Intelligent Identification of Trademark Case Precedents Using Semantic Ontology

A. S. LI [a,1], A.J.C. TRAPPEY[a], C.V. TRAPPEY[b]

[a] *Department of Industrial Engineering and Engineering Management, National Tsing Hua University, Hsinchu, Taiwan*
[b] *Department of Management Science, National Chiao Tung University, Hsinchu, Taiwan*

Abstract. A registered trademark distinctively identifies a company, its products or services. A trademark (TM) is a type of intellectual property (IP) which is protected by the laws in the country where the trademark is officially registered. TM owners may take legal action when their IP rights are infringed upon. TM legal cases have grown in pace with the increasing number of TMs registered globally. In this paper, an intelligent recommender system automatically identifies similar TM case precedents for any given target case to support IP legal research. This study constructs the semantic network representing the TM legal scope and terminologies. A system is built to identify similar cases based on the machine-readable, frame-based knowledge representations of the judgments/documents. In this research, 4,835 US TM legal cases litigated in the US district and federal courts are collected as the experimental dataset. The computer-assisted system is constructed to extract critical features based on the ontology schema. The recommender will identify similar prior cases according to the values of their features embedded in these legal documents which include the case facts, issues under disputes, judgment holdings, and applicable rules and laws. Term frequency-inverse document frequency is used for text mining to discover the critical features of the litigated cases. Soft clustering algorithm, e.g., Latent Dirichlet Allocation, is applied to generate topics and the cases belonging to these topics. Thus, similar cases under each topic are identified for references. Through the analysis of the similarity between the cases based on the TM legal semantic analysis, the intelligent recommender provides precedents to support TM legal action and strategic planning.

Keywords. Text mining, trademark infringement, semantic analysis, knowledge ontology, Latent Drichlet Allocation

Introduction

Legal firms and researchers are increasingly studying methods to improve the efficiency of monitoring and protecting the intellectual property of clients. Dabass et al. (2018) discussed the current application of AI in law. Many AI technologies have been applied in fields such as document search and retrieval and intellectual property valuation and forecasting. Surden (2014) described the problems legal practitioners face when machine learning is used to support legal decisions and provide guidelines for developing future applications.

  Intellectual property (IP) covers the domain of intangible property which is created by human intelligence, such as trademarks (TM), copyrights, designs, and patents. A TM

---

[1] Corresponding author, Mail: s107034601@m107.nthu.edu.tw.

represents a company's distinct image in the market (e.g., name and/or logo of a corporation and its product or service offered), and shows the brand value of the business. TM owners increasingly aware of TM infringement occurred and take actions to protect their IP rights. TM infringement indicates the unauthorized use of a TM or service mark in a manner that is likely to cause confusion, deception, or mistake about the source of the goods or services. Considering the rising problems of IP issues for online e-commerce settings, TM legal cases occured very often and finding case precedents becomes critical for quick and accurate actions. Nonetheless, manually searching and idetifying case precedents are time-consuming and laborious since each case can be as long as 30 to more than 100 pages. This research establishes a recommendation system for TM infringement cases. The system is based on the construction of a TM litigation ontology. Using text mining methods, features are extracted from cases to build a database which reduces case search time and increases efficiency.

The objective is to develop an intelligent decision support for identifying TM litigation precedents. In this study, US TM litigation cases were collected from the Westlaw database as the machine-readable e-documents. The system provides users (such as lawyers) with an effective tools for searching and preparing litigation. To develop an intelligent method suitable for the precedent recommendation, an intellectual ontology in the field of trademark court judgment is constructed to represent machine-readable trademark infringement legal documents. The system identifies similar precedent characteristics by writing a legal case summary including facts, issues, rationale, holdings and rules (applicable law and cited precedents). Given the connection between the applicable law and the cited case, topics are identified based on the facts and issues from the judgment cases.

1. Literature Review

In this section, literature related to the research domain are reviewed. First, trademark and US trademark litigaiton are breifly discussed and introduced. Then, the meaning of e-discovery and its application is reviewed. Finally, the concept of ontology is introduced.

1.1. Trademark and Trademark Litigation

TMs are constructed using words, phrases, symbols, designs, colors, product or packaging appearance, sounds, or a combination of these elements. The U.S. Patent and Trademark Office (USPTO) defines a trademark as a "word, phrase, symbol, or design, or a combination thereof" that used to identify and distinguish a company's goods. The main function is to identify the source of the goods or services and to distinguish it from that of others. Since a trademark is used as a mark to represent goods or services, the mark must be uniquely identifiable with symbolic features for differentiation that do not confuse or mislead the customer.

In the age of digitalization, trademarks are spread through the social media and have economic and social value protected by law if they mark is registered. In terms of the purpose of trademark legislation, there are two points related to customer and trademark

owners respectively. One is to protect the public interest so that customer can be confident when purchasing a product marked with a familiar trademark which ensures the quality and manufacturer of the product. Secondly, since the owner of a trademark has spent time and money developing the brand equity of the marked products and services, there is protection granted law to prevent the counterfeit of goods or infringement of the brand (Rogers, 1949).

The US judicial system is a two-track system, including a state court system and a federal court system. The Federal judiciary of the United States includes the Federal District Court, the Federal Circuit Court or United States Courts of Appeals, and the US Supreme Court. In the US, there are no specialized courts for trademark enforcement. The protection of US trademarks is examined under federal law, state law, and common law. It also adopts a two-track system, a trademark allegation can be brought to federal or district courts. For the federal level of litigations, it applies the federal trademark law; while the state level applies state law for each state and the common law of the unfair competition (Common Law of Unfair Competition).

1.2. E-Discovery

For US litigation, especially civil litigation, the parties involved have the obligation to provide each other with evidence related to issues and facts before the trial, which is called "discovery" (Roitblat et al., 2010; Oard, 2013). According to Conrad (2010), EDD, Electronic Data Discovery or E-Discovery, refers to a process of searching, identifying, processing, and organizing electronic legal information as a critical evidence in the litigation. As the volume of electronic legal documents continues to grow, it takes more time and costs for attorneys to affordably manage this obligation in time. As a result, attorneys are looking for efficient ways to reduce the time for e-discovery without increasing error.

Many e-discovery technologies have been developed for searching electronic evidence such as email, databases, and video files. Many applications use artificial intelligence (AI) to screen large amounts of evidence and performance improves as the applications learn as they search more data over time. For the conventional process of discovery, attorneys manually review and filter relevant evidence for litigation. New technologies have been applied to assist in searching and reviewing relevant documents. The growing volume of electric information drives the demand for technology-assisted review, such as keywords search and predictive coding (Belt, 2011). Endo (2018) reported that machine learning has been successfully used for legal decision making. Application of data collection and data processing of legal documents lowers costs when searching for and analysing the documents and cases. Technology-assisted review is trained by legal experts to recognize and create categories for the input documents. Calo (2016) discusses the use of a robot as a metaphor for assisting a judge during trial litigation.

People can apply the technical process to the analysis of large amounts of legal data. However, without well understanding of the algorithms and the analytic results that the technological system produced, it might cause misinterpretations and more bias. (Conrad, 2010). Some argue that data analytic approaches provide greater convenience rather than the reliability of results. There are increasing concerns that the validity and reasonableness of computational systems, such as machine learning, show different

qualities depend on human interpretation. (Endo, 2018; Hildebrandt, 2012). Explicit and implicit biases in the system influence the results and social relevance. According to Burk (2019), algorithmic fair use sholud be considered whether the legal regulations can be translated into the personalized use by machines. Analytic data technologies bring convinces to human, however, it also brings uncertainty due to its simplified assumptions. In order to prevent errors and improper inspection, human-assisted supervision and interpretation are necessary for the automated fair use system. Belt (2011) also concluded that technology-assisted review would be more reliable with proper support from the legal profession knowledge and appropriate process. E-discovery should be a solution for supporting the litigation process and it depends on the understanding and interpretation from the analytic data results.

1.3. Ontology

Ontology originates from a philosophy discipline, which aims to explore the nature of things. According to Gruber (1993), ontology is an explicit form of terminologies representing a domain knowledge (elements and their relations). Studer et al. (1998) defined an ontology as a formal and explicit specification of a shared knowledge conceptualization. In overall, an ontology should represent a knowledge domain comprehensively, which enable machines to communicate and exchange opinions (or understand each other's domain context). Ontology allows domain knowledge to have a common form of representation, enabling computers to achieve true semantic communication. According to Chandrasekaran et al. (1999), ontology schema enables definition, sharing, and reasoning of the representing content. It is not only about the nature and structure of knowledges, ontology can also define the relationships between elements and attributes in a knowledge domain (Staab et al., 2009). Ontology is usually constituted by classes, attributes, instances, and relations. These ontology elements (in nodes) and relations (in arcs), when comprehensively defined, are used to represent a semantic network of domain knowledge.

The ontology schema has been widely used to represent different fields of knowledge. In this research, a TM case judgment ontology schema is defined for developing a TM case precedent recommender using semantic. Recommender systems provide suggested choices based on user's preferences, often shown implicitly in his/her previous behaviors (Shishehchi et al. 2012; Cruz et al. 2015). Munir et al. (2018) introduces and discusses several ontology-based information retrieval approaches in the past research. Trappey et al. (2019) proposed an ontology-based TM litigation case recommender by retrieving case key features defined in TM ontology schema. In this research, we extend the previous research (Trappey et al., 2019) by comprehensively defining frame-based ontology schema for TM litigation knowledge representing approach. The method establishes frame-based feature templates for automatically retrieving features into machine-readable database tables. Machines can automatically generate the information following the framework structure, which can effectively improve recommender's efficiency and reliability.

2.   Research Methodology

Figure 1 shows the framework of the intelligent TM precedent recommendation system. There are two modules in the system, i.e., computer assisted knowledge extraction and the identification of precedents based on key legal issues/topics. Users, such as IP attorneys or company's IP counsels, can input a seed case into the system platform. In this research, we collected US trademark litigation judges from Westlaw.com during 1995 and 2018 as experimental data/cases. The input cases will be represented in ontology-based knowledge features, e.g., facts, issues, and rules in the litigation cases. As for the issue identification, the issues are defined by several key terms respectively. LDA topic modeling (i.e., a soft clustering) is, then, applied to find the cases with similar features for precedent recommendation. Finally, related laws, cites and cases are ranked as a prioritized recommended list for users.
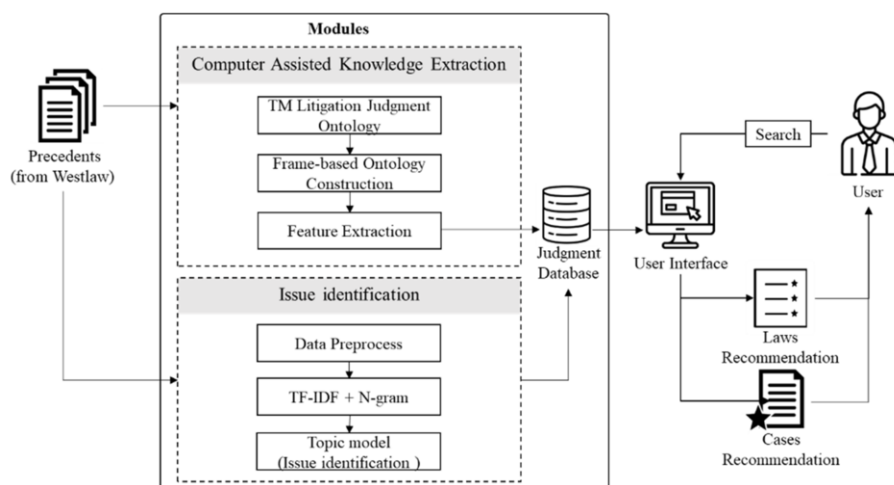

Figure 1. The framework of intelligent TM precedent recommendation system.

2.1. TM Litigation Judgment Ontology

An ontology schema of TM litigation issues is constructed in three levels of classes and their associating attributes. As the purpose of this research is to construct a computer assisted system for recommending relevant cases automatically, the key task is first to convert the judgments (documents) as machine readable data tables. The ontology schema of TM case law is defined according to the essential elements of case brief in common law system. A case brief consists of five essential categories (classes) of case information, including (1) basic data, (2) facts, (3) issues, (4) holdings, and (5) rules (i.e., relevant laws). The ontology schema is further defined into sub-levels of feature classes and their key terms of attributes (as shown in Fig. 2 – sub-classes of "issues" class and their key attributes).
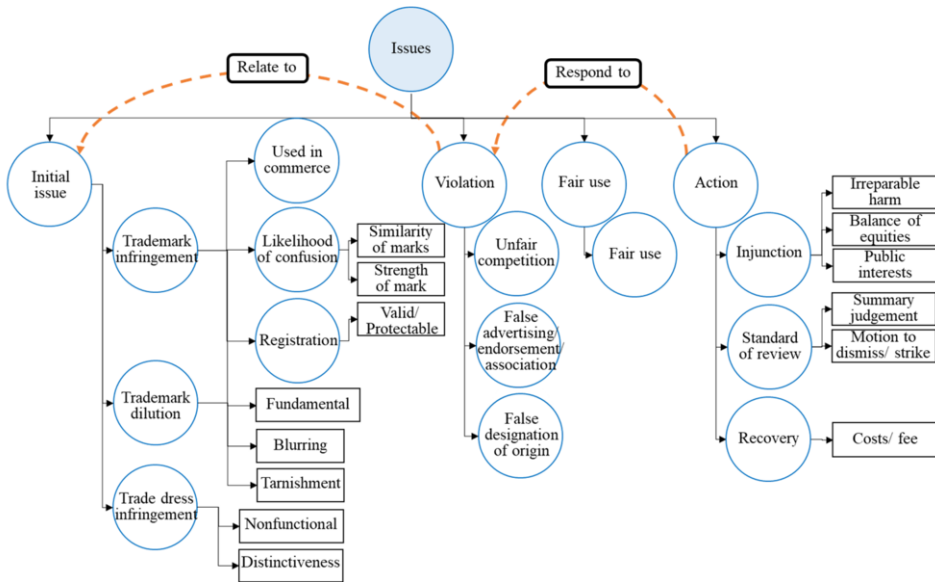
Figure 2. The definition of "issues" class for TM litigation ontology schema.

2.2. Feature Extraction Using Regular Expression

In order to make the system extracting each case judgment's features (or instance values), based on the frame-based ontology schema, there are a standard steps of feature extraction from the litigation judgment. First, to extract the information in the "basic data" class, the regular expression technique is applied according to basic data's legal expression formats (Kleene, 1951). A regular expression is a standard string complying with the required pattern/format of given formal documents and the retrieval of targeted data with regular expression in text corpus is popularly applied in information retrieval practices (Friedl, 2006; Wheeler, 2016). In legal judgments, the laws are written in a certain formats/patterns. Thus, the regular expression can extract the features automatically in large number of documents in the legal corpus.

2.3. Issue Identification

Apart from the other classes of ontology schema, the features of "issues" attributes are retrieved using term frequency–inverse document frequency (TF-IDF) and N-gram algorithms for identifying frequently appeared and highly representative key words and phrases (collectively called *key terms*) (Qaiser, 2018, Robertson, 2004). According to TM litigation ontology schema, as shown in Fig. 2, the main issues and relevant key terms are defined. The TF-IDF and N-gram key-term retrieval for any given judgment can assist the recognition of the case features and identify the most related issues of the given case. Additionally, Latent Dirichlet Allocation (LDA) is further deployed for soft clustering cases into topic models (Blei, 2003). Both identifications of issues and topics of TM cases in the corpus will support the findings of case law precedents for any given seed case. In the research, judgment documents of US TM litigation cases require pre-processed through tokenization, stop words cleaning, and lemmatization prior TF-IDF, N-gram, and LDA test mining operations.

2.4. Judgment Database

After the steps of computer assisted knowledge extraction and issue identification, we then build a database for storing and linking the key features of each case in formatted data tables for the cases (their features) similarity comparison. In the database, issues will also be linked to the most related federal laws. When users input seed cases to the system, the system will display relevant precedents (with similar issue attributes) and highlight the relevant federal laws.

3.   Trademark Judgement Discovery

Based on the TM litigation ontology constructed in Section 2, all TM litigation cases in the case corpus are automatically converted to frame-based, machine-readable knowledge representations. With the use of the system, critical features or values of instances and attributes (under 5 main classes of ontology schema) are retrieved and compared for discovering most relevant precedents. In order to verifying the system at work, fifty TM cases are randomly selected to test the case judgment recommender results. In our research, US TM litigation judgments from 1995 to 2018 are collected in the legal corpus using the search key-term "trademark infringement." We have found more than 4,800 TM infringement cases litigated in the US district and federal courts (1995~2018). Fifty cases are randomly selected as experimental judgments. The key features of these 50 judgments are retrieved for identifying similar case precedents.

3.1. Identify Precedents in Matching Issues and Their Key Terms

To identify precedents by relevant issues, issue key terms are defined and recognized before feature extraction. Table 1 lists ten TM litigation issues' key terms, their applicable rules, and the case examples. The relevant key terms are defined according to the attributes in the proposed TM ontology's "Issues" schema (in Fig. 2). The cases with frequently appeared terms, matching the key terms of specific issue(s) are listed in the last column of Table 1. The precedents, belonging to the same issue(s) of the seed case, are identified and ranked based on the matching key terms' frequencies. The system would finally identify and display the related precedents according to the related litigating issues. In addition, applicable rules are recommended according to the corresponding issues (as shown in the third column of Table 1).

Table 1. Issue key terms, applicable rules, and matching cases.

Issue	Key terms	Rules (Law)	Cases IDs
Trademark infringement	Trademark infringement, confusion, likelihood of confusion, consumer confusion, strength, similarity, valid, validity, protect, protectability, registration, Polaroid Factors	15 U.S.C. § 1114 15 U.S.C. § 1125 15 U.S.C. § 1051 15 U.S.C. § 1115(a)	All the cases
Trademark dilution	Dilution, diluted, distinctive, blurring, tarnishment, distinctiveness, famous, similarity, in commerce, recognition, actual association, secondary meaning	15 U.S.C. § 1125(a) 15 U.S.C. § 1125(c)	8, 13, 17, 20, 45
Trade dress infringement	**Trade dress infringement, trade dress, nonfunctional, distinctive, appearance, design, shape, strength, similarity**	**15 U.S.C. § 1125(a)**	**5, 12, 18, 28, 44**

Unfair competition	Unfair competition, deceptive trade practice, deceptive trade, bad faith, fraud, imitating, counterfeiting,	15 U.S.C. § 1125(a)	16, 19, 20, 27, 44
False designation of origin	False designation of origin, country of origin, false, mislead	15 U.S.C. § 1125(c)	13, 20, 26, 30, 34
False advertising/ endorsement/ association	False advertising, false endorsement, false association	15 U.S.C. § 1125(a)	13, 19, 30, 32, 41
Fair use	Fair use, fair use doctrine, good faith, affirmative defense, Nominative use, descriptive use	15 U.S.C. § 1115(b)	25, 29, 33, 40, 48, 50
Injunction	Irreparable harm, equities, balance of equities, public interest, preliminary injunction, permeant injunction, Injunctive relief	15 U.S.C. § 1116 15 U.S.C. § 1125(c) 15 U.S.C. § 1127	1, 14, 19, 24, 42
Standard of review	Summary judgement, motion to dismiss, dismiss	Lanham act is not applicable.	13, 14, 16, 20, 31
Recovery	Costs, fee, recover, remedies, profit, attorney fee, damages	15 U.S.C. § 1114 15 U.S.C. § 1117(a)	16, 21, 22, 27, 35

3.2. Identify Precedents in Maching Topic Models

LDA topic modeling is applied to identify the implicit semantic topics in the TM litigation cases. In order to ensure high coherent level of topic modeling result, Cv coherence score (Röder, 2015) is calculated to measure the topic coherence and performance. Cv measures the quality of topic model using normalized pointwise mutual information (NPMI) and the cosines similarity. A higher Cv score indicates better generalization performance. In our research experiment, total eight topics are generated using 4800 TM litigation cases as training dataset. Table 2 shows all topics' top frequently appeared terms, topic interpretation, and cases matching the topics and high-frequent terms. The LDA topics are further linked to the related issues as presented in the ontology schema in Section 2 (Fig. 2). With the trained topic models, the research identifies the dominant relevant issues and cases for precedents recommender. For example, topic 7 is closely related to the trade dress infringement issue. The cases (#5, 12, 18, 28, 44) match well with the feature extraction discovery in Table 1.

The approaches of identifying precedents based on issues' key terms (Sec. 3.1) and semantic topic modeling (Sec. 3.2) serve as valuable cross-validation for a reliable TM case precedent recommendation system. Through the analysis of the cases' similarities on features, dominant issues, and laws, the precedents can be reliably identified with legal insights and recommender's depth under the core structure of TM litigation knowledge.

Table 2. Topic key terms, interpretations, and example cases.

Topic	Key terms	Interpretation	Cases IDs
1	0.005*"descriptive" + 0.005*"strength" + 0.005*"generic" + 0.004*"secondary" + 0.003*"dilution" + 0.003*"brand" + 0.003*"secondary_meaning" + 0.003*"actual_confusion" + 0.002*"confused" + 0.002*"evidence_actual" + 0.002*"logo"	Relevant policies underlying Interpretation and definition of registered marks by the court.	6, 15, 35, 44, 50

2	0.010*"preliminary" + 0.007*"agreement" + 0.005*"irreparable" + 0.005*"preliminary_injunction" + 0.004*"franchise" + 0.003*"injunctive" + 0.003*"irreparable_harm" + 0.002*"license" + 0.002*"ownership" + 0.002*"motion_preliminary"	(Preliminary) injunction is an temporary order which prohibits the parties from doing an act before the final judgment is made.	1, 14, 19, 24, 42
3	0.012*"dilution" + 0.010*"distinctive" + 0.010*"international" + 0.008*"intellectual_property" + 0.007*"famous" + 0.006*"trademark_dilution" + 0.005*"ownership" + 0.003*"dilution_claim" + 0.003*"trade_name"	Regarding the distinctiveness of the famous mark, whether the act of trademark dilution is constituted.	5, 8, 13, 17, 20
4	0.011*"agreement" + 0.010*"contract" + 0.006*"dismiss" + 0.006*"counterclaim" + 0.005*"breach" + 0.004*"motion_dismiss" + 0.004*"amended" + 0.003*"fraud" + 0.003*"interference" + 0.003*"false_advertising" + 0.002*"pleading"	Unfair competition caused by disputes over trademark infringement.	7, 13, 22, 23, 32
5	0.010*"breach" + 0.008*"bad faith" + 0.008*"essential" + 0.008*"false_advertising" + 0.006*"" + 0.004*"essential" + 0.002*"event" + 0.002*"undisputed" + 0.002*"advertisements"	False advertising with bad faith to mislead and confused the trademark representation.	2, 10, 13, 19, 23
6	0.016*"fee" + 0.014*"award" + 0.012*"profit" + 0.008*"attorney_fee" +0.004*"willful" + 0.004*"counterfeit" + 0.004*"statutory" + 0.003*"verdict" + 0.003*"testimony" + 0.003*"appellant" + 0.003*"remedy"	Awards and remedies judgment decision made on the defendants.	15, 16, 19, 21, 27
7	**0.031*"trade_dress" + 0.009*"secondary" + 0.006*"secondary_meaning" + 0.006*"color" + 0.006*"feature" + 0.005*"functional" + 0.004*"packaging" + 0.003*"shape" + 0.003*"functionality" + 0.002*"distinctiveness" + 0.002*"dilution"**	**Trade dress infringement and dilution on the distinctives and strength of product appearance.**	**5, 12, 18, 28, 44**
8	0.020*"vendors" + 0.012*"contributory_trademark" + 0.010*"property" + 0.008*"contributory_liability" + 0.006*"infringing_activity" + 0.004*"merchandise" + 0.004*"contributory"	Contributory trademark infringement that vendors induces another to infringe other's trademark.	16, 21, 22, 38, 45

4.   Conclusion

The research aims to develop an intelligent recommendation system for discovering TM case precedents. The system integrates advanced technologies, such as knowledge ontology, regular expression, text mining, and soft clustering LDA algorithms to help users, such as trademark attorneys or relevant TM stakeholders, find the most relevant TM case precedents and provide valuable case insights, such as litigation issues, rules, and laws. The key contribution of the research is that the system combines the TM litigation key-feature extraction and judgment semantic topic recognition modules to cross-validate the discovery of precedents. The system provides the reliable results based on the similarities between legal precedents' issues, topics, and other pre-defined attributes/features. This research expects to offer a broad view of the intelligent legal analytic methodologies with the combined advanced text mining and semantic processing techniques.

References

[1]   J. Dabass, B. S. Dabass, Scope of Artificial Intelligence in Law, *preprints* 201806.0474.v1, 2018.
[2]   H. Surden, Machine Learning and Law, *Wash. L. Rev.*, 2014, pp. 89-87.

[3]   E. S. Rogers, The Lanham Act and the Social Function of Trade-Marks. *Law & Contemp. Probs.,* 14, 1949, pp.173.

[4]   USPTO, *Protecting Your Trademark*, United States Patent and Trademark Office, 2019.

[5]   H. L. Roitblat, A. Kershaw, P. Oot, Document categorization in legal electronic discovery: computer classification vs. manual review, *Journal of the American Society for Information Science and Technology*, vol. 61(1), 2010, pp. 70-80.

[6]   D. W. Oard, W. Webber, Information retrieval for e-discovery, *Foundations and Trends® in Information Retrieval*, vol. 7, 2013, pp. 99-237.

[7]   J. G. Conrad, E-Discovery revisited: the need for artificial intelligence beyond information retrieval, *Artificial Intelligence and Law*, vol. 18(4), 2010, pp. 321-345.

[8]   W. W. Belt, D. R. Kiker, D. E. Shetterly, Technology-Assisted Document Review: Is It Defensible?, *Rich. J.L. & TECH,* vol. 18, 2012.

[9]   S. K. Endo, Technological Opacity & Procedural Injustice, *B.C. L. Rev.*, 2018, 59: 281.

[10]   R. Calo, Robots as Legal Metaphors, *Harvard Journal of Law and Technology*, vol. 30(1), 2016, pp. 209- 237.

[11]   M. Hildebrandt, The Meaning and the Mining of Legal Texts, In: D.M. Berry (eds.) *Understanding Digital Humanities*, pp.145-160, 2012.

[12]   D. L. Burk, Algorithmic Fair Use, U. Chi. L. Rev., 2019, 86: 283.

[13]   T. R. Gruber, A translation approach to portable ontology specifications, *Knowledge acquisition*, vol. 5(2), 1993, pp.199-220.

[14]   R. Studer, V. R. Benjamins, D. Fensel, Knowledge engineering: principles and methods, *Data & knowledge engineering*, vol. 25(1-2), 1998, pp. 161-197.

[15]   B. Chandrasekaran, J. R. Josephson, V. R. Benjamins, What Are Ontologies, and Why Do We Need Them, *IEEE Intelligent Systems*, vol. 14 (1), 1999, pp. 20-26.

[16]   N. Guarino, D. Oberle, S. Staab, What Is an Ontology?, In: S. Staab et al. (eds.) *Handbook on Ontologies*, pp. 1-17, 2010.

[17]   S. Shishehchi, S. Y. Banihashem, N. A. Zin, S. A. Noah, S. A., K. Malaysia, Ontological approach in knowledge based recommender system to develop the quality of e-learning system, *Australian Journal of Basic and Applied Sciences*, vol. 6(2), 2012, pp. 115-123.

[18]   C. M. Cruz, C. Porcel, J. B. Moreno, E. H. Viedma, A model to represent users trust in recommender systems using ontologies and fuzzy linguistic modeling, *Information Sciences*, vol. 311, 2015, pp.102-118.

[19]   K. Munir, M. S. Anjum, The use of ontologies for effective knowledge modelling and information retrieval, *Applied Computing and Informatics*, vol. 14(2), 2018, pp. 116-126.

[20]   C. V. Trappey, A. C. Chang, A. J. C. Trappey, Building an internet based knowledge ontology for trademark protection, *Journal of Global Information Management (JGIM)*, 2019.

[21]   S. C. Kleene, Representation of events in nerve nets and finite automata, In: C. Shannon et al. (eds.) *Automata Studies*, pp. 3-41, 1956.

[22]   J. E. Friedl, Mastering regular expressions, *O'Reilly Media*, California, 2006.

[23]   L. Wheeler, *The Art of Regular Expressions*, CreateSpace Independent Publishing Platform, South Carolina, 2016.

[24]   S. Qaiser, R. Ali, Text mining: use of TF-IDF to examine the relevance of words to documents, *International Journal of Computer Applications*, vol. 181(1), 2018, pp. 25-29.

[25]   S. Robertson, Understanding inverse document frequency: on theoretical arguments for IDF, *Journal of documentation*, vol. 60(5), 2004, pp. 503-520.

[26]   D.M. Blei, A.Y. Ng, M.I. Jordan, Latent dirichlet allocation, *Journal of machine Learning research*, vol. 3, 2003, pp. 993-1022.

[27]   M. Röder et al., Exploring the space of topic coherence measures, In: *Proceedings of the eighth ACM international conference on Web search and data mining*, New York, 2015, pp. 399-408.