Implementation of Semantic Retrieval by Data Similarity of Trademark

Priya Malve[1] Prof. Megha Singh[2]
[1]M. Tech Scholar [2]Assistant Professor
[1,2]Department of Computer Science & Engineering
[1,2]Central India Institute of Technology, Indore, India

Abstract— A trademarks is a sign that you can use to distinguish your business goods or services from those of other traders. Trademark can be defined expressly in the form of any symbol, logo, titles etc. so, they need to be secure. This project deciphers the hypothetic similarities among trademarks, which happens when more than two or more trademarks hail equal or relevant semantic implant. The state-of-the-art by offering a semantic algorithm to similitude trademarks in pre conditions of hypothetic parallelism. By using data similarity, it is derived that search and indexing technique developed similarity distance. The offered reflow algorithm is confirmed using two resources: a trademark database of conflicting cases and a databases company names. Extends the conceptual model by developing and evaluating a semantic algorithm for trademark retrieval based on conceptual similarity. The conceptual comparison of text documents that share similar domain, use similar concepts, or express similar ideas has been studied extensively. The underlying technology embedded in existing trademark search systems is primarily based on text-based retrieval. Use the different domains to measures the accuracy of the algorithm which gathered different data.

Keywords: Data Mining, EDM, K-Means, Decision Tree, Students Data

I. INTRODUCTION

TRADEMARKS, as prescribed by the European Office of Harmonization in the Internal Market (OHIM). They do immaterial intellectual property (IP) wealth that permit well or service to be well valid to subscriber. Each year number of trademarks registered and used that marketplace. For example, the OHIM increases 2 percent form the last year and applications about 108000 trademarks. The trademarks registered improve by 10 percent from 2010 year to 2012 year in the word. Trademarks violation is a aspect of IP delinquency that hegemony to solemn financial issue. Few general disservice outcomes from trademarks violation is lost income, scarce benefit and extra charge of conservancy to stave off next violation. Economic Committee then inquired by number of fraud cases. That time many trademarks fraud cases filed in Courts but in this cases not involves the compromise cases, it's on another filed in courts. That year IP crime is improved by 97 percent of trademarks fraud cases. When estimate the trademark violation cases then it is a client uncertainty dissection. The dissection is an all things that embrace different auto components of parallelism trademarks. Hence, the point of parallelism has well understand in trademark fraud lawsuit. Two trademarks necessary not be same to build up an violation. Ethically to the trademark roll offered by the OHIM, conceptual similarities of trademark that implant notes or sentence checkup based on the semantic implant act by the trademarks. The roll of another point of view two trademarks are practically same if they hail as semantic contain. For example trademark that contains the word "run" is same to a trademark that uses the word "scoot" because both has same significances that two word are synonyms. The conceptual comparison of text documents that share similar domain, use similar concepts, or express similar ideas has been studied extensively. However, the conceptual comparison of trademarks is a unique problem. The conceptual comparison of trademark words and phrases is there Semantic Retrieval by Data Similarity of Trademarks for a new problem in the domain of trademark retrieval. It requires a cross-disciplinary approach involving natural language processing and external knowledge sources. This provides a mechanism via which to compare the conceptual aspects of trademarks by proposing a trademark retrieval algorithm based on their conceptual similarity. The proposed al-.

II. RELATED WORK

F. M. Anuar, R. Setchi, and Y. K. Lai [2] author proposed trademark image re trieval using an integrated shape descriptor. As the proposing innovatory trademark reflows technique to use the reform performance of expositor. Trademarks are dis tinctive visual symbols with high reputational value, due to the perception of quality and innovation associated with them. They are important reputational assets used as a marketing tool to convey a certain assurance of quality, innovation, and the standards, which the manufacturer seeks to maintain. This motivates the need for trademark protection by providing a solution to prevent infringement. This problem can be addressed by developing retrieval systems capable of comparing the visual similarity of trademarks. The visual similarity is checked during the trademark registration process where one of the steps involved is making sure that the trademark to be registered is not similar to any trademark which is already registered. This is important in order to avoid infringements as well as to protect the rights of the existing trademarks.

H. Qi, K. Q. Li, Y. M. Shen, and W. Y.Qu [3] author proposed to introduced sub- stance point of a exclusive figure and this the point used to search nook pixel from it. Number of image collections available has increased due to ease of capturing images by different acquisition systems. The storage format of image data is relatively standardized however the effective retrieval of images from such databases remains a significant challenges. For the performance evaluation of the system we use the most commonly used method namely precision-recall. From the experimental result we conclude that the Trademark Image Retrieval based on shape feature perform better and gives satisfactory result.

L. Sbattella and R. Tedesco [4], author proposed a fact and ideal for substance and listing information from main data. Use the conceptual level and lexical level for describes the main information. The stochastic model is then used, during the document indexing phase, to disambiguate word meanings. The semantic information retrieval engine we

developed supports simple keyword-based queries, as well as natural language-based queries. The engine is also able to extend the domain knowledge, discovering new and relevant concepts to add to the domain model. The validation tests indicate that the system is able to disambiguate and extract concepts with good accuracy. A comparison between our prototype and a classic search engine shows that the proposed approach is effective in providing better accuracy. The common goal of such methodologies is to automatically extract structured information from natural language documents. The used of model for knowledge extraction from natural language documents.

M.-Y. Pai, M.-Y.Chen, H.-C.Chu, and Y.-M. Chen [5] Author proposed the many data reflow systems use search information as user input data, but it is a mainly hard and complicated so use the semantic mechanism. To address this problem developed a semantic-based content mapping mechanism for an information retrieval system. This approach employs the semantic features and ontological structure of the con- tent as the basis for constructing a content map, thus simplifying the search process and improving the accuracy of the returned results. Information retrieval systems include the searching technologies and functions that can help users find the information that they need based on criteria they are given. Existing IR systems mostly perform searches based on keywords entered by the user, although keywords cannot render a complete representation of the content semantics.

F. M. Anuar, R. Setchi, and Y. K. Lai [6], author proposed to mainly focus on main fact by proposing a notation flow of the different procedure, to main at reflow the same trademarks. Trademarks are proprietary words and images with high reputational value; they are important assets, often used as a marketing tool, which require Infringement protection. One of the issues considered during infringement litigation is the visual, conceptual and phonetic similarity of different trademarks. In particular, the conceptual similarity of trademarks is an area never previously studied in information retrieval. The focuses on this important aspect by proposing a conceptual model of the comparison process, aimed at retrieving conceptual Semantic Retrieval by Data Similarity of Trademarks actually similar trademarks. The proposed model employs natural language processing and semantic technology to compute the conceptual similarity between trademarks. The model employs natural language processing techniques, knowledge sources and a lexical ontology to compute conceptual similarity between textual trademarks.

Rossitza Setchi , Qiao Tang, Ivan Stankov[8], author proposed, The semantic- based image retrieval tool tags images by first processing all significant words in the text around them, extracting all keywords and key phrases in it, ranking them ac- cording to their significance, and linking them to ontological concepts. It generates a set of concept numbers for each text, which is then used to retrieve information in a process called semantic expansion, where a keyword query is also processed semantically. The semantic-based image retrieval tool developed has demonstrated good performance and scalability, and has been integrated with keyword-based indexing and content retrieval algorithms in an industrial prototype.

Jay J. Jiang, David W. Conrath [9], author proposed, a new approach for measuring semantic similarity between words and concepts. It combines the lexical taxonomy structure with corpus statistical information so that the semantic distance between nodes in the semantic space constructed by the taxonomy can be better quantified with the computational Semantic Retrieval by Data Similarity of Trademarks evidence derived from distributional analysis of corpus data. Specifically, the measure is a combined approach that inherits the edge-based approach of the edge counting scheme, which is enhanced by the node-based approach of information content calculation.

Sr.No.	Paper	Technique	Advantage	Limitations
1	A novel semantic Information retrieval system based on a different level model.	Paper proposed a fact and ideal for substance and listing information from main data. Use the conceptual level and lexical level for describes the main Information.	Provided good precision compare to regular search engine that is a simple and well powerful system.	Not used put the bound on the accuracy of data.
2	Development of a semantic- based content mapping mechanism for information retrieval.	Paper proposed many data reflow Systems use search information as user input data, but it is a mainly hard and complicated so use the semantic mechanism.	It has semantic advantage and good flow of the listing as the increasing the precision and fast searching.	Not allow fast recognizance and documentations.
3	Trademark Image retrieval using an integrated shape descriptor.	Author proposed Trademark image retrieval using an integrated shape descriptor as the proposing in innovatory trademark reflow technique to use the reform performance of Expositor.	Used employed shape Features and descriptor matching stage	Indirect same events of minimum human understanding of parallelism.

III. DEFINITIIII. PROBLEM DEFINATION AND SCOPE

As deep Trademarks violation is an aspect of IP delinquency that hegemony to serious financial issue that is it trademark infringement is lost revenue. The search and indexing technique developed uses similarity distance. Hence, the concept of similarity has become well understood in trademark infringement litigation. A successful one of the

aspects of similarity assessed during trademark analysis, which is conceptual similarity.

A. Goals and Objectives

- To make an improvement of trademark conceptual similarity to make them more accurate and more secure against the trademark infringement.
- Propose the search and indexing technique developed uses similarity distance.
- Propose a computational approach based on semantics that can be used to compare trademarks for conceptual similarity.
- To avoid the additional cost of protection to future Infringement.

IV. METHODOLOGIES OF PROBLEM SOLVING AND EFFICIENCY ISSUES

A. Feature Extraction

Each trademark is represented by two kinds of features (i.e., the trademark tokens and the synonym list). The feature extraction step begins with a spelling correction process that corrects any spelling mistakes using a spell checker. Then, frequent words (i.e., no, and, the, etc.) are removed, and the trademarked words are extracted in the form of tokens. The trademark tokens extracted here are sets of English root words. For example, the word flying will be converted into fly. The second feature is defined as the synonym set of the tokens and is extracted from the WordNet database. The synonym set, as defined in the context of this algorithm, includes the synonyms, the direct hypernyms, and the direct hyponyms of the corresponding tokens.

B. Hash Indexing

Hash indexing takes the trademark as the key index. It is then mapped to a list of trademark features from the database using a mapping function. The mapping function is designed so that the trademark similarity distance computation is performed only on the set of trademarks that consist of at least one of the terms in the synonyms set belonging to the trademark query. The final indexing table is merely a table that maps each trademark in the database to a set of trademarks from the same database for the trademark similarity computation. In this manner, the distance computation is not conducted over the entire database, which enhances the speed of the retrieval process.

C. Trademark Distance Computation

The distance computation is based on the similarity concept introduced in Tverskys contrast theory. Tversky defines the similarity between two objects as a function of unique and shared information about the object. The first part of the equation is the ratio of the number of elements shared by the two trademark token sets and the number of elements in their set union operation the second part is the ratio of the number of elements in the intersection of the Red Bull synonyms set and Blue Bull token set. The third part is the word similarity between the difference sets of both trademarks, measured using WordNet ontology, and the final part is the summation of the three parts, which provides the conceptual similarity score between the two trademarks.

D. Trademark Design Similarity Computation

Trademark design similarity computation is based on image similarity. In proposed system, random pixels will be selected for comparison and if pixels are found as similar then message will be displayed to the user that trademark is already used by another user.

E. Outcome

The project helps to check the similarity in the trademark. If trademark is
Conceptually matching with any other trademark the system will shows that this trademark is already registered. And system will provide hint trademark.

F. User Registration

Fig. 10.1: User Registration

This is important page of application, where new users of system get registered to the system. User needs to enter all information. Here primary fields are mobile numbers and mail id.

G. User Login

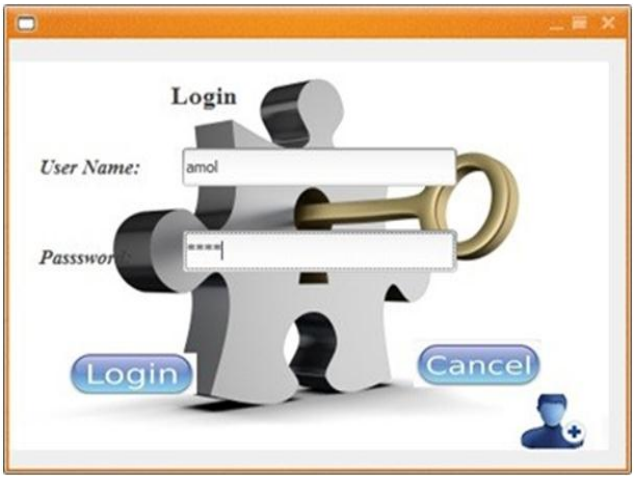

Fig. 10.2: User Login

This is login page by using this page user get access to the system. But that user needs to enter the valid user id and password.

H. Loading of Training Dataset

Fig. 10.3: Loading of Training Dataset

Here user needs to load the trademark dataset first to check the trademark ilarity later. This dataset trademarks will used to check the similarity of trademark at time of trademark validation.

I. Stop Word Removal

Fig. 10.4: Stop Word Removal

This page the stopword are get removed form the query trademark. The may contents am, is, are, the, near, to etc. This words are get removed and after this step is extraction of synonyms.

J. Feature Extraction

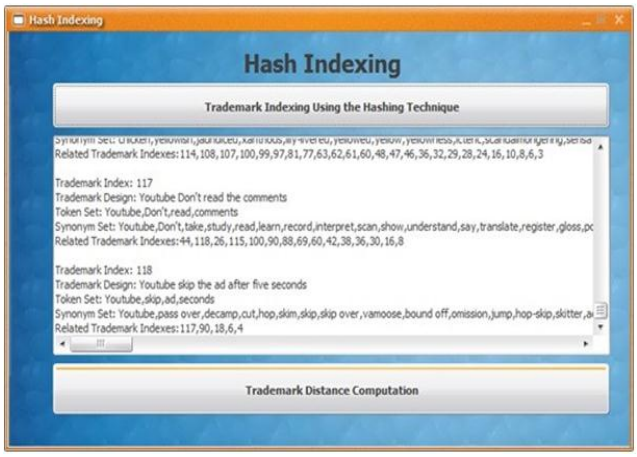

Fig. 10.5: Feature Extraction

This page content the extracted features of the query trademark.

K. Distance Computation

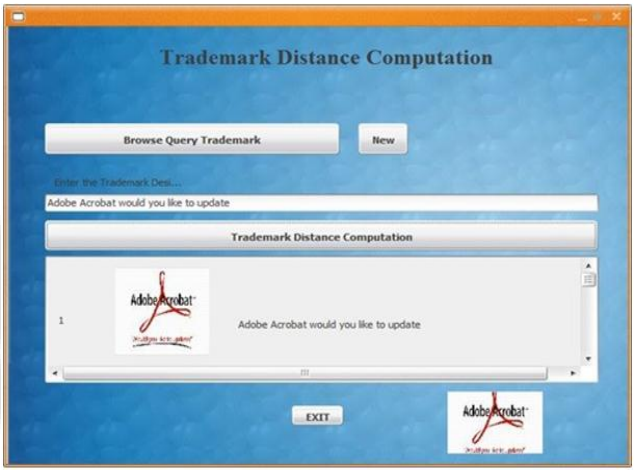

Fig. 10.6: Distance Computation

This page system calculates the distance or similarity between the query and training dataset trademarks and suggest user he/she can use this trademark or not.

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 Hu-man
Oak Tree	The Ash Tree	The Olive Tree	The walnut Tree

Table 10.1: Trademark Suggestions Table

L. Trademark Suggestion

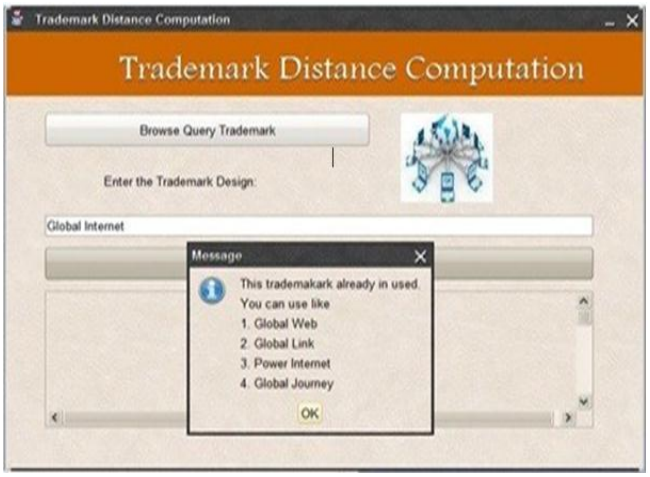

Fig. 10.7: Trademark Suggestion

If the query trademark is already used by some other company then user will get suggestion that u cannot use this trademark. And after this messages user get some trademark similar to the query trademark.

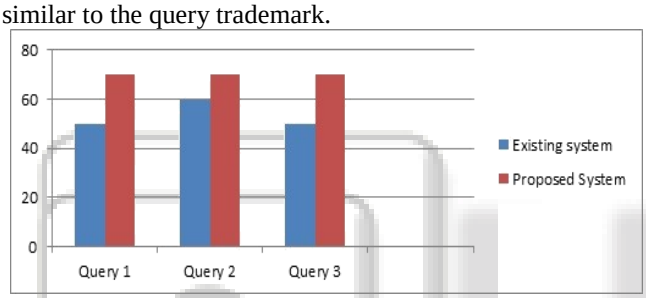

Fig. 10.8: System Analysis

V. SYSTEM ANALYSIS

A trademark reflow system using the proposed reflow algorithm is evolved, and the algorithm is tested on hypothetic similarity. Provide the conceptual similarity of trademark for avoiding the many trademark infringement. In proposed system hav- ing 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.

REFERENCES

[1] FatahiyahMAnuar, Yu-Kun Lai, R. Setchi,"Semantic Retrieval of Trademarks Based on Conceptual Similarity." 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, andW. Y. Qu, "An effective solution for trademark image retrieval by combining shape description and feature matching," Pattern Recognit., vol. 43, no. 6, pp. 2017-2027, 2010.
[4] 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.
[5] 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.
[6] F. M. Anuar, R. Setchi, and Y. K. Lai, "A conceptual model of trademark re- trieval based on conceptual similarity," in Proc. 17th Int. Conf. Knowl. Based Intell. Inf. Eng. Syst., Kitakyushu, Japan, 2013, pp. 450-459.
[7] 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.
[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, Tai- wan, 1997, pp. 19-33.

