Advanced Engineering Informatics 29 (2015) 787–798

Contents lists available at ScienceDirect

Advanced Engineering Informatics

journal homepage: www.elsevier.com/locate/aei

# Collective intelligence applied to legal e-discovery: A ten-year case study of Australia franchise and trademark litigation

Charles V. Trappey [a,b], Amy J.C. Trappey [c,*]

[a] National Chiao Tung University, Taiwan
[b] Queensland University of Technology, Australia
[c] National Tsing Hua University, Taiwan

ARTICLE INFO

Article history:
Received 13 December 2014
Received in revised form 24 April 2015
Accepted 29 April 2015
Available online 12 June 2015

Keywords:
Collective litigation intelligence
Legal e-discovery
Retail franchise
Trademark infringement

ABSTRACT

The purpose of this research is to develop a formal knowledge e-discovery methodology, using advanced information technology and decision support analysis, to define legal case evolution based on Collective Litigation Intelligence (CLI). In this research, a decade of Australia's retail franchise and trademark litigation cases are used as the corpus to analyze and synthesize the evolution of modern retail franchise law in Australia. The formal processes used in the legal e-discovery research include a LexisNexis search strategy to collect legal documents, text mining to find key concepts and their representing key phrases in the documents, clustering algorithms to associate the legal cases into groups, and concept lattice analysis to trace the evolutionary trends of the main groups. The case analysis discovers the fundamental issues for retail modernization, advantages and disadvantages of retail franchising systems, and the potential litigation hazards to be avoided in the Australian market. Given the growing number of legal documents in global court systems, this research provides a systematic and generalized CLI methodology to improve the efficiency and efficacy of research across international legal systems. In the context of the case study, the results demonstrate the critical importance of quickly processing and interpreting existing legal knowledge using the CLI approach. For example, a brand management company, which purchases a successful franchise in one market is under limited time constraints to evaluate the legal environment across global markets of interest. The proposed CLI methodology can be applied to derive market entry strategies to secure growth and brand expansion of a global franchise.



## 1. Introduction

This research focuses on developing a methodology for legal e-discovery and, most importantly, litigation evolution analysis using collective intelligence of precedential documents, court cases, and data existing in legal databases. The specific techniques of text mining, data mining, cluster analysis, and formal concept analysis are modified and applied as the Collective Litigation Intelligence (CLI) methodology for legal e-discovery. The case study of the Australia retail franchise and trademark litigation is used to demonstrate the effectiveness of the generalizable CLI processes. The last decade of legal cases (2004–2013), related to retail franchises and trademarks, are searched as the fact-base to derive collective litigation intelligence to project the litigation trends and evaluate hazards underlying the retail legal evolution. These franchise and trademark litigation cases and selected documents are used to demonstrate the significant findings of the CLI approach and the validity of the methodology. The paper strategically advises franchise trademark holders of market strategies for sustainable market development that avoids past mistakes and identifies new opportunities. The research results enables managers and legal advisors to answer critical questions such as "Has legislation stabilized the franchise environment?" and "What are the trademark, legal, and franchise hazards that restrict market development?" The research clusters the last decade of 35 precedent setting franchise and trademark cases into four homogenous groups. Briefly speaking, the clustering analysis yields four groups of cases with distinctive characteristics and features. The CLI's final step is to derive the time-varied evolution trends underlying the development of the franchise and trademark legal landscape. Applying the developed CLI processes, Clusters 1, 2, and 3 are considered bellwether case clusters whereas Cluster 4 supports legislative stability and a market place of opportunity rather than a legal hazard. Franchisers and brand managers interested in the Australian market should focus on the case clusters to avoid franchise and trademark litigation and to develop opportunities that

* Corresponding author.
E-mail addresses: trappey@faculty.nctu.edu.tw (C.V. Trappey), trappey@ie.nthu.edu.tw (A.J.C. Trappey).

http://dx.doi.org/10.1016/j.aei.2015.04.006
1474-0346/

create sustainable market plans. The methodology and the case research can be rapidly repeated in other legal domains for their litigation e-discovery.

This paper is organized using the following sections. In the literature review section, the key concepts and related research of ontology, knowledge discovery, data and text mining, and formal concept analysis are reviewed, cited, and described. We refer to these papers for the readers' further reference and study when conducting following up research for advanced theory development and applications. The "collective litigation intelligence (CLI) methodology" section describes the CLI methodology framework and logic for each detailed step. The CLI procedure is best demonstrated and described using an applied legal case study. Thus, the step-wise sub-sections including the document search strategy, the case text mining, the CLI clustering and context interpretations, and legal concept evolution lattice are presented using the Australian franchise example. Finally, the conclusion section discusses the research results, contributions, and suggested future work for a generalizable CLI methodology applicable for legal e-discovery beyond the franchise and trademark case domain.

## 2. Literature review

This research focuses on using knowledge discovery algorithms and methods to effectively extract litigation trends and insights from a huge collection of text documents. The key knowledge e-discovery techniques specifically applicable for CLI methodology development include the capturing of key entities in the domain ontology, knowledge discovery from text documents (KDT) and from databases (KDD), document clustering, and formal concept lattice analysis. From these critical components of knowledge discovery, well-defined domain entities (such as key phrases of given documents), being the essence of the ontology schema, enable the documents being machine readable, inter-operable, and auto-interpretable in the corpus [17]. Although building a comprehensive domain-specific ontology is not the focus of this research (but is a future research direction), ontology engineering, applied to define domain concepts and scopes using a set of key entities, is adopted as the principal research step [9,11].

Knowledge discovery is a process used to extract implicit, previously unknown, and potentially useful knowledge from known data which are relevant and meaningful [8]. Depending on the data type, knowledge discovery divides knowledge into two categories: knowledge discovered from a database (KDD or data mining) and knowledge derived from a text document base (KDT or text mining). KDD is the process used to automatically discover previously unknown patterns, rules, and other types of content in large volumes of data [6,5]. The steps of KDD consists of identifying the analytical objectives, creating the target data set, cleaning and preprocessing data, data reduction or projection, using data mining techniques or algorithms to search the patterns of data, and interpreting the patterns [6].

Text mining is commonly known as knowledge discovery when analyzing documents or text [7]. The framework of text mining consists of text refinement and knowledge distillation [18]. Term frequency-inverse document frequency (TF-IDF) is a statistical method using the frequency of word occurrence in text to reflect the importance of words or phrases and representing key concepts in a given document set [15]. Salton and Buckley [14] reported that the length of a document can affect the term weight. Therefore, TF-IDF was modified and called normalized term frequency-inverse document frequency (NTF-IDF). The number of words in a set of documents is used to normalize the value of term frequency. For instance, in patent analysis, text mining techniques including text segmentation, summary extraction, feature selection, term association, cluster generation, topic identification and information mapping are commonly applied [10], [25]. Trappey et al. [23] combined the techniques of ontology-based text mining and data mining to identify patent sub-technologies.

Formal concept analysis (FCA) was first proposed by Wille [27] based on lattice theory [1]. FCA is used as a method to derive a concept hierarchy from a collection of objects and their attributes. FCA uses formal context and formal concepts, as represented in Table 1 and can be expressed as a triple set {O, A, R} where O is a set of objects, A is a set of attributes, and R is the set of relations between O and A. The left side of the Table 1 is the set of objects ($O_i$, $i = 1, \ldots, 5$). The top row of Table 1 lists the set of attributes ($\mathbf{A} \in \{a_j, j = 1, \ldots 4\}$). If an object ($O_i$) consists of a given attribute ($a_j$), the relation between them will be marked ($\times$) in the matrix. Each concept in the hierarchy represents the set of objects sharing the same values for a certain set of attributes. Each sub-concept in the hierarchy contains a subset of the objects in the concepts above it.

The concept lattice is built based on the formal contexts used to define super-concepts and sub-concepts from general to specific levels of relationships. Suppose that an object $O_1 \subseteq O_2$ with common attributes $A_1 \subseteq A_2$. This means the concept ($O_1$, $A_1$) is a sub-concept of the concept ($O_2$, $A_2$) while the concept ($O_2$, $A_2$) is a super-concept of the concept ($O_1$, $A_1$). The concept lattice is drawn as a top-down graph depicting concepts from general to specific relationships. The top and bottom of the concept lattice describes supremum and infimum relationships denoted as maximum sub-concepts and minimum sub-concepts. In Fig. 1, the concepts are derived from the formal context relations of Table 1. For instance, a super-concept is composed of object $O_1$ and has a set of two attributes {$a_1$, $a_4$}. After adding an attribute $a_3$, a narrowed-down (specific) sub-concept $O_4$ is defined with a set of attributes {$a_1$, $a_3$, $a_4$}. The objects, belonging to narrower concept(s), are continuously identified with attributes added into the lattice. Finally, as shown in Fig. 1, the most specific concept has all attributes but is so narrow that it has a null set of objects.

**Table 1**
The formal context and its concept matrix example.

| | | Attributes | | | |
|---|---|---|---|---|---|
| | | $a_1$ | $a_2$ | $a_3$ | $a_4$ |
| Objects | $O_1$ | × | | | × |
| | $O_2$ | | × | | |
| | $O_3$ | × | | × | |
| | $O_4$ | × | | × | × |
| | $O_5$ | | × | × | × |

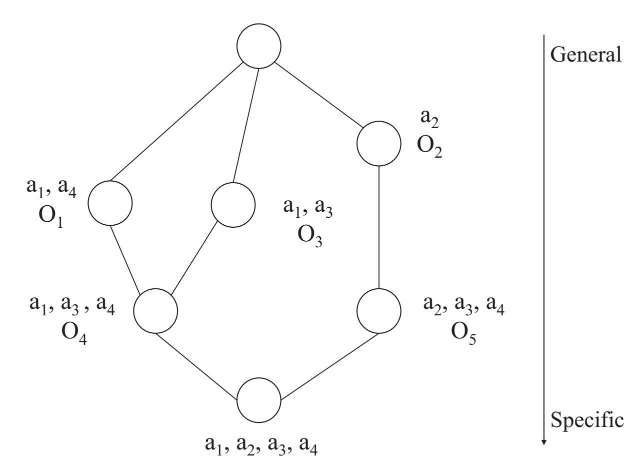


**Fig. 1.** The concept lattice derived from a formal context in Table 1.