www.ijiarec.com

ISSN:2348-2079

Volume-7 Issue-1

International Journal of Intellectual Advancements and Research in Engineering Computations

Semantic web mining trademark databases

M.N. Shri Gowtham[1], S. Kavin[2], J. Prabhakaran[2], D. Preethi[2]

[1]Assistant Professor, Department of Information Technology.

[2]UG Students, Department of Information Technology, Nandha Engineering College (Autonomous), Erode-52, Tamil Nadu, India

ABSTRACT

Semantic web is a one the concept of web mining. The aim of the project is find a duplicate logo from a large number of logo. This paper presents a method for measuring the semantic similarity between concepts in Knowledge Graphs (KGs) such as WorldNet and DBpedia. Previous work on semantic similarity methods have focused on either the structure of the semantic network between concepts (e.g., path length and depth), or only on the Information Content (IC) of concepts. We propose a semantic similarity method, namely wpath, to combine these two approaches, using IC to weight the shortest path length between concepts. Conventional corpus-based IC is computed from the distributions of concepts over textual corpus, which is required to prepare a domain corpus containing annotated concepts and has high computational cost. IC based on the distributions of concepts over instances. Through experiments performed on well-known word similarity datasets, we show that the w-path semantic similarity method has produced a statistically significant improvement over other semantic similarity methods.

INTRODUCTION

Trademark Data Mining

Research in Effective text mining is starting to produce technology data collection which can make information in Effective literature more accessible for data mining. One of the current challenges is to integrate and refine this technology dataset to support real-life technology tasks in data mining, and to evaluate its usefulness in the context of such tasks. We describe Effective Pattern Discovery for Text Mining (EPTEM) – a fully integrated text mining tool designed to support data and its risk assessment. This task is complex and time-consuming, requiring a thorough review of existing technology data on particular information.

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 separate property (IP) goods that permit well or service to be well validated to clients. Each year many trademarks registered and used to 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 defendable and compared by traders. Searching for conceptually same 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

Author for correspondence:
Department of Information Technology, Nandha Engineering College (Autonomous), Erode-52, Tamil Nadu, India

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. 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 similar domain, utilization alike notations, or demonstration same consideration has been used broadly**.**

TRADEMARK DATASET

A trademark identifies the brand owner of a particular product or service. Trademarks can be licensed to others; for example, Bully land obtained a license to produce Smurf figurines; the Lego Group purchased a license from Lucas film in order to be allowed to launch Lego Star Wars; TT Toys is a manufacturer of licensed ride-on model cars for children. The unauthorized usage of trademarks by producing and trading counterfeit consumer goods is known as brand piracy.

The owner of a trademark may pursue legal action against trademark infringement. Most countries require formal registration of a trademark as a precondition for pursuing this type of action. The United States, Canada and other countries also recognize common law trademark rights, which means action can be taken to protect an unregistered trademark if it is in use. Still, common law trademarks offer the container in general less permissible protection than registered trademarks [1, 2].

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 indorse or brand properties 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 explorations 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 symbols are necessary not same to make a contravention. The theoretical different of text files that part of similar domain, utilization same notations, or demonstration same consideration has been used broadly.

PROPOSED SYSTEM

Relational Keyword search based on WSMO K-SVM Clustering algorithms have been studied for decades, and the literature on the subject is large. Therefore, we decided to choose a WSMO K-SVM is our illustrative algorithm in order to show the possible of the planned approach, namely: the separated the cluster semantic word extraction algorithm known as K-Support Vector Machine. Trade mark judgment based on conceptual resemblances. This paper extends the conceptual model by developing and appraising a semantic algorithm for symbol retrieval based on conceptual similarity. The projected algorithm employs NLP techniques and the word similarity distance method, which was derived from the WorldNet ontology, together with a new trademark comparison measure. Word Net is employed in this algorithm due to its lexical relationships, which mirror human semantic organization, and because it has also been proven successful in many previously developed works.

These algorithms were run with different combinations of their parameters, resulting in sixteen different algorithmic instantiations. Thus, as a contribution of our work, we compare their relative performances on the studied application domain using Trademark datasets. In order to make the comparative analysis of the algorithms more realistic, two relative validity indexes related keyword result to the user that have been used to estimate the number of clusters automatically from data.

Advantages of proposed system

- Most importantly, we observed that clustering algorithms indeed tend to induce clusters formed by either relevant or irrelevant documents, thus contributing to enhance the expert examiner's job.
- This method in applications shows that it has the potential to speed up the computer inspection process.
- Better support with clustering group of data's.
- Highly efficient.
- Provide good result.

METHODOLOGY

The rapid expansion of pattern recognition and machine learning, a lot of data mining algorithms have been proposed by specialists, through which researchers can find much interesting concealed information from the observational data, classification information is one of the most important ones in it, which can be used to know, predict or classify those current unnoticed data. Generally, machine learning algorithms can be separated into three categories: Supervised Learning algorithm, Unsupervised Learning algorithm, and Semi-supervised Learning algorithm. The common supervised learning algorithms comprise regression analysis and classification analysis. Data classification process includes two stages: the first one is the learning stage, the aim of which is to shape a classifier through analyzing the labeled data; the second one is the predicting stage, which using the recognized model for forecasting. The model should have sufficient simplification ability, i.e., that the model not only has good classification presentation on the training data, but also has a high classification accuracy for the future data, which supposed has the same statistical supply as the training data. The main classification algorithms include Decision Tree, Bayes, Neural network, Support Vector Machine (SVM), etc. Support Vector Machine (SVM) is one of the most general and effective algorithms in machine learning. SVM is based on the structural risk minimization criterion and its goal is to find the optimal separating hyper plane where the separating margin should be maximized. This approach improves the generalization aptitude of the learning machine and solves about problems like non-linear, high measurement data separation

and the classification issue that lacking of prior knowledge. SVM is used in systems for face recognition [3, 4], road sign recognition and other similar application areas [5] because of its sound theoretical foundation and good oversimplification ability in practical application. SVM works well in both linear and non-linear conditions, and finding the optimal unraveling hyper plane is the key to separate data. For non-linear circumstances, SVM exploits the kernel trick to map low-dimension data into high dimension feature space. In the practical application, SVM makes use of all the labeled data to find the unraveling rule, but training on large scale data can bring with higher computation cost. In order to decrease the computational complexity, the solution that can be exploited includes two species, one is to improve the algorithm itself, such as the Least Square SVM [6, 7], the SMO [8] (Sequential Minimal Optimization) under semi positive definite kernel; the other is to reduction the number of input vectors. The main task of clustering [9] is to group the objects into clusters, objects in the same cluster are more semblable than those in dissimilar clusters. It can find the relationships amongst data objects in an unsupervised way. A lot of clustering algorithms have been proposed and better, aiming to enhance the efficiency and accuracy. According to cluster mode, clustering algorithms can be categorized into: centroid-based bunching, hierarchical clustering, distribution-based clustering and density based clustering.

METHODS

Edge Based Measure

Semantic similarity depends on the path length and on the position of the concept in the taxonomy. It employs the concept of common subsumes (i.e., the ancestor concept that subsumes the two concepts). It is simple to implement. Two concept pairs of equal length will have the same similarity. Two concept pairs that share exactly the same least common subsume and are of equal length will have the same similarity [10, 11].

Information Content

It assumes that the similarity between the two concepts can be derived based on the specificity of the concepts. The more specific a concept is in the taxonomy, the richer the information content will be. The information content calculation is derived based on the probability of the occurrence of concepts in the taxonomy. Two pairs with similar lcs and cumulative IC may have the same similarity.

Feature Based Measure

It is independent of taxonomy and the subsumes of the concepts. It assumes that each concept has specific features that can be employed to measure similarity. It is defined as the „glosses" (i.e., the definitions of concepts as the features that represent the concepts). The computational complexity is very high.

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.

CONCLUSION

The work was motivated by cumulative of fraud cases best a data comparisons, where information retrieval system do not grip this particular issue and trademark similarity. The target on resemblances during trademarks, which becomes when more than two or more symbols like equal or relevant semantic implant. The advantages and limitations of each data similarity of reflow algorithm are described. The system work, theoretical similarities among trademarks like equal or relevant semantic implant. The desire of a hypothetic model of retrieval trademark is depends on imaginary similarity. The main model language dispensation technology, data paths and

lexical resources to calculate hypothetic similarity between different trademarks. The system is enthused for improving of deception cases best on data processing similarities, where data retrieval system does not manage this specific problem. The system reforms on all ready trademarks find system by law making an 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 contrast of trademarks in terms of abstract similarity. In future work to improve the precision of the proposed semantic algorithm should include an education comparing the use of various verbal resources.

REFERENCES

[1]. P. Eagle, Yanming Shen "News on patent, trademark and design databases on the Internet," World Patent Inf., 34(2), 2012, 164–165.
[2]. L. Dodell. J M Boardman and K Shields, The Trademark Problem: Casualty Insurance's Dirty Little Secret. [Online]. Available: http://www.carriermanagement.com, accessed, 2013.
[3]. Bojan Furlan, conceptual model of trademark retrieval based on conceptual similarity," in Proc. 17th Int. Conf. Knowl. Based Intell. Inf. Eng. Syst., Kitakyushu, Japan, 2013, 450–459.
[4]. C. D. Scott, Jesús Oliva "Trademark strategy in the Internet age: Customer hijacking and the doctrine of initial interest confusion," J. Retail., 89(2), 102, 176–189.
[5]. C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," in *WordNet: An Electronic Lexical Database*, C. Fellbaum, Ed. Cambridge, MA, USA: MIT Press, 1998, 265–283.
[6]. U.S. International Trade Commission.. China: Effects of Intellectual Property Infringement and Indigenous Innovation Policies, On the U.S. Economy. [Online]. Available: http://www.usitc.gov/, accessed, 2013.
[7]. Office for Harmonization in the Internal Market. Guidelines for Examination in the Office for Harmonization in the Internal Market on Community Trade Marks, Part C Opposition, Identity and Likelihood of Confusion, Available: https://oami.europa.eu, accessed 2(3), 2014.
[8]. F. M. Anuar, R. Setchi, and Y. K. Lai, "Trademark image retrieval using an integrated shape descriptor," Expert Syst. Appl., 40(1), 2013, 105–121.
[9]. H. Qi, K. Q. Li, Y. M. Shen, and W. Y. Qu, "An effective solution for trademark image retrieval by combining shape description and feature matching," Pattern Recognit., 43(6), 2010, 2017–2027.
[10]. 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., 42(3), 2009, 386–394, 2009.
[11]. Y. S. Kim and W. Y. Kim, "Content-based trademark retrieval system using a visually salient feature," Image Vis. Comput., 16(12–13), 1998, 931–939.