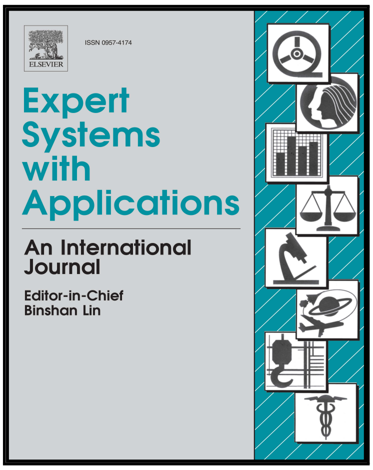
Accepted Manuscript



Multi-faceted Assessment of Trademark Similarity

Rossitza Setchi , Fatahiyah Mohd Anuar

PII: S0957-4174(16)30421-3
DOI: 10.1016/j.eswa.2016.08.028
Reference: ESWA 10821

To appear in: *Expert Systems With Applications*

Received date: 22 December 2015
Revised date: 3 August 2016
Accepted date: 4 August 2016

Please cite this article as: Rossitza Setchi , Fatahiyah Mohd Anuar , Multi-faceted Assessment of Trademark Similarity, *Expert Systems With Applications* (2016), doi: 10.1016/j.eswa.2016.08.028

Highlights

- A novel method for the assessment of trademark similarity is proposed.
- The method blends together visual, semantic and phonetic similarity.
- It produces an aggregated score based on the individual assessments.
- Evaluation using information retrieval measures and human judgment.

Multi-faceted Assessment of Trademark Similarity

Rossitza Setchi[1] and Fatahiyah Mohd Anuar[1,2],

[1] School of Engineering, Cardiff University, 14-17 The Parade, Cardiff CF24 3AA, UK

E-mail: setchi@cf.ac.uk

[2] Faculty of Engineering, Multimedia University, Cyberjaya, 63100, Malaysia

E-mail: fatahiyah@mmu.edu.my

***Abstract*—**Trademarks are intellectual property assets with potentially high reputational value. Their infringement may lead to lost revenue, lower profits and damages to brand reputation. A test normally conducted to check whether a trademark is highly likely to infringe other existing, already registered, trademarks is called a likelihood of confusion test. One of the most influential factors in this test is establishing similarity in appearance, meaning or sound. However, even though the trademark registration process suggests a multi-faceted similarity assessment, relevant research in expert systems mainly focuses on computing individual aspects of similarity between trademarks. Therefore, this paper contributes to the knowledge in this field by proposing a method, which, similar to the way people perceive trademarks, blends together the three fundamental aspects of trademark similarity and produces an aggregated score based on the individual visual, semantic and phonetic assessments. In particular, semantic similarity is a new aspect, which has not been considered by other researchers in approaches aimed at providing decision support in trademark similarity assessment. Another specific scientific contribution of this paper is the innovative integration, using a fuzzy engine, of three independent assessments, which collectively provide a more balanced and human-centered view on potential infringement problems. In addition, the paper introduces the concept of degree of similarity since the line between similar and dissimilar trademarks is not always easy to define especially when dealing with blending three very different assessments. The work described in the paper is evaluated using a database comprising 1,400 trademarks compiled from a collection of real legal cases of trademark disputes. The evaluation involved two experiments. The first experiment employed information retrieval measures to test the classification accuracy of the proposed method while the second used human collective opinion to examine correlations between the trademark scoring/rating and the ranking of the proposed method, and human judgment. In the first experiment, the proposed method improved the F-score, precision and accuracy of classification by 12.5%, 35% and 8.3%, respectively, against the best score computed using individual similarity. In the second experiment, the proposed method produced a perfect positive Spearman rank correlation score of 1.00 in the ranking task and a pairwise Pearson correlation score of 0.92 in the rating task. The test of significance conducted on both scores rejected the null hypotheses of the experiment and showed that both scores correlated well with collective human judgment. The combined overall assessment could add value to existing support systems and be beneficial for both trademark examiners and trademark applicants. The method could be further used in addressing recent cyberspace phenomena related to trademark infringement such as customer hijacking and cybersquatting.

***Keywords*—**Trademark assessment, trademark infringement, trademark retrieval, degree of similarity, fuzzy aggregation, semantic similarity, phonetic similarity, visual similarity.

1. Introduction

Trademarks are valuable intellectual property (IP) assets that identify the commercial source or origin of products or services. They are visual signs in the form of logos or brand names that allow goods or services to be easily recognized and distinguished by consumers. Similar to other intangible company assets, trademarks can be subject to legal protection. Trademark registration through an IP office provides legal protection for companies and individuals on registered marks in the jurisdiction(s) that the registration office covers. It therefore provides legal certainty and underpins the right of the trademark owner.

Trademark infringement is a form of IP crime that may lead to lost revenue, lower profits and additional costs, such as the legal fees necessary to enforce a trademark. In addition, trademark infringement is time-consuming when enforcing rights and, perhaps more importantly, can lead to severe damage of brand reputation. Recent statistics show that trademark infringement has become a serious economic and legal issue. For example, the United States International Trade Commission, as reported by the Chairman of the Joint Economic Committee, stated that the number of investigated infringement cases rose from the year 2010 to 2011 by 23.2%. A total of 3,400 trademark infringement cases were filed in the US District Courts in 2012, which excluded a presumably even larger number of cases where settlements were reached prior to the filing of cases (Scott, 2013). Some of the reported cases involve new cybercrime phenomena such as customer hijacking and cybersquatting (Scott, 2013). In another investigation conducted by the US International Trade Commission in 2011, the average annual increase of trademark litigation cases concerning US-based companies from 2002–2011 was 39.8% (US International Trade Commission, 2011). Despite these alarming trademark infringement statistics, the number of newly registered trademarks together with the existing trademarks used in the market continues to grow (Office for Harmonization in the Internal Market [OHIM], 2012; Dodell, 2013). This trend, which has been observed worldwide, has recently created administrative problems for many trademark registration offices as the registration process has become more complex and lengthy.

The trademark registration process includes a trademark similarity examination (OHIM, 2014), which requires a multi-faceted similarity assessment. One of the steps involved is making sure that the trademark to be registered is not similar to any trademark that has already been registered, as the registration of trademarks that are found to be identical or similar to any existing trademarks and provide identical or similar goods or services may potentially be opposed, as indicated in section 5 of the Trade Marks Act 1994 (Trade Marks Act, 1994). This is important in order to avoid infringements and protect the rights of existing registered trademarks.

The current practice of examining trademark similarity generally involves a search to retrieve relevant trademarks from a very large trademark database on the basis of a specific type of similarity. For example, the Industrial Property Automation System (IPAS), a support system developed by the World Industrial Property Organisation (WIPO), provides three trademark search options, namely a bibliography search based on the filing date and registration number, a phonetic search based on phonetic rules and common prefixes and suffixes, and a logo search based on the Vienna classification code for figurative trademarks (WIPO, 2014).

The research in this paper is motivated by the guidelines in the trademark examination manual, which require overall similarity assessment. From a theoretical point of view, the paper contributes to the body of knowledge in the area of intelligent human-centered decision support and in particular the use of fuzzy logic and semantics in complex evaluations and assessments related to infringement and the likelihood of confusion. Previous research has addressed some of these aspects to a certain degree. For example, the need to consider many facets or aspects in complex evaluations has been recognized by a number of researchers working in various domains. Many of them employ fuzzy logic, which is a particularly suitable reasoning technique in domains where the selection of the best alternative is highly complex and the judgement is based on subjective perceptions (Mardani et al., 2015). For example, a knowledge evaluation method aimed at estimating the quality of knowledge and its market value uses fuzzy logic to aggregate several aspects including knowledge complexity, marketable value, and the reputation of the knowledge supplier (Chen, 2011). Fuzzy numbers are also used to calculate the value of a patent and the chance of mitigation (Agliardi and Agliardi, 2011), which similar to quality of knowledge in the above example, are also parameters very difficult to measure objectively. Semantics and fuzzy logic are employed in group decision making (Gupta and Mohanty, 2016), consensus building (Li et al., 2017), opinion mining (Martínez-Cruz et al., 2016) and knowledge management (Li et al., 2011).

This paper offers an original approach to the problem of trademark infringement, which is based on multi-facet assessment and verified through human judgement. The proposed computational method for assessing trademark similarity employs multi-faceted evaluation of the three main aspects of trademark similarity: visual, semantic and phonetic. In particular, semantic similarity is a new aspect which has not been considered in any previous approaches aimed at developing decision support systems for trademark similarity assessment. Therefore, the specific scientific contribution of this paper is the innovative integration, using a fuzzy engine, of three independent assessments, which collectively provide a more balanced view on potential infringement problems. The combined overall assessment could add value to existing support systems and be beneficial for both trademark examiners and trademark applicants.

The rest of the paper is organized as follows: The next section provides an overview of existing trademark search systems and briefly discusses fuzzy logic, the inference concept employed in this research. The proposed computational method is introduced in Section 3. Section 4 describes the experimental setup and presents the results. A discussion is provided in Section 5. Section 6 concludes the study.

2. Related Work

This section reviews related work in the scope of this study. It consists of two subsections. The first subsection reviews existing trademark search systems, and the second subsection briefly discusses the concept of fuzzy inference, which inspired the development of the proposed method for the multi-faceted assessment of trademark similarity.

2.1 Existing Trademark Search Systems

Table 1 shows examples of trademarks with different types of similarity: visual, semantic and phonetic. The trademark pair NEXT and NEST possess some degree of visual similarity due to the total number of letters and the number of identical letters used. In addition, although NEXT is a figurative trademark, its style/font is similar to the typeface font of the trademark NEST, which contributes to the visual similarity between them. The second pair, MAGIC TIMES and MAGIC HOUR, are semantically similar due to the identical word that they share and the lexical relation between the non-identical words in the trademark text. The last pair, i.e. SVIZZEROTALER and SWISS TALER, are phonetically similar because although these trademarks are spelled differently, their pronunciation is similar.

<insert Table 1 here>

Many trademarks share more than one type of similarity; however, despite the existing variety in the types of similarity, most of the research in this area is focused on retrieving trademarks based on their visual similarity using low-level features. Examples of such systems include TRADEMARK (Kato et al., 1990), STAR (Wu et al., 1996) and ARTISAN (Eakins et al., 1996), which have been widely referred to by many researchers. TRADEMARK uses graphical descriptor vectors derived from shape features while STAR employs a traditional content-based image retrieval (CBIR) framework together with a set of shape-based descriptors, including Fourier descriptors, gray-level projection and moment invariants. In addition, it utilizes the spatial layout of the images although this has been found to be extremely challenging. ARTISAN also utilizes shape-based feature descriptors but includes Gestalt-based principles to retrieve abstract geometric trademark designs.

These three studies have inspired further research in trademark image retrieval focused on the visual similarity aspect of trademarks. For example, Kim and Kim (1998) employed a moment-based shape descriptor and analyzed the distribution model of 90 moment orders for all the images in their database. A closed contour shape descriptor using angle code strings was developed by Peng and Chen (1998). Jain and Vailaya (1998) proposed the use of the edge direction histogram and improved the descriptor so that it became scale and rotation invariant. Other research includes a comparative study of several common shape-based descriptors for trademark similarity comparison (Eakins et al., 2003) and a compositional shape descriptor that combines several shape descriptors (Hong & Jiang, 2008; Wei et al., 2009).

Despite the amount of work undertaken, visual similarity assessment is mainly limited to trademarks with figurative marks or logos. Notwithstanding, statistics of registered trademarks in five European countries have shown that only 30% of all trademarks employ logos as their proprietary marks (Schietse et al., 2007). The trademark similarity of the remaining 70% of registered trademarks is still insufficiently researched. For example, despite the recent advances in computational semantics, the existing trademark search systems that focus on text are primarily built around keyword-based retrieval or approximate string matching. Such systems return trademarks that match parts or entire words in query text. In Europe, OHIM recently launched a search system that allows trademark applicants and third parties to search for trademarks in different languages (OHIM, 2013). The system also provides an advanced search option that offers three search types: *word prefix*, *full phrase* and *exact match*. In the United Kingdom, the UK Intellectual Property Office (IPO) offers similar search options with an additional option that looks for similar query strings (UK IPO, 2013). The IPO search system utilises an approximate string-matching technique, which looks for fairly similar patterns in strings, together with several predefined criteria including word length and the number of similar and dissimilar letters shared by the words. Despite their usefulness, the comparison mechanism employed in such systems limits their effectiveness as it does not cover all similarity aspects that are normally assessed during the trademark examination process.

Advances in computational semantics provide an opportunity to overcome the limitations of traditional text-based retrieval by exploring semantic similarity. In the context of trademark similarity examination and analysis, it allows the comparison of trademarks based on their semantic similarity derived using external knowledge sources such as lexical ontologies. From the point of view of knowledge engineering, a lexical ontology is a framework that specifies the underlying structure and lexical relationships for knowledge representation and the organization of lexical information (Storey et al., 1998). On the other hand, advances in computational linguistics and genealogy provide a mechanism to compare trademarks based on their phonetic similarity (Convington, 1998; Kondrak, 2003; Pfeifer et al., 1996; Philips, 2000). This includes computational linguistics studies of similarities between cognates, i.e. words from different languages that share the same linguistic origin and etymology, and name-matching applications in genealogy, which retrieve similar names despite spelling variations.

This research promotes the view that the existing work on visual similarity can be extended using the recent advances in semantic retrieval, computational linguistics and computational genealogy. This approach is consistent with the requirement for holistic assessment outlined in the OHIM trademark manual (OHIM, 2014). Trademark comparison based on visual, semantic and phonetic similarity, individually, has been the paramount focus of the present authors' previous work (Anuar et al., 2013; 2014; 2016). The main contribution of this paper is that it extends previous approaches by providing a consolidated holistic assessment process. In addition, the paper introduces the concept of *degree of similarity* since the line between similar and dissimilar is not always easy to define.

2.1 Fuzzy Logic

Studies on information retrieval of music and artist recommendations (McFee & Lanckriet, 2009; Zhang et al., 2009) compute multi-faceted similarity based on low-level features and subjective criteria. Fuzzy logic has not yet been applied to multi-faceted similarity assessment but has been used in many applications that require human reasoning and decision-making. Examples include control systems in the engineering domain, doctor–patient decision-making in the medical domain, and risk analysis in e-commerce (Abou & Saleh, 2011; Fazzolari et al., 2013; Ngai & Wat, 2005). Furthermore, the concept of fuzzy logic has long been recognized in legal studies (Cook, 2001; Kosko, 1994), which is an important consideration in the area of IP rights protection. This paper promotes the use of fuzzy logic to compute the degree of similarity between trademarks due to its natural modelling capability that can mimic the very complex system underlying the human mind.

The concept of fuzzy logic was first introduced as a mathematical tool for dealing with uncertainty (Zadeh, 1965). From the point of view of set theory, the concept of fuzzy logic is an extension of the crisp set concept in which every preposition must be either *'true'* or *'false'* or in a range of values. Instead, fuzzy logic asserts that every preposition can simultaneously have a certain degree of a membership function of the *'true'* or *'false'* class. An inference system based on fuzzy logic uses fuzzy set operations and properties for reasoning and consists of a fuzzy rule base. A fuzzy rule generally has two components, the *IF* component, i.e. the antecedent, which describes a condition, and the *THEN* component, i.e. the consequent, which describes a conclusion. It follows the format:

IF <antecedent>, THEN <consequent> (1)

In the context of a human-oriented process that requires approximate human reasoning or decision-making based on experiences and insights, a human inference system tends to use verbal variables to create verbal rules in a form similar to Eq. 1. Since the terms and variables used in human inference systems are normally fuzzy rather than precise, a fuzzy inference system is highly applicable in such applications. Verbal terms and variables can therefore be expressed mathematically as membership degrees and membership functions with symbolic verbal phrases rather than numeric values. Indirectly, this provides a systematic mechanism to utilize the uncertain and imprecise information used in human judgment.

The implementation of the fuzzy inference approach in various applications commonly involves two inference models, i.e. the Mamdani inference model, which is based on a fuzzy relational model, and the Takagi–Sugeno inference model (Akgun, 2012). Both models employ slightly different approaches in the output aggregation process in that Mamdani uses defuzzification and Takagi–Sugeno employs weighted average to compute the crisp output. An alternative approach is the Tsukamoto model, which represents the consequent of the fuzzy rules with monotonical membership functions (Jang et al., 1997). A more recent approach is the inference model based on a combination of adaptive neural networks and fuzzy logic (Leng et al., 2009).

The Mamdani inference model is employed in this paper due to its intuitive and linguistic model applicability, which makes it very suitable for human-oriented applications.

3. Trademark Degree-of-Similarity Aggregation Method

This section introduces the proposed method and highlights the main steps involved in it. The method was based on a systematic analysis of 1,400 trademarks extracted from real dispute cases. This analysis revealed that the trademark cases in the collection were either real words/phrases such as 'MAGIC HOUR', out-of-vocabulary words/phrases such as 'SVIZZEROTALER' or a combination of both. In addition, the analysis also showed that in cases involving only out-of-vocabulary words, only visual and phonetic assessments were performed since such words do not carry any lexical meaning. The four different types of trademarks defined in OHIM (2014), namely word mark, figurative word mark, purely figurative mark and purely figurative mark with figurative word mark (Fig. 1), require different processing techniques and analytical approaches, hence the development of a method that facilitates the similarity comparison of both real words and out-of-vocabulary words.

<insert Fig. 1 here>

The conceptual model of the proposed system (Fig. 2) comprises four main modules. Three of these modules assess trademarks in terms of their visual, semantic and phonetic similarity while the fourth module, the fuzzy inference engine, aggregates the final score based on the three individual assessments. Each module has its individual functional requirements and uses a different approach to achieve its predefined function. For example, the visual similarity module employs visual descriptors based on the shape features of the individual letters included in the trademarks. Fig. 3 shows the flowchart of the proposed method and the four individual steps involved: (i) individual assessments, (ii) fuzzification, (iii) inference and (iv) defuzzification.

<insert Fig. 2 here>

<insert Fig. 3 here>

3.1 Step 1: Assessment of Visual, Semantic and Phonetic Similarity

This step involves the assessment of the three main aspects of similarity. The *visual* similarity assessment of purely figurative trademarks such as logos is computed using an advanced algorithm (Anuar et al., 2013) that employs global and local shape features, i.e. Zernike moment and an edge-gradient co-occurrence matrix, represented as vectors. The similarity between the trademarks is then computed using normalized Euclidean distances between their corresponding vectors. The same approach, combined with the string algorithm (Navarro, 2001) is used to compute the visual similarity of trademarks with word marks and figurative word marks (Table 2). Unlike approximate string matching that uses binary values in the letter-to-letter comparison, such as '1' and 'l' in the example shown in Fig. 4, the algorithm developed in this paper computes the visual similarity between letters using their shape descriptors. This provides a mechanism that differentiates between different letters and numbers that look similar, such as '1' and 'l', and less similar letters and numbers, such as '1' and 'X'. Table 3 shows that the proposed algorithm exhibits better discriminating power compared to approximate string matching.

<insert Table 2 here>

<insert Fig. 4 here>

<insert Table 3 here>

The trademark *semantic* similarity assessment is based on a similarity computation model (Anuar et al., in press), which utilizes a lexical ontology, i.e. WordNet, as an external knowledge source. WordNet is a large electronic lexical database of the English language that is freely available and was developed based on psycholinguistic theories that model human semantic organization. It has been extended to over 30 different languages, including Dutch, Spanish, German, Basque and Arabic (Abouenour et al., 2013; Fernandez-Montraveta et al., 2008; Gonzalo et al., 1999; Hinrichs et al., 2013; Pociello et al., 2011). The computation of semantic trademark similarity uses two sets of features to represent each trademark: the token feature set and the synonyms feature set. The token feature set consists of a set of words included in the trademark. For example, the token feature set for the trademark ‘Red Bull’ is (red, bull). The synonym feature set on the other hand comprises synonyms, direct hypernyms, i.e. words that are more general in meaning in the taxonomic hierarchy, and direct hyponyms, i.e. words that are instances of their corresponding trademark tokens. The similarity score is computed using a combination of Tversky’s contrast model of similarity (Tversky, 1977), which considers the number of shared features, together with the edge-based word similarity score between the tokens, derived using lexical ontology, i.e. WordNet.

Finally, the *phonetic* similarity assessment computes trademark similarity based on the phonological features of the phonemes in the trademark text combined with typographic mapping and a token rearrangement process (Anuar et al., 2014). The algorithm represents the phonemes in a word string as vectors with phonetic features where each vector consists of 10 binary main features and two multi-valued features extracted from the phonological properties of human speech production (Kondrak, 2003). The algorithm differentiates between more similar phoneme pairs, such as *‘m’* and *‘n’*, and less similar phoneme pairs, such as *‘m’* and *‘p’*. In addition, the algorithm converts special characters or symbols in the trademark text to their corresponding meaning. For example, the ampersand symbol ‘&’ is substituted by ‘and’. This conversion allows typographic symbols to be processed in the way regular words appearing in trademarks are handled.

3.2 Step 2: Fuzzification

The fuzzification step is the process of mapping the crisp values of the input variables to fuzzy sets. Three input variables corresponding to the visual, semantic and phonetic assessments are fuzzified in this step using five triangular-based membership functions, as defined in Eq. 2. These functions were employed in this study because of their simplicity and good performance, which have been proven theoretically (Barua et al., 2014) and used in various engineering and non-engineering applications (Gañán et al., 2012; Kaur & Kaur, 2012; Ngai & Wat, 2005). Moreover, these functions have recently been used in a court case decision-making study that included traffic violations and crime cases (Sabahi & Akbarzadeh-T, 2014). A graphical representation of the input membership functions is shown in Fig. 5.

$$f_1(x) = \begin{cases} \frac{0.25 - x}{0.25}, & 0 \le x \le 0.25 \\ 0, & x \ge 0.25 \end{cases} \qquad f_2(x) = \begin{cases} 0, & x \ge 0 \\ \frac{x}{0.25}, & 0 \le x \le 0.25 \\ \frac{0.5 - x}{0.25}, & 0.25 \le x \le 0.5 \\ 0, & x \ge 0.5 \end{cases}$$

$$f_3(x) = \begin{cases} 0, & x \le 0.25 \\ \frac{x - 0.25}{0.25}, & 0.25 \le x \le 0.5 \\ \frac{0.75 - x}{0.25}, & 0.5 \le x \le 0.75 \\ 0, & x \ge 0.75 \end{cases} \qquad f_4(x) = \begin{cases} 0, & x \le 0.5 \\ \frac{x - 0.5}{0.25}, & 0.5 \le x \le 0.75 \\ \frac{1 - x}{0.25}, & 0.75 \le x \le 1 \\ 0, & x \ge 1 \end{cases} \qquad (2)$$

$$f_5(x) = \begin{cases} 0, & x \le 0.75 \\ \frac{x - 0.75}{0.25}, & 0.75 \le x \le 1 \end{cases}$$

<insert Fig. 5 here>

3.3 Step 3: Inference

This step uses the Mamdani fuzzy inference model, a well-known inference model used in various fuzzy logic-based applications (Abou & Saleh, 2011; Akgun et al., 2012; Chatzichristofis et al., 2012). A set of fuzzy rules was first developed based on the OHIM trademark examination manual (OHIM, 2014) and an empirical study of 1,400 trademarks involved in dispute cases. The rules are expressed in tabular form using five two-dimensional fuzzy associative matrices, which correspond to a total of 125 rules. Fig. 6 shows the five associative matrices of the developed rules. Five input and output conditions are associated with each rule: very low (VL), low (L), medium (M), high (H), and very high (VH). Each cell in the associative matrices corresponds to the output condition triggered by the rules associated with the condition of the input variables. For example, the verbal rule corresponding to the first cell of matrix (c) in Fig. 6 is translated as *'IF the phonetic score IS M (medium) and the semantic score IS VL (very low) and the visual score IS VL (very low), THEN the output score IS L (low)'*. The output membership functions that correspond to the five output conditions also consist of five triangular-based functions, as in Eq. 3. A graphical representation of these functions is shown in Fig. 7.

<insert Fig. 6 here>

<insert Fig. 7 here>

$$f_1(x) = \begin{matrix} \frac{0.2 \quad x}{0.3}, & 0 \quad x \quad 0.2 \\ 0, & x \quad 0.2 \end{matrix} \qquad f_2(x) = \begin{matrix} 0, & x \quad 0 \\ \frac{x+0.1}{0.3}, & 0 \quad x \quad 0.2 \\ \frac{0.5 \quad x}{0.3}, & 0.2 \quad x \quad 0.5 \\ 0, & x \quad 0.5 \end{matrix}$$

$$f_3(x) = \begin{matrix} 0, & x \quad 0.25 \\ \frac{x \quad 0.2}{0.3}, & 0.2 \quad x \quad 0.5 \\ \frac{0.8 \quad x}{0.25}, & 0.5 \quad x \quad 0.8 \\ 0, & x \quad 0.75 \end{matrix} \qquad f_4(x) = \begin{matrix} 0, & x \quad 0.5 \\ \frac{x \quad 0.5}{0.3}, & 0.5 \quad x \quad 0.8 \\ \frac{1.1 \quad x}{0.3}, & 0.8 \quad x \quad 1 \end{matrix} \qquad (3)$$

$$f_5(x) = \begin{matrix} 0, & x \quad 0.8 \\ \frac{x \quad 0.8}{0.3}, & 0.8 \quad x \quad 1 \end{matrix}$$

The aggregation of the compositional output involves a fuzzy operation between the fuzzified input and the fuzzy relations established by the rules. It is derived using the implication–aggregation (min–max) method (Akgun et al., 2012):

$${}_0 = \max(\min({}_{i_1}(k), {}_{i_2}(k), {}_{i_3}(k))) \qquad (4)$$

where ${}_{i_1}, {}_{1_2}, {}_{i_3}$ are the mapping of the first, second and third inputs from the crisp set to the fuzzy set, i.e. the visual, semantic and phonetic similarity scores, respectively, and *k* is the k-th IF–THEN preposition, or the fuzzy rule.

3.4 Step 4: Defuzzification

This step uses the centroid or centre of mass defuzzification method to quantify the compositional output from the fuzzy set to the real output that corresponds to the degree-of-similarity value. It computes the centroid under the curve resulting from the compositional operation performed during the inference step. The centroid computation is given by the following equation:

$$\text{centroid} = \frac{f(x) \times x\,dx}{f(x)\,dx} \quad (5)$$

where *f(x)* is the membership function associated with the compositional output. Fig. 8 shows an illustrative example of the proposed aggregation process for the trademark pair SKYPINE and SKYLINE. Their degree of similarity was computed as 0.798.

<insert Fig. 8 here>

4. Experimental Setup and Results

This section describes the two experiments performed in this study and the evaluation method used to conduct them. The first experiment evaluated the proposed method from a computational point of view using information retrieval measures. The second experiment was designed to capture human perception, i.e. the way people view similarity in trademarks.

4.1 Experiment 1

The main objective of the first experiment was to test the classification performance of the proposed method when differentiating between possible cases of infringement. The developed method was compared to the traditional approach of considering the individual aspects of similarity. The experiment employed information retrieval measures such as F-score, precision score and accuracy. The scores were derived from the classification confusion matrix shown in Table 4, where *TP, FP, FN* and *TN* refer to true positive, false positive, false negative and true negative, respectively.

<insert Table 4 here>

A collection of real court cases comprising 1,400 trademarks (Schweizer, 2013) was analysed and used to create a database. An excerpt from a court case report for two disputed trademarks, AURA and AUREA, is shown below. It provides the conclusion and rationale of the experts investigating this particular case. Based on such findings, the database was then split into two groups, i.e. with degree of similarity that may or may not lead to confusion as judged by the experts.

> *On the visual level, the trademarks have a strong similarity in the sense that the length of the verbal elements is almost identical (AURA/AUREA), i.e. four against five letters. Only the vowel 'E' of the contested trademark differs from the four letters of 'AURA' trademark.* ***The overall visual impression is therefore very similar.***
>
> *Aurally, the signs are also very similar. The vowel 'E' can be easily used.* ***The overall phonetic impression is also very similar.***
>
> ***Although that there is no semantic similarity, the risk of misperception on trademarks does exist due to high visual and phonetic similarity.***
>
> *The fact that the opponent has an additional letter 'E' does not change the overall similarity finding. In view of that, the similarity of the trademarks is therefore recognized.*

For evaluation purposes, a repeated holdout evaluation procedure was performed in which the database was divided into two random disjoint training (50%) and testing (50%) sets. The training set was used to obtain a threshold score to classify the dataset employed in this experiment. Pairwise degrees of similarity scores between the trademark pairs in the training set were first computed using the proposed method. A histogram-based thresholding algorithm (Nobuyuki, 1979) was then used to estimate the threshold value of the computed degree-of-similarity scores by exhaustive searches for a value that minimized the intra-class variance of the binary classes. The threshold value obtained from the training set was then used to classify the data in the testing set. This procedure was repeated 1,000 times and in each repetition the F-score, precision and accuracy were computed using Eqs. 6–8:

$$F\ \ score = \frac{2TP}{TP + FP + TP + FN} \quad (6)$$

$$precision = \frac{TP}{TP + FP} \quad (7)$$

$$\text{accuracy} = \frac{TP + TN}{Total \quad Data} \tag{8}$$

where *TP, TN, FP* and *FN* are the true positive, true negative, false positive and false negative trademarks, respectively, as classified by the binary classification performed in this experiment, and *Total Data* (calculated as 700) is the total number of trademark pairs in the database. The average scores were then used to evaluate the overall performance of the proposed method. The procedure was repeated using the scores from the individual assessments of visual, semantic and phonetic similarity.

Table 5 shows the classification results obtained using the three individual similarity assessments and the proposed method.

<insert Table 5 here>

4.2 Experiment 2

The main objective of the second experiment was to prove the following two hypotheses:

1. The similarity ranking of the trademark pairs produced by the proposed method correlates with human collective judgment.
2. The similarity rating of each trademark pair produced by the proposed method correlates with human collective judgment.

Two significance tests were performed using the Spearman rank correlation score and the Pearson pairwise correlation score to statistically prove these hypotheses and reject the null hypotheses of this experiment. The Spearman rank correlation score, which takes values in the range of -1 to 1 (both -1 and 1 being the negative and positive perfect correlations, respectively, and 0 indicating no correlation), is a measure of statistical dependence between two ranked variables. The score indicates how strong the relationship between the ranked variable can be and is described using a monotonic function. The Pearson pairwise correlation score on the other hand measures the strength of a linear association between two variables. The Pearson correlation attempts to draw a line of best fit through the values of two variables; the score itself describes the dispersion of the data points from the line of best fit. The Pearson correlation score has the same value range as the Spearman rank correlation score.

As it involved human judgment, this experiment used a crowdsourcing platform for evaluation purposes. Crowdsourcing is an open call task recently introduced in information retrieval studies and has been proven to produce fast and reliable results in a cost-effective way (Corney, 2010; Fadzli & Setchi, 2012; Snow et al., 2008). This task, commonly known as a human intelligence task (HIT), is a small portion of an even larger task distributed among a large group of workers without any apparent contact.

A total of 25 trademarks were randomly selected from the database used in Experiment 1 as a query set in Experiment 2. The trademark similarity assessment system developed in this study was then used to rank the set of trademarks returned from each query from the highest degree-of-similarity (ds) score to the lowest. Three trademarks with high ($ds > 3.5$), medium ($2.0 < ds \leq 3.5$) and low ($ds \leq 2.0$) distribution scores were selected from the retrieved set and used in the crowdsourcing task. Table 6 shows the 25 queries used in this experiment together with the three retrieved results classified by the proposed method as having high, medium and low similarity, respectively. Fig. 9 shows one of the HITs used in the experiment.

<insert Table 6 here>

<insert Fig. 9 here>

In each HIT, the workers were presented with three different trademarks and asked to score their similarity with the query trademark using a scale from 1 to 5 (1 being the least similar and 5 being the most similar). Each query was evaluated by 20 different workers, which resulted in a total of 500 HITs. The selection of the HIT workers was based on two criteria: the number and acceptance rate of their previously completed assignments. The first criterion required the workers to have completed at least 1,000 HITs. The acceptance rate of the previously completed HITs was set to 95%, indicating the approval level of the work done as evidenced by their HITs requestors. These two criteria were introduced to ensure the quality of the collected feedback. Next, the average similarity scores for each query given by the workers were computed and compared with the normalized similarity score produced by the proposed method (Table 7). The similarity scores (Fig. 10) were used to compute the Spearman rank correlation score and the Pearson pairwise correlation score shown in Table 8.

<insert Table 7 here>

<insert Fig. 10 here>

<insert Table 8 here>

5. Discussion

The first experiment verified the classification performance of the proposed multi-faceted method, which aggregated a similarity score based on all three similarity aspects (see Table 5). The method produced an F-score of 0.911, which translated into respective improvements of 15.2%, 150% and 12.5% compared to the F-scores produced using visual, semantic and phonetic similarity individually. Among these three similarity aspects, phonetic similarity produced the best F-score (0.810) while semantic similarity showed the worst performance in terms of F-score (0.364). The proposed method also surpassed the three individual similarity aspects in terms of precision. With a precision score of 0.924, it improved the individual performance of the visual, semantic and phonetic similarity assessments by 35%, 312% and 35.4%, respectively. Similar improvements were demonstrated in terms of accuracy. The proposed method produced an accuracy score of 0.910 compared to the accuracy produced using visual, semantic and phonetic similarity (0.819, 0.610 and 0.840, respectively), which resulted in improvements of 11%, 49% and 8.3%, respectively. Overall, the results from the first experiment clearly show that the proposed degree-of-similarity aggregation method has the best classification performance compared to assessments based on individual similarity aspects. Moreover, this approach is well aligned to the recommended trademark examination procedure, which requires trademarks to be examined in a holistic way.

The second experiment was designed to investigate the performance of the proposed method in comparison with human collective judgment. Two correlation measures, the Spearman rank correlation and the Pearson pairwise correlation, were used to statistically prove the hypotheses. The proposed method obtained a perfect Spearman rank score of 1 and a Pearson pairwise correlation score of 0.92. A statistical significance test performed on both correlation scores rejected the null hypotheses of the experiment and indirectly proved that the degree-of-similarity scores produced by the proposed method correlated well with human collective judgment on trademark overall similarity. This strong correlation can be also observed in the scatter plot shown in Fig. 10, which displays a concentration of almost all points along the best-fit line (the straight black line on the graph).

5. Conclusions

A support system to assess the overall degree of similarity between trademarks is essential for trademark protection so the work presented in this paper was motivated by the need to help prevent trademark infringement by identifying existing similarities between trademarks.

This paper contributes to the body of knowledge in this area by the development of a method that measures the degree of similarity between trademarks on the basis of all three aspects of similarity: visual, semantic and phonetic. The method uses fuzzy logic to aggregate the overall assessment, which provides a more balanced and human-centered view on potential infringement problems. In addition, the paper introduces the concept of degree of similarity since the line between similar and dissimilar trademarks is not always easy to define especially when dealing with blending three very different assessments.

One of the strengths of the proposed method is its rigorous evaluation using a large, purpose-built collection of real legal cases of trademark disputes. Moreover, the experiments performed in this study examined the performance of the proposed method from two points of view. First, the relative performance of the method was investigated from an information retrieval perspective in terms of classification performance. Using a crowdsourcing platform, the second experiment investigated the performance of the method relative to human judgment. The results of the experiments confirmed that there is a significant improvement in trademark similarity assessment when all similarity aspects are carefully considered. The results also showed that the proposed method demonstrates a statistically significant correlation against human collective judgment. Therefore, the experiments convincingly validated both original hypotheses outlined in this study.

In conclusion, the proposed system can provide a support mechanism in the trademark similarity analysis performed by trademark examiners during trademark registration. Moreover, the method for assessing the trademark similarity could be extended to address recent cyberspace phenomena such as consumer hijacking and cybersquatting. A particular limitation of the proposed work is its focus on only one aspect of the concept of *likelihood of confusion*, i.e. computing the similarity between trademarks. In reality, there are several other factors influencing the perceptions of the consumers. Such factors include strength of the registered trademarks, proximity of the channels of trade, product relatedness and consumer traits (sophistication and care). Such a study, which is currently underway, requires a multi-disciplinary approach, which involves experts from business studies, marketing, psychology and engineering.

Acknowledgements—The authors wish to acknowledge the help of Christopher Harrison, Peter Evans, William Morell and Rich Corken from the UK Intellectual Property Office in finalizing some of the ideas behind this research.

References

Abou, A., and Saleh, E. (2011). A Fuzzy decision support system for management of breast cancer. *International Journal of Advanced Computer Science and Applications*, 2(3), 34-40.

Abouenour, L., Bouzoubaa, K., & Rosso, P. (2013). On the evaluation and improvement of Arabic WordNet coverage and usability. *Language Resources and Evaluation*, 47(3), 891-917.

Agliardi, E. and Agliardi, R. (2011). An Application of Fuzzy Methods to Evaluate a Patent under the Chance of Litigation. *Expert Systems with Applications,* 10, 13143-13148.

Akgun, A., Sezer, E. A., Nefeslioglu, H. A., Gokceoglu, C., & Pradhan, B. (2012). An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. *Computers and Geosciences*, 38(1), 23-34.

Anuar, F. M., Setchi, R., and Lai, Y. K. (2013). Trademark image retrieval using an integrated shape descriptor. *Expert Systems with Applications,* 40, 105-121.

Anuar, F. M., Setchi, R., and Lai, Y. K. (2014). Trademark retrieval based on phonetic similarity. *IEEE International Conference of Systems, Man and Cybernetics*, San Diego, USA, 1642-1647.

Anuar, F. M., Setchi, R., and Lai, Y. K. (2016). Semantic retrieval of trademarks based on conceptual similarity. *IEEE Transaction of Systems, Man and Cybernet*ics: System, vol. 46(2), pp. 220-233.

Barua, A., Mudunuri, L.S., and Kosheleva, O. (2014). Why trapezoidal and triangular membership functions work so well: Towards a theoretical explanation. *Journal of Uncertain Systems*, 8(3), 164-168.

Chatzichristofis, S. A., Zagoris, K., Boutalis, Y., and Arampatzis, A. (2012). A fuzzy rank-based late fusion method for image retrieval. In: Schoeffmann, K., Merialdo, B., Hauptmann, A., Ngo, C-W., Andreopoulos, Y., Breiteneder, C. (Eds). *Lecture Notes in Computer Science*, 7131, 463-472. Berlin Heidelberg: Springer.

Chen, T.-Y. (2011). Value Ontology-Based Multi-Aspect Intellectual Asset Valuation Method for Decision-Making Support in k-Commerce, *Expert Systems with Applications*, 38(5), 5471-5485.

Cook, B. B. (2001). Fuzzy logic and judicial decision making. *Judicature*, 85(2), 70-100.

Corney, J. R., Torres-, C., Jagadeesan, A. P., Yan, X. T., Regli, W. C., and Medellin, H. (2010). Putting the crowd to work in a knowledge-based factory. *Advanced Engineering Informatics*, 24(3), 243-250.

Covington, M. A. (1998). Alignment of multiple languages for historical comparison. *International Conference on Computational Linguistics*, Montreal, Canada (pp. 275-280).

Dodell, L. (2013) The trademark problem: Casualty insurance's dirty little secret. Retrieved from: http://www.carriermanagement.com (accessed 14 Dec 2015).

Eakins, J. P., Riley, K. J., and Edwards, J. D. (2003). Shape feature matching for trademark image retrieval. *International Conference of Image and Video Retrieval*, Urbana Champaign, USA (pp. 28–38).

Eakins, J. P., Shields, K., and Boardman, J. (1996). ARTISAN – A shape retrieval system based on boundary family indexing. *Storage and Retrieval for Still Image and Video Databases*, iv, 2670, 17–28.

Fadzli, S. A., and Setchi, R. (2012). Concept-based indexing of annotated images using semantic DNA. *Engineering Applications of Artificial Intelligence*, 25(8), 1644-1655.

Fazzolari, M., Alcala, R., Nojima, Y., Ishibuchi, H., and Herrera, F. (2013). A review of the application of multi objective evolutionary fuzzy systems: Current status and further directions. *IEEE Transactions on Fuzzy Systems,* 21(1), 45-65.

Fernandez-Montraveta, A., Vazquez, G., and Fellbaum, C. (2008). The Spanish version of WordNet 3.0. In: *Text Resources and Lexical Knowledge*, 8, 175-182. Berlin and New York: Mouton de Gruyter.

Gañán, C., Muñoz, J. L., Esparza, O., Mata, J., and Alins, J. (2012). Risk-based decision making for public key infrastructures using fuzzy logic. *International Journal of Innovative Computing, Information and Control*, 8(11), 7925-7942.

Gonzalo, J., Verdejo, F. and Chugur, I. (1999). Using EuroWordNet in a concept-based approach to cross-language text retrieval. *Applied Artificial Intelligence*. 13(7), 647-678.

Gupta, M. and Mohanty, B. K. (2016). An Algorithmic Approach to Group Decision Making Problems under Fuzzy and Dynamic Environment. *Expert Systems with Applications,* 55, 118-132.

Hinrichs, E., Henrich, V., and Barkey, R. (2013). Using part-whole relations for automatic deduction of compound-internal relations in GermaNet. *Language Resources and Evaluation*, 47(3), 839-858.

Hong, Z., and Jiang, Q. (2008). Hybrid content-based trademark retrieval using region and contour features. *Advanced Information Networking and Applications Workshop*, Okinawa, Japan (pp. 1163–1168).

Jain, A. K., and Vailaya, A. (1998). Shape-based retrieval: A case study with trademark image databases. *Pattern Recognition*, 31(9), 1369–1390.

Jang, J. S. R., Sun, C. T., and Mizutani, E. (1997). Fuzzy inference systems. *Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence* (pp. 73-91). Upper Saddle River, NJ: Prentice Hall.

Kato, T., Fujimura, K., and Shimogaki, H. (1990). TRADEMARK. Multimedia image database system with intelligent human interface. *Systems and Computers in Japan*, 21(11), 33–46.

Kaur, A., and Kaur, A. (2012). Comparison of Mamdani-type and Sugeno-type fuzzy inference systems for air conditioning system. *International Journal of Soft Computing and Engineering*, 2, 323-325.

Kim, Y. S., and Kim, W. Y. (1998). Content-based trademark retrieval system using a visually salient feature. *Image and Vision Computing*, 16(12–13), 931–939.

Kondrak, G. (2003). Phonetic alignment and similarity. *Computers and the Humanities*, 37(3), 273-291.

Kosko, B. (1994). *Fuzzy Thinking: The New Science of Fuzzy Logic*. New York: Hyperion.

Leng, G., Zeng, X. J., and Keane, J. A. (2009). A hybrid learning algorithm with a similarity-based pruning strategy for self-adaptive neuro-fuzzy systems. *Applied Soft Computing*, 9(4), 1354-1366.

Li, C.-C., Liu, L., and Li, C.-B. (2017). Personalized Individual Semantics In Computing With Words for Supporting Linguistic Group Decision Making. An Application on Consensus Reaching. *Information Fusion*, 33, 29-40.

Li, M., Liu, L., and Li, C.-B. (2011). An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems. *Expert Systems with Applications,* 38, 8586-8596.

Mardani, A., Jusoh, A. and Zavadskas, E. K. (2015). Fuzzy Multiple Criteria Decision-Making Techniques and Applications - Two Decades Review from 1994 to 2014. *Expert Systems with Applications, 42(8), 4126-4148.*

McFee, B., and Lanckriet, G. R. (2009). Heterogeneous Embedding for Subjective Artist Similarity. *International Society for Music Information Retrieval Conference,* Kobe, Japan (pp. 513-518).

Navarro, G. (2001). A guided tour to approximate string matching. *ACM Computing Survey*, 33(1), 31-88.

Ngai, E. W. T., and Wat, F. K. T. (2005). Fuzzy decision support system for risk analysis in e-commerce development. *Decision Support Systems*, 40(2), 235-255.

Nobuyuki, O. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics*, 9(1), 62-66.

OHIM (2012). Annual report 2012. Retrieved from: https://oami.europa.eu (accessed 10 Dec 2013).

OHIM (2013). Trademark Search. Available at: https://oami.europa.eu. (accessed 10 Dec 2013).

OHIM (2014). Guidelines for examination in the office for harmonization in the internal market on community trade marks, part c opposition, section 2 identity and likelihood of confusion, chapter 3 comparison of signs. Retrieved from: https://oami.europa.eu (accessed 1 Feb 2014).

Peng, H. L., and Chen, S. Y. (1997). Trademark shape recognition using closed contours. *Pattern Recognition Letters*, 18(8), 791–803.

Pfeifer, U., Poersch, T., and Fuhr, N. (1996). Retrieval effectiveness of proper name search methods. *Information Processing and Management*, 32(6), 667-679.

Philips, L. (2000). The double metaphone search algorithm. *C/C++ Users Journal*, 18(6), 38-43.

Pociello, E., Agirre, E., and Aldezabal, I. (2011). Methodology and construction of the Basque WordNet. *Language Resources and Evaluation*, 45(2), 121-142.

Sabahi, F., and Akbarzadeh-T, M.R. (2014). Introducing validity in fuzzy probability for judicial decision-making. *International Journal of Approximate Reasoning*, 55(6), 1383-1403.

Schietse, J., Eakins, J. P., and Veltkamp, R. C. (2007). Practice and challenges in trademark image retrieval. *International Conference on Image and Video Retrieval*, Amsterdam, The Netherlands (pp. 518–524).

Schweizer, M. (2013). Trade Marks Court Cases Database. Retrieved from: http://decisions.ch (accessed 10 Dec 2013).

Scott, C. D. (2013). Trademark strategy in the internet age: Customer hijacking and the doctrine of initial interest confusion. *Journal of Retailing*, 89(2), 176-189.

Storey, V. C., Dey, D., Ullrich, H., and Sundaresan, S. (1998). An ontology-based expert system for database design. *Data and Knowledge Engineering*, 28(1), (31-46).

Trade Marks Act 1994. (1994). Retrieved from: http://www.legislation.gov.uk/ (accessed 24 Nov 2014).

Tversky, A. (1977). Features of similarity. *Psychology Review*, 84, 327–352.

Snow, R., O'Connor, B., Jurafsky, D., and Ng, A. Y. (2008). Cheap and fast-but is it good? Evaluating non-expert annotations for natural language tasks. *Conference on Empirical Methods in Natural Language Processing*, Honolulu, Hawaii (pp. 254-263).

USITC (2011). China: Effects of intellectual property infringement and indigenous innovation policies on the U.S. economy. Retrieved from: http://www.usitc.gov/ (accessed 20 Dec 2013).

UKIPO (2013). Trademark Search. Available at: http://www.ipo.gov.uk. (accessed 10 Dec 2013).

Wei, C. H., Li, Y., Chau, W. Y., and Li, C. T. (2009). Trademark image retrieval using synthetic features for describing global shape and interior structure. *Pattern Recognition*, 42(3), 386–394.

WIPO (2014). WIPO IPAS Functional and Technical Overview. Retrieved from: https://www3.wipo.int/confluence/display/wipoimd/WIPO+IPAS+Functional+and+Technical+Overview (accessed 24 Nov 2014).

Wu, J. K., Lam, C. P., Mehtre, B. M., Gao, Y. J., and Narasimhalu, A. D. (1996). Content-based retrieval for trademark registration. *Multimedia Tools and Applications*, 3(3), 245–267.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8, 338-353.

Zhang, B., Shen, J., Xiang, Q., and Wang, Y. (2009). CompositeMap: a novel framework for music similarity measure. *32nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, MA, USA (pp. 403-410).

Table 1 Different type of trademark similarity

Trademark 1	Trademark 2	Similarity Aspect
NEXT	NEST	Visual
MAGIC TIMES	MAGIC HOUR	Conceptual
SVIZZEROTALER	SWISS TALER	Phonetic

Table 2 Pseudocode of the visual similarity computation employed in the proposed algorithm

*Pseudocode: /*comment*/*

```
/* This part of the code is performed for the visual similarity
score computation for trademarks with text*/
define Qt and Dt as the query and trademark from the database
compute Aq and Ad as new strings that produce optimal alignment between Qt and
Dt
define score as the letter-to-letter visual similarity matrix between Qt and Dt;
define m=maximum(length(Aq), length(Ad));
for i=0 until m
        if Aq(i)=Null || Ad(i)==Null
             score(i)=0;
        else
     score(i)=compute visual similarity score between Aq(i) and Ad(i)
 end
define total_score= sum(score)/m);
```

Table 3 Degree-of-similarity computed using approximate string matching and visual similarity

	Approximate String Matching	Visual Similarity
1NDEX :: INDEX	0.80	0.923
1NDEX :: XNDEX	0.80	0.861

Table 4 Confusion matrix employed for the computation of the F-score, precision score, and accuracy score.

Actual Class	Predicted Class	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Table 5 F-score, precision, and accuracy computed using visual, conceptual, and phonetic similarity and the proposed method.

	Visual Similarity	Conceptual Similarity	Phonetic Similarity	Proposed Method
F-score	0.791	0.364	0.810	0.911
Precision	0.683	0.224	0.682	0.924
Accuracy	0.819	0.610	0.840	0.910

Table 6 List of 25 queries and their corresponding results used in this experiment

Queries	Result 1	Result 2	Result 3
webautor	WEBIATOR	WebFOCUS	autoscout24
FRUIT TIGER	LION FRUIT	SMOOTH FRUIT	RED BULL
GSTAR	XSTAR	STAR FLEX	sakira
SVIZZEROTALER	SWISS TALER	SEVIKAR	SCHNEIDER
NEXT	NEST	Nexans	NR
SKYPINE	SKYLINE	SKY ROOM	DECOLINE
Prevista	PREVISA	ad-vista	BONITA
SWEETLAND	SUGAR LAND	HEIDI LAND	SWISSOLAR
AMORA	AMORE	AXARA	ARTOR
RIMOSTIL	Rivotril	REBOVIR	REFODERM
CYRA	CYREL	ara	adria
GLOBRIX	Globix	ZYLORIC	GRILON
Lifestyle	Living Style	LIFE TEX	SNOW LIFE
WOOD STONE	MOONSTONE	WILTON	SwissTron
NUTELLA	NATURE ELLA	NATURESSA	MARQUELA
ecopower	ECOPOWER	GreenPower	HARRY POTTER
Twix	TRIX	TREAC	TREAKOL
SANTHERA	SANZEZA	SALFIRA	sunirse
MUROLINO	MURINO	MONARI	MATTERHORN
MAGIC TIMES	MAGIC HOUR	Maritimer	MATCH WORLD
RED BULL	BLUEBULL	FLYING BULL	A&A Bull and Bear
Feel'n LEARN	SEE'N LEARN	FEEL GOOD	FIGUREHEAD
bonvita	BONAVITA	biovital	Botoceutical
FMH	FNH	FTG	MR
ACTIVIA	ACTEVA	ADWISTA	ACCET

Table 7 Similarity scores obtained from the hit assignments and the proposed trademark degree-of-similarity aggregation method.

No	QUERIES	Human Interactive Task Rating			Proposed Method		
		Result 1	Result 2	Result 3	Result 1	Result 2	Result 3
1	webautor	3.40	2.35	1.00	4.98	2.99	1.90
2	FRUIT TIGER	3.45	2.05	1.20	3.94	2.17	1.75
3	GSTAR	4.05	2.45	1.00	4.23	2.82	1.86
4	SVIZZEROTALER	3.70	2.10	1.15	3.84	2.77	1.82
5	NEXT	4.00	2.80	1.10	4.29	2.86	1.79
6	SKYPINE	4.20	2.65	1.60	3.99	2.84	1.93
7	Prevista	4.70	3.20	1.35	4.17	2.68	1.96
8	SWEETLAND	3.70	2.10	1.20	3.94	2.85	2.00
9	AMORA	4.50	2.35	1.85	4.28	2.67	1.05
10	RIMOSTRIL	3.95	2.30	1.65	4.04	2.22	1.76
11	CYRA	3.75	2.25	1.45	3.94	2.68	1.83
12	GLOBRIX	4.75	1.60	1.40	4.14	2.14	1.84
13	Lifestyle	4.25	2.35	1.50	3.98	2.43	1.82
14	WOOD STONE	3.60	1.70	1.45	4.32	2.30	1.91
15	NUTELLA	3.65	2.20	1.40	3.74	2.96	2.00
16	ecopower	4.45	2.80	1.10	5.00	2.96	0.87
17	TWIX	4.00	1.70	1.20	3.98	2.48	1.94
18	SANTHERA	3.20	2.05	1.15	3.86	2.96	1.96
19	MUROLINO	4.50	3.35	1.65	3.97	2.59	1.85
20	MAGIC TIMES	3.70	2.15	1.50	3.78	2.82	1.88
21	RED BULL	3.90	3.00	1.75	3.85	3.33	1.98
22	Feel'n LEARN	4.00	2.55	1.30	3.95	3.28	1.85
23	bonvita	4.90	2.65	1.55	4.20	2.69	1.85
24	FMH	4.40	2.75	1.40	4.43	2.07	1.57
25	ACTIVIA	4.25	2.00	1.65	4.20	2.22	1.98
	Average	4.04	2.38	1.38	4.12	2.67	1.80

Table 8 Spearman rank correlation and Pearson pairwise correlation

Spearman Rank Correlation	Pearson Pairwise Correlation
1.00 (p<0.05)	0.92 (p<0.0001)

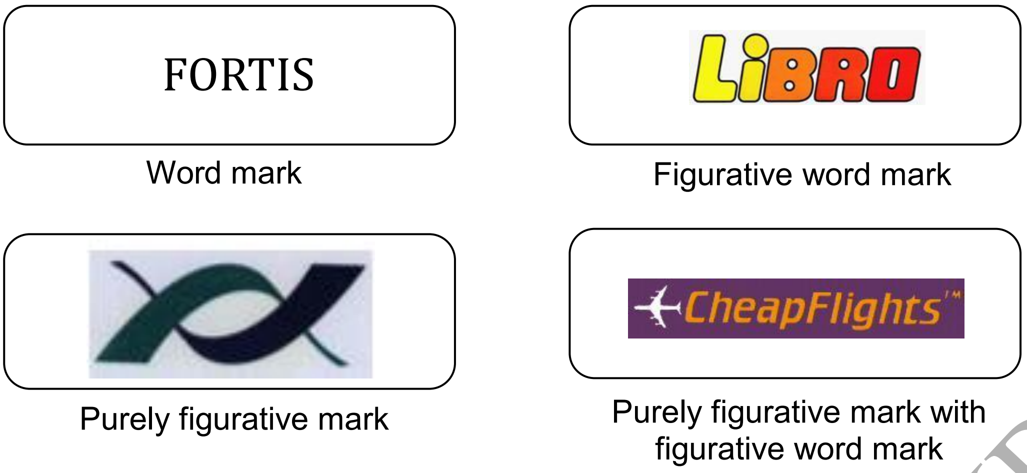


Fig. 1 Different types of trademarks (OHIM, 2014)

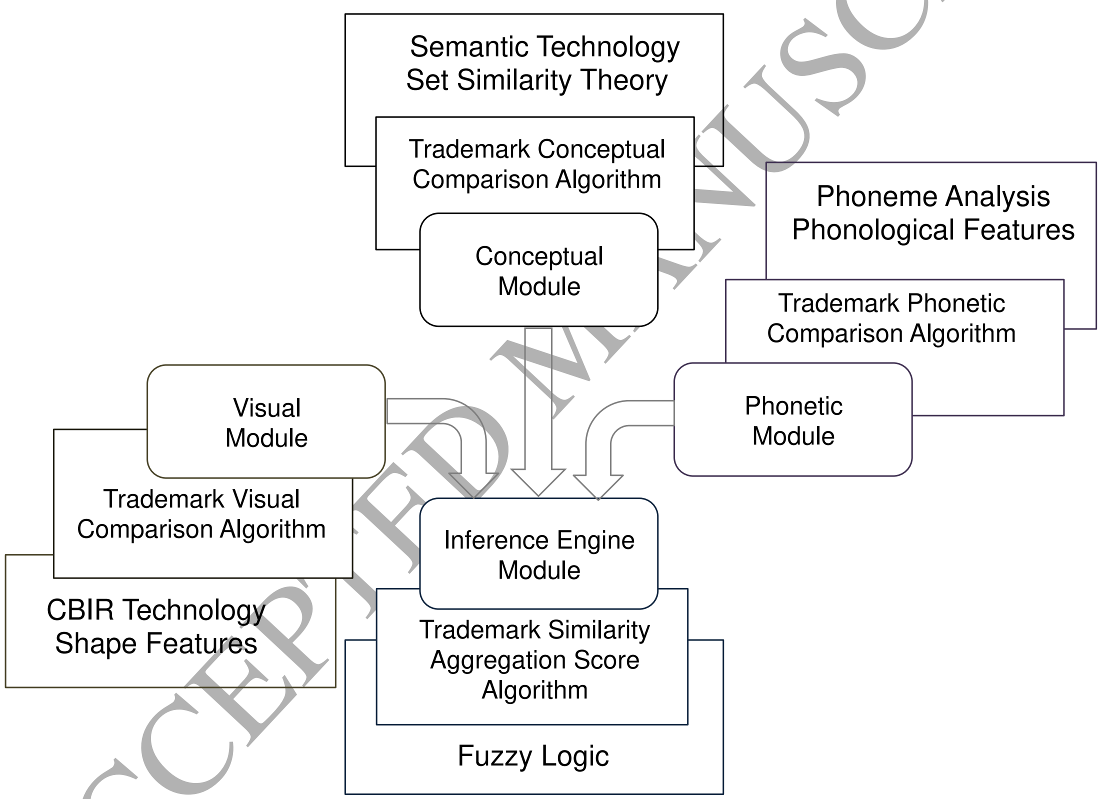


Fig. 2 Conceptual model of the proposed method for multi-faceted assessment of trademarks.

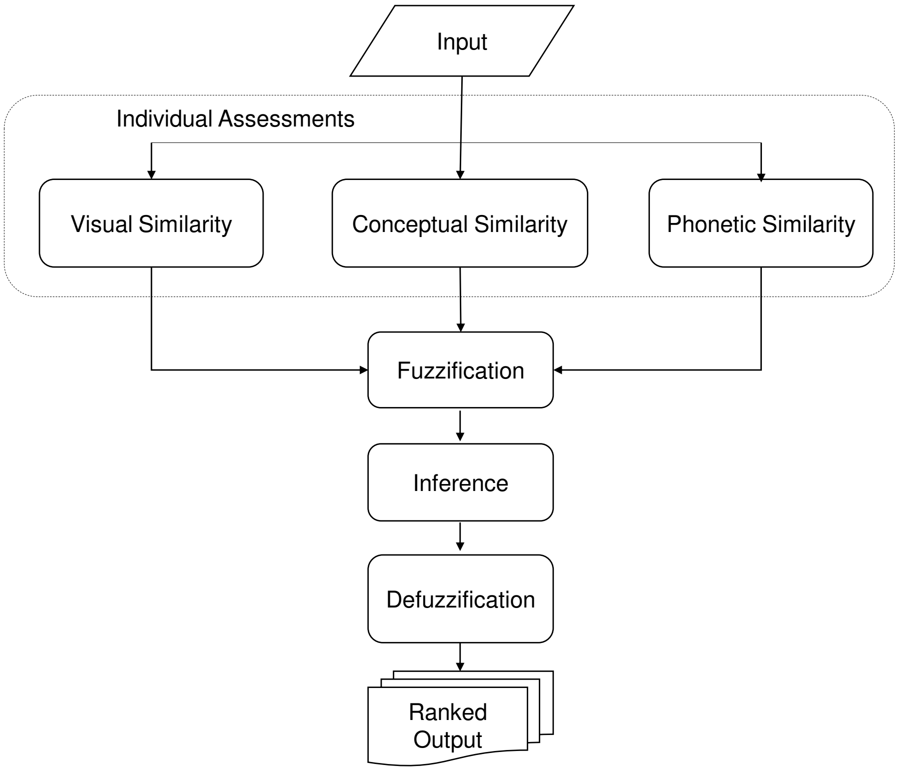


Fig. 3 Flow chart of the proposed method.

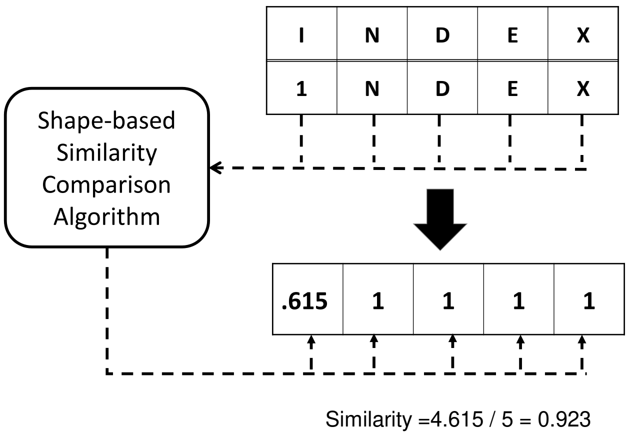


Fig. 4 Illustrative example of the visual similarity score computation employed in the proposed method.

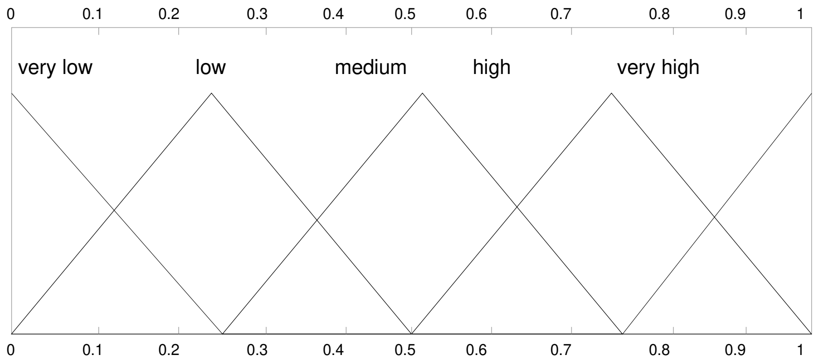


Fig. 5 Input membership functions used.

PHONETIC VL		CONCEPTUAL				
		VL	L	M	H	VH
VISUAL	VL	VL	VL	L	H	VH
	L	VL	L	L	H	VH
	M	L	L	M	H	VH
	H	H	H	H	H	VH
	VH	VH	VH	VH	VH	VH

(a)

PHONETIC L		CONCEPTUAL				
		VL	L	M	H	VH
VISUAL	VL	VL	VL	L	H	VH
	L	VL	L	L	H	VH
	M	L	M	M	H	VH
	H	H	H	H	H	VH
	VH	VH	VH	VH	VH	VH

(b)

PHONETIC M		CONCEPTUAL				
		VL	L	M	H	VH
VISUAL	VL	L	M	M	H	VH
	L	M	M	M	H	VH
	M	M	M	M	H	VH
	H	H	H	H	H	VH
	VH	VH	VH	VH	VH	VH

(c)

PHONETIC H		CONCEPTUAL				
		VL	L	M	H	VH
VISUAL	VL	H	H	H	H	VH
	L	H	H	H	H	VH
	M	H	H	H	H	VH
	H	H	H	H	H	VH
	VH	VH	VH	VH	VH	VH

(d)

PHONETIC VH		CONCEPTUAL				
		VL	L	M	H	VH
VISUAL	VL	H	H	VH	VH	VH
	L	H	H	VH	VH	VH
	M	VH	VH	VH	VH	VH
	H	VH	VH	VH	VH	VH
	VH	VH	VH	VH	VH	VH

(e)

Fig. 6 Associative matrices used for rule derivation in the inference process.

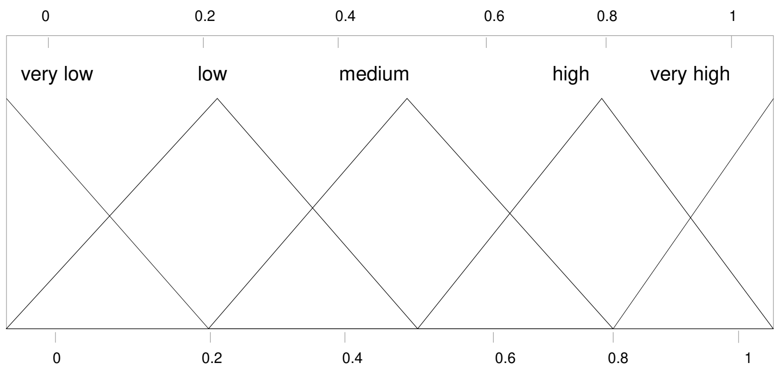


Fig. 7 Output membership functions utilized in the inference process.

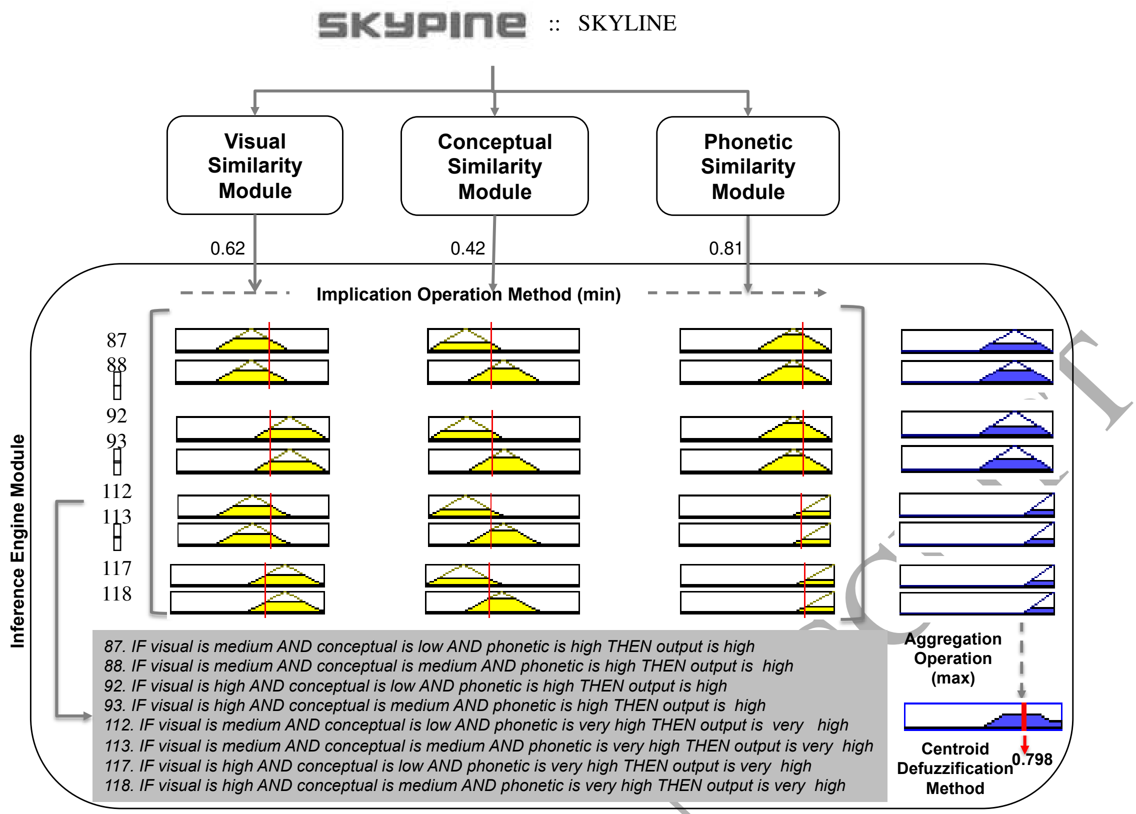


Fig. 8 Illustrative example of the proposed aggregation method for the trademark pair Skypine and SKYLINE.

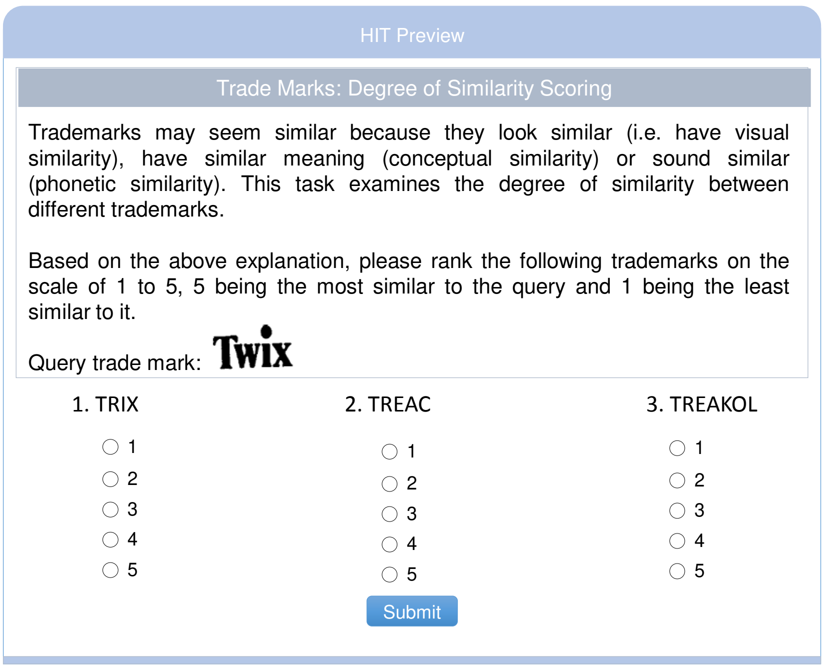


Fig. 9 An example task used in Experiment 2.

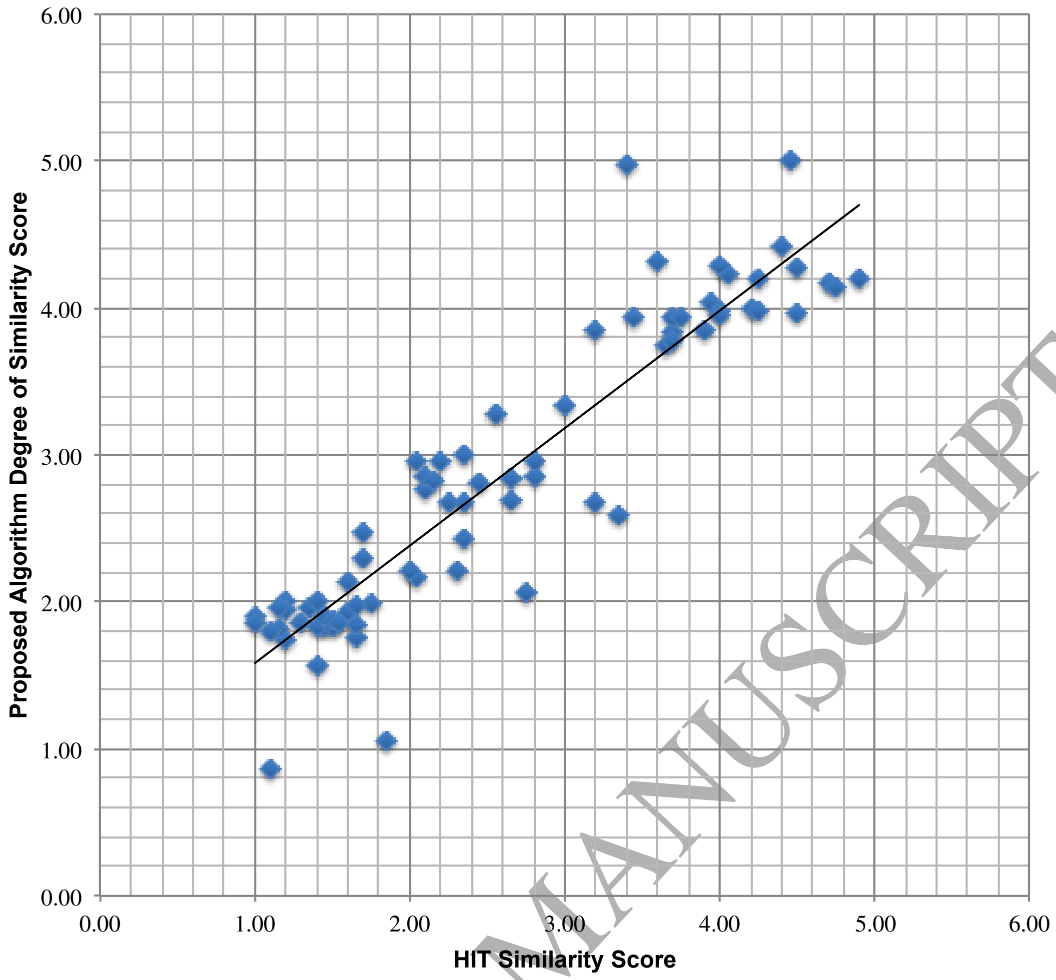


Fig. 10 Similarity scores obtained in Experiment 2.