to turn by issuing a command LEFT 30 DEGREES and then another command to see the new scene; but that is the level of operation of many office-automation tools of today!"[181] Another example is using the mouse to slide a cursor on screen and see it move in the same direction and at a proportional rate. Direct Manipulation is discussed here as it is "one of the key components to highly interactive, advanced information-seeking systems."[182]

181 Schneiderman, Ben, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Reading, Massachussets: Addison-Wesley, 1998, p 187

182 Marchionini, Gary, *Information Seeking in Electronic Environments*, Cambridge: Cambridge University Press, 1995, p 43

4.3.1 "Rules for a Journey of Mind and Eye"[183]

To summarize the design principles commonly ignored by search engines in data display, that directly impact the information seeking process:
i) overviews are important, as context is necessary for comparison and deductive reasoning
ii) focus on changes in data rather than changes in design
iii) users tend to refine their search query over time and the interface should provide mechanisms to support successive searching
iv) spatial positioning can be an effective cue in assisting search tasks
v) information seeking is fundamentally an interactive process and interaction should be maximised in user interface design.

The above heuristics can be further reduced to three rules:

a) Never quote data out of context *ie* always attempt to show context.
b) Constancy in design allows the user to focus on changes in data, rather than changes in design.
c) Always try and show cause and effect.

183 Morrison, Philip, and Phylis Morrison and, The Office of Charles and Ray Eames, *Powers of Ten: About the Relative Size of Things in the Universe*, New York: Scientific American Books, 1982

4.4 The Navigation Tool

In Chapter 2, the nature of information is discussed, where it is distinguished from data and knowledge. Raw material 'data' become information to someone at some point in time within a specific context. Information assimilated over time becomes knowledge. The different states of this entity which we call information need active representation in a system designed for information seeking. These cognitive states in the mind manifest themselves as three physical states in the proposed system - an Observer, Participant and User state - as data crystallizes into knowledge through its intermediate stage as information.

The Participant state is the most important and active state implemented in the system proposed here. In this state, as will be seen later, the user becomes a participant in the process of information seeking; the user becomes information. Consider this analogy to horticulture: the User mode is where the soil is prepared for seeds to be planted; the Participant mode involves planting the seed and nurturing it so that it can germinate; and the Observer mode is where cross pollination takes place. From a communication theory or information retrieval viewpoint, the User state is where the gathering, assimilation, processing and storage of information takes place. The Participant state involves data display and interaction. The Observer state is where data interact with each other spontaneously to establish relevant relationships, which can then refine the index file and hence the display of results in Participant mode. Both the User state and the Observer state are vast and deeply engaging subjects within the realm of Computer Science and are not implemented or discussed here, as the focus is on the intermediate 'Participant' stage of information seeking using design principles and techniques based on information design.

Information Seeking is a fundamentally complex, interactive process, involving search, browse and serendipity, where the "[users] conceptualization of the high-level information need is imperfect and evolving, their understanding of the relevant concepts and terminology is limited and their understanding of the presentation and interactions available in the interface is incomplete."[184] This highly fluid state is thus intended to be supported by a navigation tool entitled an Observer Participant User System (OPUS); the primary aim of which is to support information seeking by integrating searching and browsing.

184 Kules, Bill and Ben Schneiderman, Categorized graphical overviews for web search results: An exploratory study using U. S. government agencies as a meaningful and stable structure, *Proceedings of the Third Annual Workshop on HCI Research in MIS*, December 10 -11 2004, pp 1 - 5

4.5 Stages in the development of OPUS

The following section describes the evolution of the navigation tool. Eight distinct stages have been identified which represent significant steps in the design, development and scripting of the software application.

naming conventions:
For technical consistency and ease of comprehension, the names of User Interface (UI) elements used in the source code of the application will be used to describe them henceforth. These are as follows:

UI Element	Name
Label	*Opusdikt*
Text field	*resultsPane*
Text field	*windowPane*
Text field	*inputA*
Text field	*inputB*
Button	*Observer button*
Button	*Participant button*
Button	*User button*
RadioButton	*Occurrence radio button*
RadioButton	*Context radio button*
RadioButton	*Symantec radio button*
Button	*Add+ button*
Window	*pop-up Window*

A description of these UI elements is given in the next section. A few frequently occurring terms also need to be declared: *ActionScript* refers to the native scripting or programming language of Macromedia Flash MX, using which the software is authored, and is used here both as verb and noun. *MovieClip* refers to an object authored in Flash MX, either manually, or at run-time. *Word, term and concept*, used sometimes simultaneously and often interchangeably, refer to the actual words in a sentence, the query that a user may be searching for, or even a query that the user may have in mind. Indirectly, they all refer to the search expression. *Screen* refers to the actual desktop resolution on the monitor of a computer, which has been chosen as 1024 x 768 pixels.[185] It also refers to the total space available for 'laying out' the user interface. *Stage* refers to the virtual paper-space on which actions are performed by the user in Participant mode.

185 Section 4.5.2

4.5.1 The Data-set

A primary requirement for this navigation tool is a data-set that is reasonably large, specialist in nature, and valid in terms of its content. The data-set chosen for this particular task is an online reference website for the 'Masters in Environmental Design of Buildings' programme at the Welsh School of Architecture. The website holds course material for six modules taught on the programme, namely, Skins and Spaces, Ambient Environment, Passive Design, Framework for Sustainable Design, Efficient Building Services and Investigations in the Built Environment. The specific module chosen to further narrow the data-set is 'Skins and Spaces'. The following characteristics[186] of the data-set make it suitable for this case study:

topicality: the domain represented by the Skins and Spaces website relates to Environmental Design.

aim: the website is a primary source of information for students on the MSc course.

data type: the data on the website is in text, graphic and numerical form, with some animated illustrations.

quantity: the Skins and Spaces website consists of fifteen html web pages, which is equivalent to approximately ninety-six A4 printed pages.

quality: as a teaching and learning resource, the MSc online website is produced by staff at the Welsh School of Architecture, who are experts in their respective fields, and is updated regularly as required. The data is thus valid and up-to-date.

granularity or specificity: the data-set is very specific in nature as it discusses ventilation design in modern buildings.

186 The sub-headings that describe the characteristics of the Skins and Spaces website are adapted from Marchionini (1995), where these characteristics are used define the database of a search system.

pages are displayed here

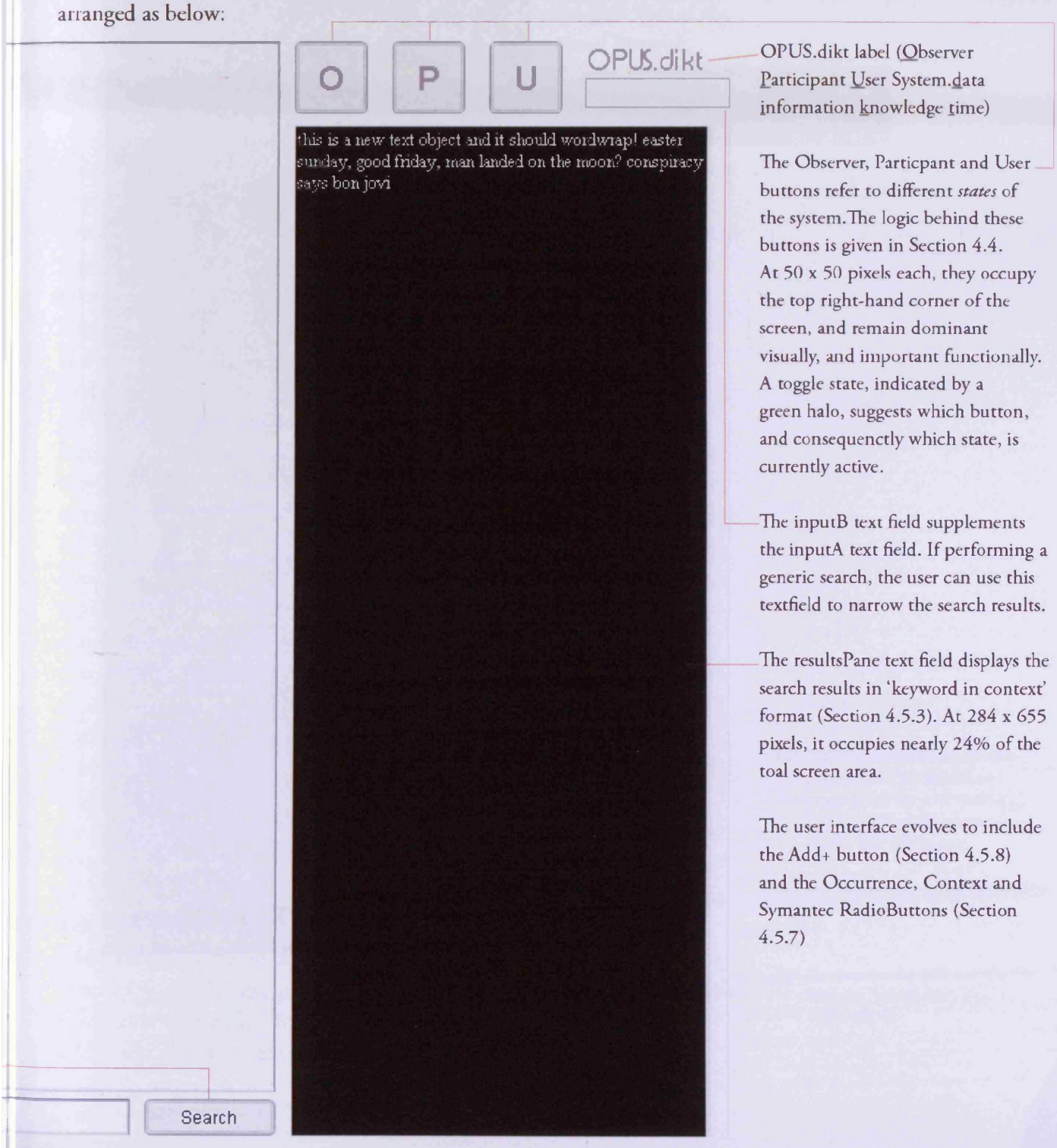
The main space of the application is occupied by the windowPane text field in User mode. The same area is occupied by an overlapping metaphor movieclip in Participant mode (Section 4.5.4). At 692 x 712 pixels, the windowPane text field occupies nearly 63% of total screen space. The spatial requirements of displaying web pages required maximizing the size of the windowPane text field, hence occupying nearly two-thirds of the screen.

The inputA textfield is the main Search field for information retrieval. The user can enter a term in the text field and trigger an action by pressing the Search button.

4.5.2 Stage 1: Laying out the User Interface

This involved considerations for the size of the stage, which in turn is derived from the screen or monitor size. The size of the screen is chosen as the current XGA (eXtended Graphics Array)[173] standard at the time of writing, which is 1024 x 768 pixels (for monitors 14" or 15" in size). This resolution offers sufficient density to arrange user interface elements for this task. The UI elements required are: Buttons (for actions), Text Fields (for input and display), Text (for title), RadioButtons (for preference selection) and a pop-up Window (for further data display). These are arranged as below:

187 XGA or Extended Graphics Array is a high resolution graphics standard introduced by IBM in 1990. The most common resolution at the time of writing is 1024 x 768 pixels.



OPUS.dikt label (Observer Participant User System.data information knowledge time)

The Observer, Particpant and User buttons refer to different *states* of the system. The logic behind these buttons is given in Section 4.4. At 50 x 50 pixels each, they occupy the top right-hand corner of the screen, and remain dominant visually, and important functionally. A toggle state, indicated by a green halo, suggests which button, and consequenctly which state, is currently active.

The inputB text field supplements the inputA text field. If performing a generic search, the user can use this textfield to narrow the search results.

The resultsPane text field displays the search results in 'keyword in context' format (Section 4.5.3). At 284 x 655 pixels, it occupies nearly 24% of the toal screen area.

The user interface evolves to include the Add+ button (Section 4.5.8) and the Occurrence, Context and Symantec RadioButtons (Section 4.5.7)

O P U OPUS.dikt

search results: "ventilation"

ventilation | process | supplying | outdoor | air | space | without | mechanical | assistance

controlled | intentionally | provided | openings | windows | doors | non -powered | ventilators

openings | level | ventilation | addition | infiltration

infiltration | sufficient | outdoor | air | ventilate | building

infiltration | component | ventilation | excessive | waste | heat | incur | high | energy | penalty | building | difficult | heat | cool | comfort | levels

mechanical | ventilation | movement | air | mechanical | means | space

'build | tight | ventilate | right'

infiltration | naturally ventilated | mechanically |ventilated | spaces

'best | practice' | minimise | infiltration | applying | sealing | measures | controllable | natural | mechanical ventilation

ventilation | effectiveness | fraction | fresh | air | delivered | space | reaches | occupied | zone

ventilation | efficiency | term | quantify | ventilation effectiveness

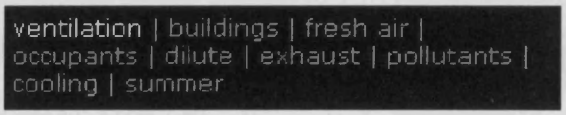
measure | ability | ventilation | system | exhaust | pollutants | generated | space

oldest | air | closest | extract | maximum ventilation effectiveness

naturally | ventilated | buildings | air | well

4.5.3 Stage 2: Deploying search results; principle at play: Context and Comparison

The resultsPane text field displays the search results. The approach devised is to present the keywords in a sentence as a grouped entity, where each keyword is separated by the '|' operator. Hence, the sentence 'Ventilation in buildings is needed to provide fresh air for occupants to dilute and exhaust pollutants and to provide cooling in summer' is presented in its *keyword in context* format as 'ventilation | buildings | fresh air | occupants | dilute | exhaust | pollutants | cooling | summer'. The term that the user is searching for is highlighted, with other words in a muted grey against the black background. Thus. the sentence above takes the visual form of,



This approach is termed here as 'keyword in context'. The most frequently occurring lemmas are removed from the sentence. In this case, the lemmas not required are: *in, is, to, far, and.* As discussed earlier, this edited form of hypertext is more suited for reading online. The preservation and presentation of (key)words within a sentence is essential to retain and display context to the user, who must then decide whether the instance of the keyword detected is the search result anticipated based on the terms surrounding it, *ie* its context. Also intended is the notion that the user may learn new facts from studying the context within which a particular term occurs by discovering hidden relationships that may have been unknown to the user hitherto. Finally, the presentation of multiple instances of keywords in their context (left) forces the user to compare the different instances or study the contexts within which a specific keyword, term or concept occurs, within an eye span. As seen in Chapter 1, colour is preattentive, allowing the user to pick-up highlighted terms in white, against a black background, with other terms in a muted grey.

ActionScript:

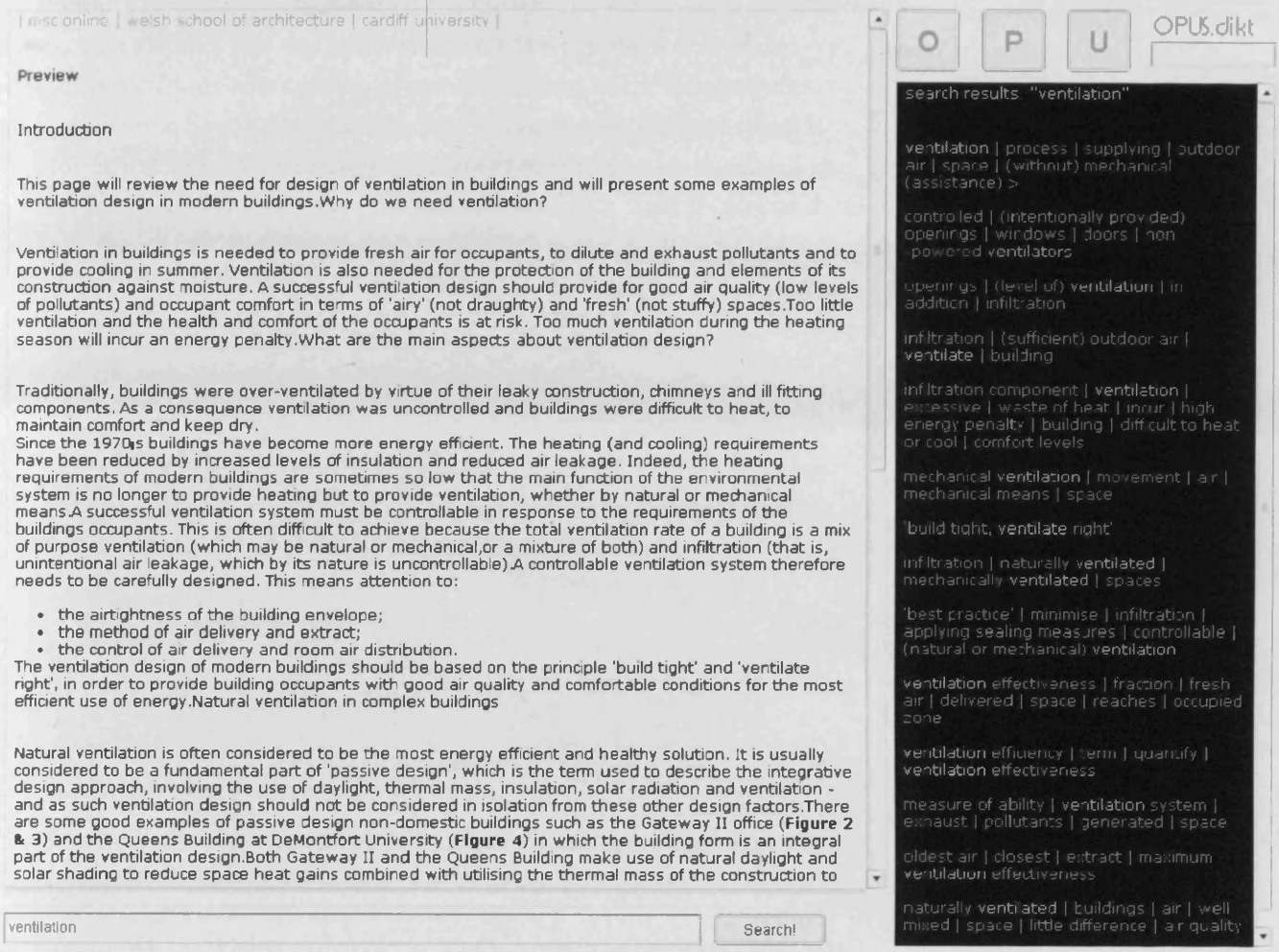
The resultsPane text field loads two files for each corresponding search: a .txt file and its corresponding .css file. The .txt file is written in html[188] language and contains the main content to be displayed. The .css file is the Cascading Style Sheet[189] which formats the text for it to be displayed using the appropriate font type, font size, font colour and spacing. As a test, the occurrences of the term 'ventilation' were extracted from a small section of the Skins and Spaces website. These results were written in a 'ventilation.txt' file in html format, and the corresponding stylesheet 'ventilation.css' was also prepared.[190] The user can now search for the term 'ventilation' and see results displayed in the left-hand resultsPane text field. Later, this data-set is extended to include the terms {air, natural,

188 HTML or HyperText Markup Language is 'the publishing language of the web' (w3c.org), and is used by web browsers to display text and images.

189 Cascading Style Sheets are used to format web documents.

190 To view the .txt and .css files for the term 'ventilation', refer to Appendix B.

pollutant, space}. As a test, the user can click on the term 'ventilation' in the first sentence of the search results, and see content loaded into the windowPane text field on the left-hand side.



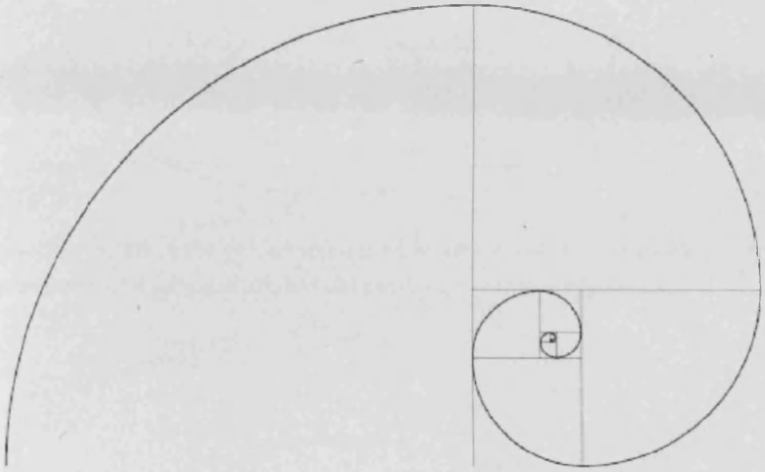
4.5.4 Stage 3: Investigations in Form, activating the Participant button, introducing the Proximity Principle, and exploring Cause and Effect.

At this stage, theoretical aspects explored earlier on probability (frequency of occurrences) and entropy (rearrangement of terms) are applied to test their validity. Chapter 2 discussed probability in language and how a statistical study of the English language reveals the most frequently occurring lemmas[191] within the corpus of the language. Related to this frequency of terms in a particular language are mathematical processes underlying the grammar of a language. A mathematical process in which a system produces a sequence of symbols according to some probability (for example, a sequence of characters or letters in a word, or words in a sentence) is called a stochastic process. "A special case of the stochastic

191 Chapter 2: Section 2.3

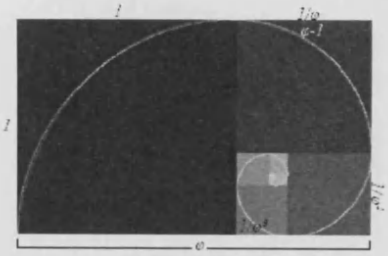
process in which the probabilities depend on the previous events is called a Markoff process."[192] The syntax of a language governs the choice of words in a phrase or sentence, and is thus a markoff process in nature. In response to the syntactic form of a language, the search for a physical form that emulates this stochastic character at a topological level, a distant but curious relationship is found in nature in the Equiangular or Logarithmic Spiral, for here too, the terminal growth of the spiral is in proportion to a previous event. The form of this spiral is emulated in several plants and animals, for example, the shell of a snail (Nautilus shell) and the florets of a sunflower.

192 Shannon, Claude E. and Warren Weaver, *The Mathematical Theory of Communication*, Urbana: The University of Illinois Press, 1962, p 102

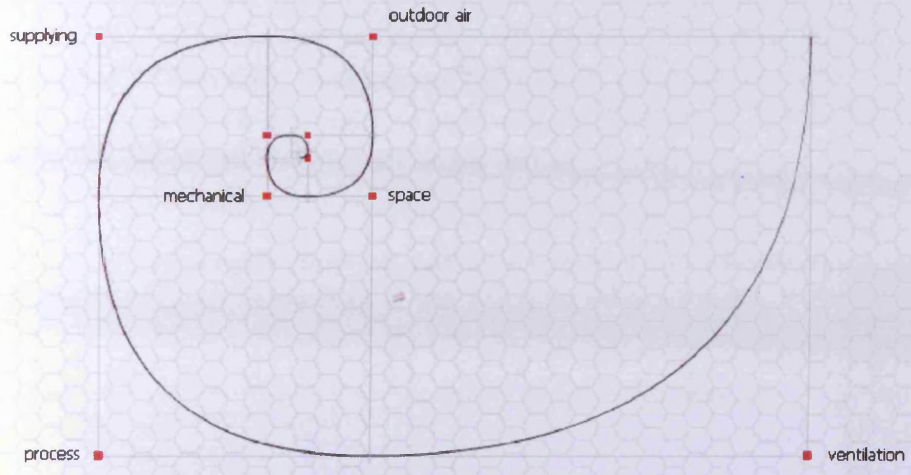


The Equiangular Spiral grows in length and width in the same unvarying proportions, thus keeping its overall form unchanged - a property unique to this spiral.

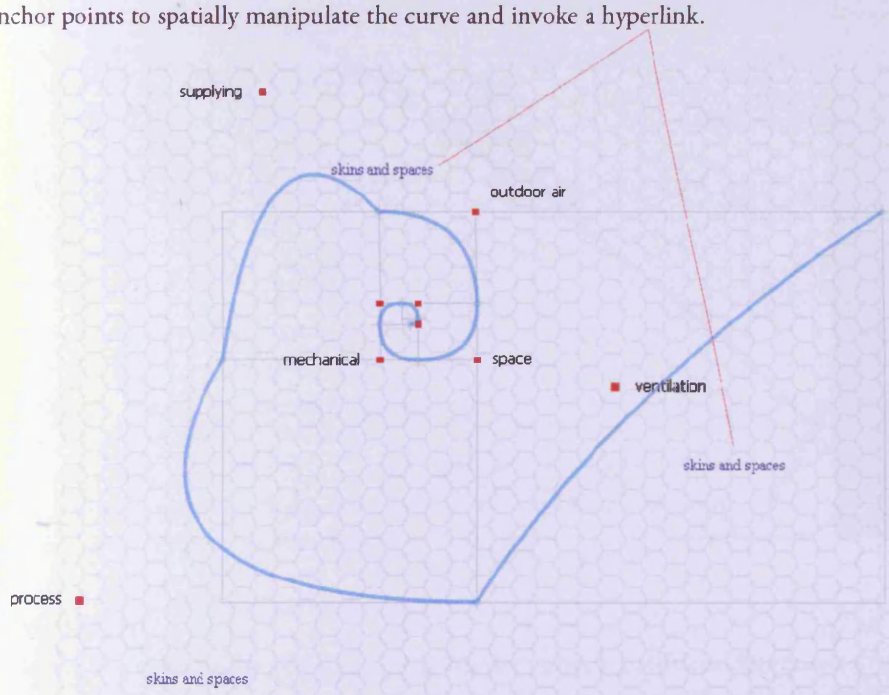
In its form, the Logarithmic Spiral is closely related to the Golden Spiral, which can be derived from the Golden Rectangle. This Rectangle in turn, is derived from the Golden Section, where each subsequent division is in the order of *phi* (Φ), where $\Phi = (1 + \sqrt{5})/2 = 1.6108339887...$ Though the analogy here between the Markoff process in language and the equiangular spiral is deliberate, the underlying forces at play which govern the form of both: syntactic in the former and physical growth in the latter are the same - broadly stochastic in nature. The intention here is to spatially represent a sentence through the geometry of the spiral and apply the principle of direct manipulation seen earlier. Each quarter-arc tangent to the interior of each square represents one division or one unit of a sentence in diminishing order of occurrence. The spiral becomes a device for spatial manipulation.



The anchor point of each arc is used a) as a 'hinge' for any term or an item of data to be attached to, and b) to spatially manipulate the curve. Dragging and dropping an anchor point displays a 'skins and spaces' hyperlink, which, upon clicking, loads the Skins and Spaces website.



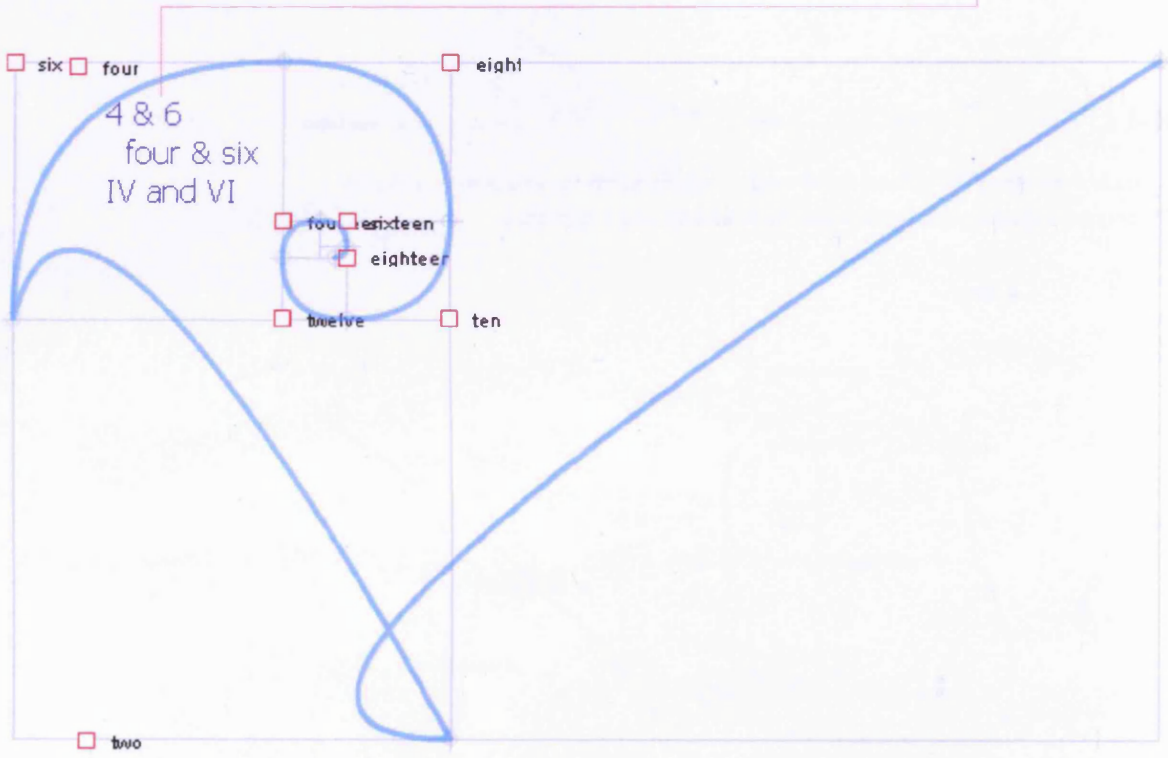
above: anchor points of the golden spiral used to attach words to; and below: using the anchor points to spatially manipulate the curve and invoke a hyperlink.



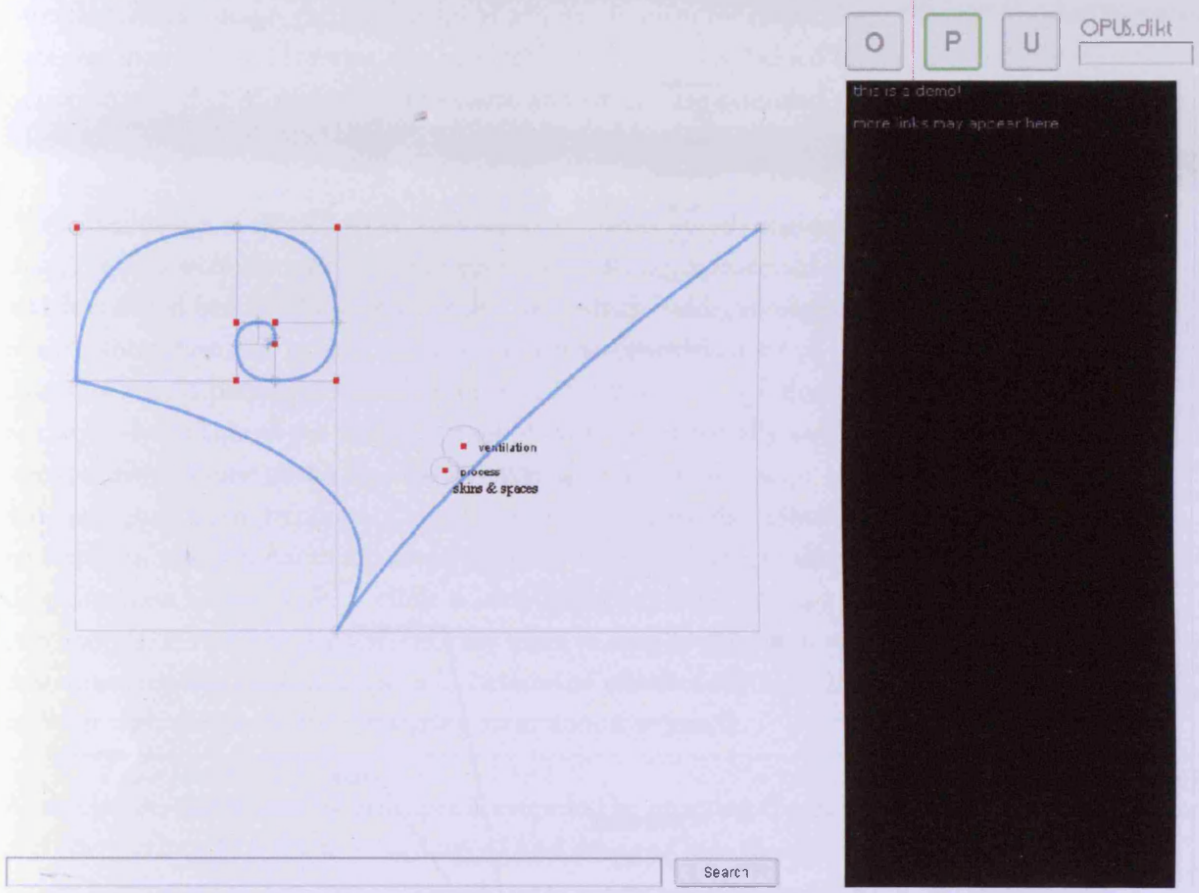
The Cause and Effect principle seen earlier also appears in user interface design as a 'heuristic' *ie* a general design principle or rule of thumb: "The system should always keep users informed about what is going on, through appropriate feedback within reasonable time."[193] At the stage of the application described so far, dragging and dropping any of the words triggers an event *ie*, loading the 'skins and spaces' hyperlink. This idea is extended further by introducing the *proximity principle*. The logic is simple: any two words can be brought in close range of each other, and

193 The first rule of thumb - *visibility of system status* - in Nielsen J., 'Ten Usability Heuristics', *www.useit.com*, Accessed <March 11, 2007>

Using the *hitTest* method in actionscript, the collision of movieclips is detected. This event displays active hyperlinks, as in the example below, where moving the anchor point 'four' next to 'six' displays



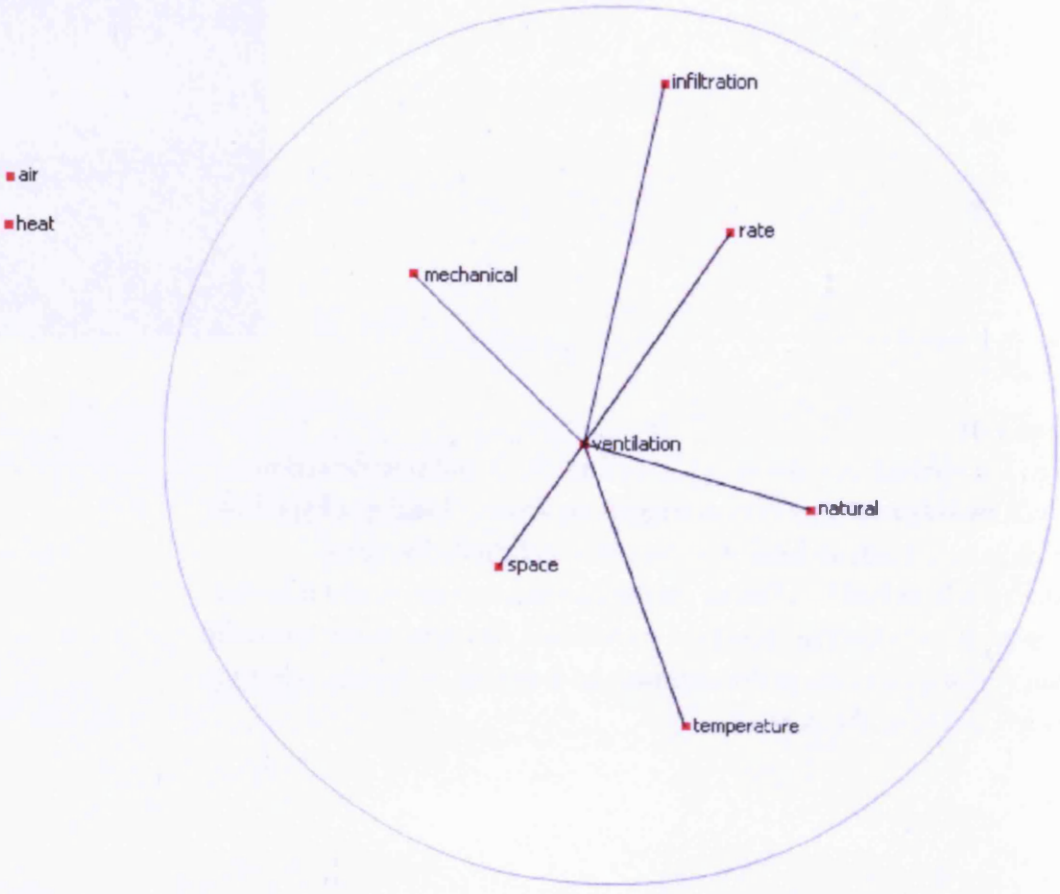
this event triggers an action. As a test, in the next version of the application (opus8, below), when the terms 'ventilation' and 'process' are brought in close proximity to each other, a hyperlink is displayed, with a corresponding message in the resultsPane text field. Moving the words away from each other removes the hyperlink. As a final test, to the anchor points of the curve are attached numbers in word form (for example, *two* instead of *2*). This time, bringing any two pairs of numbers - two and four, four and six, six and eight, or eight and ten - displays 2, 3, 1 and 7 links respectively; all of which are active (left).



ActionScript:

The *hitTest* method in actionscript is used to check collision detection between two objects. This in turn triggers an action - loading a hyperlink - which in turn loads an html web page when clicked. The spiral movieclip is launched by clicking on the Participant button and is loaded on top of the windowPane text field seen earlier. The area of the movieclip within which actions are performed such as dragging, dropping, selecting, etc is referred to as the *stage*.


ventilation
air
heat
infiltration
natural
occupant
pollutant
space
temperature


air
heat
infiltration
rate
mechanical
ventilation
natural
space
temperature

4.5.5 Stage 4: Probability and Entropy

The explorations in the previous stage were about imposing 'a form on a formless entity.' Although the diminishing curve of the golden spiral is self restricting (as the number of useful anchor points is nine, before the curve diminishes out of view), it graphically illustrates the limitations of imposing a form, be it any, on a set of data. As discussed in Chapter 1, the mapping of data variables to spatial variables can often be arbitrary and bear no direct correlation with the data-set in question or the task intended. At this stage, the golden spiral as a mechanism for displaying data was abandoned. However, the principles that the curve helped to demonstrate - that of proximity and cause and effect - are extended further.

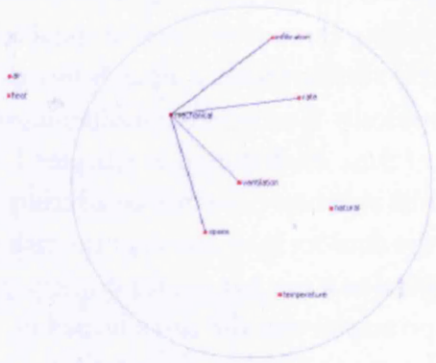
At the beginning of this Chapter, strategies by current search engines to mine for data were discussed. Common to them all is preprocessed data which is stored in a database or an 'index file', which holds, amongst other information, the statistical count of each term within a set of data. The user is presented with a set of nine[194] terms (left, top) docked to the left-hand side of the stage. This set of terms is statistically and semantically relevant to the Skins and Spaces data-set. At this stage of the software application, freedom of choice to (re)arrange words, terms or concepts on screen is handed back to the user. The user is free to drag and drop any term on the stage. A circle is introduced or placed on stage for two reasons: a) to serve as a reference for users to drag or drop their word *into*, thus enabling ActionScript to b) determine whether any term has collided with the circle, thus initiating an action, if required.

At this point, the proximity principle is extended by enacting the cause and effect principle *ie* if a term is dragged and dropped into the circle, an action is triggered in response. If a single term is dragged into the circle, the search results are displayed on the left-hand side. If two or more terms are dragged into the circle, apart from updating and displaying the results, the action is to draw a line dynamically, between the (two or more) terms dragged into the circle. This line is drawn from one term to another only if the other term is within range *ie* within a specified distance from the currently selected keyword. This distance is specified as 250 pixels which is the radius of the circle. This line is dynamically regenerated or redrawn when the user changes the position of any term within the circle, and is not drawn if a term is outside the circle even if it is within range. Also, as shown, lines are drawn from a specific word (known here as the 'keyword') to all other words within the circle. The user can indicate his or her preference for another keyword by using the middle mouse button. As will be seen in the next stage, this keyword specifies the main item of

194 The number of terms *ie* nine is based on the upper limit of short-term memory or 'immediate memory' as suggested by George A. Miller in his classic 1956 paper 'The Magical Number Seven Plus or Minus Two: Some Limits on our Capacity for Information Processing'. Seven (±2) *chunks*[195] of information still apply as a guide to limits on short-term memory to process perceptual input.

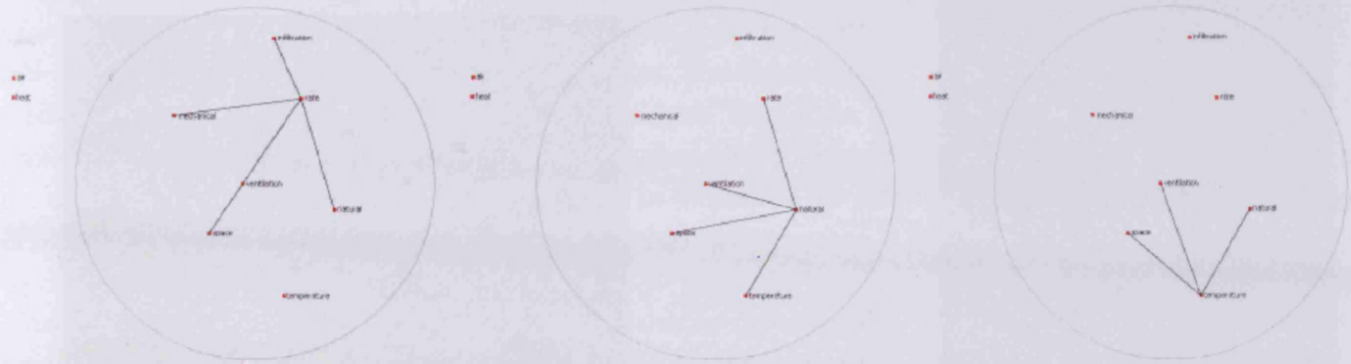
195 Schneiderman, Ben, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Reading, Massachussets: Addison-Wesley, 1998


air
heat
infiltration
mechanical
natural
temperature


air
heat
mechanical
natural
temperature


air
heat
infiltration
mechanical
natural
temperature

interest the user is searching for. The other terms form the context around the currently selected keyword.

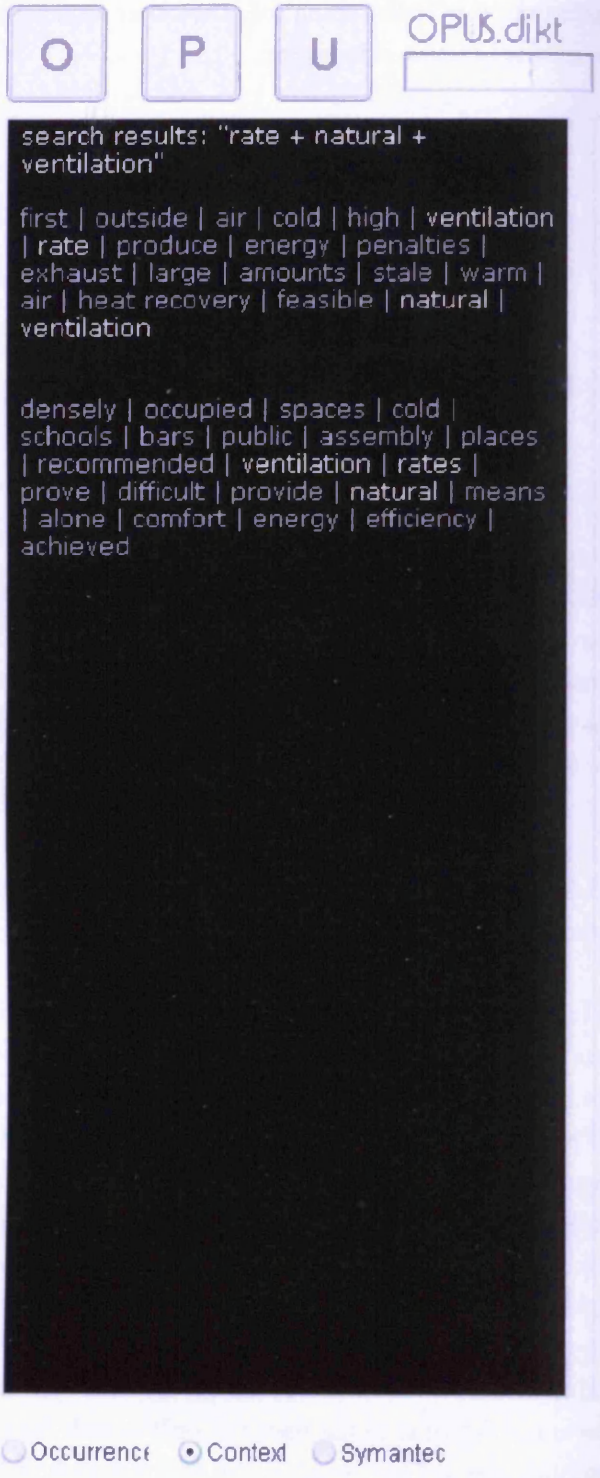


ActionScript:
The *ASnative* function is used to detect if the middle mouse button has been clicked. If a positive value (*true*) is returned, the movieclip selected becomes the current keyword of interest. This in turn redraws lines from this term to all other terms dragged into the circle. Lines are drawn using the *lineTo* method.

4.5.6 Stage 5: Making search dynamic - introducing the 'master.xml' index file

So far, the user can drag and drop terms of interest into the circle, see lines drawn dynamically from the active or chosen keyword to all other words within range; and see the line redrawn if the user changes his or her preference for a keyword by clicking the middle mouse button. In this stage, the functionality of the software is extended by deploying search results corresponding to the actions that occur on stage: searching for one item (serial search), searching for combinations of terms (conjunction search), and updating results based on the currently specified keyword. In Stage 1, deploying search results is discussed, where .txt and .css files are loaded into the resultsPane text field on the right-hand side of the stage. This is restrictive as a) the user can only search for one keyword at a time; combinations are not possible and b) the search is not dynamic *ie* there is no single index file from which the application can search and display results. Instead, it relies on having a .txt and .css file for every search result anticipated. This is a very cumbersome and inefficient approach, which may work for a very small data-set, but is impossible for large data-sets.

At this stage, a master index file is formed in the .xml format. On entering a word in the inputA search field and clicking on the Search! button or hitting the Enter or Return key, ActionScript enables the application


O
P
U
OPUS.dikt
search results: "rate + natural + ventilation"
first | outside | air | cold | high | ventilation | rate | produce | energy | penalties | exhaust | large | amounts | stale | warm | air | heat recovery | feasible | natural | ventilation
densely | occupied | spaces | cold | schools | bars | public | assembly | places | recommended | ventilation | rates | prove | difficult | provide | natural | means | alone | comfort | energy | efficiency | achieved
Occurrence
Context
Symantec

to search through the 'master.xml' index file, filter the relevant search results, and display them in the appropriate format in the resultsPane text field. The same action is triggered, using the hitTest method, when the user drags a word into the circle. At this point however, (opus14), it is still only possible to search for a single entity at a time. This is extended further so that combinations of words can be searched for. Thus, if the user drags in the keyword 'rate' into the circle, and subsequently drags in the terms 'natural' and 'ventilation', thus specifying an interest in 'the rate of natural ventilation', the application returns the results (left).

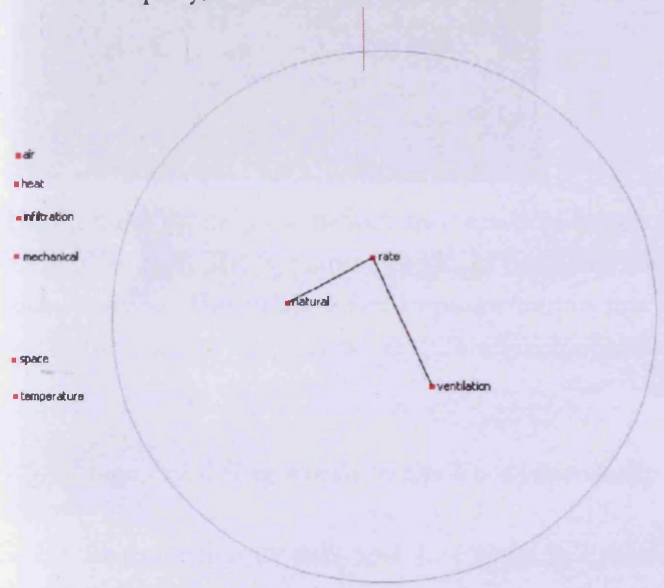
Finally, the search functionality is extended even further by including the proximity principle in the search algorithm. Hence, if *a* is the keyword, and *b* and *c* are also included in the search query; and *c* is spatially closer to *a* than *b*; the algorithm returns results in the following combinatorial order:

search results for terms *a*, *b* and *c*, based on the conditions above,

$$\frac{a + c + b}{a + c}$$

$$\frac{a + c}{a}$$

Hence the query, 'rate of natural ventilation' would be returned as >



search results: "rate + natural + ventilation"

first | outside | air | cold | high | ventilation | rate | produce | energy | penalties | exhaust | large | amounts | stale | warm | air | heat recovery | feasible | natural | ventilation

densely | occupied | spaces | cold | schools | bars | public | assembly | places | recommended | ventilation | rates | prove | difficult | provide | natural | means | alone | comfort | energy | efficiency | achieved

---

search results: "rate + natural"

first | outside | air | cold | high | ventilation | rate | produce | energy | penalties | exhaust | large | amounts | stale | warm | air | heat recovery | feasible | natural | ventilation

densely | occupied | spaces | cold | schools | bars | public | assembly | places | recommended | ventilation | rates | prove | difficult | provide | natural | means | alone | comfort | energy | efficiency | achieved

---

search results: "rate"

incorporate | filtration | heat | recovery | system | extract | heat | exhaust | air | use | pre-heat | supply | air

generated | within | space

minimum | acceptable | fresh | air | ventilation | rates | depend | occupancy | levels | space | use | for example | relation | specific | source | pollution

ventilation | rate | requirements | different | building | types | different | occupancy | levels | obtained | CIBSE | guide

term | typically | used | specifying | ventilation | rates | housing | occupancy | levels | space | volumes | within | predictable | limits

unit | l/s/p | generally | used | specifying | ventilation | rates | spaces | vary | size | known | relatively | high | occupancy | levels | often | predominantly | mechanically ventilated

first | outside | air | cold | high | ventilation | rate | produce | energy | penalties | exhaust | large | amounts | stale | warm | air | heat recovery | feasible | natural | ventilation

densely | occupied | spaces | cold | schools | bars | public | assembly | places | recommended | ventilation | rates | prove | difficult | provide | natural | means | alone | comfort | energy | efficiency | achieved

however | ventilation | rates | sometimes | reduced | recommended | levels | taking | account | duration | levels | occupancy | volume | depth | space | pattern | room | air | distribution

ActionScript:

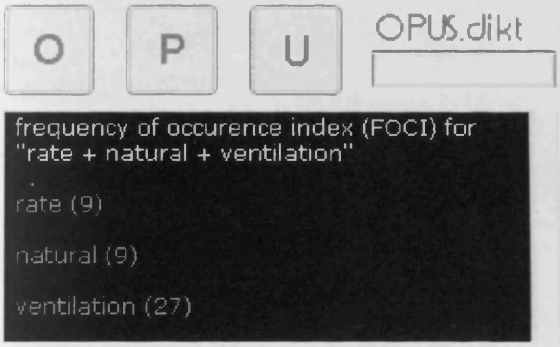
In this stage, the .txt and .css files used to display search results are replaced by a 'master.xml' index file. The search algorithm gleans this database for the current search term(s) and displays them accordingly. XML is the eXtensible Markup Language introduced by the World Wide Web Consortium (w3c.org), for the universal exchange of data over the web. As it is platform independent, xml is capable of supporting a wide

variety of applications. It is not a web authoring language like html, but is a data description language. The 'master.xml' index file can be viewed in Appendix C on the CD[196] accompanying this thesis.

196 Please refer to the 'Chapter 4_Case Study 2\OPUS\queryB' folder on the CD accompanying this thesis to view the master.xml index file

4.5.7 Stage 6: Introducing the Occurrence, Context and Symantec Radiobuttons

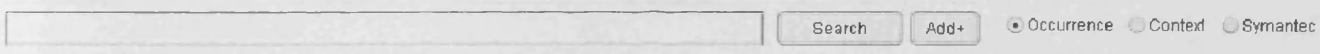
This stage sees a further refinement of the search results. Three 'states' are introduced: an Occurrence state, a Context State and a Symantec State. Each is activated by its respective Radiobutton, at the lower right-hand corner of the screen (image). In Occurrence mode, the user is presented with a statistical count of the number of times a particular word occurs in that document or data-set, and is known here as *Frequency of Occurrence Index* or *FOCI*. For example, in our previous illustration on a conjunction search for the terms 'rate, natural and ventilation', the *foci* of these terms is,



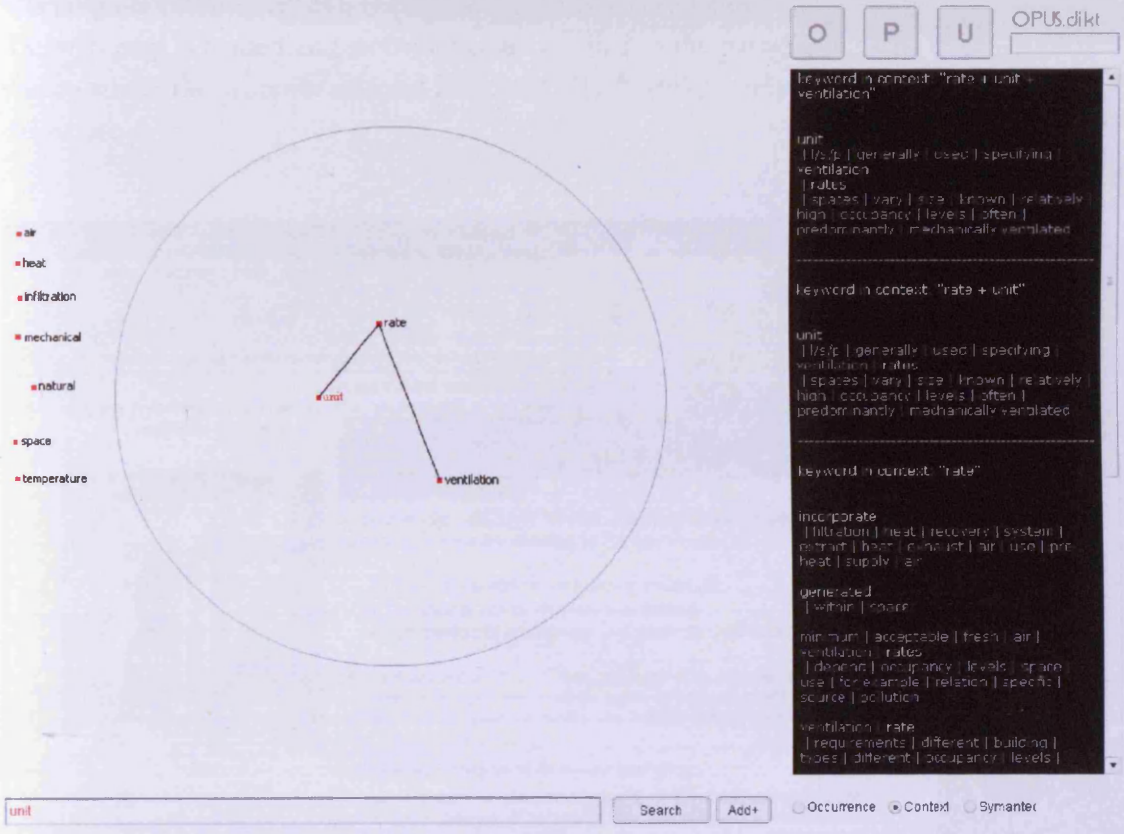
The Context mode is the default mode, where search results are presented in their context. The Symantec mode, as the name suggests, is associated with meaning. This mode is not implemented in this software application, but is discussed in 'Further Work' (Conclusion, Section 6.3).

4.5.8 Stage 7: Adding words to the list dynamically

So far the user can only drag and drop from an existing list of terms on the left-hand side of the circle. Though these terms are carefully chosen from a small section of the Skins and Spaces website, the usability is limited *ie* the user is restricted to searching for items using a pre-selected set of terms. In Stage 5, the search process using the inputA text field was made dynamic by introducing the master.xml index file, thus allowing the user to dynamically search for *any* term. This principle is extended here. An Add+ button is created and placed on stage next to the Search button.



The user can enter any query in the inputA search text field and click on the Add+ button. This event dynamically creates and places the query specified on stage, next to the existing list of terms. The user can drag this term into the circle and see it being included to refine the search query. Continuing from the previous example, where a user may be interesting in the search query: 'rate of natural ventilation'; in a *successive search*, the user may be interested in 'units' related to 'the rate of natural ventilation'. The user now enters the term 'unit' into the inputA search text field and clicks on the Add+ button, which dynamically generates and places a movieclip on stage (below). After re-arranging the terms according to the priorities required or determined by the new search interest, the user's visual query returns the results:

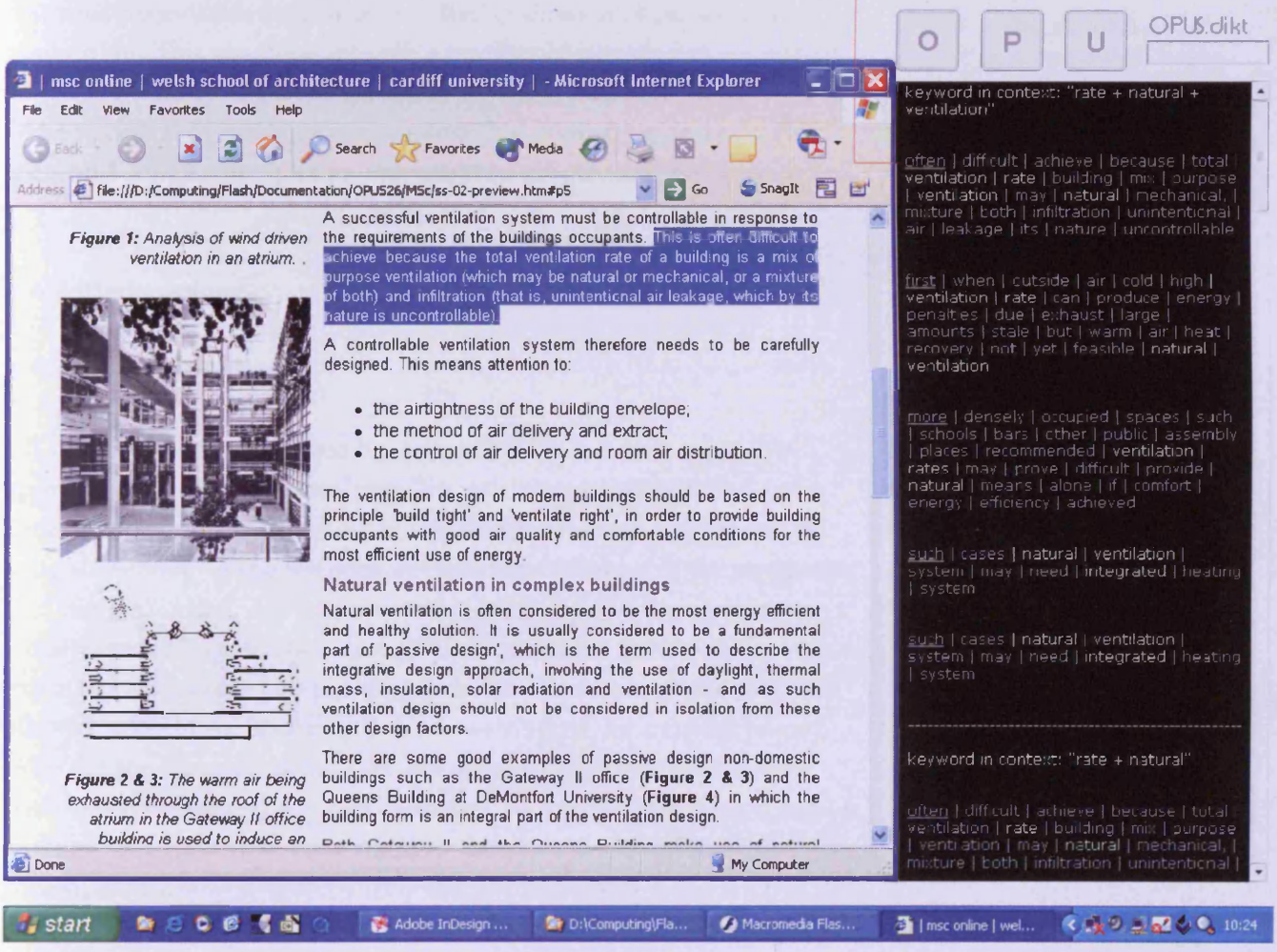


ActionScript:
The *duplicateMovieClip* method is used to create copies of the search query entered in the inputA text field. The user can add any number of terms to the stage dynamically, and use them to refine the search query. The newly generated term appears in red, so as to be able to distinguish from the default set of terms on stage, and any newly added terms.

4.5.9 Stage 8: The pop-up Window

This is the last stage in the functionality of the system and involves displaying additional data to the user. Once the resultsPane text field has been populated with search results and the user has selected an item of interest, a natural tendency of hypertext users is to click on a link, in anticipation that this action will lead to an appropriate result. At this stage, a pop-up Window is introduced which contains the corresponding set of data, in the form of a web page, to the link or data element selected. To illustrate, in our example on the query 'rate of natural ventilation', the user may be interested in the string

Clicking on 'often' invokes a pop-up Window as shown below. The web page is loaded and its contents are scrolled to the paragraph within which the sentence selected is located. The highlight below is deliberate.



ActionScript:

The 'link' attribute in xml is used to define a hyperlink. The first keyword in every passage is hyperlinked and thus appears underlined. In the

example above, the hyperlink for 'often' in the first passage is specified in its xml format as: <word link="MSc/ss-02-preview.htm?#p5">often</word>. The html reference refers to the specific web page where that particular passage is located. The question mark (?) and 'hash' (#) signs, followed by the characters p5 (*ie* passage 5) are references to the *anchor tag*, which is used to specify the location of that passage in the web page. The anchor tag in the corresponding html file is written as: <a name="h5">. The anchor tag is responsible for autoscrolling the web page to the location of the paragraph within which the passage of interest lies, so that the user does not have to browse through the entire web page.

An earlier unsuccessful attempt to deploy the functionality described in this section is as follows:

The pop-up Window is created using the *PopUpManager* class in actionscript. A movieclip named *windowContent* is created in the library, which has a text field with horizontal and vertical scrollbars. The *contentPath* variable in the Window component class allows access to the content to be loaded into the Window, which is the windowContent movieclip in the library. Thus, when the user clicks on a hyperlink, Flash creates the Window at runtime, and loads the windowContent movieclip into its body, which in turn loads the corresponding html file that corresponds to the result clicked upon.

However, due to limitations in Flash to load multiple images dynamically into a text field, the more direct approach of loading a web page using a browser window was adopted.

4.6 Evaluation

4.6.1 User Studies

The final stage of this case study involved evaluation of the software application. This was done through a number of user studies conducted to test the validity of the navigation tool. The studies were not intended to be statistical or quantitative in nature, but qualitative, to assess the ideas and design principles implemented in the software application.

4.6.2 Methodology

4.6.2.1 Test Procedure

All studies were administered by the author. Each session was approximately sixty minutes long. The participants were greeted and a general background to the author's thesis was provided, followed by a general overview of the test session. The demonstration of the navigation tool was performed, any questions were answered and the users were made aware of bugs in the software. Finally, the test questionnaire was explained. The screen capture mode was 'switched on' and the user was allowed to begin the search task. Users were asked not to speak to each other for the duration of the task and to inform the author after they had finished answering the task questions, so that the recording could be stopped. The users were then asked to complete the User Response Form. Finally, the users were asked if they had any additional comments on the navigation tool and were thanked for their participation in evaluating OPUS.

Participant:
Date & Time:
Profession:

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?

3) Does this ability make searching easier and quicker?

4) Does the display of keywords in their context help you to compare the different passages?

5) Does highlighting the terms you are searching for help?

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?

7) What was the highest number of terms you combined to search for a specific query?

8) Is there a learning effect in using and understanding this tool?

9) Comments on user interface:

4.6.2.2 Questionnaire

A questionnaire was prepared for users to complete after using the tool. For a critical evaluation of the tool, a challenging task was required. By nature, the task should obviously be search related, but should be designed to utilize the search capabilities of the tool and the functionality of its user interface elements. Design principles and ideas discussed in earlier sections shape the questionnaire. These are summarized below:
a) freedom of choice to re-arrange words, terms or concepts
b) the keyword in context format
c) the proximity principle
d) 'dynamically adding terms to a list' functionality
e) successive search capability and advantage

With the above in mind, it was decided that the following questions needed to be asked:
1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
2) Does re-arranging words, terms or concepts give you a greater 'freedom of choice' in searching for a result?
3) Do the above two characteristics of the software allow you to form a mental map or model of items of interest?
4) Does the display of keywords in their context force you to compare the different occurrences?
5) Does highlighting the terms you are searching for help?
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
7) What was the highest number of terms you combined to search for a specific query?
8) Is there a learning effect in using and understanding this tool?

Keeping in mind the skills and knowledge the users may possess, neutrality of language is important. The questionnaire in its final form appears to the left, and is entitled 'User Response Form'. However, this part of the questionnaire is answered after performing the search task, which is described below.

4.6.2.3 Search task

As mentioned before, a challenging task is required for a critical assessment of the tool. The questions asked and their search-related characteristics are described below:

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

2. What is the age of air? What is the temperature of air?

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

4. Explore the relation between background ventilation and trickle ventilators.

Q1) On what principle should the ventilation design of modern buildings be based?
This question requires a single search term as part of its search expression, although multiple terms can be combined to find the answer. This is perhaps the easiest question in the task.

Q2a) What is the age of air? Q2b) What is the temperature of air?
This question requires a minimum of two terms to be combined, for each question, to find the correct answers. The term 'air' is the common denominator, and successive searching is required to find the answers, which are located in different sections of the same web page.

Q3a) What is the minimum ventilation requirement to control body odour? Q3b) What is the air infiltration rate of a modern house?
This is perhaps the most difficult question in the task. Two separate questions, with two separate answers, in two different locations need to be found. However, this time the answer to Q3a is not in sentences which have been indexed, but in a table whose contents have not been indexed. The caption to the table however, has been indexed. This question requires carefully chosen terms as part of the search expression, involves successive searching as well as open-ended browsing.

Q4. Explore the relation between background ventilation and trickle ventilators.
This is an open-ended question which has answers in multiple locations. It integrates searching and browsing to find an answer which the user finds appropriate.

The 'task questionnaire' in its final form appears to the left.

4.6.2.4 Pilot Studies

Two separate pilot studies were conducted - one with a single user and the second one with two users. All users involved in the pilot study had specialist knowledge related to environmental design and were very familiar with the contents of the data-set *ie* the Skins and Spaces website. The pilot studies were essential to refine the search task to be performed using OPUS. After the first pilot study, the questionnaire was refined. It was also suggested that the sessions be recorded for later analysis. Also, a demonstration video could be prepared to show to the participants, which would reduce the need for verbal description of the navigation tool. All of these three suggestions were implemented in the second pilot study which consisted of two users. A video of the tool in action was shown to

the users as part of the introduction. The users performed the search task, answered the questions and the user response form. Their search process was recorded using a screen capture software.[197] The second pilot study confirmed that the questions were sufficiently challenging to answer and that the overall format of the questionnaire was appropriate. It was also suggested that a verbal description of the search tool should be kept to a minimum and that the video should demonstrate the tool in action.

197 described in Section 4.6.2.6

4.6.2.5 Participants

Twenty (20) participants were involved in the user studies. All participants were familiar with the internet and searching online. Apart from staff and students at the Welsh School of Architecture, external participants were also involved. Overall, they represented a range of ages, backgrounds, professions, education and skill levels. A generic breakdown is as follows:
Students (undergraduate) = 3
Students (graduate) = 2
Students (also part-time staff members and research students) = 5
Staff (architecture related) = 1
Staff (specialist knowledge - Librarian) = 1
Staff (non-architecture related) = 5
External (architect in practice and ex-student on MSc programme at WSA) = 1
External (Librarian - Law) = 1
External (research student - marketing) = 1

4.6.2.6 Infrastructure and Environment

User studies were usually conducted in pairs, or groups of 3 or 4 participants. The configuration of computers used in the task were:
1) Toshiba laptop: Intel Pentium 4, 2.66 GHz, 512 MB RAM, 15" monitor at 1024 x 768 pixels resolution.
2) Toshiba laptop: Intel Pentium Mobile 1.40 GHz, 240 MB RAM, 12" monitor at 1024 x 768 pixels resolution.
3) Dell laptop: Intel Pentium 4, 2.66 GHz, 512 MB RAM, 17" monitor at 1024 x 768 pixels resolution.
4) Dell workstation: Intel Pentium 4, 3.0 GHz, 2 GB RAM, 19" monitor at 1024 x 768 pixels resolution.
A digital projector was used to demonstrate the search tool. BB Flashback Recorder by Blueberry Software Ltd was used for screen capture *ie* to record the actions of the user. The search task was performed in a quiet

room, where digital projection facilities were available.

4.6.2.7 Limitations and Known 'Bugs'[198]

198 A 'bug' is a flaw in a software application - usually in its source code or design - that may cause errors or malfunctioning of the programme.

The navigation tool does not support searching using phrases or multiple words. Users are required to combine multiple 'single' search terms to find relevant results. The navigation tool has been scripted to a point where it is operational - all intended search tasks using the user interface elements can be performed. A few errors within the system that do not impact the search process, but are necessary to be made aware of to the user, so that they can be bypassed if encountered are:
1) *Lines are not drawn between search terms in the circle*: This error occasionally occurs when the user adds multiple terms to the search list using the 'Add+' button, and drags them into the circle. The user is advised to reset the system by clicking on the Participant button, and starting anew.
2) *Words are still active outside the circle and included in the search results*: This bug is related to the one above, and occurs when the user adds words to the list dynamically by using the Add+ button. If dragging a word outside the circle does not remove it from the search expression *ie* it is still 'active', the user needs to use the middle mouse button to click on the relevant term(s) inside the circle to update the search expression, and consequently the search results.
3) *Clicking on a relevant result on the right-hand side does not load the appropriate web page*: This error occurs if a web page is already open. The user is advised to close all active web pages that may have been launched by navigating through the results on the right-hand side, before another web page is launched.

Although the errors above can be resolved, they would require much advanced scripting, which was considered unnecessary for the purposes of the software prototype.

4.7 Results

Results of the analysis of the search task performed by twenty participants is summarized below under four headings: time, accuracy of results, use of occurrence mode and summary of user comments. However, before discussing the results under these headings, a summary of the above variables in tabular form is presented as four double-page spreads from pages 110 to 113. The page overleaf is a summary of results from the second pilot study.

Pilot Studies		
Users	**# 1**	**# 2**
Time minutes : seconds : milliseconds		
overall	17:43:00	19:21:00
for each question		
Q1	04:29:40	02:06:80
Q2a	01:10:60	04:42:50
Q2b	01:34:70	02:42:40
Q3a	05:25:80	05:52:90
Q3b	00:59:30	00:49:90
Q4	04:03:20	03:06:50
Terms used as part of search expression		
Q1	ventilation, principle, modern	principle, modern
Q2a	age, air	age, air
Q2b	air, temperature	temperature, what
Q3a	minimum, ventilation, infiltration, requirement, rate, body, odour	ventilation, odour, body, requirement, minimum
Q3b	modern, rate, infiltration	infiltration, rate, modern
Q4	background, trickle	ventilation, trickle
Accuracy		
number of answers wrong	Q3a incorrect	did not answer Q1
use of occurrence mode (number of times)	X	Yes (1)
comments on user interface	Nil	Nil
specific comments	The user commented that the Occurrence mode could be removed, as it provided 'statistics only', which did not seem to be useful.	The user contradicted the statement by user 1, saying that in his search, the Occurrence mode helped to establish whether a particular search term existed in the database.

User Evaluation

Users	# 1	# 2
Time minutes : seconds : milliseconds		
overall	23:03:00	45:09:90
for each question		
Q1	04:35:20	07:10:20
Q2a	01:53:60	04:20:60
Q2b	01:33:40	10:51:80
Q3a	04:52:40	09:49:00
Q3b	01:34:40	03:02:10
Q4	08:34:00	09:06:70
Terms used as part of search expression		
Q1	ventilation, design, principles	ventilation, principle, modern, principles
Q2a	age, air	age, air
Q2b	air, temperature, measure	air, temperature, define, definition, type
Q3a	ventilation, body, odour, requirement, minimum	ventilation, body, 'body odour', minimum, requirement, odour
Q3b	infiltration, ventilation, house	air, infiltration, modern
Q4	background, ventilation, trickle, natural, age	ventilation, background, ventilator, trickle, relation
Accuracy		
number of answers wrong	Nil	Q2b not perfectly right
use of occurrence mode (number of times)	X	Yes (3)
comments on user interface	interface could highlight searched terms in the web page and a 'find next' tool could be incorporated.	Nil
specific comments	user commented that the formation of a map at the end of the process was very useful and that the idea could be extended to build up a personal dictionary of relevant items as part of regular search routines.	

	User Evaluation		
	# 3	# 4	# 5
	18:51:10	24:26:40	13:50:60
	01:11:40	04:04:00	02:38:50
	01:30:40	01:55:00	01:48:20
	03:10:00	00:47:40	01:16:20
	02:39:90	12:05:40	03:11:30
	01:37:30	01:31:50	00:52:60
	08:42:10	04:03:30	04:03:80
	ventilation, modern	ventilation, modern	ventilation, principle, modern
	age, air	age, air	age, air
	air, temperature	air, temperature	air, temperature
	ventilation, body, rate, minimum	ventilation, odour, body, temperature, mechanical, control, requirement, minimum	ventilation, odour, body, minimum
	air, infiltration, house	infiltration, rate, modern	air, infiltration, modern
	ventilation, background, trickle	ventilation, background, trickle	ventilation, background, trickle
	Nil	Nil	Nil
	Yes (1)	X	X
	the terms on the left-hand side could perhaps be labelled in alphabetical order; an auto-align function for words placed outside the circle would make things clearer.	nice, simple and clear interface.	would possibly be easier if words in sentences - and, if, but, a, the, etc were left in, as sometimes it is not possible to know exactly which result is best; could be turned on or off? the text could be larger - easier to read and move about.
	user commented that he had discovered new items of interest during the search task, and found that very useful.		

User Evaluation

Users	# 6	# 7
Time minutes : seconds : milliseconds		
overall	23:56:20	17:16:80
for each question		
Q1	03:50:40	03:18:00
Q2a	05:02:20	01:51:80
Q2b	04:47:50	01:27:60
Q3a	05:50:50	05:03:80
Q3b	01:44:40	02:18:40
Q4	02:41:20	03:17:10
Terms used as part of search expression		
Q1	ventilation, principle	ventilation, principle, building
Q2a	air, age	air, age
Q2b	air, temperature	air, temperature
Q3a	ventilation, odour, minimum, comfort, rate	ventilation, body, odour, requirement, minimum
Q3b	air, infiltration	air, infiltration, modern
Q4	ventilation, ventilators	ventilation, trickle
Accuracy		
number of answers wrong	Q3a incorrect	Nil
use of occurrence mode (number of times)	X	Yes (1)
comments on user interface	it seems very simple and quick to understand	highlighting the results in the web page or document would help; a mechanism to delete words in the circle, but keep them in the list; and the possibility of building a personal 'dictionary' and saving it.
specific comments		

	User Evaluation		
	# 8	# 9	# 10
	25:14:60	27:38:80	15:10:00
	07:18:00	03:51:20	03:51:20
	02:19:80	02:08:80	01:21:20
	03:29:00	00:58:20	00:47:30
	03:18:30	09:21:60	05:09:70
	02:36:00	08:08:00	00:58:80
	06:13:50	03:11:00	03:01:80
	ventilation, 'modern buildings'	ventilation, design, modern, buildings	ventilation, modern, principle
	air, age	age, air	age, air
	air, temperature	air, temperature	air, temperature
	ventilation, 'body odour', rate	ventilation, minimum, body, odour, requirement, control	ventilation, temperature, body, odour
	air, infiltration, rate	air, infiltration, rate, modern, house	air, infiltration, modern, rate
	ventilation, background, trickle	background, ventilation, trickle, ventilators, relation, between	ventilation, background, trickle
	Nil	check	Q3a answered but actions not recorded in video
	Yes (3)	Yes (1)	X
	very user friendly, simple to understand and use.	a bit confused on how to gather groups of text together as linking did not always work.	having the ability to change search process is very useful especially with the 'add' functionality; passages of text with searched words contained in them is extremely useful for identifying passages/ results quickly and efficiently.

User Evaluation

Users	# 11	# 12
Time minutes : seconds : milliseconds		
overall	24:15:50	30:53:30
for each question		
Q1	02:51:40	09:36:20
Q2a	01:57:40	01:59:70
Q2b	01:09:30	01:34:70
Q3a	02:58:10	07:18:00
Q3b	04:43:20	01:00:70
Q4	10:36:10	05:14:30
Terms used as part of search expression		
Q1	ventilation, modern, principle	ventilation, space, building, principles, modern
Q2a	air, age	air, age
Q2b	air, temperature	air, temperature
Q3a	ventilation, body, odour, minimum, control. requirements	ventilation, 'body odour', body, odour, minimum, rate
Q3b	air, infiltration, modern	infiltration, rate
Q4	ventilation, background, trickle, relation	ventilation, background, trickle, ventilators
Accuracy		
number of answers wrong	Q3a: table identified, but not answered	Nil
use of occurrence mode (number of times)	Yes (2)	Yes (1)
comments on user interface	Nil	generally user friendly and appears to give more accurate results; the right-hand list of terms can get very long and nearly requires a search engine to search within search results; perhaps colour would help to highlight main terms; could you click on words from occurrence list to get those sentences?
specific comments		

User Evaluation

# 13	# 14	# 15
23:19:00	22:05:00	17:59:00
03:23:40	06:08:60	02:05:70
02:10:90	02:53:80	01:56:20
01:00:00	01:29:10	04:21:90
05:06:10	05:48:30	02:28:40
02:41:60	01:44:80	00:53:50
08:57:00	04:00:40	06:13:30
ventilation, principle, building	ventilation, principles, modern, principle	ventilation
air, age	air, age	air, age, temperature, measure
air, temperature	air, temperature	
body, odour, minimum, ventilation, requirement	ventilation, body, odour, control, minimum	odour, ventilation, control, rate
infiltration, air, rate, modern	infiltration, house	infiltration, rate
ventilation, background, trickle, ventilators	ventilation, background, trickle	ventilation, natural, mechanical, trickle background
Nil	Nil	Nil
Yes (2)	Yes (2)	Yes (1)
very good and useful; the spatial arrangement of terms is very effective; more detailed sensitivity in the positioning of words and the ranking of results, perhaps a series of orbits, would really improve speed; perhaps there is also need for a quarantine area for unwanted words.	the user interface needs to be more intuitive if it is to be used generally; right and left mouse click options and drop down list on search terms in the circle would be helpful as well as integrating web-based searched in page results.	it made more sense than other efforts I've seen to represent searches and results visually. I recall search tools that put results in bubbles spreading out according to relevant - these were largely gimmicky and didn't (I believe) take off. This interface is visual without extraneous gizmo-bits. I liked very much that it employed Boolean syntax without overtly stating so, and think that those who struggle to grasp this concept will accept a visual representation far more easily.

User Evaluation

Users	# 16	# 17
Time minutes : seconds : milliseconds		
overall	18:15:30	18:37:10
for each question		
Q1	02:53:40	02:33:80
Q2a	02:20:50	01:58:50
Q2b	01:43:20	01:58:80
Q3a	07:04:60	05:09:80
Q3b	02:02:70	00:43:30
Q4	02:10:90	05:51:10
Terms used as part of search expression		
Q1	ventilation, 'modern buildings'	ventilation, design, principle
Q2a	air, age	air, age
Q2b	air, temperature	air, temperature, definition
Q3a	ventilation, 'body odour', control, odour, requirement	ventilation, body, odour, minimum
Q3b	air, infiltration, rate	-
Q4	ventilation, background	ventilation, background, trickle
Accuracy		
number of answers wrong	Q1, Q3a & Q3b incorrect	Q3a incorrect
use of occurrence mode (number of times)	X	Yes (2)
comments on user interface	easy to use; appeared user friendly and fast; the bug of not always linking the words could prove frustrating?; to keep the words you added on the left-hand side could be useful, rather than losing them each time.	I like the spatial/graphical interface of the searching, in particular to be able to use a 'central' term.
specific comments	Q1 answered in Q3a	Answering Q3b by chance *ie* opportunistic

	User Evaluation		
	# 18	# 19	# 20
	21:56:60	19:06:80	20:36:80
	03:25:80	02:41:30	02:07:40
	01:16:70	01:06:50	05:35:10
	02:31:40	01:07:60	00:43:40
	08:09:30	04:36:20	06:44:20
	01:41:80	01:03:80	01:25:80
	04:51:60	08:31:40	04:00:90
	ventilation, 'modern building'	ventilation, principle	ventilation, principle
	air, age	air, age	air, age
	air, temperature	air, temperature	air, temperature
	ventilation, 'body odour control', 'body odour', body, heat, minimum	ventilation, odour, minimum, control	ventilation, requirement, odour, minimum, control, body, 'body odour'
	infiltration, air, rate	air, infiltration	air, infiltration, rate
	ventilation, background	ventilation, 'trickle ventilator', trickle	background, ventilation, trickle
	Nil	Q3a incorrect	Nil
	X	X	X
	Nil	I think it would be interesting to think about a way of identifying words (highlight differently, underline maybe) in each of the selected sentences of the 'keyword in context box' that could inform which kind of info related to the keyword is contained in the sentence.	the tool is easy to use and I found it useful in many instances, for example, in academic life to help in the selection of articles, literature review, etc.
	Answering Q3a - ventilation + minimum in a successive search	answered Q1, Q2a, Q2b and Q3b without opening web pages. User commented that she was used to searching using pairs of words.	user commented that the tool goes beyond the knowledge/data-set represented.

4.7.1 Time

The table below represents the minimum, maximum and average times for different categories of the search task.

Summary of Results: Time minutes: seconds: milliseconds			
	Minimum	Maximum	Average
Overall	13:50:60	45:09:90	22:13:00
Q1	01:11:40	09:36:20	03:55:10
Q2a	01:06:50	05:35:10	02:25:50
Q2b	00:43:40	10:51:80	02:19:30
Q3a	02:28:40	12:05:40	05:49:30
Q3b	00:43:30	08:08:00	02:00:60
Q4	02:10:90	10:36:10	05:29:70

Since this is not a statistical comparative study, the individual time required for each question is not relevant. Attention is drawn however, to the minimum, maximum and average 'overall' times required by the users:

• The minimum time required by one of the participants (user 5) to perform the search task was 13 minutes, 60 seconds. This indicates that the participant is skilled at searching online and used appropriate terms as part of the search expression to find the relevant answers within a very short period of time.[199]

• In contrast, the maximum time required to complete the search task (user 2) was 45 minutes, 10 seconds; twice the average overall time required by all users. Lack of proper articulation of the search expression, and more importantly, the user's mis-preference for browsing and scanning of hypertext are responsible for this extended time required.[200]

• The average time required to answer all questions in the search task was 22 minutes, 13 seconds.

199 User 5, p 110. Refer to the video 'user2.exe' in the 'Chapter 4: Case Study 2' folder on the CD accompanying this thesis.

200 User 2, p 110 facing

4.7.2 Accuracy of Results

The table below summarizes the number of incorrect answers for each question, from a sample of twenty participants.

Summary of Results: Accuracy		
Question	Incorrect answers (out of 20)	Accuracy
Q1	1	95%
Q2a	0	100%
Q2b	1	95%
Q3a	5	75%
Q3b	1	95%
Q4	0	100%

The character of each question has been explained in Section 4.6.2.3 and the terms used by each participant for each question are indicated in the User Evaluation tables (pages 110 - 113). As explained earlier, the content of the answer to Q3a has not been indexed. Only the table heading within which the answer was located had been indexed. This search required successive searches with careful consideration to the search expression, and in most cases browsing, to find the right answer. Out of the five participants who did not answer Q3a, three did not persist with the search task (a behaviour commonly seen as part of search behaviour); one was unable to find the result, whereas one of the users identified the table during browsing, but failed to recognise it as the source of the answer. However the majority of users - 75% - answered Q3a correctly. Seven out of twenty participants switched to Q3b after failing to answer Q3a in the first instance, and then returned to it later, and answered it correctly.

4.7.3 Use of Occurrence mode

Twelve (12) out of twenty participants (60%) used the occurrence mode to confirm whether a term used as part of a search expression existed in the index, subsequently allowing them to modify the search expression.

4.7.4 Summary of user comments

Below is a summary of users' comments after completing the search task. Completed questionnaires can be found in Appendix D. Comments are summarized in response to each question in the User Response Form.

1) *Do you understand the utility of the 'principle of proximity'? Is it intuitive?*

Most participants agreed that arranging terms based on their proximity to a keyword was intuitive and worked well. Some users thought that the concept was easily grasped after it was explained. It was also suggested that the spatial arrangement could be used or extended to refine search results further.

2) *Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?*

Most participants agreed that it did allow them to refine their search results. One participant liked the ability to redefine relationships between terms by using the middle mouse button, whereas another user commented that the act of 'playing'

3) *Does this ability make searching easier and quicker?*

Most users commented that spatially rearranging terms did make searching quicker and easier, especially for complex sequences of words. It also appeared to allow for a higher accuracy of search.

4) *Does the display of keywords in their context help you to compare the different passages?*

Users concurred that the 'keyword in context' format did enable them to 'scan' results quickly.

5) *Does highlighting the terms you are searching for help?*

Highlighting was useful, but most users suggested that this facility be extended to highlight passages or sentences in the web page as well.

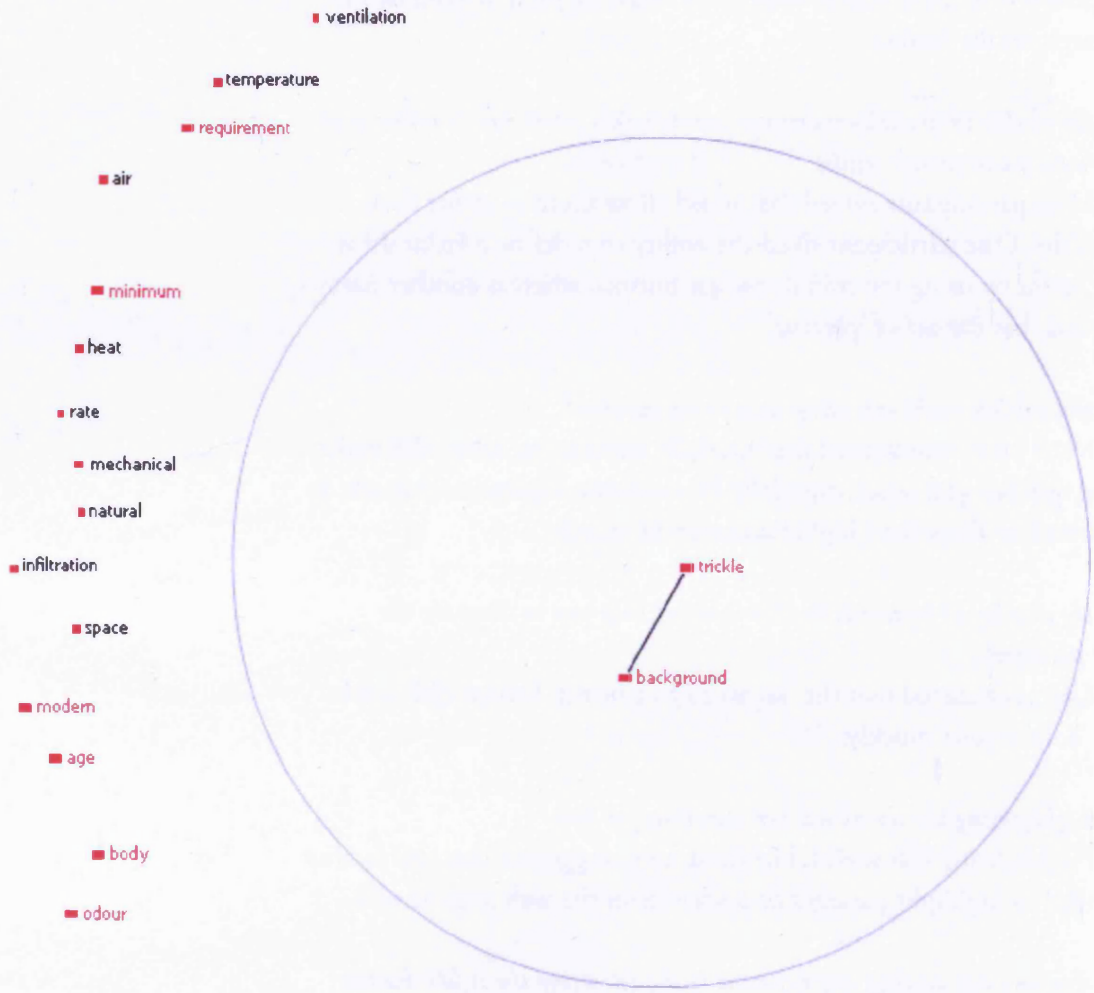
6) *Does dynamically adding any keyword to the list using the Add+ button assist in the search task?*

Yes.

7) *What was the highest number of terms you combined to search for a specific query?*

The average number of terms used as part of a search expression, as mentioned in the user response form was four.

8) *Is there a learning effect in using and understanding this tool?*



background

Search!

Add+

Some users thought that there was a short learning curve to understanding the logic and design rationale behind the tool. Others thought that it was intuitive and easy to understand.

9) *Comments on user interface.*

Comments under this question are summarized in pages 110-113.

4.8 Discussion

The user studies conducted on the navigation tool revealed positive results. The application was seen to be intuitive and the interface was regarded as simple and easy to use. The ability to spatially manipulate terms or concepts had a positive impact on the search results as users felt they could visually define the relationships between terms, and for some users, achieve a higher accuracy of results. This direct spatial manipulation enabled fast searches and, more importantly, the terms in the search expression could be redefined if a suitable result was not found. This rapid successive searching capability is seen as one of the main advantages of the navigation tool. The keyword in context format proved highly successful in enabling users to scan for relevant results, and in some cases, answer the questions without having to visit the web pages. Another strong characteristic of the tool has been its ability to integrate the tasks of searching and browsing during an information seeking process. This was particularly demonstrated when users were trying to answer Q3a and Q4, both of which required redefining search terms, scanning relevant results and browsing web pages as part of a cyclical thought process.

The formation of a 'map' of items of interest, as shown to the left, was considered to be a useful feature that could extend the existing list of terms to form a dictionary of most frequently used search terms during information seeking sessions. More than half of the participants used the occurrence radio button to determine whether a particular term existed in the index, and this led them to refine their search query. It must also be noted that no attempt to 'rank' results was made, and this did not seem to impact the users abilities to answer the task questions. Also, Boolean syntax (and/+, not/-) is subtly supported by visual means, and as one of the users, who is a Librarian by profession commented (user 15, p 101)[201], 'those who struggle to grasp this [Boolean] concept will accept a visual representation far more easily.'

201 User 15, p 113. Refer to the video 'user15.exe' in the 'Chapter 4: Case Study 2' folder on the CD accompanying this thesis.

As mentioned earlier, information seeking is a complex process involving search, browse and serendipity. It can be open-ended or heuristic, in which case it is commonly known as browsing. On the other hand, it can

be goal oriented or strategic. A well designed navigation tool can support both opportunistic and heuristic states of searching or information seeking. The navigation tool has been successful in three specific areas:
a) By using techniques that 'enforce comparisons within an eyespan' (*ie* the keyword in context format), it has enabled rapid scanning of results which have led to decision-making and browsing.
b) Defining the problem by formulating a query using appropriate search terms is a critical stage of the information seeking process, and remains active as long as the search progresses. The ability to visually redefine relationships between terms using direct manipulation has supported this stage of the search process, and enabled efficient searches.
c) The direct spatial manipulation of terms, combined with the keyword in context format, the occurrence mode showing 'statistics' and the formation of a map of terms used - they all facilitate the ability to search in rapid successions during the information seeking process, as displayed in the videos on the CD accompanying this thesis.

4.9 Future Work

Two key recommendations that emerge from the user studies are:
a) Highlighting of sentences or passages in web pages would greatly increase the efficiency of search, and b) An option to save the keywords that have been used during a search session, and perhaps building a personal dictionary of most frequently used terms for search sessions within a specific data-set. Apart from the above two recommendations which would enhance the functionality of the system, an advanced prototype of the navigation tool would require implementing both the 'User' state and the 'Observer' state, which, as explained in Section 4.4, are deeply engaging subjects in Computer Science.

5.0 Discussion

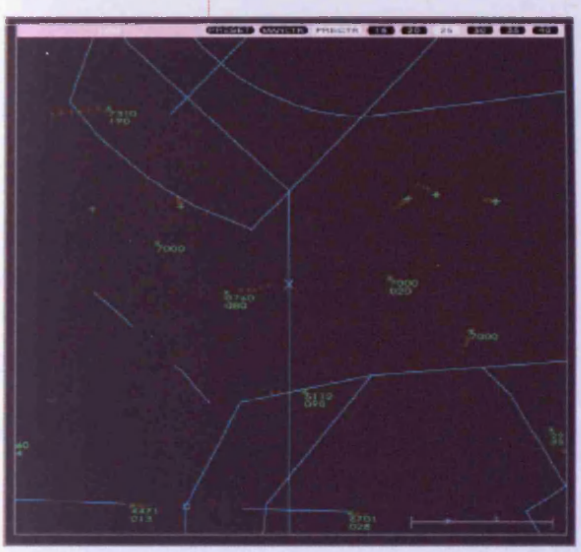
5.0 Discussion

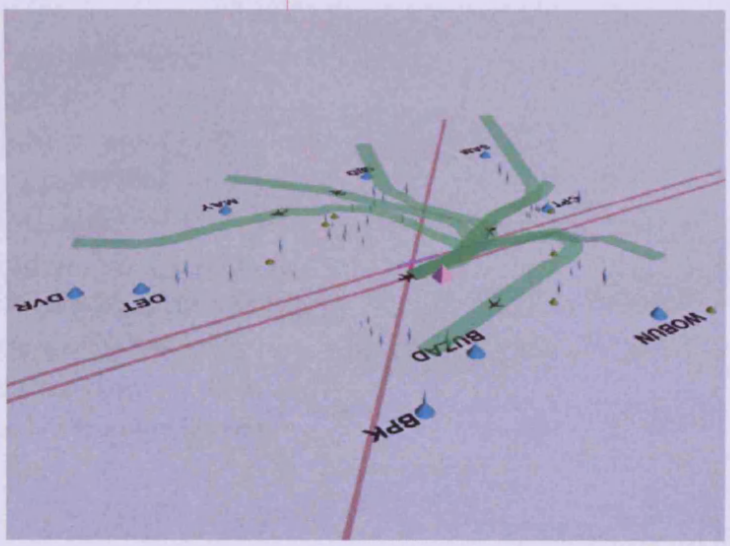
5.1 Principles applied

The case studies in the second part of this thesis *apply* design principles inherent in Data Graphics to construct two visualizations. The three-dimensional visualization of air traffic control is based on data that is inherently spatial, whereas the navigation tool to search text within a hypertext document is based on data that has no spatial bearings *ie* it is abstract. The principles employed in each visualization are discussed below.

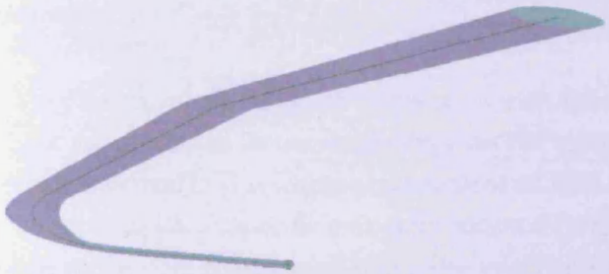
5.1.1 A three-dimensional radar display for air traffic control

A fundamental rule of data design rather than a design principle governs the visualization of air traffic control data: *data integrity.* It is much more meaningful to represent or map data that has three-dimensional spatial bearings to a familiar cartesian world on computer screen rather than limit it to a plan view.





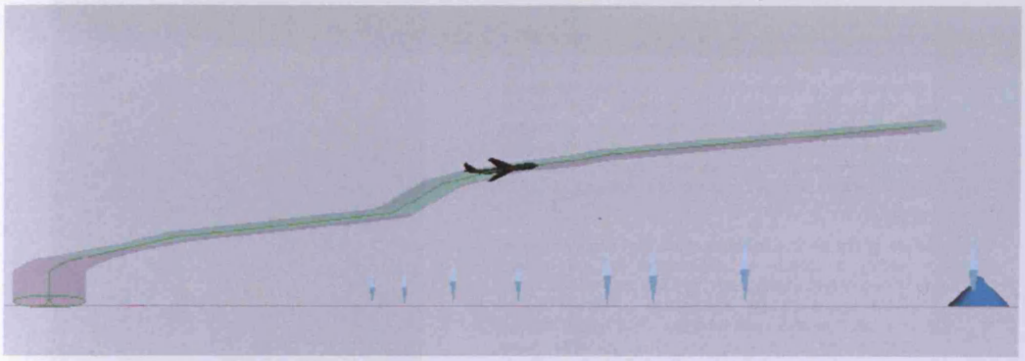
In doing so, the data becomes obvious; it speaks for itself. The same principle guides the construction of elliptical tunnels that represent the Standard Instrument Departure routes (SID's) or flightpaths that guide the movement of aircraft departing from Heathrow. The three-dimensional elliptical tunnels of



airspace are important as apart from structuring the airspace, in the actual simulation[202], they place the data (the flight, its x, y and z co-ordinates, its speed, height, velocity and bearing) in its context; and allow relative comparisons of their changing positions over time.

202 Appendix E: Chapter 3_Case Study 1 folder

A generic data graphic principle applied to visualize ATC data is 'multiples and variables'. The symbols used to represent the altitude constraints (multiples of) repeat themselves, but vary in size according to the height represented by the constraint. These small multiples reveal an overall picture currently absent from two-dimensional radar displays, as is evident in the images below and overleaf.

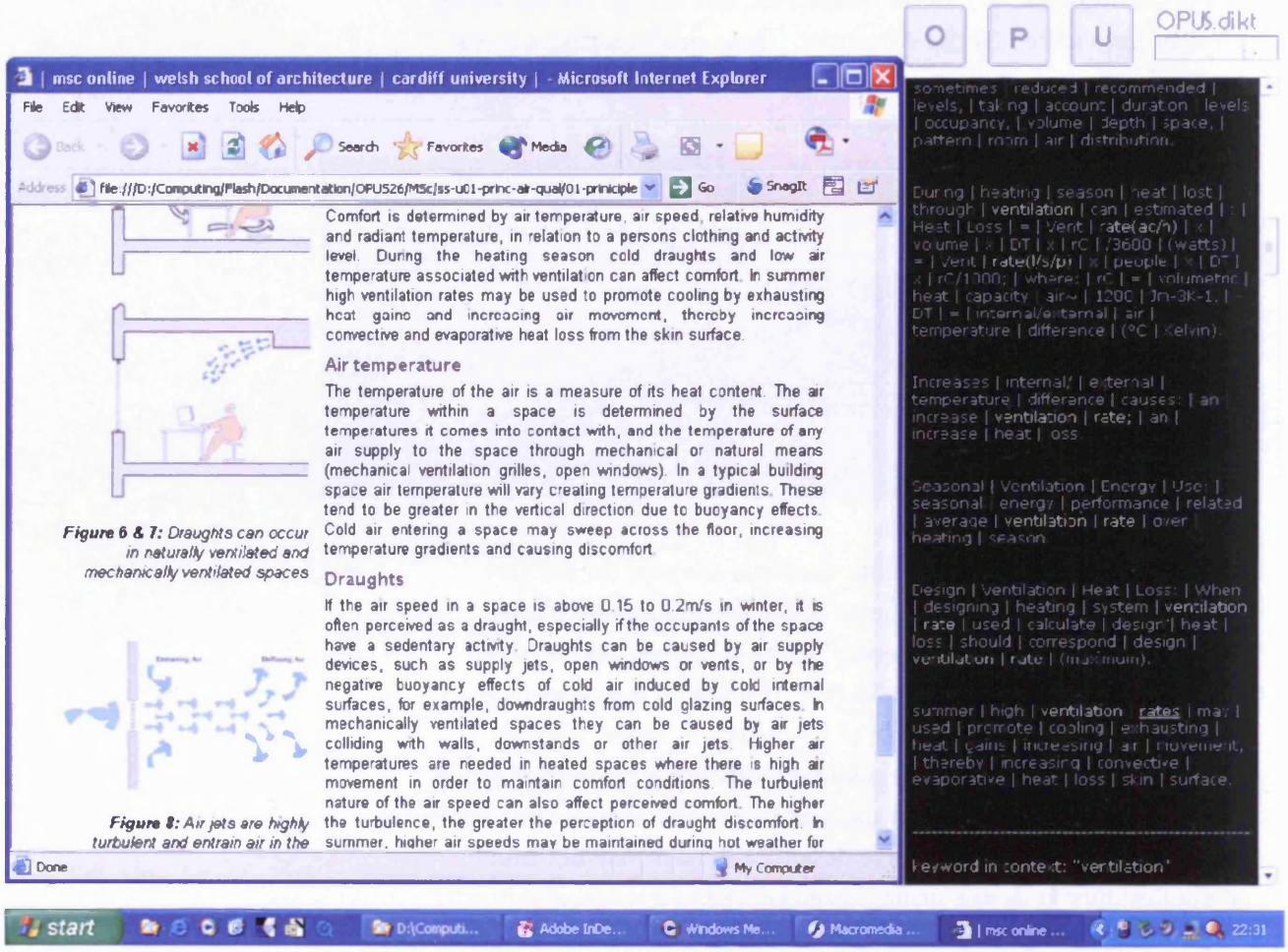


A specific design principle from colour theory is adopted for the grey background against which the elements of the foreground are placed. As discussed earlier (Chapter 3, Section 3.5.4), the green colour of the SID's is based on its established use in ATC displays as it maintains good brightness contrast throughout most of its saturation range. Another specific design principle grounded in scientific theory is that of a 'moving and flashing object', used to attract the user's attention outside the Useful Field of View (UFOV). Both of these design principles may not find general applicability as design principles employed in any visualization depend on the data-set, the context and desired goal of the visualization.

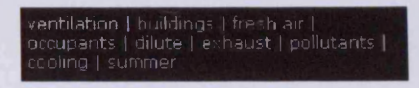
5.1.2 A navigation tool for information seeking

This visualization implements all three generic data graphic principles discussed in Part 1 of this thesis: *Multiples and Variables*, *Context and Comparison* and *Cause and Effect*.

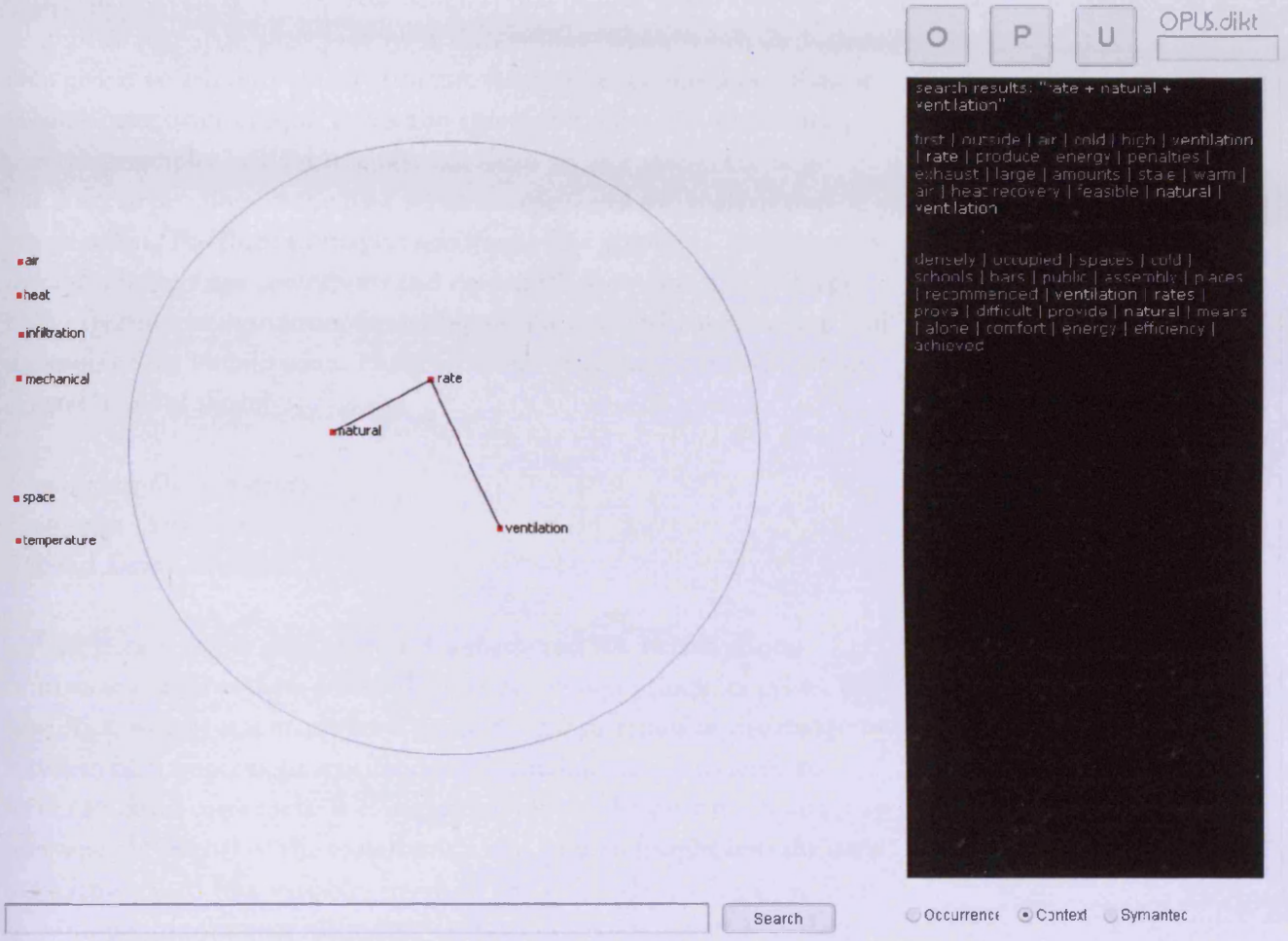
Data is presented in its context - sentences with (common) lemmas omitted are presented in the resultsPane on the right-hand side of the interface (overleaf). This linear arrangement of data enforces comparisons within an eyespan. A specific principle adopted from colour theory is applied to the data represented here. The small multiples of words or


O
P
U
OPUS.dikt
msc online | welsh school of architecture | cardiff university | - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Search Favorites Media
Address file:///D:/Computing/Flash/Documentation/OPUS26/MSc/ss-u01-princ-air-qual/01-principle
Go SnagIt
Comfort is determined by air temperature, air speed, relative humidity and radiant temperature, in relation to a persons clothing and activity level. During the heating season cold draughts and low air temperature associated with ventilation can affect comfort. In summer high ventilation rates may be used to promote cooling by exhausting heat gains and increasing air movement, thereby increasing convective and evaporative heat loss from the skin surface.
Air temperature
The temperature of the air is a measure of its heat content. The air temperature within a space is determined by the surface temperatures it comes into contact with, and the temperature of any air supply to the space through mechanical or natural means (mechanical ventilation grilles, open windows). In a typical building space air temperature will vary creating temperature gradients. These tend to be greater in the vertical direction due to buoyancy effects. Cold air entering a space may sweep across the floor, increasing temperature gradients and causing discomfort.
Figure 6 & 7: Draughts can occur in naturally ventilated and mechanically ventilated spaces
Draughts
If the air speed in a space is above 0.15 to 0.2m/s in winter, it is often perceived as a draught, especially if the occupants of the space have a sedentary activity. Draughts can be caused by air supply devices, such as supply jets, open windows or vents, or by the negative buoyancy effects of cold air induced by cold internal surfaces, for example, downdraughts from cold glazing surfaces. In mechanically ventilated spaces they can be caused by air jets colliding with walls, downstands or other air jets. Higher air temperatures are needed in heated spaces where there is high air movement in order to maintain comfort conditions. The turbulent nature of the air speed can also affect perceived comfort. The higher the turbulence, the greater the perception of draught discomfort. In summer, higher air speeds may be maintained during hot weather for
Figure 8: Air jets are highly turbulent and entrain air in the
Done
My Computer
sometimes | reduced | recommended | levels, | taking | account | duration | levels | occupancy, | volume | depth | space, | pattern | room | air | distribution.
During | heating | season | heat | lost | through | ventilation | can | estimated | : | Heat | Loss | = | Vent | rate(ac/h) | x | volume | x | DT | x | rC | /3600 | (watts) | = | Vent | rate(l/s/p) | x | people | x | DT | x | rC/1000; | where: | rC | = | volumetric | heat | capacity | air~ | 1200 | Jm-3K-1, | DT | = | internal/external | air | temperature | difference | (°C | Kelvin).
Increases | internal/ | external | temperature | difference | causes: | an | increase | ventilation | rate; | an | increase | heat | loss.
Seasonal | Ventilation | Energy | Use: | seasonal | energy | performance | related | average | ventilation | rate | over | heating | season.
Design | Ventilation | Heat | Loss: | When | designing | heating | system | ventilation | rate | used | calculate | design | heat | loss | should | correspond | design | ventilation | rate | (maximum).
summer | high | ventilation | rates | may | used | promote | cooling | exhausting | heat | gains | increasing | air | movement, | thereby | increasing | convective | evaporative | heat | loss | skin | surface.
Keyword in context: "ventilation"
start
D:\Computi...
Adobe InDe...
Windows Me...
Macromedia ...
msc online ...
22:31

terms vary in colour - white for the search term and grey for other terms part of the sentence - leading to the pop-out effect[203], an important perceptual feature that is preattentively processed. Once the user scans this body of text and attempts to retrieve an item of interest by clicking on a hyperlink, the cause and effect principle comes into action by ensuring that relevant data is displayed in the windowPane within an appropriate timeframe (left).



203 Triesman, Anne, Preattentive Processing in Vision, *Computer Vision, Graphics and Image Processing*, Vol 31, 1985, pp 156 - 177



A specific cognitive principle based on preattentive processing is termed here as the 'principle of proximity'. Line length is preattentively processed. In the visual interface of the navigation tool, whenever a term is dragged and dropped into the circle, a line is dynamically drawn from the keyword to other terms in the circle which are within range (above). The lines and their lengths help to infer the 2d spatial positioning of terms in relation to the keyword, and also help in intuitively rearranging the configuration to alter the results (right). The cause and effect principle is also in action here as dragging and dropping terms into the circle or refining the search expression by altering the arrangement of terms leads to relevant search results being displayed within an appropriate time span.

5.2 Generality and Specificity

It emerges from discussions[204] on form in the thesis that trying to impose a form on a data-set in the process of mapping it to an appropriate visual structure is counterproductive to the principles of data design. This was apparent during the development of the navigation tool, when a spiral form was used to represent data and also as a device to seek information (left). An overriding rule of data design is that of data integrity. Any attempt to map data arbitrarily in 'n' dimensions (where n = 1, 2, 3...), or even give it an arbitrary visual structure is bound to be less successful, or in most cases unsuccessful. It has also emerged from both visualizations that the principles utilised to construct them are not always the same but differ depending on the data-set in question and the goals of the visualization. The three principles specific to data graphics - *multiples and variables, context and comparison* and *cause and effect* - find applicability under specific circumstances depending on the data to be represented and the goals of the visualization. However, both visualizations observe three general 'rules of thumb':

1. maintain *Data Integrity*
2. attempt *Data Density*
3. avoid *Data Overload*

All the data graphics discussed in this thesis and the visualizations constructed observe these rules. They are not design principles per se, but heuristics, which, at a macro level guide the construction of visualizations, which in turn implement specific design principles at micro level, to form successful constructs. It is important for the designer to maintain an overview of the goal of the visualization and gain an insight into the data to be represented (the variables involved and their relationships), so that an appropriate construct ultimately leads to the extraction of knowledge from data.

204 Chapter 2, Sections 2.4, 2.5 and 2.6; and Chapter 4, Section 4.5.4

5.3 Reliability

The design principles discussed in this thesis are independent of the medium on which they are generated and presented, and the source of content. The application of these principles depends on the data-set in question and the context of the visualization. There is no recipe or formula to apply these principles, but as the two visualizations have shown, their selective application produces reliable results. If the visualization constructed is effective, which, in this study the validations have confirmed, then the reliability of the design principles is also

confirmed. In general, the design principles to be applied depend on the mental tasks to be performed. If the cognitive task requires comparing alternatives, then 'show comparisons'; if the cognitive task requires reasoning about causality, 'show cause and effect'; if the cognitive task is to show variation, 'show multiples and variables'. If applying these design principles does not yield appropriate results, one should deviate from these principles. These design principles have been derived from cognitive tasks - "understanding causality, multivariateness and comparison" - [which are universal] and tied to nature's laws; indifferent to language, culture, gender, or the particular mode of information that is provided."[205]

205 Zachry, Mark and Charlotte Thralls, Cross-Disciplinary Exchanges: An Interview with Edward R. Tufte, *Technical Communications Quarterly*, Vol 13, No 4, 2004, pp 447 - 462

5.4 Implications

The comprehension of data is at the heart of producing exemplary data graphics. In the contemporary digital world, the same concern guides the design of visualizations where users can interact with the data and gain an understanding of phenomenon inherent in data. Simple design principles and rules of thumb guide the design of data graphics and visualizations. Some find general applicability and others need to be applied with caution. An awareness of these principles and knowledge of how to apply them to create effective applications for visual reasoning can yield great results. The effectiveness of the visualizations in this thesis have been confirmed through domain experts and user studies. The applications demonstrate how design principles inherent in data graphics, which have been historically represented on paper, are still valid to create effective, contemporary, computer-based, interactive visualizations of data.

6.0 Conclusion

6.0 Conclusion

6.1 Principles uncovered

The aim of historic data graphics and modern computer-based visualizations is to reveal data. Visualizations need to be effective. The three heuristics that *need* to be applied to create successful visualizations are:

1. maintain Data Integrity
2. attempt Data Density
3. avoid Data Overload

Under this umbrella, the three data graphic principles which have been explored in this thesis that *can* be applied to create effective visualizations are:

1. Multiples and Variables
2. Context and Comparison
3. Cause and Effect

Furthermore, specific cognitive principles derived from domains such as colour theory, vision theory, etc need to be applied selectively. Their application depends on the design of the visualization which remains open-ended, but bound by the framework of visualization (Chapter 1, Section 1.4) where data is transformed into information through an appropriate visual structure, and over a period of time this information crystallizes into knowledge.

6.2 Contribution

The contributions of this thesis lie in the two visualizations constructed. Both visualizations are based on large, complex, multivariate data-sets - one with, and the other without spatial bearings.

The three-dimensional visualization of air traffic control data presents a very simple and novel interactive interface to visualize ATC data. The background colour, symbology used to represent altitude constraints and landmarks, multiple orthographic and perspective views, and real-time rotation are techniques that, in the opinion of domain experts, greatly enhanced interaction and increased comprehension of data. The willingness of the Organisation to commission a design guide is indicative of its potential as an effective tool for future air traffic control.

The navigation tool entitled an Observer Participant User System or OPUS, presents a novel direct manipulation interface which enables information retrieval and seeking within a hypertext document. This is based on cognitive principles that support intuitive interaction to refine the search query, rather than relying on algorithms to rank the data in order of relevance. This 'drag, drop and draw line' method saliently supports Boolean syntax without users realising so. The tool also supports

the formation of a 'mental map', which is an important part of the information seeking process. Users also used statistical information to ascertain the presence of a term in the database and subsequently refine their search query. The simplicity of the user interface, its ease of use and efficiency in search-related tasks was confirmed through user studies.

6.3 Further Work

In the case of the three-dimensional radar display, it is worthwhile investigating the visualization of STandard Arrival Routes or STAR's, where incoming flights are stacked horizontally above each other. Since the structure of airspace created by STAR's is very different from that created by SID's, it may be worth exploring the potential of technologies such as virtual reality, which may be more suitable to visualize STAR's.

As explained before, the navigation tool for information seeking named an Observer Participant User System or OPUS requires implementing both the Observer and User states, apart from the Participant state implemented here. Highly effective information retrieval systems of the future would require a deeper investigation into indexing. The 'Symantec' mode introduced in OPUS suggests an IR system that brings 'meaning' into the picture by retrieving search results which the system 'thinks' are closer in meaning to the results anticipated by the user. It is also suggested that in the case of indexing, perhaps the grammar of the specific language in which the documents are required to be indexed, guides the structure of the index file.

Finally, information visualization applications would benefit from applying data graphic design principles, and significantly aid our understanding of large and complex multivariate data sets, which in turn may be immensely beneficial for sectors of the economy such as healthcare, education, defense, retail, agriculture, to name a few.

6.4 Closing

The case studies explored in Part 2 are both examples of visualizations where "the *information becomes the interface*."[206] Applying principles of data graphics to two types of data - spatial (air traffic control) and non-spatial (a hypertext document) has yielded high-density multivariate pieces of work. The three-dimensional representation of data in the radar display and the two-dimensional interactive interface of the navigation tool are both escapes from what Tufte calls 'flatlands', where the user is forced to ask relevant questions, query the system, receive feedback and make appropriate judgements. In both cases, perceptual features that 'pop-out', such as colour, shape, position or length, have been utilised by the human perceptual system to rapidly recognise elements of interest

206 Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997, p 146, emphasis in original

and subsequently led to "discovery, decision-making and explanation"[207] which, as explained earlier, are the main goals of visualization. Design principles derived from the study of data graphics in Part 1 trascend the medium in which they are produced and presented. These principles are often ignored in our contemporary digital world, where the task of revealing data often leads to the unintended consequence of obscuring it. In the digital age, the cost of a "grapho (the graphical equivalent of a typo)"[208] is, in critical applications (such as ATC), unacceptable.

The exponential proliferation of digital technologies such as personal computers, mobile phones and personal diaries, and their interconnectivity via fixed or wireless networks will continue to greatly extend the electronic environment around us. The term 'ubiquitous computing' vividly captures a vision where computers and peripheral devices integrate seamlessly in the environment surrounding us, and in some cases, adapt and augment our physical bodies as well. Our capacities to generate and disseminate digital data will grow many fold. In trying to navigate through this dense cloud of data, human beings are likely to take "the path of least cognitive resistance."[209] Against this background, designing highly effective interactive information environments, which can "amplify our natural abilities and proclivities rather than create new methods and tools that themselves must be learned"[210] and where users can extract knowledge from data, remains a key challenge in this newly established field of information visualization.

207 Schneiderman B., S.K. Card and J.D. Mackinlay (Editors), *Readings in Information Visualization: Using Vision To Think*, San Francisco: Morgan Kaufmann, 1999, p 6

208 Wainer, Howard, 'Making Newspaper Graphs Fit to Print', in Kolers, Paul A., Merald E. Wrolstad and Herman Bouma (Editors), *Processing of Visible Language: Volume 2*, New York; London: Plenum Press, 1980, p 135

209 Sutcliffe, A.G., M. Ennis and S.J. Watson, Empirical Studies of End-User Information Searching, *Journal of the American Society for Information Science*, Vol 51, No 13, 2000, pp 1211 - 1231

210 Marchionini, Gary, *Information Seeking in Electronic Environments*, Cambridge: Cambridge University Press, 1995, p ix

Bibiliography

Bibliography

Books

Abbott, Edwin A., *Flatland: A Romance of Many Dimensions by A Square*, Oxford: Basil Blackwell, 1978

Allingham, Michael, *Choice Theory: A Very Short Introduction*, New York: Oxford University Press, 2002

Arnheim, Rudolph, *Entropy and Art: An Essay on Disorder and Order*, Berkeley: University of California Press, 1971

Arnheim, Rudolph, *Visual Thinking*, Berkeley: University of California Press, 1997

Bateson, Gregory, *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution and Epistomology*, St Albans: Paladin, 1973

Bateson, Gregory, *Mind and Nature: A Necessary Unity*, London: Fontana, 1980

Bacyer, Hans Christian von, *Information: The New Language of Science*, London: Phoenix, Orion Books, 2003

Berlin, Brent, and Paul Kay, *Basic Color Terms: Their Universality and Evolution*, London: University of California Press, 1969

Burnett, Ron, *How Images Think*, Cambridge, Massachusetts: MIT Press, 2004

Coward, H. G. and K. K. Raja (Editors), *Encyclopedia of Indian Philosophies: Vol 5: The Philosophy of the Grammarians*, Princeton, New Jersey: Princeton University Press, 1990

Calvino, Italo, translated from The Italian by William Weaver, *Invisible Cities*, London: Vintage, 1997

Card, Stuart K., Jock D. Mackinlay and Ben Schneiderman (Editors), *Readings in Information Visualization: Using Vision To Think*, San Francisco: Morgan Kaufmann, 1999

Chomsky, Noam, *Syntactic Structures*, The Hague: Mouton & Co., 1957

Cleveland, William S., *The Elements of Graphing Data*, Monterey, California: Wadsworth, 1985

Culler, Jonathan, *Saussure*, Second Edition, London: Fontana Press, 1985

Drake, S., *Discoveries and Opinions of Galileo*, New York: DoubleDay, 1957

Drake, Stillman, *Galileo Studies: Personality, Tradition and Revolution*, Ann Arbor: University of Michigan Press, 1970

Drake, Stillman, *Galileo*, Oxford: Oxford University Press, 1980

Denning, Peter J. and Robert M. Metcalfe, *Beyond Calculation: The Next Fifty Years of Computing*, New York: Copernicus, Springer-Verlag, 1997

Denning, Peter J. (Editor), *Talking Back to the Machine: Computers and Human Aspiration*, New York: Copernicus, Springer-Verlag, 1999

Dretske, Fred I., *Knowledge and the Flow of Information*, Oxford: Basil Blackwell, 1981

Fayyad, Usama, Georges G. Grinstein and Andreas Wierse (Editors), *Information Visualization in Data Mining and Knowledge Discovery*, San Francisco: Morgan Kaufmann, Academic Press, 2002

Foley, James D., Andries van Dam, Steven K. Feiner and John F. Hughes, *Computer Graphics: Principles and Practice*, Reading, Massachussets: Addison-Wesley, 1996

Forty, Adrian, *Words and Buildings: A Vocabulary of Modern Architecture*, London: Thames & Hudson, 2000

Feynman, Richard P., *QED: The Strange Theory of Light and Matter*, Princeton, New Jersey: Princeton University Press, 1985

Grayling, A. C., *Wittgenstein: A Very Short Introduction*, New York: Oxford University Press, 2001

Galilei, Galileo, (translator) Stillman Drake, *Discoveries and Opinions of Galileo*, Garden City, New York: DoubleDay, 1957

Gowers, Timothy, *Mathematics: A Very Short Introduction*, New York: Oxford University Press, 2002

Harris, Robert L., *Information Graphics: A Comprehensive Illustrated Reference: Visual Tools for Analyzing Managing and Communicating*, New York: Oxford University Press, 1999

Hinton, Charles Howard, (Editor) Rudolph v.B. Rucker, *Speculations on the Fourth Dimension: Selected Writings of Charles H. Hinton*, New York: Dover Publications, 1980

Herdeg, Walter (Editor), *Graphis Diagrams: The Graphic Visualization of Abstract Data*, Zurich: The Graphis Press, 1974

Holtzman, Steven R., *Digital Mantras: The Languages of Abstract and Virtual Worlds*, Cambridge, Massachusetts: MIT Press, 1994

Hopkin, V. D., *Human Factors in Air Traffic Control*, London: Taylor & Francis, 1995

Imhof, Eduard, (Editor) H. J. Steward, *Cartographic Relief Presentation*, Berlin: Walter de Gruyter & Co., 1982

Kaufman, Lloyd, *Sight and Mind: An Introduction to Visual Perception*, New York; London: Oxford University Press, 1974

Kaku, Michio, *Hyperspace: A Scientific Odyssey through Parallel Universes, Time Warps and the Tenth Dimension*, Oxford: Oxford University Press, 1994

Koffka, Kurt, *Principles of Gestalt Psychology*, London: Routledge & Kegan Paul, 1962

Kolers, Paul A., Merald E. Wrolstad and Herman Bouma (Editors), *Processing of Visible Language: Volume 1*, New York; London: Plenum Press, 1979

Kolers, Paul A., Merald E. Wrolstad and Herman Bouma (Editors), *Processing of Visible Language: Volume 2*, New York; London: Plenum Press, 1980

Lidwell, William, Katrina Holden and Jill Butler, *Universal Principles of Design: 100 ways to enhance usability, influence perception, increase appeal, make better design decisions, and teach through design*, Rockport Publishers, 2003

March, Lionel, *Architectonics of Humanism: Essays on Number in Architecture*, Chichester; New York: Academy Editions, 1998

Marchionini, Gary, *Information Seeking in Electronic Environments*, Cambridge: Cambridge University Press, 1995

Marey, E. J., *La Methode Graphique: Dans Les Sciences Experimentales Et Principalement En Physiologie Et En Medecine*, Paris: G. Masson, 1885

McLuhan, Marshall, *Understanding Media: The Extensions of Man*, Cambridge, Massachusetts: MIT Press, 1999

Matilal, Bimal Krishna, *The Word and the World: India's Contribution to the Study of Language*, New Delhi: Oxford University Press, 2001

Mitchell, William J., *City of Bits: Space, Place, and the Infobahn*, Cambridge, Massachusetts: MIT Press, 1995

Moggridge, Bill, *Designing Interactions*, Cambridge, Massachussets: MIT Press, 2007

Monmonier, Mark, *How to Lie with Maps*, Chicago: University of Chicago Press, 1996

Morrison, Philip, and Phylis Morrison and, The Office of Charles and Ray Eames, *Powers of Ten: About the Relative Size of Things in the Universe*, New York: Scientific American Books, 1982

Nielsen, Jakob, *Usability Engineering*, Boston: AP Professional, 1993

Norman, Donald A., *Emotional Design: Why We Love (or Hate) Everyday Things*, New York: Basic Books, 2004

Ostwald, Wilhelm, (translator) J. Scott Taylor M.A., *Colour Science: Vol I: Colour Theory and Standards of Colour*, London: Winsor & Newton, 1931

Owen, William, (Editor) Roger Fawcett-Tang, *Mapping: An illustrated guide to graphic navigational symbols*, Mies, Switzerland: RotoVision, 2002

Pascoe, David, *Airspaces*, London: Reaktion, 2001

Robinson, Arthur H., Randall D. Sale and Joel L. Morrison, *Elements of Cartography*, New York; Chichester: Wiley, 1978

Saussure, Ferdinand de, *Course in General Linguistics*, (Editors) Charles Bally, Albert Sechehaye, Albert Riedlinger, (translator) Wade Baskin, New York: McGraw-Hill, 1966

Schneiderman, Ben, *Designing the User Interface: Strategies for Effective Human-Computer Interaction*, Reading, Massachussets: Addison-Wesley, 1998

Schrodinger, Erwin, *What Is Life?: with Mind and Matter and Autobiographical sketches*, Cambridge: Cambridge University Press, 1967

Shannon, Claude E. and Warren Weaver, *The Mathematical Theory of Communication*, Urbana: The University of Illinois Press, 1962

Silver, Mike, and Diana Balmori (Editors) *Mapping in the Age of Digital Media: The Yale Symposium*, London: Wiley, Academy Editions, 2003

Stewart, Ian, *Flatterland: Like Flatland Only More So*, London: Macmillan, 2001

Thompson, D'Arcy Wentworth, (Editor) John Tyler Bonner, *On Growth and Form*, Cambridge: Cambridge University Press, 1992

Tukey, John Wilder, *Exploratory Data Analysis*, Reading, Massachussets: Addison-Wesley, 1977

Tufte, Edward R., *The Visual Display of Quantitative Information*, Cheshire, Connecticut: Graphics Press, 1983

Tufte, Edward R., *Envisioning Information*, Cheshire, Connecticut: Graphics Press, 1990

Tufte, Edward R., *Visual Explanations: Images and Quantities, Evidence and Narrative*, Cheshire, Connecticut: Graphics Press, 1997

Tufte, Edward R., *The Cognitive Style of PowerPoint*, Cheshire, Connecticut: Graphics Press, 2003

Tufte, Edward R., *Beautiful Evidence*, Cheshire, Connecticut: Graphics Press, 2006

Veltman, Martinus, *Diagrammatica: The Path to Feynman Diagrams*, Cambridge: Cambridge University Press, 1994

Ware, Colin, *Information Visualization: Perception for Design*, San Francisco: Morgan Kaufman, 2000

Wainer, Howard, *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*, Princeton: Princeton University Press, 2005

Wiener, Norbert, *Cybernetics: Or, Control and Communication in the Animal and the Machine*, Cambridge, Massachussetts: MIT Press, 1982

Wildbur, Peter, and Michael Burke, *Information Graphics: Innovative Solutions in Contemporary Design*, London: Thames and Hudson, 1998

Wildbur, Peter, *Information Graphics: A Survey of Typographic, Diagrammatic and Cartographic Information*, London: Trefoil, 1989

Wilkinson, Leland, *The Grammar of Graphics*, New York: Springer-Verlag, 1999

Wyszecki, Judd, and Deane B. Gunter, *Color in Business Science and Industry*, New York: Wiley, 1975

Young, Paul, *The Nature of Information*, New York: Praeger, 1987

Zusne, Leonard, *Visual Perception of Form*, New York; London: Academic Press, 1970

Journals

Azuma, Ronald, Mike Daily and Jimmy Krozel, Advanced Human-Computer Interfaces for Air Traffic Management and Simulation, *Proceedings of 1996 AIAA Flight Simulation Technologies Conference*, 29-31 July 1996, pp 656 - 666

Bartholomew, David J., What is Statistics?, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, Vol 158, No 1, 1995, pp 1 - 20

Bayes, Thomas, An Essay towards solving a Problem in the Doctrine of Chances, *Philosophical Transactions of the Royal Society of London*, 53, 1763, pp 370 - 418

Benford, Steve *et al*, Three-Dimensional Visualization of the World Wide Web, *ACM Computing Survey*, Vol 31, No 4, December 1999, pp 1 -16

Bentley, R. *et al*, Enthnographically-informed systems design for air traffic control, *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, November 1992, pp 123 - 129

Biederman, Irving and Peter C. Gerhardstein, Recognizing Depth-Rotated Objects: Evidence and Conditions for Three-Dimensional Viewpoint Variance, *Journal of Experimental Psychology: Human Perception and Performance*, Vol 19, No 6, 1993, pp 1162 - 1182

Bland, J. Martin and Douglas G. Altman, Statistics notes: Bayesians and frequentists, *British Medial Journal*, Vol 317, October 24, 1998, pp 1151 - 1160

Brown, Mark A. and Professor Mel Slater, Some Experiences with Three-Dimensional Display Design: An Air Traffic Control Visualization, *IEEE International Workshop on Robot and Human Communication*,

1997, pp 296 - 301

Campbell, Stuart, 4D, or not 4D: that is the question, *Ultrasound Obstet Gynecol*, Vol 19, 2002, pp 1 - 4

Campbell, Stuart *et al*, Ultrasound Scanning in pregnancy: the short-term psychological effects of early realtime scans, *J Psychosom Obstet Journal*, 1 - 2, pp 57-61

Chen, Chaomei and Roy Rada, Interacting with Hypertext: A Meta-Analysis of Experimental Studies, *Human-Computer Interaction*, Vol 11, 1996, pp 125 - 156
Chernoff, Hermnan, The Use of Faces to Represent Points in K-Dimensional Space Graphically, *Journal of the American Statistical Association*, Vol 68, No 342, June 1973, pp 361 - 368

Cilibrasi, Rudi and Paul Vitanyi, Automatic Meaning Discovery Using Google, *arXiv:cs*, March 15, 2005

Cleveland, William S., Graphs in Statistical Publications, *The American Statistician*, Vol 38, No 4, November 1984, pp 261 - 269

Cleveland, William S. and Robert McGill, Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods, *Journal of the American Statistical Association*, Vol 79, No 387, September 1984, pp 531 - 554

Conklin, Jeff, Hypertext: An Introduction and Survey, *Computer*, Vol 20, No 9, September 1987, pp 17 - 41

Daub, Edward E., Probability and Thermodynamics: The Reduction of the Second Law, *Isis*, Vol 60, No 3, Autumn 1969, pp 318 - 330

de Vries, J.I.P.; G.H.A. Visser and H.F.R. Prechtl, The emergence of fetal behaviour: I: Qualitative aspects, *Early Human Development*, Vol 7, 1982, pp 301 - 322

de Vries, J.I.P., G.H.A. Visser and H.F.R. Prechtl, The emergence of fetal behaviour: II: Quantitative aspects, *Early Human Development*, Vol 12, 1985, pp 99 - 120

Dumais, Susan and Edward Cutrell, Optimizing Search by Showing Results in Context, *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2001, pp 277 - 284

Eaton J.J. and D. Bawden, What Kind of Resource is Information?, *International Journal of Information Mangement*, Vol 11, 1991, pp 156-165

Eick, Stephen G., Information Visualization at 10, *IEEE Computer Graphics and Applications*, Vol25, No 1, January/February 2005, pp 12 - 14

Ellis, Stephen R., On the Design of Perspective Displays, *Proceedings, IEA 2000 / HFES 2000 44th Annual Meeting, Human Factors & Ergonomics Society*, 2000

Funkhouser, H. Gray, Historical Development of the Graphical Representation of Statistical Data, *Osiris*, Vol 3, 1937, pp 269 - 404

Gardner, Martin, Mathematical Games, On the fanciful history and the creative challenges of the puzzle game of tangrams, *Scientific American*, 1974, pp 98 - 103B

Goldstein, Donald A. and Jerry C. Lamb, Visual Coding Using Flashing Lights, *Human Factors*, Vol 9, No 5, 1967, pp 405 - 408

Grenander, Ulf and Michael I. Miller, Representations of Knowledge in Complex Systems, *Journal of the Royal Statistical Society, Series B (Methodological)*, Vol 56, No 4, 1994, pp 594 - 603

Haskell, Ian D., and Christopher D. Wickens, Two- and Three- Dimensional Displays for Aviation: A Theoretical and Empirical Comparison, *The International Journal of Aviation Psychology*, Vol 3, No 2, 1993, pp 87 - 109

Healey, Christopher G., Kellog S. Booth and James T. Enns, High-Speed Visual Estimation Using Preattentive Processing, *ACM Transactions on Computer-Human Interaction*, Vol 3, No 2, June 1996, pp 107-135

Healey, Christopher G., Choosing Effective Colours for Data Visualization, *Proceedings, Seventh IEEE Visualization 1996 (VIS '96)*, 1996, pp 263 - 270

Hendley, R.J., N.S. Drew, A.M. Wood and R. Beale, Case Study: Narcissus: Visualizing Information, *Proceedings, IEEE Symposium on Information Visualization*, 1995, pp 90 - 96

Hibbard, Bill, Top Ten Visualization Problems, *ACM SIGGRAPH Computer Graphics*, Vol 33, No 2, May 1999. pp 21 - 22

Hill, Austin Bradford, The Environment and Disease: Association or Causation, *Proceedings of the Royal Society of Medicine*, Vol 58, 1965, pp 295 - 300

Hogg, Robert V., Statistical Education: Improvements Are Badly Needed, *The American Statistician*, Vol 45, No 4, November 1991, pp 342 - 343

Jansen, Bernard J. and Amanda Spink, How are we searching the World Wide Web?: A comparison of nine search engine transaction logs, *Information Processing and Management*, Vol 42, No 1, January 2006, pp 248 - 263

Johnson, Chris, Vizualization Viewpoints: Top Scientific Visualization Research Problems, *IEEE Computer Graphics and Applications*, July/August 2004, pp 13 - 17

Kosara, Robert, Christopher G. Healey, Victoria Interrante, David H. Laidlaw and Colin Ware, Thoughts on User Studies: Why, How and When, *IEEE Computer Graphics and Applications*, Vol 23, No 4, July/August 2003, pp 20 -25

Kules, Bill and Ben Schneiderman, Categorized graphical overviews for web search results: An exploratory study using U. S. government agencies as a meaningful and stable structure, *Proceedings of the Third Annual Workshop on HCI Research in MIS*, December 10 -11 2004, pp 1 - 5

Larkin, Jill H. and Herbert A. Simon, Why a Diagram is (Sometimes) Worth Ten Thousand Words, *Cognitive Science*, Vol 11, 1987, pp 65 - 99

Lorensen, Bill, On The Death of Visualization, *Position Papers NIH/NSF Proceedings, Fall 2004 Workshop Visualization Challenges*, 2004

Mackay, Wendy E., Anne-Laure Fayard, Laurent Frobert and Lionel Medini, Reinventing the Familiar: Exploring an Augmented Reality Design Space for Air Traffic Control, *Proceedings of the SIGCHI conference on Human factors in computing systems*, April 18 - 23, 1998, pp 558 - 565

Mackay, Wendy E., Is Paper Safe?: The Role of Paper Flight Strips in Air Traffic Control, *ACM Transactions on Computer-Human Interaction*, Vol 6, No 4, December 1999, pp 311 - 340

Maunder, E. Walter, Note on the Distribution of Sun-spots in Heliographic Latitude, 1874 - 1902, *Monthly Notices of the Royal Astronomical Society*, Vol 64, June 1904, pp 747 - 761

Meek, Steve and Francis Morris, ABC of clinical electrocardiography: Introduction: 1: Leads, rate, rhythm, and cardiac axis, *British Medical Journal*, Volume 324, 16 February 2002, pp 415 - 418

Miller, George A., The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Information Processing, *The Psychological Review*, Vol 63, No 2, March 1956, pp 81 - 97

Nielsen, Jakob, The Art of Navigating Hypertext, *Communications of the ACM*, Vol 33, No 3, March 1990, pp 296 - 310

Nielsen, Jakob, User Interface Directions for the Web, *Communications of the ACM*, Vol 42, No 1, January 1999, pp 65 - 72

Nygren, Else, Between the clicks: Skilled Users Scanning of Pages, *Proceedings of Designing for the Web: Empirical studies*, October 1996

Peterson, Harold E. and Doris J. Dugas, The Relative Importance of Contrast and Motion in Visual Detection, *Human Factors*, Vol 14, No 3, 1972, pp 207 - 216

Ribarsky, William *et al*, Glyphmaker: Creating Customized Visualizations of Complex Data, *Computer*, July 1994, pp 57 - 64

Savant, Marilyn vos, Letters to the Editor, *The American Statistician*, November 1991, Vol 45, No 4, p 347

Saracevic, Tefko, Paul Kantor, Alice Y. Chamis and Donna Trivison, A Study of Information Seeking and Retrieving: I: Background and Methodology, *Journal of the American Society for Information Science*, Vol 39, No 3, 1988, pp 161 - 176

Saracevic, Tefko and Paul Kantor, A Study of Information Seeking and Retrieving: II: Users, Questions and Effectiveness, *Journal of the American Society for Information Science*, Vol 39, No 3, 1988, pp 177 - 196

Saracevic, Tefko and Paul Kantor, A Study of Information Seeking and Retrieving: III: Searchers, Searches and Overlap, *Journal of the American Society for Information Science*, Vol 39, No 3, 1988, pp 197 - 216

Schneiderman, Ben, Direct Manipulation: A Step Beyond Programming Languages, *Computer*, August 1983, pp 57 - 69

Sen, Amartya, Freedom of Choice: Concept and Content, *European Economic Review*, Vol 32, No 2 - 3, March 1988, pp 269 - 294

Spink, Amanda, Judy Bateman and Bernard J. Jansen, Searching the Web: a survey of EXCITE users, *Internet Research: Electronic Networking Applications and Policy*, Vol 9, No 2, 1999, pp 117 - 128

Spink, Amanda, T.D. Wilson, Nigel Ford, Allen Foster and David Ellis, Information-Seeking and Mediated Searching: Part 1: Theoretical Framework and Research Design, *Journal of the American Society for Information Science and Technology*, Vol 53, No 9, 2002, pp 695 - 703

Spink, Amanda, T.D. Wilson, Nigel Ford, Allen Foster and David Ellis, Information-Seeking and Mediated Searching: Part 2: Uncertainity and its Correlates, *Journal of the American Society for Information Science and Technology*, Vol 53, No 9, 2002, pp 704 - 715

Spink, Amanda, T.D. Wilson, Nigel Ford, Allen Foster and David Ellis, Information-Seeking and Mediated Searching: Part 3: Successive Searching, *Journal of the American Society for Information Science and Technology*, Vol 53, No 9, 2002, pp 716 - 727

Stigler, Stephen M., Thomas Bayes's Bayesian Inference, *Journal of the Royal Statistical Society, Series A (General)*, Vol 145, No 2, 1982, pp 250 - 258

Sutcliffe, A.G., M. Ennis and S.J. Watson, Empirical Studies of End-User Information Searching, *Journal of the American Society for Information Science*, Vol 51, No 13, 2000, pp 1211 - 1231

Tilling, Laura, Early Experimental Graphs, *British Journal for the History of Science*, Vol 8, No 30, 1975, pp 193 - 211

Tufte, Edward R., Dynamic Graphs for Data Analysis, *Statistical Science*, Vol 2, No 4, November 1987, pp 389 - 392

Triesman, Anne, Preattentive Processing in Vision, *Computer Vision, Graphics and Image Processing*, Vol 31, 1985, pp 156 - 177

Walsum, Theo van, *et al*, Feature Extraction and Iconic Visualization, *IEEE Transactions on Visualization and Computer Graphics*, Vol 2, No 2, June 1996, pp 111 - 119

Watson, J.D. and F.H.C. Crick, Genetical Implications of the Structure of Deoxyribonucleic Acid, *Nature*, Vol 171, May 30, 1953, pp 964 - 967

Wattenberg, Martin, Visualizing the Stock Market, *Conference on Human Factors in Computing Systems*, 1999, pp 188 - 189

Wickens, Christopher D., Steve Fadden, David Merwin and Patrica May Ververs, Cognitive Factors in Aviation Display Design, *Proceedings of The 17th AIAA/IEEE/SAE Digital Avionics Systems Conference (DASC)*, Vol 1, 31 October - 7 November, 1998, pp E32/1 - E32/8

Wijk, Jarke J. van, Views on Visualization, *IEEE Transactions on Visualization and Computer Graphics*, Vol 12, No 4, July/August 2006, pp 421- 432

Wittenbrink, C.M., A.T. Pang and S.K. Lodha, Glyphs for Visualizing Uncertainity in Vector Fields, *IEEE Transactions on Visualization and Computer Graphics*, Vol 2, No 3, September 1996, pp 266-279

Zachry, Mark and Charlotte Thralls, Cross-Disciplinary Exchanges: An Interview with Edward R. Tufte, *Technical Communications Quarterly*, Vol 13, No 4, 2004, pp 447 - 462

Zhang, Jiajie, The Nature of External Representations in Problem Solving, Cognitive Science, Vol 21, No 2, 1997, pp 179 - 271

Reports

McCormick, Bruce H., Thomas A. Defanti and Maxine D. Browne, 'Visualization in Scientific Computing', New York, ACM Siggraph, 1987

Wickens, Christopher D. and Patricia May, 'Terrain Representation for Air Traffic Control: A Comparison of Perspective with Plan View Displays', Prepared for Federal Aviation Administration, Systems Operations & Engineering Branch, Washington DC, September 1994

Wickens, Christopher D., M. Campbell, C.C. Lian and D.H. Merwin, 'Weather Displays for Air Traffic Control: The Effect of 3D Perspective', Prepared for Federal Aviation Administration, Systems Operations & Engineering Branch, Washington DC, January 1995

Xing, Jing, 'Color and Visual Factors in ATC Displays', Final Report, Civil Aerospace Medical Institute, Federal Aviation Administration, Oklahoma/Washington DC, June 2006

Department for Transport, 'White Paper: The Future of Air Transport', Presented to Parliament by the Secretary of State for Transport by command of Her Majesty, December 2003

NATS document, Section 2: Command Entry and Display Symbology, Vol 1, Issue 11

NATS document, Section 3: General MMI Facilities, Vol 1, Issue 11

NATS document, Section 4: Command Palettes, Vol 1, Issue 11

Other

Schmidt, Eric, 'Don't bet against the internet', *The Economist: The World in 2007*, p 130

'How do Internet search engines work?', Scientific American, January 2003, p 82

Internet

http://www.hawkeyeinnovations.co.uk

http://www.pcquest.com/content/cricket/103030404.asp

www.answers.com/radar

http://www.ilo.org/public/english/protection/safework/stress/trafctr.htm

http://www.guardian.co.uk/uk_news/story/0,,1348184,00.html

http://news.bbc.co.uk/1/hi/uk/3772022.stm

http://kosara.net/research/infovis/index.html

http://www.cybergeography.org

http://www.infovis-wiki.net

http://www.useit.com

http://www.edwardtufte.com

http://www4.ncsu.edu/~healey/PP/index.html

http://www.cs.brown.edu/people/ifc/hci/shneiderman.html

Video

'Life Before Birth', Channel 4, broadcast <April 07, 2005 at 21:00 hrs>

Appendix A: Transcript and Email

Transcript

Excerpts of conversation between the author and Mark Green, ATC Planner, National Air Traffic Services Ltd on June 30, 2004

JK: We talked last time about two approaches, the most interesting was the air tunnel one where the plane moves through its own airspace, so you always see the airspace as it moves.

MG: That is one of the concepts we are currently developing for deployment. Operational deployment will probably be around 2015 for that concept. But we're already looking at it and doing fast-time modelling at the moment and that's why we're very interested in the 3d visualization capability that you have with this tool.

JK: This is what we (I) came up with....(demo of model)

MG: That's very incredible, I mean, the ability to rotate in three-dimensions in real-time is very impressive.

JK: One of the things I want to talk to you (about), probably towards the end is setting up a real case scenario within this system.

MG: Well, its funny you should say that because having seen this some of the visualization techniques you are using here would merge very well with some airspace development work that we are doing at the moment on something called the TC Northwest Sector, which is if you imagine the airspace around London we tend to divide the terminal areas of all the TMA sectors are divided typically into quadrants, so with those quadrants what we're doing at the moment is, I mentioned we're looking at the concept of looking at ARNAF to separate tracks so slightly like depicted here tunnels and sky, so the TC Northwest sector we have the Heathrow departures, and also departures from Stansted and from Luton and from London city; and all of this traffic is taking off and merging onto parallel routes to fly north. So we've already developed the airspace but what I would be very interested in doing is seeing if you can then take some real traffic scenarios that we've got developed and whether you could run it through an airspace visualization model for us.

MG: Its very useful just to visualize what the aircraft are doing, I mean, I'm extrapolating what the tool is showing us here and putting it in context of the development we're doing at the moment. I mean when we come to looking at the actual work that we are doing which is a project called SMART, which is Simulation, Modelling, Analysis of Arnav and the TMA, which is these 3d profiles, a visualization tool like this would be very very useful for us.

JK: for example the possibility of using icons, for example if there is a helicopter, then you see a helicopter; if there is a Boeing, then you see a Boeing.

MG: certainly for this future work, it would be interesting to see the aircraft divided into large jets, small jets, turbo props, piston aircraft, just to see a generic breakdown or a breakdown into generic categories.

MG: already what has been depicted there is very useful or has potential to be very useful from an airspace design point of view.

JK: way ahead...

MG: Well certainly if you're going to be available for another two and a half years, I think we can certainly provide you with lots of information that would benefit both yourself and the PhD, but we would also draw benefit for ourselves as the air traffic service provider, because I think these sorts of visualization techniques are the way ahead for us. I'm already investigating purchasing at the moment a tool that will enable us to look at airspace in a 3d perspective. But at the same time even though we are investigating that tool, by the time we purchase the tool and we look at making the necessary software changes to it, I think its certainly worthwhile pursuing the investigation here, and the area I think I'd like you to get involved with, especially if you're available for two and a half years, is working on this ARNAV concept of future airspace. What we need to do is basically identify sort of a roadmap. I mean I know where we would like to get to at the end - at the end I would like to see the whole of the TC Northwest sector and the Midlands sectors incorporated into this tool so you can look at the interactions between the arrivals and the departures flying through these 3d tunnels and I think that this tool would be ideal to demonstrate that. Already I'm impressed how you've taken the SID's and you've been able to model the SID's as they exist at the moment obviously incorporating the 3d profiles. They are certain areas which you've identified which we need to tidy up, things like the departures separations and so on to make it more operational acceptable, but I think it has the potential to be a very powerful tool. Especially this sort of view here - the Camera1 perspective - that is ideal.

MG: I've been very impressed by the performance, I mean the graphical processing is really great. All in all the demonstration today has been very powerful, and I would certainly like you to be involved in this airspace development work that we are doing.

From: "LAKHANI, Devang" <Devang.LAKHANI@nats.co.uk>
To: Jignesh Khakhar <KhakharJ@Cardiff.ac.uk>
Date: 28/09/2004 13:54:07
Subject: RE: Funding

Jig,

Please have a look at this for background info.

Dev

-----Original Message-----
From: Jignesh Khakhar [mailto:KhakharJ@Cardiff.ac.uk]
Sent: 24 September 2004 11:53
To: LEWIS Alison <Alison.Lewis@nats.co.uk
Cc: LAKHANI, Devang
Subject: Funding

Hi Alison,

The following is a breakdown of the funding required to undertake the 3d visualization project for NATS:

We discussed the requirement of employing an asssitant for the 3d modelling. On a salary basis of £12, 000 - £15, 000 per annum, this would yield a final estimate of:

£19, 656 - £ 24, 670 (includes NI + overheads)
+
£5, 000 (maximum) for equipment purchases (if required)
+
travel and subsistence (as required)

The School would like to know how you would like us to put this proposal forward in a formal manner ie writing?

Please do not hesitate to contact me for any further queries.

Looking forward to a response from you soon.

Many thanks.

Regards,

Jig

Meeting with Roy Byres

27/09/04

Agenda
To find out more about Luciad ATC Playback

- ATC Playback is used to replay traffic movements from fast time and real time simulators.
- The fast time simulators use software algorithms for simulating traffic movements whilst real time simulators use pseudo pilots for traffic movements
- The inputs to Luciad are
    - Navaid Data
    - Airport Data
    - Airspace/ground data
    - Airspace rules
    - Traffic Files – flight plans etc
    - Not radar data

  All the above are in ASCII format

Useful features in the existing software.

- Distance measuring tool; used for track lengths etc.
- Different colours for sectors based on traffic loads
- Future path "prediction" lines/stalks
- Cumulative trajectories, whisker plots for track data
- Trajectory profile for aircraft; requested vs. actual flight profile
- Filtering the display based on flight level and times
- Focusing in on areas of interest

Benefits of Jignesh's tools

- Can provide more graphical detail on tracks, aircraft, buildings, terrain etc
- Faster processing and allows for a multiple of camera views
- Cylindrical tracks whilst Luciad has lines only
- A more detailed simulation of an aerodrome, London TMA etc.
- More emergent and realistic aircraft behaviour.

Appendix B: ventilation.txt & ventilation.css files

Appendix B

Below are descriptions of the ventilation.txt and ventilation.css files described in Section 4.5.3, Chapter 4. Since the parameters of the ventilation.css file are repetitive, only part of the entire file is printed here. Both files can be viewed in their entirety on the CD accompanying this thesis.

"Ventilation.txt" file:

<html><title>search results: "ventilation"</title>

<body><1.a>ventilation<1.a><1.b> | process | supplying | outdoor | air | space | without | mechanical | assistance<1.b>

<2.a>controlled | intentionally | provided | openings | windows | doors | non-powered | <2.a><2.b> ventilators<2.b>

<3.a>openings | level | <3.a><3.b>ventilation<3.b><3.c> | addition | infiltration<3.c>

<4.a>infiltration | sufficient | outdoor | air | <4.a><4.b>ventilate<4.b><4.c> | building<4.c>

<5.a>infiltration | component | <5.a><5.b>ventilation<5.b><5.c> | excessive | waste | heat | incur | high | energy | penalty | building | difficult | heat | cool | comfort | levels<5.c>

<6.a>mechanical | <6.a><6.b> ventilation<6.b><6.c> | movement | air | mechanical | means | space<6.c>

<7.a>'build | tight | <7.a><7.b>ventilate<7.b><7.c> | right'<7.c>

<8.a>infiltration | naturally <8.a><8.b>ventilated<8.b><8.c> | mechanically |<8.c><8.d>ventilated<8.d><8.e> | spaces<8.e>

<9.a>'best | practice' | minimise | infiltration | applying | sealing | measures | controllable | natural | mechanical <9.a><9.b>ventilation<9.b>

<10.a>ventilation<10.a><10.b> | effectiveness | fraction | fresh | air | delivered | space | reaches | occupied | zone<10.b>

<11.a>ventilation<11.a><11.b> | efficiency | term | quantify | <11.b><11.c>ventilation<11.c><11.d> effectiveness<11.d>

<12.a>measure | ability | <12.a><12.b>ventilation<12.b><12.c> | system | exhaust | pollutants | generated | space<12.c>

<13.a>oldest | air | closest | extract | maximum <13.a><13.b>ventilation<13.b><13.c> effectiveness<13.c>

<14.a>naturally | <14.a><14.b>ventilated<14.b><14.c> | buildings | air | well | mixed | space | little | difference | air quality | <14.c><14.d>ventilation<14.d><14.e> | efficiency | space<14.e>

<15.a>ventilation<15.a><15.b> efficiency E = Ce - Cs / Cl - Cs<15.b>

</body></html>

"Ventilation.css" file:

```
title {
  color: #FFFFFF;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

1.a {
  color: #FFFFFF;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
```

```
  margin-right: 5px
}

1.b {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

2.a {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

2.b {
  color: #FFFFFF;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

3.a {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

3.b {
  color: #FFFFFF;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

3.c {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
```

```
  margin-right: 5px
}

4.a {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

4.b {
  color: #FFFFFF;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

4.c {
  color: #AAAAAA;
  font-family: Verdana;
  font-size: 12px;
  margin-left: 5px;
  margin-right: 5px
}

5.a {
...
}
```

The css file continues in the manner above, until the format of all tags upto <15.b> have been described.

Appendix C: master.xml index file

Appendix C

This section represents the 'master.xml' index file described in Section 4.5.6, Chapter 4. Including the entire xml file in print would unnecessarily add weight to this thesis. Hence, the 'Preview' web page of the Skins and Spaces website is presented here as seen online, and part of its corresponding translation into xml format in the master.xml file. The Skins and Spaces website, with all its web pages and the master.xml file can be viewed in their entirety on the CD accompanying this thesis.

Skins and Spaces 'Preview' web page

SKINS AND SPACES

PREVIEW

Introduction

This page will review the need for design of ventilation in buildings and will present some examples of ventilation design in modern buildings.

Why do we need ventilation?

Ventilation in buildings is needed to provide fresh air for occupants, to dilute and exhaust pollutants and to provide cooling in summer. Ventilation is also needed for the protection of the building and elements of its construction against moisture. A successful ventilation design should provide for good air quality (low levels of pollutants) and occupant comfort in terms of 'airy' (not draughty) and 'fresh' (not stuffy) spaces. Too little ventilation and the health and comfort of the occupants is at risk. Too much ventilation during the heating season will incur an energy penalty.

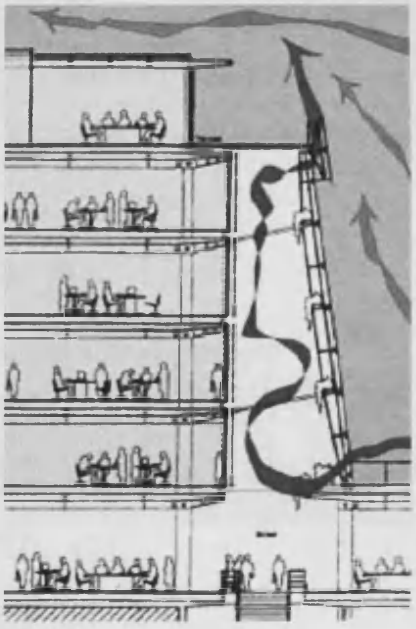


Figure 1: *Analysis of wind driven ventilation in an atrium. .*

What are the main aspects about ventilation design?

Traditionally, buildings were over-ventilated by virtue of their leaky construction, chimneys and ill fitting components. As a consequence ventilation was uncontrolled and buildings were difficult to heat, to maintain comfort and keep dry.

Since the 1970's buildings have become more energy efficient. The heating (and cooling) requirements have been reduced by increased levels of insulation and reduced air leakage. Indeed, the heating requirements of modern buildings are sometimes so low that the main function of the environmental system is no longer to provide heating but to provide ventilation, whether by natural or mechanical means.

A successful ventilation system must be controllable in response to the requirements of the buildings occupants. This is often difficult to achieve because the total ventilation rate of a building is a mix of purpose ventilation (which may be natural or mechanical, or a mixture of both) and infiltration (that is, unintentional air leakage, which by its nature is uncontrollable).

A controllable ventilation system therefore needs to be carefully designed. This means attention to:

- the airtightness of the building envelope;
- the method of air delivery and extract;

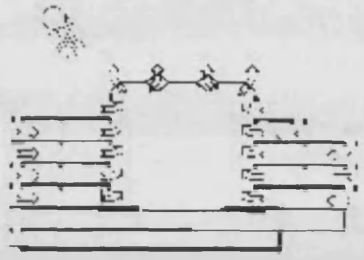


Figure 2 & 3: *The warm air being exhausted through the roof of the atrium in the Gateway II office building is used to induce an airflow through the adjoining offices thus providing natural ventilation to a deep plan space.*

Figure 4: *The chimneys of the Queens building at DeMontfort University provide the exhaust for the naturally ventilated lecture theatres and other teaching spaces*

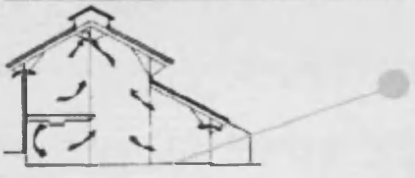


Figure 5 & 6: *'Seasonal hybrid' ventilation system at Trinity College Library, Carmarthen*

- the control of air delivery and room air distribution.

The ventilation design of modern buildings should be based on the principle 'build tight' and 'ventilate right', in order to provide building occupants with good air quality and comfortable conditions for the most efficient use of energy.

Natural ventilation in complex buildings

Natural ventilation is often considered to be the most energy efficient and healthy solution. It is usually considered to be a fundamental part of 'passive design', which is the term used to describe the integrative design approach, involving the use of daylight, thermal mass, insulation, solar radiation and ventilation - and as such ventilation design should not be considered in isolation from these other design factors.

There are some good examples of passive design non-domestic buildings such as the Gateway II office (**Figure 2 & 3**) and the Queens Building at DeMontfort University (**Figure 4**) in which the building form is an integral part of the ventilation design.

Both Gateway II and the Queens Building make use of natural daylight and solar shading to reduce space heat gains combined with utilising the thermal mass of the construction to absorb the peak heat gains. These measures eliminate the need for mechanical ventilation and cooling in summer.

Natural ventilation can be successfully applied in the design of complex buildings- but such solutions are not easy!

Supplementing natural ventilation

Many buildings have some form of mechanical ventilation system in combination with natural ventilation. There is currently a great deal of speculation and innovation in the design of such 'hybrid' ventilation systems. For example, ventilation for some spaces may need mechanical assistance at different times of the year. Trinity College Library (**Figure 5 & 6**) is naturally ventilated using roof turrets and opening windows in summer. However, it has mechanical ventilation with heat recovery in winter, with low level supply ducts and high level extract ducts. This reduces the heat loss due to the high ventilation rates required for the high occupancy levels in the reading areas, and avoids the potential discomfort due to cold draughts caused by opening windows.

Hybrid or mixed ventilation

Spaces that are predominantly naturally ventilated may require some mechanical ventilation assistance in certain areas. For example, some spaces within buildings or zones within a space may need to be air conditioned to deal with high heat gains in summer or from high internal loads. The Exhibition Hall in Linz (**Figure 7 & 8**) is a daylit space which is air conditioned in the occupied zone through a floor displacement system, but which also uses natural ventilation through side (supply) and top (exhaust) louvres to exhaust high level glazing heat gains in summer.

When mechanical ventilation is necessary

In recent years there has been a reaction against total mechanically ventilated or air-conditioned buildings, largely on the grounds of energy efficiency. The result has been the trend to natural ventilation

Figure 7 & 8: *Exhibition Hall Linz with a 'spatial hybrid' ventilation system.*

Figure 9: *London Ark*

solutions. However, an overreaction to mechanical solutions should be avoided. For many buildings, their form, their function and site constraints require that they be mechanically ventilated. Such solutions need not be inefficient in energy use. However, the same 'holistic' approach needs to be applied to ventilation design in mechanically ventilated buildings as generally applied to the more innovative type of naturally ventilated buildings, if an energy efficient solution is to be achieved. Indeed, for some buildings mechanical ventilation can provide the most energy efficient design as well as providing the best option for satisfying the health and comfort needs of the building occupants.

The London Ark (**Figure 9**) is considered to represent an efficient air-conditioned building. It provides passive cooling through chilled fins in the open ceiling space. The amount of air supplied mechanically is therefore reduced to that required to satisfy the ventilation requirements of the occupants. This allows an efficient 100% fresh air supply system, and a reduction in plant space requirements. Duct sizes are reduced and space for mechanical HVAC services is minimised.

No longer is the choice between natural ventilation and mechanical ventilation or air conditioning. An appropriate solution needs to be designed in response to the building, its pattern of use, the site and its constraints, and the requirements of the occupants.

There is a spectrum of possible solutions for ventilation design that range from pure natural ventilation, through the various 'hybrid' solutions, to total mechanical ventilation and finally to the fully air-conditioned building. A modern energy efficient and environmental friendly building is more likely to lie within this spectrum than at either the extremes of pure natural ventilation or total air conditioning.

Ventilation design is difficult! Whether it is achieving a good natural solution through the building design or an efficient mechanical solution though the selection of equipment and controls, the solution must be appropriate to the needs of the user. The end product is the design of a building that people can feel good in - which must be worthwhile.

The best chance of achieving an optimised solution is when the architect and services engineer work together from the early stages of the building design, where decisions about the building form, planning, provision for openings, ducts and equipment are still to be finalised.

master.xml **index file (for web page above)**

```
<?xml version="1.0" encoding="utf-8"?>
<opus>
        <line>
                <word link="MSc/ss-02-preview.htm?#h1">skins</word>
                <word>and</word>
                <word>spaces</word>
        </line>
        <line>
```

```
        <word link="MSc/ss-02-preview.htm?#h2">preview</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#h3">introduction</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p1">page</word>
        <word>review</word>
        <word>need</word>
        <word>design</word>
        <word>ventilation</word>
        <word>buildings</word>
        <word>present</word>
        <word>some</word>
        <word>examples</word>
        <word>ventilation</word>
        <word>design</word>
        <word>modern</word>
        <word>buildings</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#h4">why</word>
        <word>do</word>
        <word>we</word>
        <word>need</word>
        <word>ventilation?</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p2">ventilation</word>
        <word>buildings</word>
        <word>needed</word>
        <word>provide</word>
        <word>fresh</word>
        <word>air</word>
        <word>occupants,</word>
        <word>dilute</word>
        <word>exhaust</word>
        <word>pollutants</word>
        <word>provide</word>
        <word>cooling</word>
        <word>summer</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p2">ventilation</word>
        <word>also</word>
        <word>needed</word>
        <word>protection</word>
        <word>building</word>
```

```
        <word>elements</word>
        <word>construction</word>
        <word>against</word>
        <word>moisture</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p2">successful</word>
        <word>ventilation</word>
        <word>design</word>
        <word>provide</word>
        <word>good</word>
        <word>air</word>
        <word>quality</word>
        <word>low</word>
        <word>levels</word>
        <word>pollutants</word>
        <word>occupant</word>
        <word>comfort</word>
        <word>terms</word>
        <word>'airy'</word>
        <word>not</word>
        <word>draughty</word>
        <word>'fresh'</word>
        <word>not</word>
        <word>stuffy</word>
        <word>spaces</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p2">too</word>
        <word>little</word>
        <word>ventilation</word>
        <word>health</word>
        <word>comfort</word>
        <word>occupants</word>
        <word>risk</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p2">too</word>
        <word>much</word>
        <word>ventilation</word>
        <word>during</word>
        <word>heating</word>
        <word>season</word>
        <word>incur</word>
        <word>energy</word>
        <word>penalty</word>
</line>
<line>
```

```
        <word link="MSc/ss-02-preview.htm?#h5">what</word>
        <word>main</word>
        <word>aspects</word>
        <word>about</word>
        <word>ventilation</word>
        <word>design?</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p3">traditionally</word>
        <word>buildings</word>
        <word>were</word>
        <word>over-ventilated</word>
        <word>virtue</word>
        <word>their</word>
        <word>leaky</word>
        <word>construction,</word>
        <word>chimneys</word>
        <word>ill</word>
        <word>fitting</word>
        <word>components</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p3">consequence</word>
        <word>ventilation</word>
        <word>was</word>
        <word>uncontrolled</word>
        <word>buildings</word>
        <word>were</word>
        <word>difficult</word>
        <word>heat</word>
        <word>maintain</word>
        <word>comfort</word>
        <word>keep</word>
        <word>dry</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p4">since</word>
        <word>1970's</word>
        <word>buildings</word>
        <word>become</word>
        <word>more</word>
        <word>energy</word>
        <word>efficient</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p4">heating</word>
        <word>cooling</word>
        <word>requirements</word>
```

```
        <word>been</word>
        <word>reduced</word>
        <word>increased</word>
        <word>levels</word>
        <word>insulation</word>
        <word>reduced</word>
        <word>air</word>
        <word>leakage</word>
</line>
<line>
        <word link="MSc/ss-02-preview.htm?#p4">indeed</word>
        <word>heating</word>
        <word>requirements</word>
        <word>modern</word>
        <word>buildings</word>
        <word>sometimes</word>
        <word>so</word>
        <word>low</word>
        <word>main</word>
        <word>function</word>
        <word>environmental</word>
        <word>system</word>
        <word>no</word>
        <word>longer</word>
        <word>provide</word>
        <word>heating</word>
        <word>but</word>
        <word>provide</word>
        <word>ventilation,</word>
        <word>whether</word>
        <word>natural</word>
        <word>mechanical</word>
        <word>means</word>
</line>
<line>
....
</line>
```

The xml file continues in the manner above, until all the sentences in the Skins and Spaces 'Preview' webpage have been described in this format.

As mentioned ealier, the xml file described in this Appendix covers only the *Preview* web page of the *Skins and Spaces* website. The web pages that follow have also been indexed and form the basis for the remainder of the master.xml index file.

Principles of Ventilation and Air Quality web page

SKINS AND SPACES

PRINCIPLES OF VENTILATION AND AIR QUALITY

Teaching Aims

On the completion of this unit the user should be able to:

- define the basic principles and methods associated with natural and mechanical ventilation;
- define the basic requirements for health and comfort;
- describe the energy impact of ventilation in broad terms;
- relate methods of ventilation to types of built form.
- understand the principles of air quality

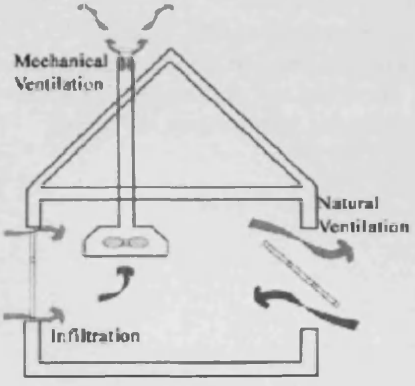


Figure 1: *Domestic housing utilising the fundamental forms of ventilation*

This section of unit 1 will cover the following topics in relation to ventilation

- definitions
- energy use
- comfort

Ventilation

Ventilation is the process of supplying and removing air by natural or mechanical means to and from any space. Ventilation is a combination of infiltration and purpose ventilation. Purpose ventilation can be either natural (opening windows), or mechanical (turning on a fan), or a combination of both.

Infiltration

Infiltration is the fortuitous leakage of air through a building due to imperfections in the structure, such as:

- cracks around doors, windows, infill panels;
- service entries - pipes, ducts, flues, ventilators;
- through porous constructions - bricks, blocks, mortar joints, joist connections within intermediate floors.

Natural ventilation

Natural ventilation is the movement of outdoor air into a space without mechanical assistance. It can be controlled by intentionally provided openings, such as windows, doors and non-powered ventilators. These openings provide a level of ventilation **in addition** to infiltration. In many cases for much of the year infiltration alone will provide sufficient outdoor air to ventilate the building. However, infiltration is uncontrollable, being dependent on the prevailing wind speed and the temperature difference between the internal and external air, and on

the size and type of the leakage areas. If the infiltration component of ventilation is excessive it can result in a waste of heat which will incur a high energy penalty and/or make the building difficult to heat (or cool) to comfort levels.

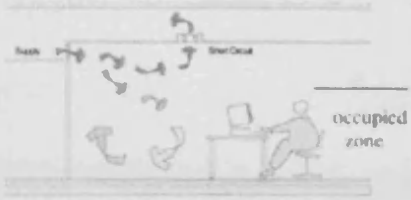


Figure 2: *Short circuiting between the supply and extract reduces ventilation effectiveness and efficiency.*

Mechanical ventilation

Mechanical ventilation is the movement of air by mechanical means to and from a space. It is controllable and can be localised, using individual wall or roof fans, or centralised, with ducted distribution. It can incorporate filtration and a heat recovery system to extract heat from exhaust air and use it to pre-heat supply air.

Build tight ventilate right!

Infiltration occurs in both naturally ventilated and mechanically ventilated spaces. It is considered 'best practice' to minimise the infiltration by applying sealing measures and then to provide **controllable** natural or mechanical ventilation.

Figure 3: *The youngest air is nearest the supply and for efficient room air distribution the oldest air should be nearest the exhaust.*

Room air distribution and fresh air mixing

When fresh air is supplied to a space it mixes with the air already there. The proportion of fresh air that is delivered to the occupied zone depends on the room air distribution which is determined by the location of the air delivery and extract points and the geometry of the space. **Ventilation effectiveness** describes the fraction of fresh air delivered to the space that reaches the occupied zone. Ideally 100% fresh air should be delivered to the occupied zone. However, air often 'short-circuits' between the supply and extract points. **Ventilation efficiency** is the term used to quantify ventilation effectiveness. It is a measure of the ability of a ventilation system to exhaust the pollutants generated within a space. For a specific pollutant it is the ratio of its concentration at the point of extract to the mean concentration of the pollutant throughout the occupied zone. The **age of air** is a measure of the amount of time air has been in a space. The youngest air is found nearest the air supply and the oldest within the space. Ideally, the oldest air should be closest to the extract for maximum ventilation effectiveness. In naturally ventilated buildings the air is usually well mixed throughout the space and there is little difference in air quality, or ventilation efficiency across the space.

Ventilation Efficiency E

$$E = \frac{C_e - C_s}{C_l - C_s}$$

where:

C_e = Concentration of pollutant at exhaust

C_s = Concentration of pollutant in supply air

C_l = Concentration of pollutant at a location

Demand ventilation

Ventilation systems may supply varying amounts of fresh air in response to occupant demand. This may be achieved simply by the manual opening of windows or by some automatic sensing device such as a CO_2 sensor which can indicate occupancy level. **'Shock ventilation'** is the ventilation purging of a space prior to, or after, occupancy. During occupancy, ventilation levels may be kept relatively low, perhaps to avoid cold draughts in winter.

Ventilation rates

All occupied spaces need to be ventilated. Minimum acceptable fresh air ventilation rates depend on occupancy levels and space use, for example, in relation to any specific source of pollution. Ventilation rate requirements for different building types and different occupancy levels can be obtained from the CIBSE Guide. Rates are generally based on health, safety and comfort requirements. They are usually stated either in units of air changes per hour (ac/h) or litres per second per person (l/s/p).

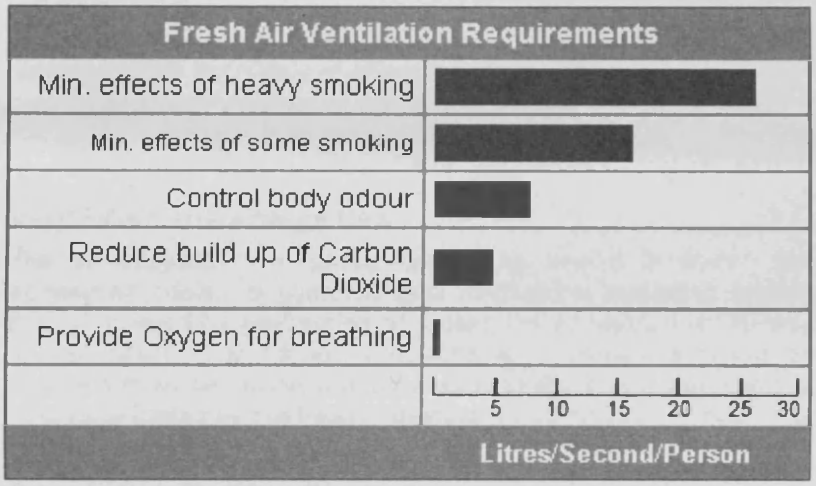


Table 1: *Fresh air requirements in relation to activity.*

Units of ventilation rate

The unit of ac/h is building specific. It relates to space volume - for example, 1ac/h means a complete volume change of air in one hour. The term is typically used for specifying ventilation rates in housing where the occupancy levels and space volumes are within predictable limits. However, with significant reductions in domestic space volumes in recent years care should be given to consideration of ac/h alone as there is a danger of under-ventilating. The unit l/s/p is generally used for specifying ventilation rates in spaces that may vary in size but have known (and often relatively high) occupancy levels and are often predominantly mechanically ventilated.

High occupancy spaces

The recommended fresh air ventilation requirements for spaces with high occupancy levels cause problems in naturally ventilated buildings in winter. First, when the outside air is cold a high ventilation rate can produce energy penalties due to the exhaust of large amounts of stale, but warm, air (heat recovery is not yet feasible with natural ventilation). Second, the ingress of large amounts of cold air from outside can result in thermal discomfort. For more densely occupied spaces such as schools, bars and other public assembly places, recommended ventilation rates may prove difficult to provide by natural means alone if comfort and energy efficiency are to be achieved. In many cases air quality is compromised to avoid discomfort due to cold draughts. However, ventilation rates can sometimes be reduced from recommended levels, by taking account of the duration and levels of occupancy, volume and depth of space, and pattern of room air distribution.

Building type	l/s/person	AC/H
Domestic: Habitable rooms	-	1
Domestic: Kitchens, bathrooms	-	3
Offices	8	1 to 2
Schools	8.3	5
Bars	15	10 to 15

Table 2: *The table summarises some of the main ventilation requirements for different spaces. These requirements are for people alone. Additional ventilation may be needed for any process that produces pollutants or heat.*

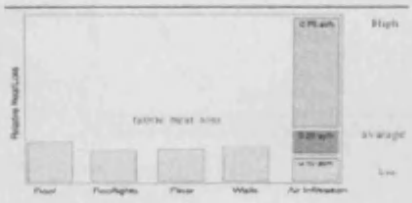


Figure 4: *Variation in the ventilation heat loss for a typical factory in relation to transmission losses through the fabric and infiltration*

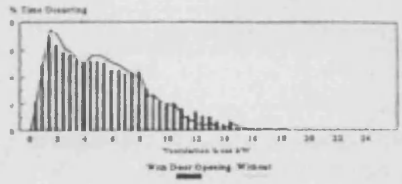


Figure 5: *Seasonal variation in ventilation heat loss for 800m² factory, shown as a frequency distrbution, with and without loading door open.*

Ventilation and energy use

The air supplied to a space has to be heated in winter, and sometimes cooled in summer. In a mechanical ventilation system this is achieved by pre-heating or cooling the air before it is delivered to the space. For natural ventilation it is usually achieved by incoming fresh air mixing with air already in the space and then this mixture is heated by the heating system, for example by contact with 'radiator' surfaces.

The air that is exhausted from the space, through natural or mechanical means, contains heat energy. For a mechanical ventilation system this heat is sometimes recovered through a heat exchanger - otherwise it is wasted. The ventilation component of heat loss can be a significant and sometimes major proportion of the total building heat loss. It can also be very variable (especially in naturally ventilated buildings) as it depends on external wind velocity and air temperature.

Ventilation heat loss

During the heating season the heat lost through ventilation can be estimated from:

Heat Loss = Vent rate(ac/h) x volume x ΔT x ρC /3600 (watts)

= Vent rate(l/s/p) x people x ΔT x ρC/1000

where:

ρC = volumetric heat capacity of air~ 1200 $Jm^{-3}K^{-1}$

ΔT = internal/external air temperature difference (°C or Kelvin)

Increases in internal/ external temperature difference causes:

- an increase in ventilation rate;
- an increase in heat loss.

Seasonal Ventilation Energy Use:

The seasonal energy performance will be related to the average ventilation rate over a heating season.

Design Ventilation Heat Loss:
When designing a heating system the ventilation rate used to calculate the design heat loss should correspond to a design ventilation rate (maximum).

	Ventilation heat loss	
	kWh/m²/yr	% of total heat loss
House	130	25-35%
Offices	100	30-40%
Factory	140	40-50%

Table 3: *Typical ventilation heat loss for different building types*

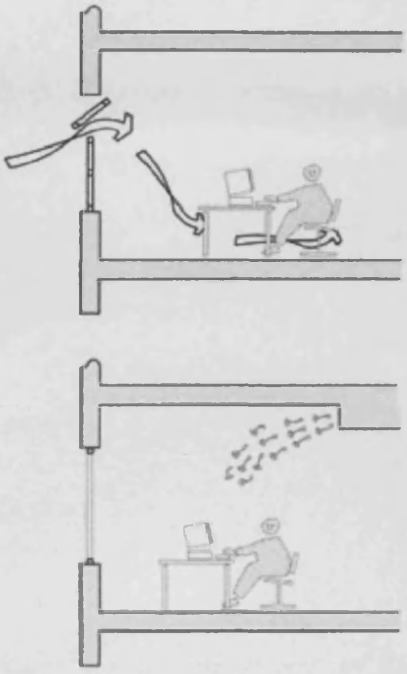


Figure 6 & 7: *Draughts can occur in naturally ventilated and mechanically ventilated spaces*

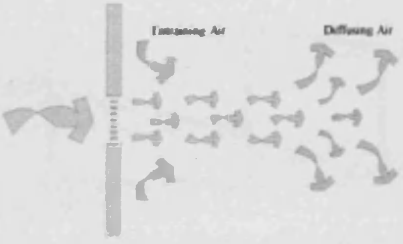


Figure 8: *Air jets are highly turbulent and entrain air in the high speed region and diffuse air in the low speed region.*

Ventilation and comfort

Comfort is determined by air temperature, air speed, relative humidity and radiant temperature, in relation to a persons clothing and activity level. During the heating season cold draughts and low air temperature associated with ventilation can affect comfort. In summer high ventilation rates may be used to promote cooling by exhausting heat gains and increasing air movement, thereby increasing convective and evaporative heat loss from the skin surface.

Air temperature

The temperature of the air is a measure of its heat content. The air temperature within a space is determined by the surface temperatures it comes into contact with, and the temperature of any air supply to the space through mechanical or natural means (mechanical ventilation grilles, open windows). In a typical building space air temperature will vary creating temperature gradients. These tend to be greater in the vertical direction due to buoyancy effects. Cold air entering a space may sweep across the floor, increasing temperature gradients and causing discomfort.

Draughts

If the air speed in a space is above 0.15 to 0.2m/s in winter, it is often perceived as a draught, especially if the occupants of the space have a sedentary activity. Draughts can be caused by air supply devices, such as supply jets, open windows or vents, or by the negative buoyancy effects of cold air induced by cold internal surfaces, for example, downdraughts from cold glazing surfaces. In mechanically ventilated spaces they can be caused by air jets colliding with walls, downstands or other air jets. Higher air temperatures are needed in heated spaces where there is high air movement in order to maintain comfort conditions. The turbulent nature of the air speed can also affect perceived comfort. The higher the turbulence, the greater the perception of draught discomfort. In summer, higher air speeds may be maintained during hot weather for comfort cooling. However air speeds greater than about 0.5m/s can cause mechanical problems of papers moving, etc.

The mechanisms for internal air movement

Internal air movement is caused by:

- external forces of wind - the pressure variation over the building envelop will cause an inflow and outflow of air at openings which will influence internal air movement;

- internal forces including stack effect, surface temperatures and air jets - internal heat sources cause air to warm and rise, exhausting through openings and replaced by cooler air entering at lower levels. Cold downdraughts will be produced at cold internal surfaces (external walls and glazing). Mechanically supplied air is jetted into the space with a particular momentum and velocity.

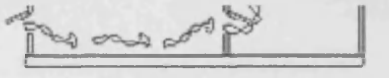
Internal air movement is also affected by:

- entrainment of air - air being drawn into jets and plumes due to drag and viscous forces;
- diffusion of air in plumes and jets due to turbulence.

Away from the immediate influence of jets, internal air movement is mainly induced by buoyancy effects. In naturally ventilated spaces buoyancy driven air movement usually dominates over any source of air movement from ventilation openings and air infiltration.

Turbulence

Air flow can be laminar, turbulent or 'transitional'. Laminar flow occurs when the flow is steady. Fluctuations in either flow direction or speed occur when the flow becomes turbulent. The flow is said to be transitional when it is between laminar and fully turbulent. An air jet is usually fully turbulent, whereas airflow in a naturally ventilated space is often in the transitional mode.



Winter - Ventilation preheat

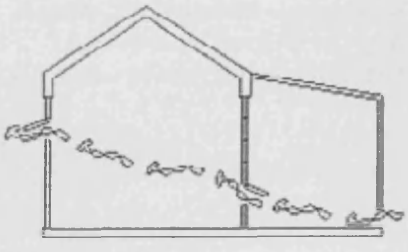
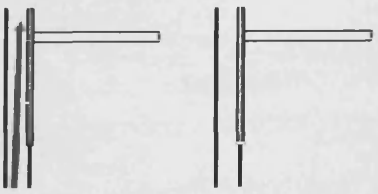


Figure 9: *Summer and winter effects on conservatory ventilation.*



Foreground and background ventilation

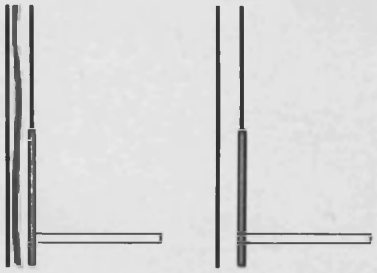
Naturally ventilated buildings generally have a mix of foreground and background ventilation. Foreground ventilation relates to opening devices that are used to produce relatively high ventilation rates usually over short periods to provide relief when the air is excessively stale, humid or warm. Background ventilation is active continuously to provide sufficient fresh air for normal occupancy requirements.

Foreground ventilation

These are general purpose ventilation openings that can be controlled to respond to the ventilation needs of the space. During warm periods in summer, windows and vents can be opened and the ventilation rate increased without incurring an energy penalty or resulting in cold draughts. In winter however, spaces that require high ventilation rates (eg. school classrooms) can be a problem. Firstly, there must be good control to avoid excessive ventilation and an energy penalty. Secondly, the cold external incoming air must be heated to avoid cold draughts. In such cases the natural ventilation system may need to be integrated with the heating system.

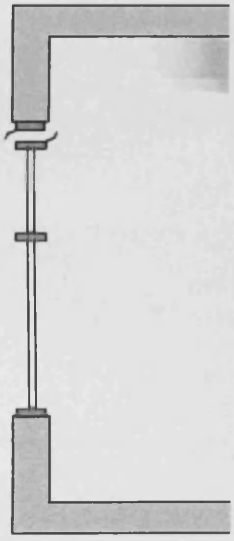
Background ventilation

Background ventilation is usually achieved by a mix of fortuitous (air leakage) and purpose (trickle ventilation) means.



Integrated double skin
Separated double skin

Figure 10: *Integrated and separated double skin ventilation systems*
Click to view illustrations



BACKGROUND + FOREGROUND

Figure 11: *Domestic example of background (trickle) vs rapid (foreground) ventilation. Click the*

Air leakage

Air leakage is usually termed air infiltration. In buildings fitted with trickle ventilators, fortuitous air leakage can account for between 50% and 90% of the background ventilation, depending on the relative 'tightness' of the building envelope. In many building types where trickle ventilators are not installed, air leakage provides the sole means of background ventilation and is often sufficient to support occupancy for most of the year. In order to have good control over ventilation the fortuitous air leakage should be minimised by designing a well sealed construction.

Controlling background ventilation

For naturally ventilated buildings in the UK, background ventilation is required in response to this general tightening up of the building envelope. This is usually achieved through some form of trickle ventilation. Trickle ventilators provide a constant ventilation to the occupied spaces to ensure good air quality and in particular to reduce levels of humidity and the associated risk of condensation.

Sizing of trickle ventilators

Trickle ventilators are intended to maintain a constant background ventilation rate of about 0.25ac/h. Their contribution to the overall background air leakage is dependent on the general leakiness of the building, they are therefore most effective in relatively well-sealed buildings. They are only intended as background ventilation and can therefore **only provide ventilation for relatively low occupancy levels**. With higher levels of occupancy other ventilation methods may be required.

Control of trickle ventilators

Most trickle ventilators are controllable in that they can be opened or closed. Some are wind controlled, providing a constant airflow for varying wind conditions. Some are humidity controlled which ensures maximum opening at times of high internal relative humidity. Mostly they are located in window or door frames. However, there may be advantages in locating them close to heating fins or radiators, incorporating them into the heating system design in order to avoid cold draughts.

Trickle ventilators often incorporate a deflector on the inside in order to direct the incoming air downwards or upwards to avoid direct draughts into the space or the perception of draughts, for example, due to the movement of curtains.

SKINS AND SPACES

DOMESTIC VENTILATION

This section of unit 2 will cover the following topics in relation to air quality:

- air leakage
- mechanical ventilation
- trickle ventilation and passive stack ventilation

This section presents ventilation design strategies and case studies for domestic buildings. Natural ventilation and mechanical ventilation systems are considered. However, it should be remembered that houses with mechanical ventilation will still have air infiltration and houses that are considered to be naturally ventilated will often have fans for local extract.
Natural ventilation design in domestic buildings can be divided into three components, as discussed below :

Background ventilation

Background ventilation is required to provide a continuous ventilation rate to meet the long term ventilation demands of the space, that is, the provision of fresh air for breathing, and for the dilution and exhaust of pollutants. Background ventilation was traditionally provided by infiltration, that is, the natural leakage of air through the building fabric. However, as buildings are constructed to be more 'airtight' in order to reduce energy use, infiltration rates are becoming too low to meet background ventilation requirements. Additional openings are therefore needed to ensure adequate background ventilation. Trickle ventilators are now commonly used in UK houses to provide a controllable background ventilation rate.

Rapid ventilation

Rapid ventilation is the purpose ventilation that can be controlled to respond to the ventilation needs of the space. During warm periods in summer, windows and vents can be opened and the ventilation rate increased without incurring an energy penalty or resulting in cold draughts. In winter however, spaces that require high ventilation rates (eg. school classrooms) can be a problem. Firstly, there must be good control to avoid excessive ventilation and an energy penalty. Secondly, the cold external incoming air must be heated to avoid cold draughts. In such cases the natural ventilation system may need to be integrated with the heating system. UK building regulations prescribe a minimum area of opening in habitable rooms as a function of floor area (1/20 of floor area in the form of an openable window).

Extract ventilation

Extract ventilation is required in specific rooms that are likely to be thesource of pollution or odour, for example, the bathroom, WC and kitchen. Extract ventilation can be achieved by mechanical or natural means. Natural extract systems involve the use of passive stack ventilation (PSV), through which air is continuously exhausted from the space via a duct to the outside. In contrast, mechanical systems generally provide a high ventilation rate over a short period.

Extract ventilation

Extract ventilation is required in specific rooms that are likely to be thesource of pollution or odour, for example, the bathroom, WC and kitchen. Extract ventilation can be achieved by mechanical or natural means. Natural extract systems involve the use of passive stack ventilation (PSV), through which air is continuously exhausted from the space via a duct to the outside. In contrast, mechanical systems generally provide a high ventilation rate over a short period.

Whole house mechanilcal ventilation with heat recovery

Whole house mechanical ventilation, though still relatively uncommon in the UK, is finding increasing application, especially when combined with heat recovery. The usual strategy for domestic mechanical ventilation is to extract air from the bathroom, WC, utility room and kitchen. The air is then passed through a heat exchanger where a proportion of its heat is transferred to the incoming fresh air supply before it is exhausted. The fresh air is delivered to the main living spaces, that is, the living rooms and bedrooms. Whole house ventilation systems ensure minimum ventilation rates which can eliminate high humidity and condensation problems. The incorporation of a heat recovery system can maintain a reasonable level of energy efficiency. However, for the most effective performance, the air infiltration (air leakage) rate of the building needs to be reduced to a minimum, which can prove difficult with the standards of airtightness typical of UK domestic construction.

Occupant control

It is important that whatever ventilation system is used, it is clearly understood by the occupants and that they can control the system to respond to their needs. The design of a ventilation system should therefore begin with a clear understanding of the ventilation requirements of the occupant. In particular there should be clear 'feedback' on any ventilation controls occupants are expected to use.

Element	Air Leakage(m^3/h) at 50Pa (% in brackets)		
	Low	Medium	High
Ceiling Brickwork Service pipe Ceiling hatch	212 (10.5) 231 (11.5) 15 (0.7) 96.5 (4.8)	439 (17) 250 (9.7) 17.3 (0.7) 102 (4.0)	710 (17.9) 281 (7.1) 19.9 (0.5) 113 (2.9)
Ceiling/ wall joints - upper - lower	 542 (26.9) 271 (13.4)	590 (22.9) 295 (11.4)	 638 (16.1) 319 (8.1)
Wall/ floor joint Window frame/ wall Door frame/ wall Window opening Door opening	271 (13.4) 104 (5.2) 24 (1.2) 214 (10.6) 37 (1.8)	295 (11.4) 120 (4.6) 28 (1.1) 324 (12.5) 122 (4.7)	319 (8.1) 131 (3.3) 30 (0.8) 1018 (25.7) 379 (19.6)
Total	2018	2582	3958
Air changes per hour at 50Pa (pressure test)	7.5	9.6	14.7

Table 1: *Typical breakdown of domestic air leakage*

Air infiltration through leakage

Modern houses are relatively well sealed by comparison with their older counterparts. A modern house would typically have an average background air infiltration rate of about 0.5 ac/h. This might be considered too low for normal occupancy, especially as it is often the case that the majority of the infiltration takes place in the circulation and highly serviced spaces, such as halls, stairwells, kitchens and bathrooms. In modern well sealed houses, some living rooms may receive very low (practically zero) fresh air infiltration, which is why the requirement for additional background ventilation has been introduced in the building regulations. Older houses were generally well ventilated by comparison. They often had chimneys associated with solid fuel heating, together will relatively leaky windows and doors, with no draught stripping. Older houses therefore had air infiltration rates typically as much as 2ac/h, and so whole house heating was difficult to achieve and energy inefficient. Older houses therefore benefit considerably from the application of draught sealing measures. Also, in older houses care must be taken when mixing open chimney fires with central heating systems. In fact, negative efficiencies can be experienced with open fires when operated with central heating, because the draught an open fire induces up the chimney can exhaust more heat from the house than is actually supplied by the open fire!

Two types of air leakage

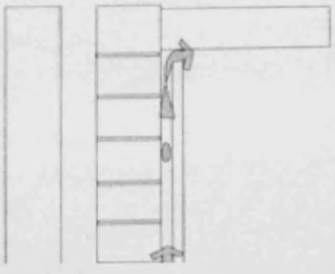
Air leakage can occur in two main areas:

- air leakage through the construction, such as through walls, construction details and service entries.
- air leakage through components, such as doors, and windows.

Although modern components are usually reasonably airtight, modern construction methods (for example, internal dry lining of walls) are not conducive to a good standard of airtightness. In practice, it is very difficult to predict the degree of air tightness that will be achieved in the finished building.

Room	Background (area of opening)	Rapid	Extract
Habitable	8000mm²	1/20 floor area opening window (no minimum size)	30l/s at hob 60l/s elsewhere or PSV
Kitchen	4000mm²		
Bathroom	4000mm²	opening window (no minimum size)	30l/s or PSV
Utility	4000mm²	opening window (no minimum size)	15l/s or PSV
Toilet	4000mm²	1/20 floor area or 6l/s mechanical extract	

Table 2: *Domestic ventilation requirements (Part F1 Building Regulations 1991)*



Appendix D: User Studies

PILOT TEST

Participant: SIMON LANNON
Date & Time: 23/05/07
Profession: RESEARCH ASSOCIATE

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?

yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?

yes

3) Does this ability make searching easier and quicker? yes

4) Does the display of keywords in their context help you to compare the different passages?

yes

5) Does highlighting the terms you are searching for help?

no

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?

oh yes

7) What was the highest number of terms you combined to search for a specific query?

4

8) Is their a learning effect in using and understanding this tool?

yes

9) Comments on user interface:

Very interesting idea can I try shares.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the navigation tool:

1. On what principle should the ventilation design of modern buildings be based? Use the main search box only.

build tight and ventilate right

2. What is the age of air? What is the temperature of air?

amount of time air has been in space / temperature of air is measurement of heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

1 4l/h 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

trickle vents provide background ventilation

constant ventilation to ensure good air q

reduced level of humidity + condensation

Only provide for low occupancy

PILOT TEST

Participant: DYLAN DIXON
Date & Time: 23/05/07
Profession: RESEARCH ASSISTANT

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? a) YES, (I THINK!) b) PROBABLY!

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? YES

3) Does this ability make searching easier and quicker? YES

4) Does the display of keywords in their context help you to compare the different passages? YES

5) Does highlighting the terms you are searching for help? YES

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? YES

7) What was the highest number of terms you combined to search for a specific query? 4 or 5 !

8) Is their a learning effect in using and understanding this tool? YES

9) Comments on user interface:

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the navigation tool:

1. On what principle should the ventilation design of modern buildings be based? Use the main search box only.

2. What is the age of air? What is the temperature of air? Measure of the amount of time has been in a space. Measure of heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house? 7L/S/P 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

background ventilation = mix of controlled (leakage) & (trickle)

Participant: ASTON ROBERTS
Date & Time: 11/06/07 15:30
Profession: STUDENT

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
It is intuitive but its effect on results is not evedent clear.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?
Yes but only when looking for 3+words in reference to a key wo
3) Does this ability make searching easier and quicker?
for complex sequences of words yes.
4) Does the display of keywords in their context help you to compare the different passages?
Yes
5) Does highlighting the terms you are searching for help?
Yes but problems when 'learning' is highlited when seeding for 'a
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
Yes, but problems when accidently pressing return instead of add
7) What was the highest number of terms you combined to search for a specific query?
6 i think
8) Is their a learning effect in using and understanding this tool?
Yes
9) Comments on user interface:
If the web page higlated the searched terms like the windows 'find now' tool and use of a find next option was encorporated?

otherwise a goo
style search sea
just as practi

Aston.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

'build tight' and 'ventilate right'

2. What is the age of air? What is the temperature of air?

age – a measure of the amount of time air has been a space

temp – a measure of its heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

7.5 litres/second/Person

0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Trickle ventilators are commonly used in UK houses to provide a controllable background ventilation rate.

Provide sufficient background ventilation to exhaust pollutants from 'low level' activities

Participant: BOBBY RIZVI
Date & Time: 11/06/07 15:30
Profession: STUDENT

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
yes but the distances are only clear when explained.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? I just ended up clearing the screen as 'middle mouse button' had no effect on deselecting words
3) Does this ability make searching easier and quicker?
only if it's use is fully understood - but it is quite a deep process to have many words on screen.
4) Does the display of keywords in their context help you to compare the different passages?
yes, I scrolled down when I needed words in different contexts
5) Does highlighting the terms you are searching for help?
Yes, shows me where to look on the page
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
yes I can add words I don't see on screen
7) What was the highest number of terms you combined to search for a specific query?
3, maybe 4
8) Is their a learning effect in using and understanding this tool?
seems fairly intuitive based on other search engines I've used
9) Comments on user interface:
the terms on the side could be perhaps clearly labled in alphabetical order. I wasn't sure if they were when I was adding words to the list, so words may have already been there! looking back I see they are, but maybe an auto align function for words outside the circle would make this clearer

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

build tight & build right

2. What is the age of air? What is the temperature of air?

age of air is measure of amount of time air has been within a space

temp of air is a measure of its heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

a) ≈ 7 l s p

b) 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Traditionally background ventilation was provided by infiltration through natural leakage. However, now buildi are more air tight & additional openings are needed in the form of trickle ventilation.

Participant: Nikos Refinitis
Date & Time: 11/06/07 16:00
Profession: Student

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?

Yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?

Yes

3) Does this ability make searching easier and quicker?

Yes

4) Does the display of keywords in their context help you to compare the different passages?

Yes

5) Does highlighting the terms you are searching for help?

Yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?

Yes

7) What was the highest number of terms you combined to search for a specific query?

4

8) Is their a learning effect in using and understanding this tool?

Maybe a paragraph

9) Comments on user interface:

No problem

An Observer Participant User System (OPUS)

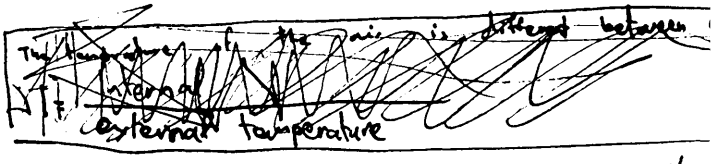
Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

The ventilation design on modern building should be based on the principle "bui tight" and ventilate right in order to to provide building occupant with a indoo quality and comfortable conditions for the most efficient use of energy (

2. What is the age of air? What is the temperature of air?



The age of air ~~concentration~~ is a measure of the amount of time air has been in a space.

The temperature of the air is different between internal and external air te

$$\Delta T = \frac{\text{internal temperature}}{\text{external temperature}}$$

3. What is the ~~minimum ventilation requirement to control~~ body odour? What is the air infiltration rate of a modern house?

A modern house would typically have an average background infiltration rate of about 0,5 ach

To Minim Control body odours - 10 ~ 7

4. Explore the relation between background ventilation and trickle ventilators.

To provide continuous ventilation rate to meet the long te ventilation demand; of the space, that is the provison of fre air for breathing and for the dilation and exhaust of pol Also to reduce energy use.

Participant: Vassilis Petropanagiotakis
Date & Time: 12th June 2007.
Profession: Student

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
After the few instruction given to me, It was easy understant.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes!.

3) Does this ability make searching easier and quicker?
Yes, very much
4) Does the display of keywords in their context help you to compare the different passages?
Yes
5) Does highlighting the terms you are searching for help? Yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? Yes

7) What was the highest number of terms you combined to search for a specific query? 5 but I shouldn't have ha

8) Is their a learning effect in using and understanding this tool?
Not really its obvious since you start it.
9) Comments on user interface:
Nice ~~dont~~, simple and
Clear

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

the ventilation design of modern building should be based on the principle "build tight" and "ventilate right" in order to provide building occupants with good air quality and comfortable temperature for most efficient use of energy

2. What is the age of air? What is the temperature of air?

The age of air is a measure of the amount of time air has been in a space

The temperature of the air is a measure of its heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

7 l/time/second/Person.

Infiltration rate of a modern house is about 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is usually achieved by a mix of infiltration (air leaks and purpose (trickle ventilation) means.

Air leakage is usually termed air infiltration. In buildings fitted with trickle ventilators, infiltration air leakage can account between 50% and 90%...

Participant: ROWAN PARNELL
Date & Time: 12th JUNE 2.30pm
Profession: STUDENT.

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
Yes. But it seems slightly difficult to grasp at first as is unusual. But it works well
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?
Yes
3) Does this ability make searching easier and quicker?
Yes.
4) Does the display of keywords in their context help you to compare the different passages?
Yes
5) Does highlighting the terms you are searching for help?
yes
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
yes
7) What was the highest number of terms you combined to search for a specific query?
Not sure - 3 or 4
8) Is their a learning effect in using and understanding this tool?
Yes - like any new software - but easy to pick up.
9) Comments on user interface:
Possibly would be easier if words in sentences, - and, if but, a, the, etc. were left in as just from some key words it is sometimes not possible to know exactly which result is best.
- could be turned on or off?

- I would like the text to be larger.
- easier to read & move about.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

'build tight' and 'ventilate right' in order to provide building occupants with good air quality and comfortable conditions ...

2. What is the age of air? What is the temperature of air?

- A measure of the ammount of time air has been in a space.
- A measure of its heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

- 7 liters / second / person
- typically 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Trickle ventilators provide a constant ventilation to the occupied spaces to ensure good air quality — and a constant background ventilation.

- intended to maintain a constant background ventilation rate of about 0.25 ac/h.
- provide ventilation for relatively low occupancy levels.

Participant: LUCIANO DE FARIA
Date & Time: 13/06/07, 11:30
Profession: ARCHITECT / PhD STUDENT

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
YES, AND IT IS INTUITIVE.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?
YES. IT HAPPENS AUTOMATICALLY.
3) Does this ability make searching easier and quicker?
YES, MOSTLY IF YOU CAN ISOLATE IT WITH CITATION INFORMATION
4) Does the display of keywords in their context help you to compare the different passages?
YES.
5) Does highlighting the terms you are searching for help?
YES.
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
YES, IT IS NECESSARY.
7) What was the highest number of terms you combined to search for a specific query?
3
8) Is their a learning effect in using and understanding this tool?
YES, MOSTLY ISOLATED TO HOW TO COMBINE KEY WORDS
9) Comments on user interface:
IT SEEMS VERY SIMPLE AND QUICK TO UNDERSTAND

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

IT HAS TO PROVIDE FRESH AIR FOR OCCUPANTS, TO DILUTE AND EXHAUST POLLUTANTS AND TO PROVIDE COOLING IN THE SUMMER. BESIDES, IT IS ALSO NEED FOR THE PROTECTION OF THE BUILDING AGAINST MOISTURE.

2. What is the age of air? What is the temperature of air?

- THE AGE OF THE AIR IS A MEASURE OF THE AMOUNT OF TIME AIR HAS BEEN IN A SPACE.
- THE TEMPERATURE OF THE AIR IS A MEASURE OF ITS HEAT CONTENT.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

0.1 ACH . 0.5 ACH

4. Explore the relation between background ventilation and trickle ventilators.

1/20 OF FLOOR AREA IN THE FORM OF AN OPENABLE WINDOW

Participant: Qian Li
Date & Time: 13/Jun/2007
Profession: PHD candidate.

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? Yes. Yes.

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes.

3) Does this ability make searching easier and quicker? Yes.

4) Does the display of keywords in their context help you to compare the different passages? Yes.

5) Does highlighting the terms you are searching for help? Yes.

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? Yes.

7) What was the highest number of terms you combined to search for a specific query? Three.

8) Is their a learning effect in using and understanding this tool? Yes.

9) Comments on user interface:

* Highlighting the results in the web page or document.
* How to delete the words in the circle but keep it in the coloum.
* is it able to build a personal "dictionary" or and save. their search.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Particpant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

"build tight"
"ventilation right"

2. What is the age of air? What is the temperature of air?

Air. A measure of the amount of time. air has been in a place.

r: A measure of its heat content. odour is in graph.

(searched minimum, ventilation)

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

8 L/S/P.

0.5 ach.

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is usually achieved by a mix of purpose (air leakage) and purpose (trickle ventilation) means.

Participant: AMIT GUPAN
Date & Time: 17/06/07 14:30
Profession: ARCHITECT

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? YES

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? YES

3) Does this ability make searching easier and quicker? YES

4) Does the display of keywords in their context help you to compare the different passages? YES

5) Does highlighting the terms you are searching for help? YES

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? YES

7) What was the highest number of terms you combined to search for a specific query? 3

8) Is there a learning effect in using and understanding this tool? YES

9) Comments on user interface:

VERY USER FRIENDLY, SIMPLE TO UNDERSTAND AND USE.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

The ventilation design of modern buildings should be based on the principle of 'build tight' & 'ventilate right' in order to provide building occupants with good air quality & comfortable conditions for the most efficient use of energy

2. What is the age of air? What is the temperature of air?

- The age of air is a measure of the amount of time air has been in space.
- The temperature of the air is a measure of its heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

- Minimum ventilation requirement to control body odour - 8 litre/se /pe
- Modern house would typically have an average background air infiltration rate of about 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is required to provide a continuous ventilat rate to meet the long term ventilation demands of the space that is provision of fresh air for breathing & for the dilution & exhaust of pollutants. Background ventilation is traditionally provided by infiltration however as buildings are getting more air tight, in order to reduce energy use infiltration rates are becoming too low to meet the requirements. Additional openings are therefor needed to ensure adequate background ventilation.

Trickle ventilators are now commonly used in the Uk to provide a controllable background vent. rate.

Participant: Janice Coyle
Date & Time: 18th June '07
Profession: Graphic Designer

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? yes

3) Does this ability make searching easier and quicker? yes

4) Does the display of keywords in their context help you to compare the different passages? yes

5) Does highlighting the terms you are searching for help? yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? yes

7) What was the highest number of terms you combined to search for a specific query? 6 or 8

8) Is there a learning effect in using and understanding this tool? yes

9) Comments on user interface:

A bit confused on how to gather groups of text together. as linking didn't always work.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

principle 'build tight' and 'ventilate right', in order to provide building occupants with good air quality & comfortable conditions.

2. What is the age of air? What is the temperature of air?

age of air is a measure of amount of time air has been in a space.
temp. is measure of its heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

Control body Odour approx 8 (Litres / Second / Person)
infiltrate rate of a modern house 0.5 ac/h.

4. Explore the relation between background ventilation and trickle ventilators.

both intended to maintain a constant background ventilation rate

Participant: MELANIE JONES
Date & Time: 18/06/07 15:00 pm
Profession: WEB DESIGNER.

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? Yes, Yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes

3) Does this ability make searching easier and quicker? Yes

4) Does the display of keywords in their context help you to compare the different passages? Yes

5) Does highlighting the terms you are searching for help? Yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? Yes

7) What was the highest number of terms you combined to search for a specific query? 4/5

8) Is there a learning effect in using and understanding this tool? Yes

9) Comments on user interface:

Having the ability to change search processes is very useful especially with the 'add' functionality.

Passages of text with searched words contained within them is extremely useful for identifying passages/results quickly and efficiently.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

The ventilation design of modern buildings should be based on the principle 'build tight'.

2. What is the age of air? What is the temperature of air?

The age of air is a measure of the amount of time air has been in a space.
The temperature of air is a measure of its heat cont

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

Via a duct to the outside / Body odour not found

Air ventilation of about 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is required to provide a continuous ventilation rate to meet the long term demands of space.

Trickle ventilation is intended to maintain a constant background ventilation rate of 0.25ac/h

Background = continuous
Trickle = specified rate.

Participant: Ann Smith
Date & Time: 18 June 2007
Profession: Research Assistant

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? Yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes.

3) Does this ability make searching easier and quicker? Yes

4) Does the display of keywords in their context help you to compare the different passages? Yes / no

5) Does highlighting the terms you are searching for help? Yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? Yes

7) What was the highest number of terms you combined to search for a specific query? 5

8) Is there a learning effect in using and understanding this tool? No. Yes.

9) Comments on user interface:

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

build tight and ventilate right, in orders to. provide building occupants with good air quality and comfortable conditions.

2. What is the age of air? What is the temperature of air?

→ age of air is measure of the amount of time air has been in space

→ The air temperature within a space is determined by the surface temperature it comes into contact with

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

→ ? Fresh air requirements in relation to activity

→ 0.5 ac/h.

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is usually achieved by a mix of fortuitous (air leakage) and purpose (trickle ventilation) means.

Participant: PAOLA SAISI
Date & Time: 19/06/07 17:00
Profession: LECTURER

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? yes quite - not clear wh some sentences are listed above

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? yes

3) Does this ability make searching easier and quicker? not sure - it appear to allow for higher accuracy of search

4) Does the display of keywords in their context help you to compare the different passages? yes

5) Does highlighting the terms you are searching for help? yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? yes very

7) What was the highest number of terms you combined to search for a specific query? 6 ? not sure

8) Is there a learning effect in using and understanding this tool? yes

9) Comments on user interface:

Generally user-friendly and as mentioned in 3 it appears to give more accurate results.
The right hand list of terms can get very long and nearly requires a search engin to search within results.
Perhaps colour would help to highlight main terms.
Could you ~~also~~ click on words from occurrence list to get those sentences.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

build tight + ventilate right

2. What is the age of air? What is the temperature of air? — measure of heat content

↳ amount of time it has spent in the room

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

7 ltr/sec 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

trickle vent for backgrd vent.

" " can account between 50 - 90 % of backgrd

alternative ventilators possible

Participant: SAM
Date & Time: 19/06/2007 16:30
Profession: RESEARCHER (ARCH)

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
yes yes — but could be more s
if position/
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? arrange
also inf
yes — results
3) Does this ability make searching easier and quicker?
certainly could do with additional functions
4) Does the display of keywords in their context help you to compare the different passages?
yes, but it would be more helpful if the order/
5) Does highlighting the terms you are searching for help? arrangement of words had
more influe
yes on ra
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
yes
7) What was the highest number of terms you combined to search for a specific query?
n 4
8) Is there a learning effect in using and understanding this tool?
- it is easier to understand how a search may be composed.
9) Comments on user interface:

Very good & useful. The spatial arrangement of terms is very effective. With more detailed sensitivity in the positioning of words & the ranking of ~~the~~ results this would ~~be~~ really improve speed ~~effective helpful~~. Perhaps there is also the need for a quarantine area too, for unwanted words.

→ perhaps a series of orbits....

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

'build tight' and 'ventilate right'

2. What is the age of air? What is the temperature of air?

Age of air: a measure of the amount of time air has been in a space

temperature of air: a measure of its heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

min ventilation:- ~8 l/s/person

infiltration rate:- ~0.5 ac/hr.

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation was traditionally provided by ~~background~~ infiltration through the leaky building fabric. Now that buildings are more airtight, this must be provided by trickle ventilators to ensure adequate rates of air change. These can then account for 50 - 90% of background ventilation, providing a rate of about 0.25 ac/hr. The rest of the fabric should then be as airtight as possib

Participant: EDWARD WAINEWRIGHT
Date & Time: 19/06/07 17:00
Profession: RESEARCHER (ARCHI)

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
Yes, but its relationship to more than 2 terms needs to be explored
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes, but to see it in more context would be helpful.
3) Does this ability make searching easier and quicker?
Yes, if words were highlighted in text.
4) Does the display of keywords in their context help you to compare the different passages?
Very much
5) Does highlighting the terms you are searching for help?
Yes - but add it to the web-page count
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
Very much, but needs to be more intuitive
7) What was the highest number of terms you combined to search for a specific query?
3
8) Is there a learning effect in using and understanding this tool?
Yes, I found it altered my method of seeking answers and allowed a greater degree of exploration in find answers quickly
9) Comments on user interface:

The user interface needs to be more intuitive if this is to be used generally. Right + left mouse click options + drop down lists on search terms in the circle would be helpful + integrated

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

Controllable in response to requirements of occupants
Build Tight Ventilate Right

2. What is the age of air? What is the temperature of air?

A measure of the amount of time air has been in a space.
The temperature of air is a measure of its heat conte

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

0·5 AC/H air infiltration rate
6/7 litres/second/person — why does odour not bring the one [illegible] up? + Bo... get a... 0 o... Sea... vent

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is required to provide a continuous ventilation rate to meet long term ventilation demands of space.
Provided trad. through natural infiltration.
Trickle ventilators used to provide a controllable background vent.

Participant: Sarah Nicholas
Date & Time: 19 June 07 17:00 approx
Profession: Librarian

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? Yes. It is intuitive. However it appeared that words not proximal to others were ignored in the second group of search results – I don't know if I'm assuming incorrectly
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? I liked being able to redefine relationships between the terms by clicking on the middle button. Otherwise, I wasn't overly excited by the spatial rearrangements as I could see no obvious change in my search results
3) Does this ability make searching easier and quicker?
It felt like it did.
4) Does the display of keywords in their context help you to compare the different passages?
Absolutely – I tend to always rely on quickly scanning results rather than rephrasing my searches
5) Does highlighting the terms you are searching for help?
Yes. It's a practise I am used to, & expect from a search interface.
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
I liked that very much, though felt those with little search experience would add irrelevant terms.
7) What was the highest number of terms you combined to search for a specific query?
Three (But I always tend towards limiting search terms – experience has taught me that this is the best route to success)
8) Is there a learning effect in using and understanding this tool?

9) Comments on user interface:

It made more sense than other efforts I've seen to represent searches and results visually. I recall search tools that put results in bubbles spreading out according to relevance – these were largely gimmicky & didn't (I believe) take off.
This interface is visual without extraneous gizmo-bits.
I liked very much that it employed Boolean syntax without overtly stating so, and think that those who struggle to grasp this concept will accept a visual representation far more easily.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

"Build tight" and "ventilate right"

2. What is the age of air? What is the temperature of air?

Air is as old as the time it's been in a space.
The temperature of air is a measure of its heat content
etc etc.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

To control body odour is around 8 litres per second per pers
The air infiltration rate of a modern house is about 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Trickle ventilators allow background ventilation to be controlled. (Background ventilation is necessary for continuous ventilation)
[There's lots more on this, but it's a bit beyond me, I'm afraid]

Participant: Fran
Date & Time: 4.07.07
Profession: Teaching Executive Officer.

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
Yes. Yes.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?
I don't think I did re-arrange the words spatially.
3) Does this ability make searching easier and quicker?
Yes.
4) Does the display of keywords in their context help you to compare the different passages?
Yes.
5) Does highlighting the terms you are searching for help?
Yes.
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
Yes - used alot.
7) What was the highest number of terms you combined to search for a specific query?
3.
8) Is there a learning effect in using and understanding this tool?
Could quickly highlight the key themes of a subject area
9) Comments on user interface:

- Easy to use. Appeared user friendly & fast.
- The way of not always linking the words - could provide frustration?
- To keep the words you added on the left hand side could be useful - rather than losing them each time?

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

to provide fresh air

2. What is the age of air? What is the temperature of air? — measure of heat content

a measure of the amount of air supply & the odors within the space.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

- build how & ventilate right?
- whole house ventilation

4. Explore the relation between background ventilation and trickle ventilators.

trickle ventilators commonly used to provide controllable background V. rate.

Participant: WOUTER POORTINGA
Date & Time: 04/07/07
Profession: RESEARCH FELLOW

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? YES

YES

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? SOMETIMES.

3) Does this ability make searching easier and quicker? YES

4) Does the display of keywords in their context help you to compare the different passages? YES

5) Does highlighting the terms you are searching for help? YES

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? YES

7) What was the highest number of terms you combined to search for a specific query?

3 (I think)

8) Is there a learning effect in using and understanding this tool?

TOOL IS QUITE INTUITIVE; IS ALREADY easy to use in the beginning

9) Comments on user interface:

I like the spatial/graphical interface of the search eg. In particular to be able to use a 'central' term.

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

The principles
— Build tight
— Ventilate right

2. What is the age of air? What is the temperature of air?

Age of air = measure for amount of time air has been in a space

Temperature of air = measure of heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

— Extract ventilation ?

— Air infiltration rate of modern house: ± 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is needed to provide the long term ventilation demand of a space.
Often this was delivered by leaky infiltration.
Nowadays trickle ventilators are used to provide controllable Background ventilation

Participant: Zainah Zain
Date & Time: 8th July 2007
Profession: Librarian

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive? Yes

2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes

3) Does this ability make searching easier and quicker? Yes

4) Does the display of keywords in their context help you to compare the different passages? No

5) Does highlighting the terms you are searching for help? Yes

6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task? Yes

7) What was the highest number of terms you combined to search for a specific query? 5

8) Is there a learning effect in using and understanding this tool? Yes

9) Comments on user interface:

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

Based on the principle of 'build tight' and 'ventile right'.

2. What is the age of air? What is the temperature of air?

The age of air is a measure of the amount of time air has been in space.
Temperature of air is a measure of its heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

8 L/S/P is the minimum ventilation requirement to control body odour
Infiltration rate of a modern house is 0.5 ac/h

4. Explore the relation between background ventilation and trickle ventilators.

A reduction in background ventilation due to the construction of airtight buildings has seen an increase in the use of trickle ventilators which provides a controllable background ventilation rate.

Participant: Clarice Blair de Souza
Date & Time: 13th July 2007 / 11:30 am.
Profession: architect

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
Sort of depends on the context
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results?
I work in pair of words most of the time, so I will first try different combination of pairs before using spatial solu
3) Does this ability make searching easier and quicker?
it helps, when I cannot use pair of words anymore
4) Does the display of keywords in their context help you to compare the different passages?
yes
5) Does highlighting the terms you are searching for help?
yes
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
I used it all the time because I tend to work in pairs of words
7) What was the highest number of terms you combined to search for a specific query?
3 as far as I remember
8) Is there a learning effect in using and understanding this tool?
yes ~~[illegible]~~
9) Comments on user interface:

Might be a very stupid comment, but I think it would be interesting to think about a way of identifying words (highlight differently, underline maybe) in each of the selected sentences of the 'key word in context box' that could inform ~~[illegible]~~ which kind of info related to the key word is contained in the sentence.

example: key word: ventilation.

(usually | achieved | through | some | form | trickle | ventilation)

↑ info contained in

↑ key word

Clarice

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

- principles: - build tight, ventilate right.

2. What is the age of air? What is the temperature of air?

- age of air → [redacted] the amount of time the air has been in space
- temperature of air → [redacted] a measure of the air heat content

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

- min vent requirements for body odor control = extract ventilati
- air infiltration rate: 0.5ach (typical avg.)

4. Explore the relation between background ventilation and trickle ventilators.

- trickle ventilators are used to provide controllable background ventilation rates. they assure good air quality & reduce humidity levels associated with condensation risk.
- most effective in relatively well sealed buildings
- provide ventilation for relatively low occupancy levels.
- can be opened & closed
- wind controlled (some)
- humidity controlled (some)
- contain deflectors
- located in window & door frames.

Participant: TANIA OLIVEIRA
Date & Time: 13/07/07, 12.30 hs
Profession: PHD CANDIDATE / [redacted]
AREA: MARKETING

An Observer Participant User System (OPUS)

User Response

Please answer the following questions.

1) Do you understand the utility of the 'principle of proximity'? Is it intuitive?
2yes. (it puts "similar" things together). 2yes.
2) Does the ability to spatially rearrange words/terms allow you to refine your previous/subsequent search results? Yes. (I like to "play" with the words and terms by myself.
3) Does this ability make searching easier and quicker?
Yes.
4) Does the display of keywords in their context help you to compare the different passages?
Yes, but it is OK if I need to create new words
5) Does highlighting the terms you are searching for help?
yes.
6) Does dynamically adding any keyword to the list using the Add+ button assist in the search task?
Yes! Very much!
7) What was the highest number of terms you combined to search for a specific query?
Four.
8) Is there a learning effect in using and understanding this tool?
2yes.
9) Comments on user interface:
The tool is easy to use and I found it useful in many stances, for example, in the academic life [redacted] to help the selection of articles, literature review etc

An Observer Participant User System (OPUS)

Task Questions

Please answer the following questions using the Participant mode in the navigation tool:

1. On what principle should the ventilation design of modern buildings be based?

The ventilation design of modern buildings should be based on the principle of "build tight" and "ventilate right".

2. What is the age of air? What is the temperature of air?

It is a measure of the amount of time air has been in a space (age of air).

The temperature of the air is a measure of heat content.

3. What is the minimum ventilation requirement to control body odour? What is the air infiltration rate of a modern house?

trol Fresh air ventilation: ± 5 litres per second per person
dy odour:

Air infiltration rate of a modern house: 0.5 ac/h.

4. Explore the relation between background ventilation and trickle ventilators.

Background ventilation is usually achieved by a mix of fortuitous (air leakage) and purpose (trickle ventilation) means.

Appendix E: CD

Appendix E: CD

The cd accompanying this thesis contains two folders:
'Chapter 3_Case Study 1' and 'Chapter 4_Case Study 4' folders.

The *Chapter 3_Case Study 1* folder has two executable (.exe) files which can be launched by double clicking the file names.

The *Chapter 4_Case Study 4* folder has three executable (.exe.) files which can be launched by double clicking them. The reader is advised to view the 'demo.exe' file. The two remaining files - user5.exe and user15.exe have been provided as a record of the actual task performed by users as part of the evaluation of the navigation tool.

Also provided in this folder is a sub-folder named OPUS. The user is advised to copy this folder onto the local hard drive of the computer they are using. After copying the folder, the user can launch the 'opus.exe' file by double clicking it. To initiate a search, the user can drag any word from the set of terms on the left-hand side and attempt to familiarize themselves with the tool as demonstrated in the 'demo.exe' file.