OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

SEMANTIC RETRIEVAL OF TRADEMARKS BASED ON TEXT AND IMAGES CONCEPTUAL SIMILARITY

Vishakha Chilpipre[1] , Prof. Varsha R. Dange[2]
Student, Department of Computer Science Dhole Patil College of Engineering, Pune, Maharashtra, India[1]
Assistant Professor, Department of Computer Science Dhole Patil College of Engineering, Pune, Maharashtra, India[2]

Abstract: **A trademarks is a mark that you can use to recognize your business products or services from those of other vendors. It can be represented graphically in the form of any Symbol, logo, words etc. so, they need to be protection. The conceptual similarities among trademarks, which happens when more than two or more trademark similar. Trademarks are possessory words and images with high reputation they are main assets, often used as a application, which need infringement protection. The problems considered until infringement cases is the aspects, hypothetic and phonetic similarity of various trademarks. This paper focuses on important aspect by proposing a conceptual similarity of trademarks that can be provide distance computation and suggestions of input retrieving conceptually similar trademarks. The search and indexing technique developed uses similarity distance, which is derived using of similarity trademark. Propose a computational approach based on semantics that can be used to suggest the input of trademarks for conceptual similarity and to avoid the additional cost of protection to future infringement. A trademark retrieval system is performing with the massive number of semantic trademark of the conceptual similarity.**

Keywords: Conceptual similarity, trademark infringement, trademark retrieval, trademark similarity.

----------∴∴∴----------

I INTRODUCTION

The rapid development of simple ways has created new challenges in these regions for lots of companies who use the Internet to trade and employ trademarks as sell-out equipment. Trademarks, as prescribed by the European Office of Harmonization in the Internal Market (OHIM). They do insignificant intellectual property (IP) goods that permit well or service to be well validated to clients. Each year many trademarks registered and used that outlet. Trademarks are exclusive words or figures with advance reputational significance, used in commerce to comparison between products and services.

They allow products or tasks to be goods tenable and compared by traders. Searching for conceptually similar trademarks is a text retrieval problem. However, traditional text retrieval systems based on keywords are not capable of retrieving conceptually related text. This limitation motivates research into semantic technology, which addresses this problem by using additional knowledge sources. Few common disservice outcomes from trademarks infringement is lost income, low benefits, and need extra money of conservancy to stave off next infringement. The trademarks registered improve by 20 percent from last many years in the word. Trademark similarity problems for the other 70 percent stay deficiently researched in more that content-based retrieval goes from different limitations. When assessing trademark infringement cases then analysis several separate components, such as the same of the goods, the especial and main points of the different trademarks, and the similarity of the trademarks.

A trademark may be designated by the following symbols: is “trademark symbol”, which is the letters “TM”, for an unregistered trademark, a mark used to promote or brand goods is the letter “R” surrounded by a circle, for a registered trademark. Infringement may occur when one party, the “infringer”, uses a trademark which is identical or confusingly similar to a trademark owned by another party, in relation to products or services which are identical or similar to the products or services which the registration covers having existence trademark look for systems as a general rule use text-based acts to get back technology. These searches look for trademark that matches some or all words in a question line wording. As indicated in their latest printing on trademark knowledge-bases and look for systems. Two trademarks are necessary not same to make an infringement. The conceptual different of text files that part of same domain, utilization same notations, or demonstration same consideration has been used broadly.

II LITERATURE SURVEY

In this paper [3], The recent trademark reflow system of working with reformed reflow execution for the unification of global and local expositors. The global expositors are using the Zernike moment's coefficients and the local expositors are the edge-gradient co-occurrence matrix, defines as outline data that means it's mainly significance in human cognition of estimation equality. The defined reflow system is tested use the standard MPEG-7 shapes. The results reformation in the case of the MPEG-7 shape databases. The bonding during two proximate factors is hold on by usage the co-occurrence matrix on incline data. The research in the round of offered a novel system for trademark reflow that increase the execution.

Author proposed [9], A recent system for counting short-text and sentence semantic similarity. The method is depends on the concept that the sense of a statement is create of nope mere the sense of its particular words, but also the anatomical path the words are concatenated. Thus hold on and connects syntactic concatenated. Thus hold on and connects syntactic and semantic data to count the semantic similarity of two phrases. Semantic data is given from lexical resources. Syntactic data is get from a strong parsing procedure that searches the sentences in every phrase. A syntax-based providence to calculate the semantic similarity between phrases or short texts. The concept on which the system is based on the sense of phrases is creating of nope mere the senses of its particular words, but as well the different words are concatenated.

Author introduces [5] a method and a model for detracting and listing information from main language data. The main domain prototype depends on a hypothetic scale that is of a domain ontology, which define the domain information, and a lexical scale based on WordNet, that's defines the domain glossary. The semantic data retrieval engine that created justification easy keyword-based problems, as well as natural language-based problems. The engine is also ability to develop the domain information, searching recent and same facts added to domain model. The induration probe suggests that the method is efficient to many forms and define nations with accurate purity.

This paper presents [6], The data reflow technique utilizes keywords passed by the user as the find measurement to find documents. Nevertheless, the language used in files is mostly hard and unclears, and hereby the outcomes obtained by using keywords are mostly not good. The way of this issue, created a semantic-based content mapping mechanism for a data reflow technique. These views simplify the find process and improving the purity of the returned results. A semantic-based content mapping mechanism uses files different keywords as the input, which substances the semantic characteristics and fabrication of the documents.

This paper [7] The problems define during infringement litigation is the visible, hypothetic and phonetic similarity of different trademarks. This is focuses on this important fact by defining a hypothetic model of the comparison process, target at retrieving hypothetic similar trademarks. The proposed model normal language accessing and semantic technology to get the hypothetic similarity between trademarks. Proposes a hypothetic model of trademark retrieval based on hypothetic similarity. The proposed model improves on already trademark finding models by providing find to hypothetically related trademarks.

III PROBLEM DEFINATION

The trademark infringement assessing cases then analysis several separate components, like similarity of the goods, the especial and main essence of the different trademarks, and the similarity of the trademarks. The system used to proportion trademarks for hypothetic similarity and suggestions to user input for avoid trademark infringement. The search and indexing technique developed uses similarity distance. Hence, the concept of similarity has become well understood in trademark infringement litigation.

IV WORK OF TRADEMARK SYSTEM

The propose system is to make a retrieval of trademark hypothetical similarity to make them more accurate and more secure against the trademark infringement. Also the systems are competent of retrieving the conceptual similarity of trademarks and manage the conventional data retrieval system.

The proposed model can then be unified into a reflow system that considers the other two phases of similarity, sight and phonetic, and will then procedure a more extensive trademark comparison. The system used to proportion trademarks for conceptual similarity. Finding for conceptually same trademarks is a text retrieval problem. The system defines the nearly string matching which is used to text searches. The proposed algorithm use techniques for the word similarity gap of that method, which was already use from the WordNet ontology and it's together with a new trademark comparison measure. In data retrieval, data is per performed for exhibit classifications, as well as realizing potentially useful information from documents.

The analysis of the trademarks is needed to comprehend the main of conceptual similarities coming from different factors. The focuses on this main fact by proposing a hypothetical model of the comparison process, purposed at retrieving conceptually similar trademarks. The hash indexing accept the token key and synonym key to pre-processing and use the indexing in that key and create a new trademark for the user, its similar to that user requirement trademark. The feature extractions are defined the token and synonyms. The excerption compatible conceptual features, which are then used to proper manage the database. The spelling corrector corrects any spelling mistakes in the trademark text, and can be adapted from any existing spell checker. The spelling corrector is mainly work for conceptual similarity of trademark that is not generating the wrong trademark in the system.

The stop words remover removes frequent words e.g. no, and, the, etc. for improve the efficiency of the trademark. The pre-processing is made a require changes then that trademark text provide the token for unique

identifications and also provided synonyms word from lexical resource for create new trademark as user conceptual similarity. In point of extractions, the provider of these steps two features: 1) Token set and 2) Synonym set. The token set is defined as a particular word has a token set for used in the trademark retrieval system. After stop word removal, every single word has a token set.

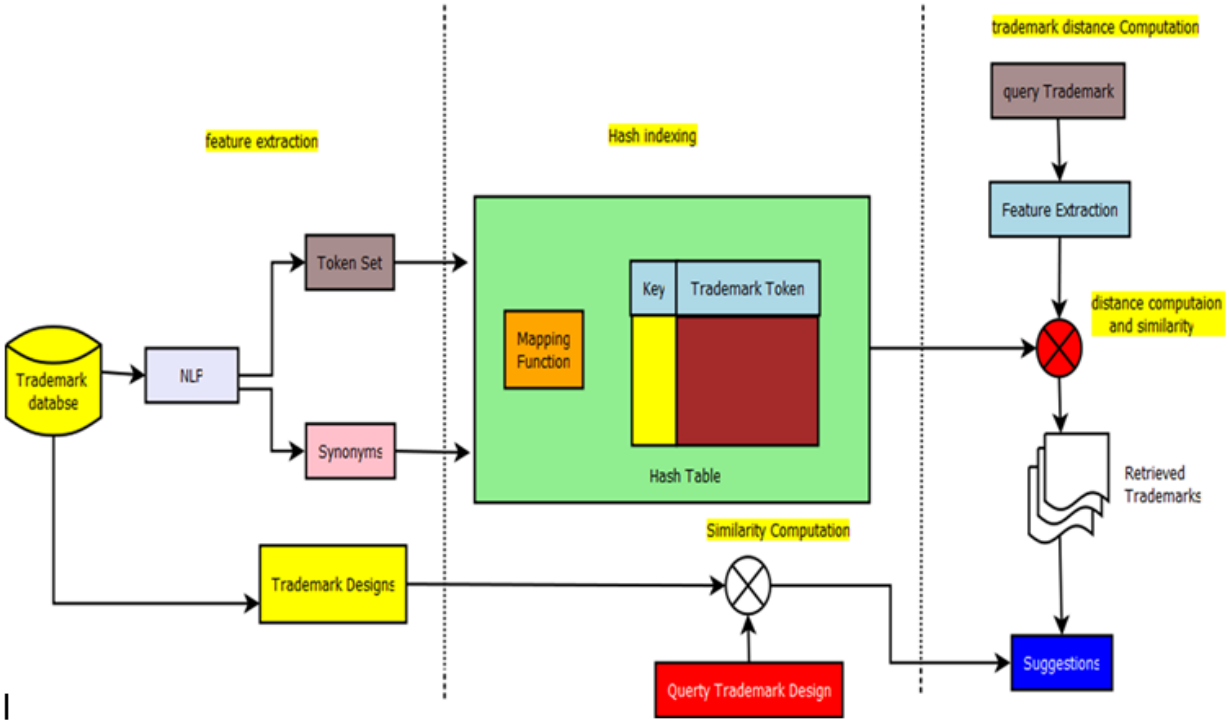

Figure 1: System Architecture

The synonym set is defined as direct token set is extracted of every word has synonyms used wordnet. This is use to provide a every word of token has a number of synonyms find out using wordnet, if not then keep as it token word. To remove hypothetic time throughout the find process, the factors are listed using a hashing technique. The hash indexing accepts the trademark as the key index. The feature extraction process stores an additional set of array features in the first feature vector, i.e. a set of synonyms and antonyms corresponding to the query tokens. The hash table is use the hash indexing for provide accurate timing for searching the process for conceptual similarity of trademark. In hash table use the token key as well as synonyms key for unique identification. The token key and synonym key as used indexing for provide new retrieval trademark. The retrieval trademark list is content of number of trademarks of the conceptual similarity. Using retrieval algorithm for the similarity between two words. The hash indexing develops the new trademark using token key and synonym key in the hash table.

A trademark reflow technique using the proposed retrieval algorithm is evolved, and the algorithm is tested on conceptual similarity. The retrieval trademark list is stored in database for next future trademark use in the next trademark retrieval concepts. To remove extra required time throughout the find procedure, the factors are listed using a hashing technique. The hash indexing is taken the trademark as the key index. Through trademark retrieval process user can enter a text which he wants to trademark. If trademark is already exist in system then it sent to trademark matching and return the similar documents to the user. If trademark is not existed in system then trademark is stored in database. The return document is send to user is the use lexical resource and apply the hash indexing to that trademark for create new trademark to get the user .

V MATHAMATICAL MODEL

A. Set Theory

Let, S be a system, S= (F, ft, fs, H, Mf, Ht, Q, D, Tr) where,

1) Feature Extraction
   F={ft, fs},
   F is set of feature extraction.
   ft is set of token set.
   fs is set of synonyms set.

2) Hash Indexing
   H={Mf, Ht},
   H is set of Hash Indexing.
   Mf is set of mapping function.
   Ht is set of hash table.
3) Query Trademark
   Q={Q1,Q2,...,Qn} ,
   Q is set of query trademark.
4) Distance Computation and suggestions
   D={D1, D2,...,Dn},
   D is set of distance computation and suggestions.
5) Retrieval list

Tr={Tr1, Tr2,...,Trn},
Tr is set of retrieval list.

B .Mathematical model for proposed system

For trademark distance computation,

$$sim(Q,T) = \left|\frac{Qf_t \cap Tf_t}{Qf_t \cup Tf_t}\right| + \frac{|Qf_s \cap Tf_t|}{D} + \frac{\sum_{i=1}^{I}\sum_{j=1}^{J}\max\left(wordsim\left(x_{i,y_i}\right)\right)}{|Qf_t \setminus Tf_t|.|Tf_t \setminus Qf_t|}$$

Qf_t & Qf_s :Token set and synonyms set of the query.
Tf_t:Token set of one of trademark from database.

C. Retrieval Algorithm

Algorithm 1 Pseudo code of proposes system is:

1: define ft as the token set of a trademark.
2: define fs as a set of synonyms that correspond to the token set.
3: define as a list of unique token extracted from the database.
4: for each trademark in the database,
5: do (extract ft ,extract fs)
6: for each token in ft,
7: if (token does not exist in ft)
8: update token into ft.
9: define hash table as index table that maps token to trademarks in the database.
10: define Q as the query of trademark
11: if(Query is present in database)
12: then(display query suggestions of trademark)
13: else(Use this as a trademark).

VI ANALYSIS AND RESULTS

The focuses on this main fact by proposing a hypothetical model of the comparison process, purposed at retrieving conceptually similar trademarks to avoid infringement. The analysis of the trademarks is needed to comprehend the main of conceptual similarities coming from different factors.

A trademark reflow system using the proposed reflow algorithm is evolved, and the algorithmis tested on hypothetic similarity. Provide the conceptual similarity of trademark for avoiding the many trademark infringement. In proposed system having trademark database which is used to find the available trademarks as well as the distance computation for dataset similarity , display the trademark with image and its provide suggestions for the search query if given search query is found in database.

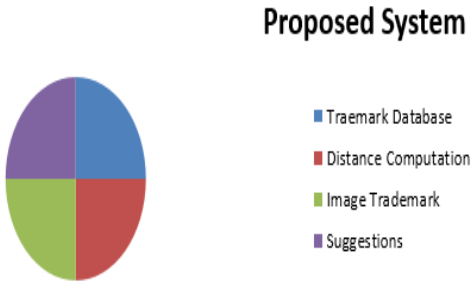

Figure 2: System Graph

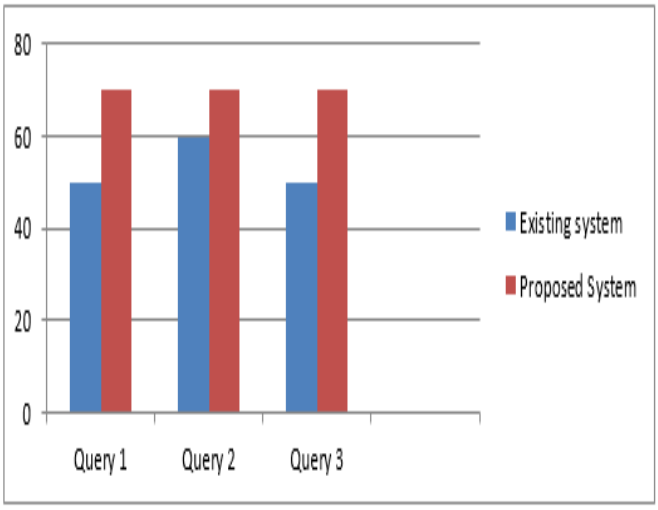

Figure 3: Efficiency comparison graph

Table: Trademark Suggestions Table

Trademarks	Suggestion 1	Suggestions 2	Suggestions 3
Image Fast	Instant Image	Smart Image	Image set
The Car Doctor	Specialist Cars	The Car House	Car Medic
LandLook	Landcare	Land Surveys	Landmark
PC AID	Computer Aid	PC Support	PC Help Center
Magic Kingdom	Magic Word	Magic City	Magic Man
Bodytone	Body To Burn	Build Tone	Body Zone
Party King	The Party Man	Party Land	Party Link
Global Internet	Global Web	Global Link	Power Internet
Computerman	PC Man	Computer Guys	Computer Human
Oak Tree	The Ash Tree	The Olive Tree	The walnut Tree

Figure 4: Load Trademark Database

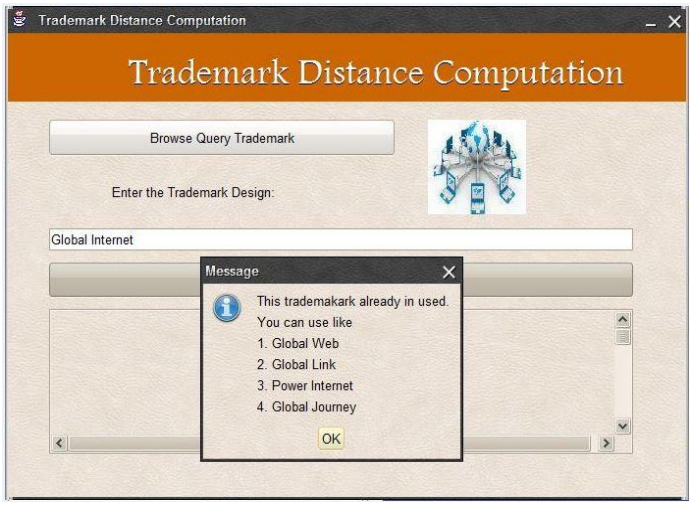

Figure 5: Input Query Suggestions

VII CONCLUSION

The work was motivated by increasing of fraud cases best an data similarities, where information retrieval system do not handle this particular issue and trademark similarity. The target on similarities during trademarks, which becomes when more than two or more trademarks like equal or relevant semantic implant. The advantages and limitations of each data similarity of reflow algorithm are described. The system work, conceptual similarities among trademarks like equal or relevant semantic implant. The desire of a hypothetic model of retrieval trademark is depends on hypothetical similarity. The main model language processing technology, data paths and lexical resources to calculate hypothetic similarity between different trademarks. The system is stimulated for improving of fraud cases best on data processing similarities, where data retrieval system does not manage this particular problems. The system reforms on all ready trademarks find system by legislation a implementing of rectification the find to hypothetic same trademarks. The system employs natural language processing techniques, knowledge sources and a lexical resource to compute conceptual similarity between trademarks. Also confirm that the comparison of trademarks in terms of conceptual similarity. In future work to improve the precision of the proposed semantic algorithm should include a study comparing the use of various lexical resources.

ACKNOWLEDGEMENT

It is pleasure for me to thank lot of people who have support me in completing to this paper in different ways. Starting with, I would like to thank my Guide, **Prof. Varsha Dange** for her support during the entire work. Her support for helping me to get helpful content in the topic. I am thankful to HOD, Principal, teachers and collage to providing us the facilities in the department .

REFERENCES

[1]F. M. Anuar, Yu-Kun Lai, R. Setchi,"Semantic Retrieval of Trademarks Basedon Conceptual Similarity." IEEE Transactions on Systems, Man, and Cybernetics: Systems IEEE permission, 2015.

[2] F. M. Anuar, R. Setchi, and Y. K. Lai, "Trademark image retrieval using an integrated shape descriptor,"Expert Syst. Appl., vol. 40, no. 1, pp. 105-121, 2013.

[3] H. Qi, K. Q. Li, Y. M. Shen, and W. Y.Qu, "An effective solution for trademark image retrieval by combining shape description and featurematching," Pattern Recognit., vol. 43, no. 6, pp. 2017-2027, 2010.

[4] C. H. Wei, Y. Li, W. Y. Chau, and C. T. Li, "Trademark image retrieval using synthetic features for describing global shape and interior structure," Pattern Recognit., vol. 42, no.3, pp. 386-394, 2009.

[5]] L. Sbattella and R. Tedesco, "A novel semantic information retrieval system based on a three-level domain model," J. Syst. Softw., vol. 86, no. 5, pp. 1426-1452, 2013.

[6] M.-Y. Pai, M.-Y.Chen, H.-C.Chu, and Y.-M. Chen, "Development of a semantic-based content mapping mechanism for information retrieval," Expert Syst. Appl., vol. 40, no. 7, pp. 2447-2461, 2013.

[7] S. A. Fadzli and R. Setchi, "Concept-based indexing of annotated images using semantic DNA," Eng. Appl. Artif. Intell., vol. 25, no. 8, pp. 1644-1655, 2012.

[8] R. Setchi, Q. Tang, and I. Stankov, "Semantic-based information retrieval in support of concept design," Adv. Eng. Inf., vol. 25, no. 2, pp. 131-146, 2011.

[9] J. J. Jiang and D.W. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," in Proc. Int. Conf. Res. Comput.Linguist, Taipei, Taiwan, pp.19-33.

[10]J. Oliva, J. I. Serrano, M. D. del Castillo, and A. Iglesias, "SyMSS: A syntax-based measure for short-text semantic similarity," *Data Knowl.Eng.*, vol. 70, no. 4, pp. 390–405, 2011.

[11]B. Furlan, V. Batanovic, and B. Nikolic, "Semantic similarity of short texts in languages with a deficient natural language processing support," *Decis. Support Syst.*, vol. 55, no. 3, pp. 710–719, 2013.

[12]J. P. Eakins, J. M. Boardman, and K. Shields, "Retrieval of trademark images by shape feature—The ARTISAN project," in *Proc. IEEEColloq. Intell. Image Databases*, London, U.K., 1996, pp. 9/1–9/6.