Context-based understanding of food-related queries using a culinary knowledge model

Abstract. Dietary practices are governed by a mix of ethnographic aspects such as social, cultural and environmental factors. These aspects need to be taken into consideration during an analysis of food related queries. Queries are usually ambiguous. It is essential to understand, analyze and refine the queries for better search and retrieval. The work is focused on identifying the explicit, implicit and hidden facets of a query, taking into consideration the context - Culinary Domain. This paper proposes a technique for query understanding, analysis and refinement based on a domain specific knowledge model. Queries are conceptualized by mapping the query term to concepts defined in the model. This allows an understanding of the semantic point of view of a query and an ability to determine the meaning of its terms and their interrelatedness. The knowledge model acts as a backbone providing the context for query understanding, analysis and refinement and outperforms other models such as Schema.org, BBC Food Ontology and Recipe Ontology.

Keywords: query understanding, contextualization, knowledge model, information retrieval

1.Introduction

Culinary studies have gained in popularity, due to its influence on culture, its impact on society and its dominance within life-styles. Cultural diversity exists in the food domain as no homogeneity can be found between Northern and Southern or Eastern and Western cuisines of the world. The ability to understand a food related query and infer the context is always a challenge, as interpretations vary from person to person. The same user might have different needs at different times and in different spaces. Let us consider a search query - "tuna recipe" - it can have two interpretations in the culinary world: (1) a recipe for a dish with tuna (*saltwater fish*); as the main ingredient; (2) a recipe for a dish with the main ingredient of tuna (*fruit of prickly pears*), popular in Mexico. In such cases domain knowledge plays a pertinent role in order to understand how each domain related component interacts with each other; how a specific ingredient becomes predominant in a cuisine; how geographic location influences food habits; and how certain events impact the culinary world. This knowledge and experience is essential to be able to identify the explicit, implicit and hidden components in a query. Given a query, we are interested to understand it from a semantic point of view rather than from a term level. One must identify aspects like concepts (a.k.a. classes), related concepts and properties to suit the most dominating purpose prevailing among the users. This paper proposes a framework to expose different facets in a user query and align the facets with a context-dependent knowledge model.

Contextualizing and analyzing a query paves the way towards *query intent* where "the individual is actively involved in finding meaning which fits in what he or she already knows, which is not necessarily the same answer for all, but sense-making within a personal frame of reference" [28]; *query representation* involves modeling the query intent/user needs based on a particular representation in order to make the context explicit; *query refinement* is "the incremental process of transforming a query into a new query that more accurately reflects the user's information need" [54]. These aspects help to optimize queries for performance. It is a significant aspect of query understanding and processing as it enhances the overall performance and increases precision. These approaches have been used in various domains such as e-commerce [66], biomedicine [67], geoscience [68], law [69], social media [70], etc. to improve search effectiveness. This paper attempts to propose a framework for understanding of food-related queries by amalgamating two approaches:

(1) **Explore explicit, implicit and hidden facets of a query.** Considering the context for a given query, explicit, implicit and hidden facets are identified thereby resolving the ambiguity in a query and expanding the query for clarity and potential retrieval. This technique is motivated from Ranganathan's approach of studying a document title to generate class numbers [42], where he proposed a set of steps to convert a title of any document into its corresponding class number.

(2) **Mapping facets with knowledge model.** Models written in natural language are used in every sphere of life. However, an informal model leads to multiple interpretations, ambiguity, and abuse in every possible way. This paper employs a formalized knowledge model that helps to provide the explicit meaning of a term, hence minimizing ambiguity. The knowledge model acts as the backbone of the system, to enhance communication between machine and users.

The significance of the study is to explore the semantic aspect of a query and determine the meanings of query terms. Each query term is mapped to the knowledge model, to understand the relevant facets - concepts, their properties and associated concepts. The query enrichment method, using the proposed knowledge model in this paper, achieves much higher precision than other models, such as Schema.org, BBC Food Ontology and Recipe Ontology. The main advantage of this approach allows a more accurate analysis of query terms, which is difficult to achieve with traditional knowledge organization systems for query understanding and enrichment.

The work is organized as follows: Section 2 discusses the state of the art, exploring the different query understanding and enrichment strategies using knowledge models - controlled vocabularies (dictionaries, taxonomies, ontologies) and the influence of Ranganathan's concept of facets for efficient information retrieval; Section 3 elaborates the steps - query understanding, analysis and refinement approach; Section 4 presents experimental findings; Section 5 concludes the paper by providing an insight into future research.

2. Literature Review

2.1. Different Query Enrichment Approaches

Several researches, based on query understanding and analysis, addressed challenges of efficient information retrieval. Query understanding and expansion techniques aimed at meeting the user's satisfaction involve several approaches – web-based, search-log-based, corpus-based, relevance and pseudo-relevance-based feedback and linguistics-based [62]. Various approaches are incorporated by the Information Retrieval community to improve linguistic properties such as morphological, lexical, syntactic and semantic to leverage query terms. Research communities have attempted to explore linguistically motivated techniques. [46] classifies a query into four categories: descriptive, relational, structural and concepts-of-interest. *Concepts-of-Interest* is recognized as the core concept that represents exactly what the users are searching for, whereas the remaining query components serve as an auxiliary concept which enhances the specificity of a query. [8] categorizes a query in three parts - common entities (entities re-used across all domains), core entities (entities returned by queries) and auxiliary entities (entities required to filter results). [45] proposes a personalized query expansion approach which considers the user, the domain and task-specific preferences pertaining to user queries to resolve ambiguity. This work suggests that most of the query terms are domain-specific and context-dependent. Specific rules are coded to identify missing information and tasks in a query that might be domain independent.

Domain-specific assistance may be provided via the inclusion of dictionaries, thesauri, classification schemes, databases, ontologies and other controlled vocabularies for query understanding and contextualization. Table 1 provides a comparative analysis of different approaches on query understanding and analyses addressed by the research and scholarly community.

1. Dictionary Based

A dictionary-based technique [33] is implemented where subject codes are used from Longman's *Dictionary of Contemporary English*, to establish semantics within a text. [5] classifies Biomedical texts with three different dictionaries - BioCreative [21], NLPBA [10] and an ad hoc subset of the UniProt database – Protein [2]. Additionally, query translations based on dictionaries have provided efficient results. This is particularly useful in cross-language information retrieval where dictionary-based techniques are incorporated in query translation architecture [31], [9].

2. Thesaurus Based

Thesaurus assisted query understanding, analysis, expansion techniques gained immense popularity as they obtain precise queries and indexing. [25] states that it is advantageous to use a thesaurus for query expansion if, and only if, there is a good match between the query term and the thesaurus term. An original query is expanded by introducing terms from lexical databases like WordNet, a popular approach employed by [51]. A thesaurus-like expansion of the original term increases the scope and hence recall is increased by adopting a traditional information retrieval approach. Tudhope and Binding [59] argues the relevance of a thesaurus in modern information retrieval and discusses the use of ‘thesauri in three key areas of infrastructure underpinning advanced retrieval functionality today: metadata enrichment, vocabulary mapping and web services’ (p. 174).

[40] proposed a thesaurus-based query expansion model. Their work emphasized a selection of terms that are closely associated with the concepts of the query. [48] studied the user’s query refinement behavior for a search system. It was concluded that a thesaurus helped to assist users in query formulation and expansion. [35], [36] experiment with different types of thesauri. Each thesaurus has strengths and weaknesses. Taking advantage of their strengths, a technique is proposed for efficient query expansion. [26] makes use of thesaurus relations and natural language processing techniques to expand user queries. [18] states that hierarchical

relations like (BT/NT) and equivalent relations like (USE/UF) are helpful for automatic query expansion, whereas associative relation (RT) is useful for interactive query expansion. Wordnet is further used to identify synonyms and related terms of a query word [37]; it is used to expand queries by following the typed links available within it [56]. [47] uses MeSH heading as it is advantageous for query expansion. Motivated by this work [43] uses MeSH Headings for query enrichment to enhance efficient retrieval of resources belonging to MEDLINE, called Cystic Fibrosis. Keywords are extracted from the query and automatically matched with the Entry terms in MEDLINE. If there is a hit, the associated MeSH heading is searched, and the query is reformulated by adding the descriptors from MeSH fields. Getty Art and Architecture Thesaurus (AAT) is used to index the National Museum of Science and Industry database [52]. Thesaurus relationships (hierarchical relations and associative relations) are involved to compute semantic closeness. The authors claim that query understanding, enrichment and expansion can be assisted by exploring semantic relationships in a faceted thesaurus. It is evident that a thesaurus provides terminology support in the process of query formulation, and selection of terms is crucial for IR success [50]. Wu [60] enriches thesaurus (related-terms (RT)) with semantic relations to assist in question-answering and promoting information retrieval. The author advocates that an enriched thesaurus improves the performance of modern information retrieval systems more than the original thesaurus.

3. Ontology-Based

Domain-specific and domain-independent ontologies are used for query understanding, analysis and refinement. [4] provides an elaborate account of the advantages of using ontology for query expansion. [15] proposes an ontology-based query expansion technique. This technique aims to resolve queries related to spatial terms by employing SPIRIT ontology database (composed of a domain ontology and a geo-ontology) [25]. A domain-specific ontology is used for query expansion and translation. This ontology includes domains and sub-domains regarding Stockholm University [38]. Domain ontology is used to resolve semantic ambiguity and efficient retrieval of engineering documents during the product development process. [19] proposes a framework to search documents in the engineering domain and cater to the personalized needs of the engineers. A personalized query expansion process is executed by analyzing the users' profile. Users' interests are identified by studying the user profile and the documents relevant to them. Two novel personalized query expansion techniques are proposed [61] based on enriched user profiles with tags and annotations. Moreover, query expansion also takes the users' search intent into consideration. The Semantic Query

Expansion technique is implemented to identify related terms of a query and refine it. The structure of an ontology is used to the best of its ability to identify related concepts like 'neoplasm' and 'tumor'. A method is proposed which includes a concept analysis technique, FCA (Formal Concept Analysis), and SQE (Semantic Query Expansion), to refine health-related outcomes using plain terminologies [11].

4. Knowledge-Based

Knowledge models based on faceted classification and facet analysis have an enormous impact on the IR community. According to the ANSI/NISO Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies, facet analysis is an efficient way to organize knowledge by incorporating a bottom-up approach, identifying and analyzing the features of concepts and assembling them based on the domain knowledge. Potential usage of facet analysis are as follows:

> Facet analysis is particularly useful for:
>
> - new and emerging fields where there is incomplete domain knowledge, or where relationships between the content objects are unknown or poorly defined;
> - interdisciplinary areas where there is more than one perspective on how to look at a content object or where combinations of concepts are needed;
> - vocabularies where multiple hierarchies are required but can be inadequate due to difficulty in defining their clear boundaries; or,
> - classifying electronic documents and content objects where location and collocation of materials is not an important issue [22].

[23] emphasizes the potential of a faceted structure to user queries. The '*structural approach*' in Ranganathan's principles influences the structural characterization of knowledge-based information retrieval. As predicted by Ingwersen and Wormell, the approach has an enormous impact, mainly in two fields:

(i) *Knowledge Representation* (KR) - a new faceted KR approach to develop descriptive ontology, DERA, which enhances automatic reasoning [16]. DERA (Domain, Entity, Relation, Attribute), a KR approach, is built based on the faceted approach employed for descriptive ontology. The methodology is inspired by category-based systems by [42] and [3]. The framework is based on the following ideas: Universe of Knowledge should be organized as domains; each domain consists of several facets; Universe of Knowledge, its corresponding domains and the identified facets are built following analytico-synthetic [42].

(ii) *Information Retrieval* (IR) - faceted classification is a key to several IR approaches. According to [6], the faceted classification scheme is one which "takes the idea of analytico-synthesis to its logical conclusion and 'deconstructs' the vocabulary of the classification into its simplest constituent parts". The use of faceted classification to enhance efficiency in IR is elaborated by [12]:

- the display of useful generic relationships;
- full and accurate cross-referencing;
- accurate application of principles of division;
- a clear citation order;
- established rules for compounding; and
- an appropriate notation.

Faceted Information Retrieval (FaIR) showcases the formalization of facets, where the system provides an interface that allows users to interact with facets by selecting concepts under different facets, thereby modifying the retrieved results [39]. [64] demonstrates an approach to publishing recipe datasets in Linked Data format, thereby enabling users (non-experts in Ontology) to produce recipe-linked-data using recipe-based ontology. [14] addresses the lack of communication between web-designers and users by bringing web-developers and potential web-users into one common platform of understanding, and by organizing a hypertext concept through facet analysis, thereby generating a hypertext knowledge base for designers and a knowledge base interface for users.

This paper attempts to identify the domain-dependent facets in a query and establish relationships within the facets using a knowledge model. The knowledge model acts as a backbone for query understanding, analysis and refinement.

Table 1. Comparative analysis of approaches related to Query Understanding, Analysis and Refinement

Article	Approach	Types of Resource	Applications	Techniques Used
[33]	Dictionary Based	Longman's Dictionary of Contemporary English	Document filtering	Lexical, Syntactic, Semantic & Discourse Linguistic Processing Technique
[5]		BioCreative[21], NLPBA [10],and an ad-hoc subset of the UniProt database named Protein [2]	Biomedical text classification	Dictionary based Entity Recognizer (Abner-Tagger)
[31]		Bilingual dictionary resources [31, p.532]	Cross-lingual IR	Lexical, Morphological and Syntactic Technique
[9]		Multilingual resources		Weighted Probabilistic technique
[25]	Thesaurus	CILKS Thesaurus	Document retrieval system	Semantic networks
[40]		Similarity Thesaurus	Probabilistic Query expansion	Concept similarity technique
[26]		EuroVoc Thesaurus	Improve recall of zero-hit queries	NLP techniques
[48],[18],[60]	Domain-specific Thesaurus	CAB Thesaurus; ProQuest Thesaurus; ERIC Thesaurus	Automatic Query expansion	Hierarchical and Associated Relationships used
[35],[36]	Heterogeneous thesauri	Wordnet, co-occurrence-based thesaurus, predicate-argument-based thesaurus	Query expansion	Weighting method
[52]	Subject-based Thesaurus	Faceted Thesaurus (Art and Architecture Thesaurus)	Faceted Query expansion	Semantic Relations
[51], [56],[37]	Database	Wordnet	Query expansion	Level of specificity and abstraction, Lexical-semantic relations

[47],[43]	Subject heading	MeSH	Abstract retrieval system; query expansion	Vector space model
[4]	Ontology Based	Ontology	Query expansion	Semantic and contextual analysis
[15]		Domain & Geographical Ontology	Spatial Query expansion	Footprint-based spatial query expansion technique
[45]		Domain Ontology	Personalized QE	Context analysis
[38]		Domain Ontology	Query expansion and translation	Linguistic approach
[19]		Domain Ontology	Engineering document retrieval	User profile
[61]	Profile based	User profile & folksonomy data	User profile enrichment	Word embeddings
[11]	Data analysis & concept analysis	Formal Concept Analysis (FCA) and Semantic Query Expansion	Query formulation and refinement	Data analysis technique

3.Methodology

Users provide their information needs via queries. It is a challenge to capture users' interests due to domain related semantic ambiguity. Hence it is significantly important to model the query based on a context. The essential aspect is to assign semantics to user queries, to enhance the appropriateness of the relevant documents. van Rijsbergen [53, pp. 84-85] rightly stated:

> Information is ultimately dependent on the interpretation the user puts on a meaning, logic is a tool that a user can use to get at the information...In IR we do not seek an answer to the 'meaning' of language, instead, we seek a model that will enable the user to find information, that is, discover something she did not already know. Simply retrieving meanings is not enough; these meanings must carry information; hence this form of retrieval is inherently uncertain…
>
> When a set of sentences about a document does not imply the query, it is not the case that a user will not want to retrieve that document. It may indeed contain the information that the user seeks. It may be that a small modification to its representation will lead to the query being inferred. It is the extent of this modification that gives a measure to the uncertainty of the implication. If the change is great, the uncertainty is great; if the change is small the uncertainty is small.

This section illustrates the strategies developed to understand user queries, analyzing and refining them and enriching the query by making the implicit context explicit, thereby expressing them in search terms. Every step of interpreting a query needs to consider the context [1]. A knowledge model [8] is used for query understanding, query analysis and query representation. In an information system, good output depends upon good knowledge representation [58], as representation of knowledge enables the establishment of relationships between real-world entities [16]. The pivotal job of the knowledge model is: i) identify the different facets in a user query; ii) identify the different properties associated with the facets; iii) establish relationships between different facets. The knowledge model is classified into three categories:

1. ***Core Entity Type***: The core entity type takes the central position of the model. These entity types are those that are returned by the generalized query. Recipe is the core entity type of the model. It is to be noted that it extends the common entity type Work. The core entity type, Recipe describes the steps that need to be followed in the process of preparing a particular food item. The process includes some important components such as ingredient information (a set of ingredients along with their corresponding quantities) and instruction necessary to make the food item. Such description is clearly a product of the human intellect. For this reason, in our model Recipe extends Work and leads to the preparation of Dishes. The properties defined for Recipe include and extend those in Schema.org.

2. ***Auxiliary Entity Types***: Auxiliary entity types are those that one comes across in the generalized query and are used to filter the results. They include Cuisine, Dish and Event. Notice that the common entity type Event plays the role of auxiliary for recipes. This model defines three auxiliary entity types: Cuisine (Cuisine captures the diversity in culture, religion and geography by introducing these attributes in the recipe model); Dish (Dish represents the final product of a recipe. According to WordNet, dish is defined as a particular item of prepared food. It is a final product of a recipe that would be served for a Menu); Event (Event is the only auxiliary entity type having a corresponding type in Schema.org. Dishes are prepared for specific occasions. In our model, this is captured via the *hasOccasion* relation).

3. ***Common Entity Types***: Common entity types are inspired from common isolates in library and information science discipline as proposed by Ranganathan [42]. He defines

common isolates as “an isolate idea in the idea plane denoted by some isolate term in the verbal plane and represented by the same isolate number in the notational plane, to whatever basic class it is attached”. This denotes that the common entity types have their own identity which remains intact no matter which conceptual model they get attached to. They give a sense of completeness to the model. Within a domain, common entity types can become core or auxiliary. The common entity types playing a role in recipes are Person, Location, (Creative) Work and Event.

Figure 1 provides a graphical representation of the recipe model developed. The recipe model includes six entity types and three structured attributes. Common entity types are represented as white boxes with dashed boundary lines. Core entity types are represented as blue boxes with solid boundary lines. Auxiliary entity types are represented as blue boxes with dashed boundary lines. Moreover, the structured attributes are represented as yellow boxes with solid boundary lines.

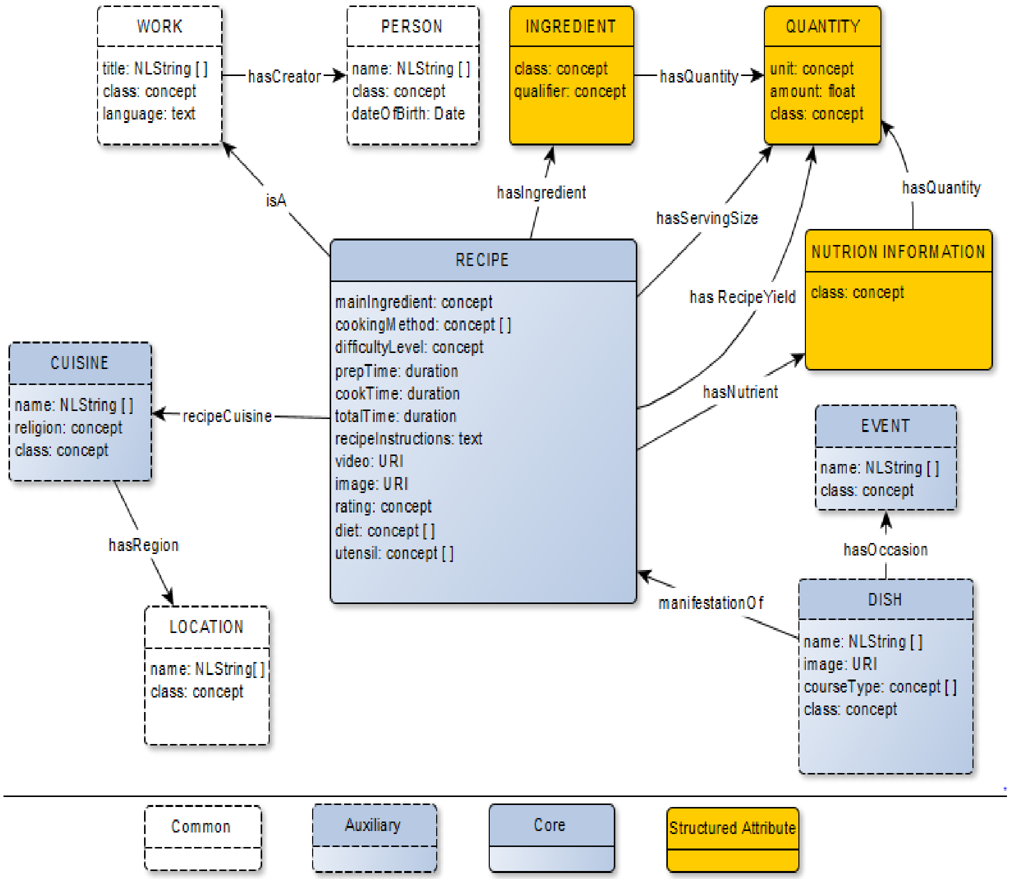

Figure 1. Recipe Model [8]

The following subsections present a detailed account of the query understanding, analysis, and refinement by employing the contextual knowledge model, to expose explicit, implicit and hidden facets of a user's query. Figure 4 provides a pictorial representation of the proposed methodology.

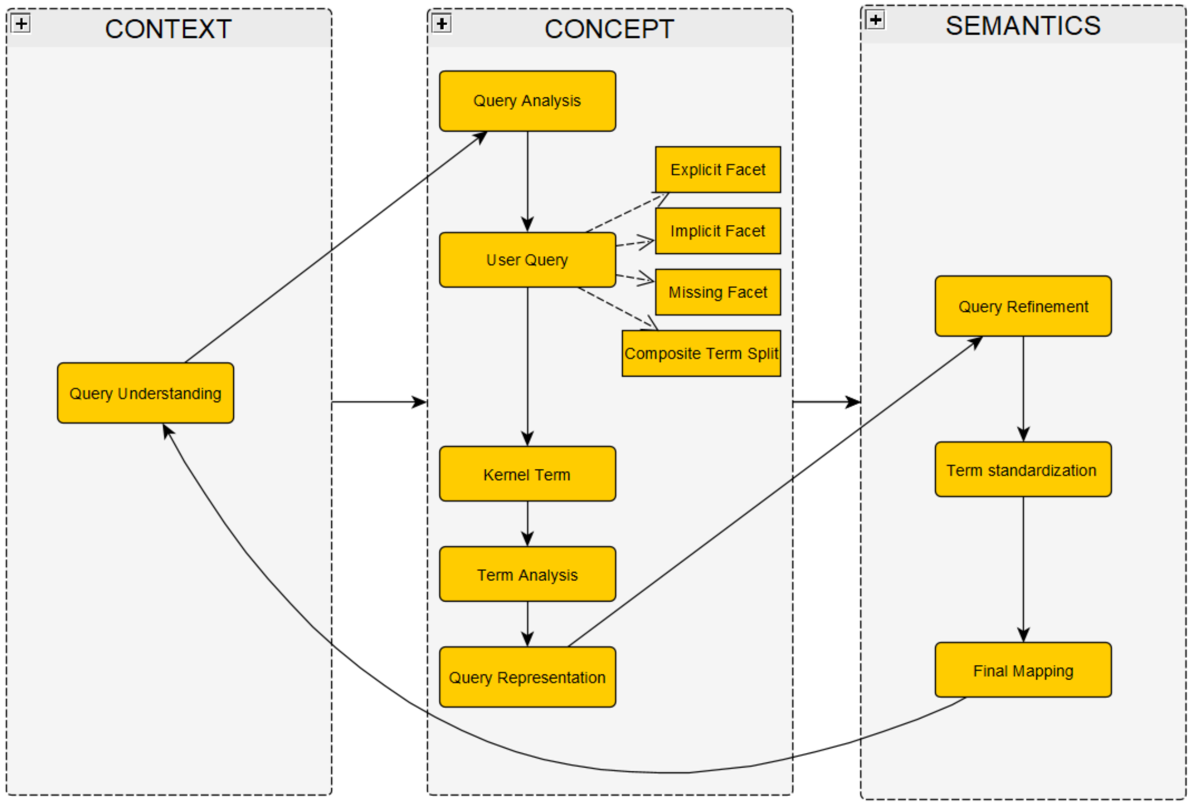

Figure 4. A pictorial representation showing the methodology proposed for context-based understanding of food-related queries.

***Step 1: Query Understanding*:**

Each query has a different interpretation based on the context. To predict intent, [49] discusses the use of geospatial features, users' demographic features such as age and gender, ethnographic features, religion, culture, and users' query logs to gain contextual knowledge. Identification of the concept based on context leverages the intent for an ambiguous query. The task is more complicated than mere text matching with query terms. This work attempts to understand the query from a semantic point of view, rather than from a term level.

This section considers several aspects associated with a query such as related location, event, cuisine, etc. for the context - Recipe. To justify this activity, an interview is conducted with people from different origins to gain their perspective on various scenarios, during which online recipe searches are made. This contextual information provides an understanding of user requirements in the context of a recipe search. A set of informal queries are collected and studied by conducting the interview sessions and by complimenting them with questions identified in the relevant state of the art works. The collected queries are studied to understand the query-pattern and user needs and requirements. This leads to a fair understanding of the context and contextual concepts, and is an incentive to identify explicit, implicit and hidden facets in user queries, thereby minimizing query ambiguity (as discussed in the following sections).

Step 2: Query Analysis

In this section, a close relevance is derived between query analysis and facet analysis. According to [55], "the essence of facet analysis is the sorting of terms in a given field of knowledge into homogeneous, mutually exclusive facets, each derived from the parent universe by a single characteristic of division". There are different components of a query which can be treated as a facet. A facet in a query can be categorized as follows:

(1) *Explicit Facet:* An explicit facet is one which is usually explicated/explained in a query. Information is clearly provided in the query, reducing ambiguity. The query may not require further processing as the concepts are semantically aligned with the context. For instance, a query on *chicken recipe* clearly indicates the users' search intent, i.e. the user is searching for different recipes to prepare a dish with chicken.

(2) *Implicit Facet:* An implicit facet is one which is suggested by the query. There is no clear evidence of the facet but it is left to interpretation. For instance, if a person is allergic to nuts, it implies that the ingredient, 'nut', should be excluded from his dish, although there is no clear evidence in the statement on his dietary requirements, i.e. to avoid nuts. Such information is retrieved from the domain knowledge.

(3) *Hidden/Absent Facet*: There might be facets hidden within a derived composite term. Identification of these facets is often very tricky. According to [42], experience helps to sense the absence of a necessary facet. The following steps elaborate the facet identification process of a query and refinement of the query based on a knowledge model [8] to make the concepts explicit.

a. User Query

A user query is mostly ambiguous and frequently consists of only a few words [63]. It can be compared to a title of a book that is left for a subject-experts interpretation. On analyzing the facets in a query, it is evident that facets in a query may be isolate facets of its subject which may be either (1) explicit, (2) implicit in the context, or (3) hidden within a derived composite term. Let us attempt to analyze a query by considering one of the competency queries. Here the following steps are to be undertaken to make the following query fully expressive to resolve ambiguity:

Query: *Recipe of Roasted Turkey for Thanksgiving*

a. Explicit Facet: The given query illustrates the user's search intent, i.e. the user is searching for different recipes to prepare roasted turkey.
b. Implicit Facet: Often the context is not explicitly provided in the query. Here the context of the above example is an Online Recipe Search. Recipe searches on the Web have gained popularity as they provide opportunities to explore and experience global cuisines. In the Information Science discipline such contexts are defined as the Main Class or Basic Class.
c. Filling up the Missing Facet: In a query, a facet may be hidden/absent. [42] defines it as Latent Facet. It is further exemplified when, if the title of a book is "wheat storage", storage is a secondary operation as it cannot be taken before harvesting is done. So, the facet "harvesting" is a hidden or latent facet. A hidden facet can be revealed only if one has enough domain knowledge. To gain domain knowledge, the Query Understanding stage is critical. Generally, when a facet is absent, one must have a complete understanding of the context to reveal the Hidden/Latent Facet. According to [17], it is not enough to have background knowledge, but it should be context-sensitive and should have the ability to capture the diversity of the world.

   In this query, Thanksgiving is celebrated by preparing Roasted Turkey in North America. So, it is mainly a North American cuisine. Employing our extensive diversity-aware knowledge base [17] we attempt to fill up the missing facet.

   Recipe of Roasted Turkey for Thanksgiving in North America.

d. Breaking of Composite Terms: In some cases, the query may contain a composite term, which may be broken into components to make the query fully expressive. This is achieved by filling up all the implied words (ellipses) in the query. This is attained by breaking up the derived composite terms into isolated ideas. In the case of the query above, Roasted Turkey is a Derived Composite term. If the composite term is broken into distinct ideas, then we have "Roasting a Turkey to prepare Roasted Turkey". Here

Roasted Turkey involves two isolated ideas - “Roasting” (preparation method) and “Turkey” (the key ingredient of the recipe).

The above step leads to a fully expressive query:

Online Recipe for Roasted Turkey by Roasting a Turkey for Thanksgiving in North America.

b. Kernel Term

Kernel Terms are obtained when all the auxiliary words or apparatus words are omitted from an Expressive Query and every composite term broken into fundamental terms. In natural language processing, a set of words like *an, of, the, and* occur frequently in texts without providing useful information. These set of words are called stopwords. Elimination of the stopwords reduces the size of the indexed terms and enhances the retrieval process. Gerard Salton and Chris Buckley provide a list of stopwords (http://www.lextek.com/manuals/onix/stopwords2.html) for the experimental SMART Information Retrieval System.

In this context, stopwords like *to, by, a, for,* etc. are removed. Each Kernel term is put in the nominative case singular form. The first letter of each Kernel term is capitalized, and each term is separated by a period. On employing the discussed strategy, the following kernel terms are obtained:

Online Recipe. Roasted Turkey. Roasting. Turkey. Thanksgiving. North America

c. *Analyzed Query*

The Kernel terms are analyzed and hence analyzed terms are obtained. The following steps elaborate upon the process of query analysis:

1. Identifying the context: as the context is the first facet it is important to identify the context. Here the context is ‘Online Recipe’.
2. Identifying the categories: the skeleton form separates the terms in the name of the concept following the trains of characteristics of classification to which each of them relate. Here, considering the context as Online Recipe, “Roasted Turkey” relates to the “Dish” characteristic; “Roasting” to the “Cooking Method” characteristic; “Turkey” to the “Ingredient” characteristic; “Thanksgiving” to the “Event/Occasion” characteristic

and "North America" to the "Location" characteristic. The skeleton form of the name of the query suggests the appropriateness of saying that it has five "Facets".

3. The outcome of this step is known as the Linguistic Syntax. It is also known as the Syntax of Words, i.e. the sequence in which the words are arranged in a sentence or a language of a given domain in natural language form. The Linguistic Syntax varies from language to language. The outcome of this step exemplifies the Linguistic Syntax of English; however, for the same example, the Linguistic Syntax of Italian might be different. The number of variations in Linguistic Syntax depends on the number of kernel terms in a given query, which is the same as the number of facets in it.

    Online Recipe [Roasted Turkey]: [Roasting]: [Turkey]: [Thanksgiving]: [North America]

d. Query Representation

Information science professionals have observed that certain concepts occur frequently in all subjects. S.R. Ranganathan has found that terminologies related to materials appear frequently; so is the case with activities, operations, and processes. Also, objects, entities, and systems are repetitive in every subject. He named them as "fundamental categories". His five isolate facets were necessary and enough to characterize all documents in existence. Ranganathan's popular PMEST formula consists of five fundamental categories: Personality, Matter, Energy, Space and Time. The arrangement of these five categories provides the facet sequence or so-called facet order. [55] proposed a list of 13 fundamental categories that will be helpful in the field of science and technology. According to [7], one of the editors of BC2, the thirteen fundamental categories proposed by Vickery were adequate for "the analysis of vocabulary in almost all areas of knowledge". [3] proposed four main categories: Discipline (D), Entity (E), Property (P) and Action (A) popularly known as DEPA, which includes one special category called Modifier (m) by studying different subject indexing languages. A similar idea is employed in this work. As participants are interviewed, certain concepts appear to be occur frequently [8]. The kernel terms are mapped with the recipe model proposed in our previous work.

Figure 1 provides a graphical representation of the recipe model developed. *Common entity types* are inspired by common isolates in library and the information science discipline as proposed by [42]. Common isolates can be defined as "an isolated idea in the idea plane denoted by some isolate term in the verbal plane and represented by the same isolate number in the notational plane, to whatever basic class it is attached". This denotes that the common entity types have their own identity which remains intact no matter which conceptual model they get attached to. They give a sense of completeness to the model. Core entity types are those returned by the queries. As can be noted from the general pattern, Recipe is the core entity type of the model. The core entity type, Recipe, describes the steps that need to be followed in the process of preparing a food item. The process includes some important components, such as ingredient information (a set of ingredients along with their corresponding quantities) and instructions necessary to make the food item. Such a description is a product of the human intellect. Auxiliary entities are those that, in the general pattern, determine how to filter results. They include Cuisine, Dish, and Event.

Online Recipe [Roasted Turkey]: [Roasting]:[Turkey]:[Thanksgiving]:[North America]

Online Recipe (Context) = Domain

Roasted Turkey (Dish) = Product

Roasting (Cooking Method) = Technique

Turkey (Ingredient) = Substance

Thanksgiving (Occasion/Event) = Event

North America (Location) = Geographic Location

It is observed each of the isolate terms on the left-hand side (L.H.S) belongs to a corresponding category in on the right-hand side (R.H.S). Here, *Product* means any food product and *Technique* means any cooking technique. It, therefore, follows that each of the above terms is a category of ideas and can be regarded as a facet of an Online Recipe. Therefore, one can discern this type of organization of ideas in any domain of interest. [41] discerned that although different subjects have facets particular to them, there is an underlying unity of ideas when these facets are examined in depth. In each one of the subjects, there is a core set of ideas that are central to every aspect of the study of that subject.

Step 3: Query Refinement

a. Query in Standard Terms

Term standardization involves adapting the terms through a standard terminology accepted across the domain to maintain consistency and avoid ambiguity. Here, controlled vocabulary (two subject-specific thesauri) is used to identify the standard terms particularly used by different expert professionals in the given field. Each kernel term is replaced by its equivalent standard term as per a standard vocabulary. As the context is related to food and recipe, two domain-specific thesauri (AGROVOC and LanguaL) are referred to for query refinement. (1) AGROVOC, a multilingual structured thesaurus/controlled vocabulary, by the Food and Agriculture Organization (FAO) of the United Nations, covers concepts and terminologies related to food and agriculture. It has more than 35,000 concepts in 35 languages arranged in hierarchical order (http://aims.fao.org/agrovoc). It uses SKOS (Simple Knowledge Organization System) to support different knowledge organization systems - classification schemes, taxonomies, thesauri, and every structured vocabulary. (2) LanguaL™ (http://www.langual.org/) stands for "Langua aLimentaria" or "language of food". It provides an automated approach to describe, capture and retrieve information about food. It is a multilingual thesaurus which uses a faceted classification system to classify food products. Food components are described with the help of standard terminologies, controlled vocabulary chosen from different facets of food such as nutrition, hygiene, cooking methods, etc. For instance, the thesaurus may include the biological origin, methods of cooking and conservation, and technological treatments, etc. covering different aspects related to food and agriculture.

When the term 'Turkey' is searched in AGROVOC, as many as six senses of the term 'Turkey' is found – Turkey, Turkey's blue comb disease virus, Turkey herpesvirus, Turkey meat, Turkey poults, Turkeys. In this context, through referring to AGROVOC, Turkey Meat is chosen as an appropriate standard term to Turkey.

b. Final mapping

The facets are mapped to the knowledge model as described below:

1) Online Recipe: Online Recipe is the context for a given domain. The context is derived from the domain knowledge to solve a task [17]. For instance, domain food

can have different contextual references like nutrition, diet, food science, food policy, food photography, etc. These diverse contexts are connected by one domain - food.

2) Roasted Turkey: This is the name of the dish. Hence, this term is considered as the standard term.
3) Turkey: AGROVOC is referred to and it is realized that Turkey Meat is the standard term commonly referred to as Turkey, since Turkey Meat is a popular poultry product used for culturally significant events like Thanksgiving.
4) Roasting- As per AGROVOC, under Cooking Method the term Roasting is available. Similarly, LanguaL Thesaurus classifies the term Roasted as a subclass of Cooking Method. See Figure 2.

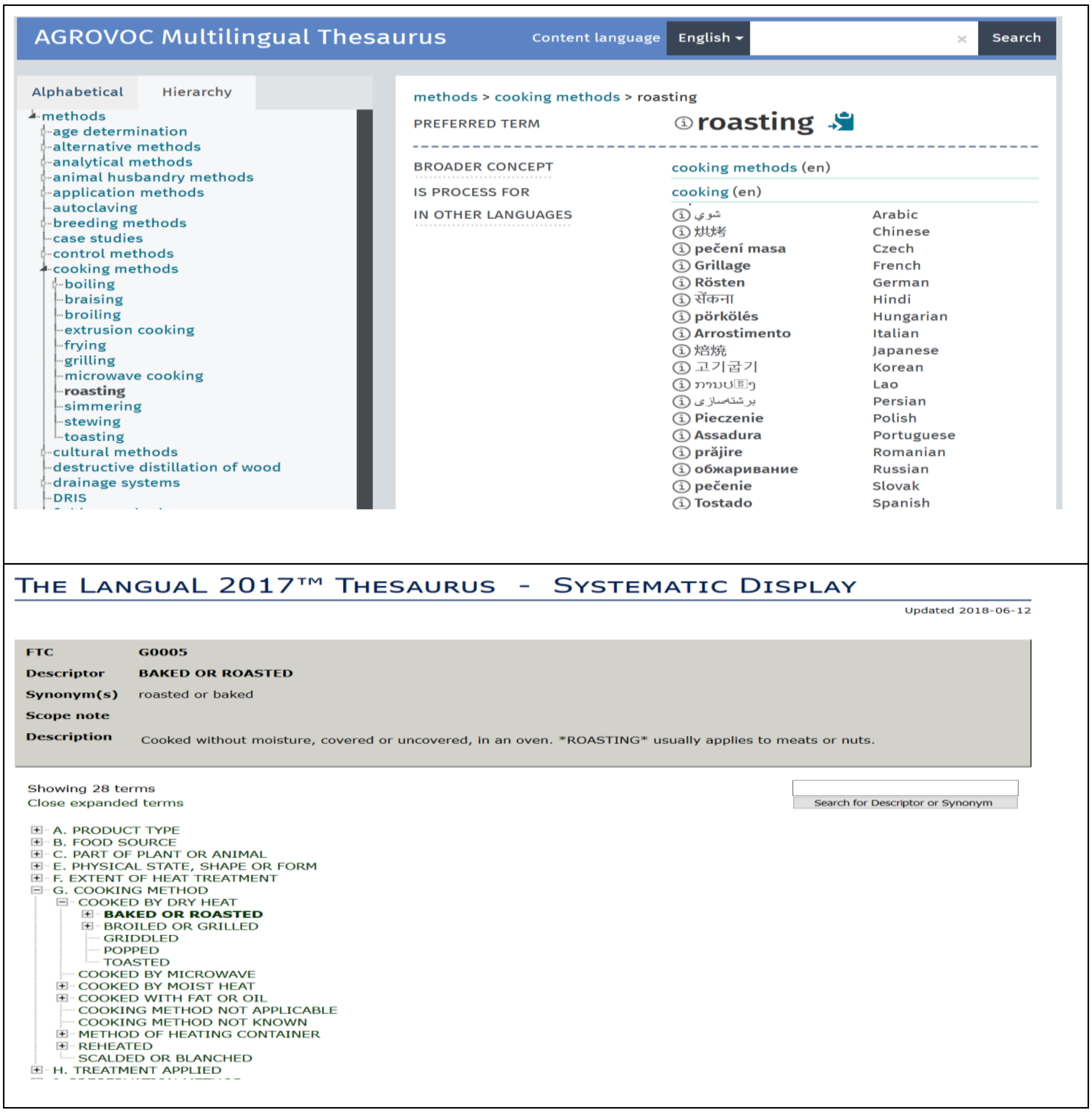

Figure 2. Identification of standard vocabulary in AGROVOC and LanguaL

5) Thanksgiving- Thanksgiving Day is celebrated in North America, in some of the Caribbean islands and in Liberia. It was originally celebrated as a festival for the harvest from the preceding year.
6) North America- The term North America is considered as a space isolate, so it can be considered as a standard term.

Hence the outcome of this step is as follows:

Online Recipe [Roasted Turkey]: [Turkey Meat]: [Roasting]: [Thanksgiving Day]: [North America]

4. Experiment

In this section the quality of the knowledge model is evaluated with three other knowledge models - Schema.org, BBC Food Ontology and Recipe Ontology - to gauge their performances on query understanding, analysis and refinement tasks.

Schema.org: Schema.org offers a schema to represent recipes. Each type is identified by a specific URL, for example, https://schema.org/Recipe for type Recipe. Each URL takes us to a page which provides definitions of the type and a list of properties, expected types and descriptions. There are as many as ten core properties enlisted with their corresponding description and their data types. Properties such as cookTime, prepTime and totalTime describe the time taken to cook the dish, prepare the dish and the total time taken to plate up the dish, respectively. They share the duration type which uses the ISO 8601 duration format. Additionally there are other sets of properties like cookingMethod (cooking technique), recipeCategory (specific set of food items served together during a meal), recipeCuisine (specific style of preparing food inherited from a country or region), recipeIngredient (food item used in a recipe) and recipeYield (total number of servings).

BBC Food Ontology: BBC Food Ontology is an ontology formalizing and extending Schema.org. Although the ontology originated from specific BBC use-cases, the applicability of the ontology reaches to the publishing of a wide range of recipe data across the globe. It publishes data about *Recipes* (a combination of ingredients and a method, created by a chef that produces certain dishes). The ontology has identified 17 classes and 22 properties. The ontology has identified several classes associated with Recipe, such as *Course* (sequence of a dish within a meal), *Cuisine* (a particular style of food, often based on a region), *Diet* (a way of selecting food that achieves a particular effect, such as eating in a way that is pregnancy-friendly or avoiding foods that contains shellfish) and *Method* (the way in which ingredients are combined, using particular techniques, in order to produce a food).

Recipe Ontology: Recipe Ontology [64] is an ontology-based recipe repository. The ontology includes several properties like *hasImage*, *hasPreparationDuration*, *hasOrigin*, *hasCalorie* and *hasServingSize*. A recipe is composed by its ingredients and instructions. An ingredient is defined as any edible substance. It has properties such as *hasSubstance, hasQuantity* and *hasUnit*. An instruction is a cooking-process. Each process is further categorized by four subclasses- *ProcessWithObject* (a process having one or more input objects), *ProcessWithDuration* (a process, which has a certain amount of time between its start and finish.), *ProcessWithHeating* (a process, which heats food and causes it to thicken and reduce in volume) and *ProcessWithPreposition* (a process having a preposition that links the input object of the process to one or more other objects, namely, prepositional objects). Additionally, the authors have reused concepts from several ontologies in Food and Agriculture domain. Recipe Ontology has mapped four of its classes - 'Recipe', 'Ingredient', 'CookingProcess' and 'EdibleThing' to its corresponding classes of BBC Food Ontology - 'Recipe', 'Ingredient', 'Step' and 'Food' respectively, using the 'owl:equivalentClass' or 'owl:subClassOf' constructs.

As a part of the evaluation process, 66 recipe related queries (Appendix A) are collected from participants, weblogs and relevant state of the art papers [44], [13]. Twenty-five participants, 9 female and 16 males, were chosen for the study. All participants were Computer Science and Information Science professionals. 15 of the participants were studying towards their Ph.D, 6 participants were post-doc researchers and the reminding 4 were lecturers and readers. Their ages ranged from 21 to 55 years old. The criteria based on which participants were selected was their familiarity with online recipe sites; namely, to actively have been searching for food and drink recipes. Further to this, the participants needed to have a minimum understanding of their regional cuisine and their experience in celebrating holidays and special events with food and drink. They were then asked in the interview to provide examples of recipe related queries typically used by them to search on the Web. Each interview lasted for around 20-30 minutes. The queries exemplify different user needs and searches posted by them on recipe websites. Many queries were alike, expressing similar needs across them (mapped in Appendix A). 16 distinct queries are identified from the user queries and analyzed. As reported in Table 2, each natural language query is converted into a structured query form to explicitly understand the query-patterns and identify entities and properties associated to the query.

Table 2. Query Analysis and Identification of Entity Types (Etypes) and their properties.

Queries	Query Analysis	Identification: Etypes and Properties
1. Beef recipes for 2	Give me all Recipe WHERE mainIngredient = "beef" AND recipeYield = 2	**Etypes**: RECIPE **Property**: Recipe.mainIngredient Recipe.recipeYield
2. Recipes using 200gm brown rice, 1litre skimmed milk.	Give me all Recips WHERE recipeIngredient = X AND X.type = "rice" and "milk" AND X.qualifier = "brown" and "skimmed" AND X.unit = "gm" and "litre" AND X.amount = 200 and 1.	**Etypes**: RECIPE **Property**: Recipe.recipeIngredient recipeIngredient.type recipeIngredient.qualifier recipeIngredient.unit recipeIngredient.amount
3. Vegan recipes.	Give me all Recipe WHERE diet = "vegan"	**Etypes**: RECIPE **Property**: Recipe.diet
4. Recipes of sugarless dessert for diabetic person	Give me all Recipe WHERE diet = "diabetic" AND Dish = X AND nutritionInformation.type != "sugar" AND X.courseType = "dessert"	**Etypes**: RECIPE, DISH **Property**: Recipe.diet nutritionInformation.type Dish.courseType
5. Low-fat main course recipes.	Give me all Recipe WHERE diet = "low-fat" AND Dish = X AND X.courseType = "main course"	**Etypes**: RECIPE, DISH **Property**: Recipe.diet Dish.courseType
6. Spicy Korean noodles recipe.	Give me all Recipe belonging to recipeCuisine X WHERE X.name = "Korean" AND recipeIngredient = Y AND Y.type = "noodles" AND Y.qualifier = "spicy"	**Etypes**: RECIPE, CUISINE **Property**: recipeIngredient.type recipeIngredient.qualifier Cuisine.name
7. Kerala recipes.	Give me all Recipe belonging to recipeCuisine = X WHERE X.region = "Kerala"	**Etypes**: RECIPE, CUISINE **Property**: Cuisine.region
8. Recipes with egg containing less than 50 calories .	Give me all Recipe WHERE mainIngredient = "egg" AND nutritionInformation.unit = calorie AND nutritionInformation.amount <= 50	**Etypes**: RECIPE **Property**: Recipe.mainIngredient nutritionInformation.unit nutritionInformation.amount
9. Recipe to make Italian pizza using microwave oven	Give me all Recipe belonging to recipeCuisine = X WHERE X.name = "Italian" AND mainIngredient = "pizza" AND utensil = "microwave oven"	**Etypes**: RECIPE, CUISINE **Property**: Cuisine.name Recipe.mainIngredient Recipe.utensil

10. Recipes for 4 people which can be prepared in less than 20 minutes.	Give me all Recipe WHERE recipeYield = 4 AND prepTime <= 20	**Etypes**: RECIPE **Property**: Recipe.recipeYield Recipe.prepTime
11. Recipes with grilled vegetables.	Give me all Recipe WHERE cookingMethod = "grilled" AND recipeIngredient = Y AND Y.type = "vegetables"	**Etypes**: RECIPE **Property**: Recipe.cookingMethod recipeIngredient.type
12. Recipes from Wolfgang Puck which is highly rated.	Give me all Recipe WHERE Work =X AND creator = "Wolfgang Puck" AND rating = "high"	**Etypes**: RECIPE, WORK **Property**: Work.creator Recipe.rating
13. Butter Chicken recipes.	Give me all Recipe for Dish X WHERE X.name = "butter chicken"	**Etypes**: RECIPE, DISH **Property**: Dish.name
14. Give me a recipe to prepare roasted turkey for Thanksgiving celebration	Give me all Recipe for Dish X AND Event Y WHERE X.name= "roasted turkey" AND Y.name = "Thanksgiving" AND mainIngredient = "turkey" AND cookingMethod = "roasting"	**Etypes**: RECIPE, DISH, EVENT **Property**: Dish.name Event.name Recipe.mainIngredient Recipe.cookingMethod
15. Give me a recipe of halal food prepared for Iftar party.	Give me all Recipe for Event X WHERE X.name = "Iftar" AND cookingMethod = "halal".	**Etypes**: RECIPE, EVENT **Property**: Event.name Recipe.cookingMethod
16. Lunch recipes that are easy to cook.	Give me all Recipe for Event X WHERE X.meal = "lunch" AND difficultyLevel = "easy"	**Etypes**: RECIPE, EVENT **Property**: Event.type.meal Recipe.difficultyLevel

Furthermore, a comparative study of Schema.org, BBC Food Ontology, Recipe Ontology and Recipe Entity Model is presented in Table 3, against the queries which are collected as part of the interview process, as part of the analysis process. The query terms are identified and mapped manually to the corresponding concepts available in four knowledge models – Schema.org, BBC Food Ontology, Recipe Ontology and Recipe Entity Model. For instance, Beef is mapped to *Recipe.recipeIngredient* (Schema.org), *Ingredient.food* (BBC Food Ontology), *Ingredient.hasSubstance* (Recipe Ontology) and *Recipe.mainIngredient* (Recipe Model); Vegan is mapped to *Recipe.suitableforDiet* (Schema.org), *Diet.diet* (BBC Food Ontology), *Recipe.diet* (Recipe Ontology) and *Recipe.diet* (Recipe Model); Butter Chicken is mapped to *Food.food* concept (BBC Food ontology), *Recipe.output* concept (Recipe Ontology) and *Dish.name* concept (Recipe Model), respectively. Moreover, the table showcases concepts absent in Schema.org framework (Dish, Event and Meal), BBC Food Ontology (Preparation Time and Meal) and Recipe Ontology (Nutritional Information, Meal) respectively.

Table 3. Comparative analysis of schema.org, BBC food ontology, recipe ontology and recipe model

Queries	SCHEMA.ORG		BBC FOOD ONTOLOGY		RECIPE ONTOLOGY		RECIPE MODEL		COMMENT
1. Beef recipes for 2	Recipe.recipeIngredient Recipe.recipeYield	√	Ingredient.food Recipe.serves	√	Ingredient.hasSubstance Recipe.hasServingSize	√	Recipe.mainIngredient Recipe.recipeYield	√	Our model distinguishes between ingredient and main ingredient.
2. Recipes using 200gm brown rice, 1 litre skimmed milk.	Recipe.recipeIngredient		Ingredient.food Ingredient.quantity Ingredient.imperial_quantity Ingredient.metric_quantity	√	Ingredient.hasSubstance Ingredient.hasQuantity Ingredient.hasUnit Ingredient.notes	√	Recipe.recipeIngredient recipeIngredient.type recipeIngredient.qualifier recipeIngredient.unit recipeIngredient.amount	√	Schema.org cannot support it as ingredient is not structured.
3. Vegan recipes.	Recipe.suitableforDiet	√	Diet.diet	√	Recipe.diet	√	Recipe.diet	√	All models can support such query as diet is included in the model.
4. Recipes of sugarless dessert for diabetic person	Recipe.nutrition Recipe.recipeCategory Recipe.suitableforDiet	√	Course.serve_as Diet.diet		Recipe.dessert Diet.diet =diabetic		nutritionInformation.type Recipe.diet Dish.courseType	√	BBC Food Ontology and Recipe Ontology cannot support it as it does not cover nutritional information.
5. Low-fat main course recipes.	Recipe. suitableforDiet Recipe.recipeCategory	√	Diet.diet Course.serve_as	√	Recipe.diet Recipe.mainCourse	√	Recipe.diet Dish.courseType	√	All models can support such query as diet and courses are included in the model.
6. Spicy Korean noodles recipe.	Recipe.recipeCuisine Recipe.recipeIngredient		Cuisine.cuisine Ingredient.food		Recipe.hasOrigin Ingredient.hasSunstance Ingredient.notes	√	recipeIngredient.qualifier Cuisine.name recipeIngredient.type	√	Recipe ontology and our model can answer such queries as we have ingredient qualifiers.
7. Kerala recipes.	Recipe.recipeCuisine	√	Cuisine.cuisine	√	Recipe.hasOrigin	√	Cuisine.region	√	All models can support such query as cuisine is included in the model.
8. Recipes with egg containing less than 50 calories.	Recipe.recipeIngredient Recipe.nutrition.calorie		Ingredient.food		Ingredient.hasSubstance Recipe.hasCalorie	√	Recipe.mainIngredient nutritionInformation.unit nutritionInformation.amount	√	BBC Food Ontology cannot support it as it does not cover nutritional information (i.e. number of calories).
9. Recipe to make Italian pizza using microwave oven	Recipe.recipeCuisine Recipe.recipeIngredient		Cuisine.cuisine Ingredient.food		Recipe.hasOrigin Ingredient.hasSubstance Ingredient.kitchenAccessory	√	Cuisine.name Recipe.mainIngredient		Kitchen appliances are only supported by Recipe Ontology.
10. Recipes for 4 people which can be prepared in less than 20 minutes.	Recipe.recipeYield Recipe.prepTime	√	Recipe.serves		Recipe.serving_Size Recipe.preparation Duration	√	Recipe.recipeYield Recipe.prepTime	√	BBC Food ontology cannot support it as it does not define any time constrain.

11. Recipes prepared with grilled vegetables.	Recipe.cooking Method Recipe.recipeIngredient	√	Method.method Ingredient.food	√	Instruction.process Ingredient.hasSubstance Ingredient.Cooking Process.haspreposit ionalObject	√	Recipe.cookingMethod recipeIngredient.type	√	All models can support such query as cooking method and ingredient information are included in the model.
12. Give me recipes of Wolfgang Puck which is highly rated.							Work.creator Recipe.rating	√	Only our model can support such queries as we have defined the creator of the recipe and the rating of a recipe.
13. Butter chicken recipes.			Food.food	√	Recipe.output	√	Dish.name	√	Schema.org does not define the name of the dish in their model.
14. Give me a recipe to prepare roasted turkey for Thanksgiving celebration	Recipe.recipeIngredient Recipe.cooking Method		Food.food Occasion.occasion Ingredient.food Method.method	√	Ingredient.Cooking Process.haspreposit ionalObject Ingredient.hasSubstance Recipe.occasion	√	Dish.name Event.name recipeIngredient.type Recipe.cookingMethod	√	Such queries cannot be supported by Schema.org as they do not define Dish and Event/Occasion which is related to Recipe.
15. Give me a recipe of halal food prepared for Iftar party.	Recipe.cooking Method		Method.method Occasion.occasion	√	Ingredient.Cooking Process.haspreposit ionalObject Recipe.occasion	√	Recipe.cookingMethod Event.name	√	Such queries cannot be supported by Schema,org as they do not define Event which is related to Recipe.
16. Lunch recipes that are easy to cook.							Event.type.meal Recipe.difficultyLevel	√	Only our model captures the concept of meal (as a type of an event). Only our model captures the concept "difficulty level" of cooking.
		7		**9**		**13**		**15**	

Additionally, the relevancy of the concepts in the queries are computed with respect to the context - Recipe. As part of the evaluation process, P@n is computed (Precision at n, where n=16 queries) [57]. This metric compute number of relevant results among the 16 identified queries:

$$P@n = \frac{(\#of\ recommended\ queries\ @n\ that\ are\ relevant)}{(\#of\ recommended\ queries\ @n)}$$

For Recall (R@n), a set of all possible properties for each n (where n=16 queries) are successfully retrieved. Here, a set of properties labelled in Table 2 are considered and recall is computed with respect to this set [29].

$$R@n = \frac{\frac{(\#retrieved\ very\ typical\ attributes\ for\ every\ n)}{(\#\ very\ typical\ attributes)}}{(\#of\ recommended\ queries\ @n)}$$

Figure 3 demonstrates the result of overall precision, recall and f-measure of the identified queries. Results show that the precision achieved by knowledge model (0.93), while the precision of Schema.org, BBC Food Ontology and Recipe Ontology is 0.43, 0.56 and 0.81, respectively. Similarly, the recall achieved by knowledge model (0.95), while the recall of Schema.org, BBC Food Ontology and Recipe Ontology is 0.63, 0.71 and 0.82, respectively. In summary, the reason for the effective and improved query understanding using Knowledge Model over Schema.org, BBC Food Ontology and Recipe Ontology is that they have not included multiple concepts associated to the domain recipe. For instance, Schema.org framework fails to cover concepts like *Dish*, *Event* and *Meal* and properties associated to them. Similarly, BBC Food Ontology does not include concepts and properties related to *Preparation Time* and *Meal*. However, it is to be noted that the recall and precision of Recipe Ontology is relatively higher than the former models. One of the main reasons is that it covers many concepts pertinent in recipe-related queries like Kitchen accessories, Cooking Process. Moreover, as Recipe Ontology reuses many ontologies in food domain, its effectiveness over the other models is justified. However, Recipe Ontology fails to address concepts like Nutritional Information and Meal. At this stage, the Recipe Model performs better than the other three models as it has considered diverse properties associated to the domain; hence the precision of the model is the highest at n=16.

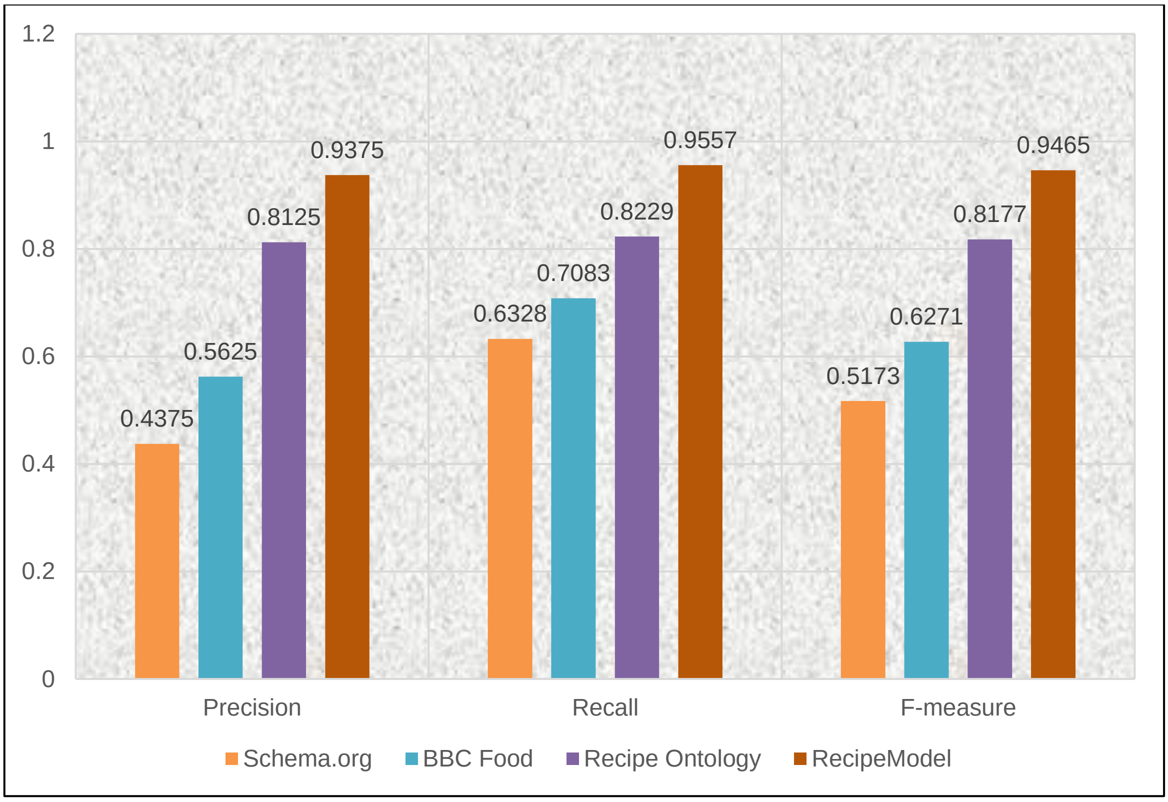

Figure 3. Knowledge model effectiveness

5. Conclusions, limitations and future work

Understanding a food related query is a challenging task as dietary preferences vary significantly, since ethnographic factors like cultural [27]; socio-economic [32]; religious [20]; or psychological factors [34] influence food choices [30]. In this paper, a knowledge model is used, which provides "... a pragmatic view of what knowledge is - it is that which can be represented as information in a structured form suitable for representation in a computer system". The knowledge model provides the precise intent of the user query. Each query term is mapped to the knowledge model to understand the relevant concepts, their properties and their relatedness with other concepts. The domain knowledge is essential in order to understand how the components interact with each other; how a specific ingredient becomes predominant in a region; how cuisines are influenced by geographic locations; how events impact the culinary world, etc. This knowledge is essential to identify the explicit, implicit and hidden components in a query. Experiments on user queries reveal that our knowledge model achieved greater importance over the three other existing models, Schema.org, BBC Food Ontology and Recipe Ontology [64] for query understanding. Applying this approach in a new domain will not be a challenge, as the creation of a knowledge model is well-defined, and the components of the model can be easily designed with the help of domain experts who have expertise in the domain. The results revealed the positive impact of the knowledge model on query expansion. This approach could be further used by search engines and organizations with specific domains.

The limitation of this work is the models have been built based on food domain. This work has been restricted to one domain and has not experimented with other domains. Moreover, the proposed methodology and evaluation are manual approaches. Although manual evaluation is one of the means by which one can assess how well the model has met a set of predefined criteria, standards and requirements [65], yet the models lack automated support. Such an approach can be used to integrate multiple ontologies that are geographically distributed, but contextually related with respect to a common domain; for example, *a restaurants database that manages clients' food preference records and another that manages clients' health and diet history records.* In such cases, an integrated knowledge model needs to be built that will capture the common, auxiliary and core entities as well as the different metadata and relevant semantics of the original distributed databases. As part of future work, the current work can be further extended to answer questions such as: How can queries be evolved automatically with partial domain information? How can knowledge models be automatically extended within the exiting framework?

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author received no financial support for the research, authorship and/or publication of this article.

References

1. Akrivas G, Wallace M, Andreou G, Stamou G, Kollias S. Context-sensitive semantic query expansion. InProceedings 2002 IEEE International Conference on Artificial Intelligence Systems (ICAIS 2002) 2002 Sep 5 (pp. 109-114). IEEE.
2. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ. UniProt: the universal protein knowledgebase. Nucleic acids research. 2004 Jan 1;32(suppl_1):D115-9.
3. Bhattacharyya G. POPSI Its fundamentals and procedure based on a general theory of subject indexing languages. SRELS Journal of Information Management. 1979 Jan 1;16(1):1-34.
4. Bhogal J, MacFarlane A, Smith P. A review of ontology based query expansion. Information processing & management. 2007 Jul 1;43(4):866-86.
5. Borrajo L, Romero R, Iglesias EL, Marey CR. Improving imbalanced scientific text classification using sampling strategies and dictionaries. Journal of integrative bioinformatics. 2011 Dec 1;8(3):90-104.
6. Broughton V. Essential classification. Facet Publishing; 2004, p. 34.
7. Broughton V. Facet analytical theory as a basis for a knowledge organization tool in a subject portal. In Lopez-Huertas, M.J. and Munoz-Fernandez, F.J. (Eds), Challenges in Knowledge Representation and Organization for the 21st Century. Integration of Knowledge across Boundaries. Proceedings of the Seventh International Conference of the International Society for Knowledge Organization, Granada, Spain, 10-13 July, Advances in Knowledge Organization, Vol. 8, Ergon, Wurzburg, pp. 135-41, available at: www.ucl.ac. uk/fatks/paper2.htm
8. Chatterjee U, Giunchiglia F, Madalli DP, Maltese V. Modeling Recipes for Online Search. In OTM Confederated International Conferences On the Move to Meaningful Internet Systems" 2016 Oct 24 (pp. 625-642). Springer, Cham.
9. Clinchant S, Renders JM. Query translation through dictionary adaptation. InWorkshop of the Cross-Language Evaluation Forum for European Languages 2007 Sep 19 (pp. 182-187). Springer, Berlin, Heidelberg.
10. Collier N, Ruch P, Nazarenko A. JNLPBA 2004: Proceedings of the International Joint Workshop on Natural Language Processing in Biomedicine and its Applications. ACL, Morristown. 2004;18.
11. Curé OC, Maurer H, Shah NH, Le Pendu P. A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest. BMC medical informatics and decision making. 2015 Dec;15(1):S8.

12. Dawson A, Brown D, Broughton V. The need for a faceted classification as the basis of all methods of information retrieval. InAslib proceedings 2006. Emerald Group Publishing Limited.
13. DeMiguel J, Plaza L, Díaz-Agudo B. ColibriCook: A CBR System for Ontology-Based Recipe Retrieval and Adaptation. InECCBR Workshops 2008 (pp. 199-208).
14. Ellis D, Vasconcelos A. The relevance of facet analysis for World Wide Web subject organization and searching. Journal of Internet Cataloging. 2000 May 31;2(3-4):96-114.
15. Fu G, Jones CB, Abdelmoty AI. Ontology-based spatial query expansion in information retrieval. InOTM Confederated International Conferences" On the Move to Meaningful Internet Systems" 2005 Oct 31 (pp. 1466-1482). Springer, Berlin, Heidelberg.
16. Giunchiglia F, Dutta B, Maltese V. From knowledge organization to knowledge representation. KO KNOWLEDGE ORGANIZATION. 2014 Feb 14;41(1):44-56.
17. Giunchiglia F, Maltese V, Dutta B. Domains and context: first steps towards managing diversity in knowledge. Web Semantics: Science, Services and Agents on the World Wide Web. 2012 Apr 1;12:53-63.
18. Greenberg J. Optimal query expansion (QE) processing methods with semantically encoded structured thesauri terminology. Journal of the American Society for Information Science and Technology. 2001;52(6):487-98.
19. Hahm GJ, Yi MY, Lee JH, Suh HW. A personalized query expansion approach for engineering document retrieval. Advanced Engineering Informatics. 2014 Oct 1;28(4):344-59.
20. Heiman A, Gordon B, Zilberman D. Food beliefs and food supply chains: The impact of religion and religiosity in Israel. Food Policy. 2017 Jul 31.
21. Hirschman L, Yeh A, Blaschke C, Valencia A. Overview of BioCreAtIvE: critical assessment of information extraction for biology.2005:S1.
22. Hudon M. NISO: Guidelines for the Construction, Format, and Management of Monolingual Thesauri: ANSI/NISO Z39. 19-1993. KNOWLEDGE ORGANIZATION. 1995;22:180-.
23. Ingwersen P, Wormell I. Ranganathan in the perspective of advanced information retrieval. Libri. 1992;42(3):184-201.
24. Jones C, Purves R, Ruas A, Sanderson M, Sester M, Van Kreveld M, Weibel R. Spatial information retrieval and geographical ontologies: An overview of the SPIRIT project. InProceedings of 25th ACM Conference of the Special Interest Group in Information Retrieval 2002 (pp. 389-390). ACM.
25. Jones S. A thesaurus data model for an intelligent retrieval system. Journal of Information Science. 1993 Jun;19(3):167-78.
26. Kapidakis S, Mastora A, Peponakis M. Query expansion of zero-hit subject searches: using a thesaurus in conjunction with NLP techniques. InInternational Conference on Theory and Practice of Digital Libraries 2012 Sep 22 (pp. 433-438). Springer, Berlin, Heidelberg.
27. Kittler PG, Sucher KP. Food and culture in America: a nutrition handbook. Wadsworth Publishing Company, Inc.; 1998.
28. Kuhlthau CC. Inside the search process: Information seeking from the user's perspective. Journal of the American society for information science. 1991 Jun;42(5):361-71.
29. Lee T, Wang Z, Wang H, Hwang SW. Attribute extraction and scoring: A probabilistic approach. In2013 IEEE 29th International Conference on Data Engineering (ICDE) 2013 Apr 8 (pp. 194-205). IEEE.

30. Leung G, Stanner S. Diets of minority ethnic groups in the UK: influence on chronic disease risk and implications for prevention. Nutrition Bulletin. 2011 Jun;36(2):161-98.
31. Levow GA, Oard DW, Resnik P. Dictionary-based techniques for cross-language information retrieval. Information processing & management. 2005 May 1;41(3):523-47.
32. Lewin K. Forces behind food habits and methods of change. Bulletin of the national Research Council. 1943 Oct;108(1043):35-65.
33. Liddy ED, Paik W, Yu ES, McVearry KA. An overview of DR-LINK and its approach to document filtering. In Proceedings of the workshop on Human Language Technology 1993 Mar 21 (pp. 358-362). Association for Computational Linguistics.
34. Lyman B. A psychology of food: More than a matter of taste. Springer Science & Business Media; 2012 Dec 6.
35. Mandala R, Tokunaga T, Tanaka H. Combining multiple evidence from different types of thesaurus for query expansion. InSIGIR 1999 Aug 15 (Vol. 99, pp. 15-19).
36. Mandala R, Tokunaga T, Tanaka H. Query expansion using heterogeneous thesauri. Information Processing & Management. 2000 May 1;36(3):361-78.
37. Miller GA, Beckwith R, Fellbaum C, Gross D, Miller KJ. Introduction to WordNet: An on-line lexical database. International journal of lexicography. 1990 Dec 1;3(4):235-44.
38. Nilsson K, Hjelm H, Oxhammar H. SUiS–cross-language ontology-driven information retrieval in a restricted domain. InProceedings of the 15th Nordic Conference of Computational Linguistics (NODALIDA 2005) 2006 May (pp. 139-145).
39. Priss U. Lattice-based information retrieval. KO KNOWLEDGE ORGANIZATION. 2000 Sep 14;27(3):132-42.
40. Qiu Y, Frei HP. Concept based query expansion. InProceedings of the 16th annual international ACM SIGIR conference on Research and development in information retrieval 1993 Jul 1 (pp. 160-169). ACM.
41. Ranganathan SR. Colon classification: basic classification. Asia Pub. House; 1963.
42. Ranganathan SR. Prolegomena to Library Classification. 1967.
43. Rivas AR, Iglesias EL, Borrajo L. Study of query expansion techniques and their application in the biomedical information retrieval. The Scientific World Journal. 2014;2014.
44. Sam M, Krisnadhi A, Wang C, Gallagher JC, Hitzler P. An Ontology Design Pattern for Cooking Recipes-Classroom Created. InWOP 2014 Oct 19 (pp. 49-60).
45. Seher I, Ginige A, Shahrestani SA. A personalized query expansion approach using context.2007: 383-390.
46. Selvaretnam B, Belkhatir M. A linguistically driven framework for query expansion via grammatical constituent highlighting and role-based concept weighting. Information Processing & Management. 2016 Mar 1;52(2):174-92.
47. Shin K, Han SY. Improving information retrieval in MEDLINE by modulating MeSH term weights. InInternational Conference on Application of Natural Language to Information Systems 2004 Jun 23 (pp. 388-394). Springer, Berlin, Heidelberg.
48. Shiri A, Revie C. Query expansion behavior within a thesaurus ‐ enhanced search environment: A user ‐ centered evaluation. Journal of the American Society for Information Science and Technology. 2006 Feb 15;57(4):462-78.

49. Shokouhi M. Learning to personalize query auto-completion. InProceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval 2013 Jul 28 (pp. 103-112). ACM.
50. Sihvonen A, Vakkari P. Subject knowledge, thesaurus-assisted query expansion and search success. InCoupling approaches, coupling media and coupling languages for information retrieval 2004 Apr 26 (pp. 393-404). LE CENTRE DE HAUTES ETUDES INTERNATIONALES D'INFORMATIQUE DOCUMENTAIRE.
51. Smeaton AF, Kelledy F, O'Donnell R. TREC-4 experiments at Dublin City University: Thresholding posting lists, query expansion with WordNet and POS tagging of Spanish. Harman [6]. 1995:373-89.
52. Tudhope D, Binding C, Blocks D, Cunliffe D. Query expansion via conceptual distance in thesaurus indexed collections. Journal of Documentation. 2006 Jul 1;62(4):509-33.
53. van Rijsbergen CJ. Towards an information logic. InACM SIGIR Forum 1989 Jun 25 (Vol. 23, No. SI, pp. 77-86). ACM.
54. Vélez B, Weiss R, Sheldon MA, Gifford DK. Fast and effective query refinement. InIN PROC. OF THE 20TH INTL. ACM SIGIR CONF. ON RESEARCH AND DEVELOPMENT IN INFORMATION RETRIEVAL 1997.
55. Vickery BC. Faceted classification: A guide to construction and use of special schemes (London: ASLIB 1960).
56. Voorhees EM. Query expansion using lexical-semantic relations. InSIGIR’94 1994 (pp. 61-69). Springer, London.
57. Wang Z, Zhao K, Wang H, Meng X, Wen JR. Query understanding through knowledge-based conceptualization. InTwenty-Fourth International Joint Conference on Artificial Intelligence 2015 Jun 27.
58. Woods WA. Knowledge Representation: What’s Important About It?. InThe Knowledge Frontier 1987 (pp. 44-79). Springer, New York, NY.
59. Tudhope D and Binding C. Still quite popular after all those years – the continued relevance of the information retrieval thesaurus. Knowl Organ 2016; 43(3): 174–179.
60. Wu Y. Enriching a thesaurus as a better question-answering tool and information retrieval aid. Journal of Information Science. 2018 Aug;44(4):512-25.
61. Zhou D, Wu X, Zhao W, Lawless S, Liu J. Query expansion with enriched user profiles for personalized search utilizing folksonomy data. IEEE Transactions on Knowledge and Data Engineering. 2017 Feb 13;29(7):1536-48.
62. Azad HK, Deepak A. Query expansion techniques for information retrieval: A survey. Information Processing & Management. 2019 Sep 1;56(5):1698-735.
63. Keyword (2020) https://www.keyworddiscovery.com/keyword-stats.html
64. Öztürk Ö, Özacar T. A case study for block-based linked data generation: Recipes as jigsaw puzzles. Journal of Information Science. 2020 Jun;46(3):419-33.
65. Lozano-Tello A, Gómez-Pérez A. Ontometric: A method to choose the appropriate ontology. Journal of Database Management (JDM). 2004 Apr 1;15(2):1-8.
66. Skinner M, Kallumadi S. E-commerce Query Classification Using Product Taxonomy Mapping: A Transfer Learning Approach. IneCOM@ SIGIR 2019.
67. Livingston KM, Bada M, Baumgartner WA, Hunter LE. KaBOB: ontology-based semantic integration of biomedical databases. BMC bioinformatics. 2015 Dec;16(1):1-21.
68. Li W, Wu L, Xie Z, Tao L, Zou K, Li F, Miao J. Ontology-based question understanding with the constraint of Spatio-temporal geological knowledge. Earth Science Informatics. 2019 Dec;12(4):599-613.

69. Kumar S, Politi R. Understanding User Query Intent and Target Terms in Legal Domain. InInternational Conference on Applications of Natural Language to Information Systems 2019 Jun 26 (pp. 41-53). Springer, Cham.
70. Nakade V, Musaev A, Atkison T. Preliminary research on thesaurus-based query expansion for Twitter data extraction. InProceedings of the ACMSE 2018 Conference 2018 Mar 29 (pp. 1-4).

Appendix A

Set of 16 Questions	
1. Beef recipes for 2	Question 1
2. Recipes using 200gm brown rice, 1litre skimmed milk.	Question 2
3. Vegan recipes.	Question 3
4. Recipes of sugarless dessert for diabetic person	Question 4
5. Low-fat main course recipes.	Question 5
6. Spicy Korean noodles recipe.	Question 6
7. Kerala recipes.	Question 7
8. Recipes with egg containing less than 50 calories .	Question 8
9. Recipe to make Italian pizza using microwave oven	Question 9
10. Recipes for 4 people which can be prepared in less than 20 minutes.	Question 10
11. Recipes with grilled vegetables.	Question 11
12. Recipes from Wolfgang Puck which is highly rated.	Question 12
13. Butter Chicken recipes.	Question 13
14. Give me a recipe to prepare roasted turkey for Thanksgiving celebration	Question 14
15. Give me a recipe of halal food prepared for Iftar party.	Question 15
16. Lunch recipes that are easy to cook.	Question 16
Questions collected from Interview	Mapping
1. Search based on what is available in pantry and fridge.	Question 1, 2, 8
2. During some family visits or birthday celebrations some new dish.	Question 14, 15

3.	Recipes some main course/side dish.	Question 5
4.	During reunion a favourite dish prepared for all. [like baked item, barbeque, rice/bread]	Question 14, 15
5.	Search for recipes with rice or zucchini	Question 1, 2, 8
6.	New type of cheese cake or vegan alternative or traditional dish	Question 3, 13
7.	Recipe without using microwave and oven and cooks easily.	Question 9, 16
8.	Search a bunch of recipes and select the most suitable one.	General Question
9.	Heard of a dish, search for recipe instructions as it is an experimental process.	Question 13
10.	From an available list of ingredients choose whether it can be fried provided in less time and it should be very tasty. Preparation time is most significant here.	Question 2,10
11.	Sometimes depend on available ingredients. If cooking 1^{st} course then look for rice or pasta recipe.	Question 2, 5
12.	Search for recipes with following combinations: a. Rice with egg b. Rice with spinach c. Not (Rice with egg, potato, ...)	Question 1, 2, 8
13.	Search for cooking process when the recipe is already known.	Question 13
14.	To prepare cakes search for ingredients, instruction.	Question 13
15.	Search based on special occasion like Christmas Recipe	Question 14,15
16.	Recipe of a typical risotto	Question 13
17.	Search for some complicated dish based on time, special occasion, based on which time in the year then decide main course.	Question 14,15, 5
18.	Search is based on ingredient/course(1^{st} course or 2^{nd} course).	Question 2, 5
19.	Potato is available in his grocery. So look for recipes on Potatoes.	Question 2, 8
20.	Recipe of FalsoMago, a dish that you might have tasted in a restaurant.	Question 13
21.	For a known dish, search for several recipes, and select the one which suits us the best.	Question 13
22.	Known recipe, but search online to make variations.	Question 13
23.	Browse some recipes based on some suggestions.	Very General
24.	Search for a typical Antipasti where available ingredients are prawn, tomato and bread	Question 2, 5
25.	Recipes with NO milk	Question 4
26.	Recipes with 2 potatoes, 1 egg and less than 100 gr flour	Question 2
27.	Recipes with eggs that are easy to cook	Question 8, 16

28. Popular recipes with chicken"	Question 1, 12
29. Recipes of ... for 6 people	Question 1
30. Recipes by such and such famous person	Question 12
31. Recipe of crunchy spicy Japanese Noodle for dinner	Question 6,16
32. Indian recipes prepared with fried cauliflower and baked beans without spices.	Question 7,11,4
33. Recipes for Iftar party	Question 14,15
34. Give me a rice recipe of Gordon Ramsay prepared for lunch in less than 20minutes and which has high rating.	Question 10,12
From SOA: An Ontology Design Pattern for Cooking Recipes - Classroom Created [44]	
35. Question 1 "Breakfast dishes I can prepare with 2 potatoes and 100 grams of wheat flour."	Question 2, 16
36. Question 2 "Gluten-free desserts with less than 100 calories."	Question 4
37. Question 3 “Mexican dishes which do not use chili."	Question 4, 6
38. Question 4 “Easy Gordon Ramsey breakfast dishes.”	Question 12, 16
39. Question 5 “Grilled meat in less than 1 hour."	Question 10, 11
40. Question 6 “Spicy Korean beef dishes."	Question 6, 7
41. Question 7 “Crunchy brownie recipes."	Question 13
42. Question 8 ‘Cold appetizers."	Question 16
43. Question 9 “Baked/Mashed potatoes."	Question 11
44. Question 10 “Easy desserts with less than 100 calories."	Question 4, 8, 16
From SOA: ColibriCook:A CBR System for Ontology-Based Recipe Retrieval and Adaptation [13]	
45. Exercise 1: Cook a main dish with meat and cauliflower.	Question 2,5
46. Exercise 2: I would like to have a nut-free cake.	Question 4
47. Exercise 3: Prepare a Chinese dessert with fruit.	Question 9
48. Exercise 4: Cook a main dish with turkey, pistachio, and pasta.	Question 2, 5
49. Exercise 5: I would like to cook eggplant soup.	Question 13
50. Exercise 6: I want to have a salad with tomato but I hate garlic and cucumber.	Question 2, 4