

# Using GHSOM to construct legal maps for Taiwan's securities and futures markets

Jen-Ying Shih <sup>a,\*</sup>, Yu-Jung Chang <sup>b</sup>, Wun-Hwa Chen <sup>c</sup>

<sup>a</sup> Department of Business Administration, Chang Gung University, 259, Wen-Hwa 1st Road, Taoyuang 333, Taiwan, ROC

<sup>b</sup> Institute of Information Science, Academia Sinica, Taipei, Taiwan

<sup>c</sup> Graduate Institute of Business Administration, National Taiwan University, Taipei, Taiwan, ROC

## Abstract

A good legal knowledge representation system, capable of effectively providing investors with comprehensive legal knowledge, is needed for investors to prevent erratic behavior before investment decisions. This is especially important in Taiwan's securities and futures markets because the majority of market participants are individual investors who have limited access to legal knowledge about markets. Besides, the construction of the knowledge representation has to be automatic in order to efficiently handle the fast-growing and changeable legal information. Thus, we use the GHSOM algorithm to present a content-based and easy-to-use map hierarchy for Chinese legal documents in the securities and futures markets in the Chinese language. Meanwhile, an enhanced topic selection module and a web-based user interface are also proposed. The maps can be browsed on the web site (<http://synteny.iis.sinica.edu.tw/legalmap/>). To evaluate the legal maps, we apply two approaches, namely a validity test and a task experiment.

© 2006 Elsevier Ltd. All rights reserved.

*Keywords:* Legal informatics; Knowledge representation; Self-organizing maps; Securities markets

## 1. Introduction

In legal informatics, successful searching of huge legal digital libraries depends a great deal on the user's ability to master the legal terminology (Schweighofer, Rauber, & Dittenbach, 2001). Due to cost considerations, traditional approaches to legal knowledge representation based on thesauri and classification by professional persons have gradually developed into semi-automatic or automatic approaches. Thus, how to automatically categorize, index, organize, present, and summarize the enormous amounts of legal documents to enable quick and efficient retrieval of accurate legal information is a key issue (Merkl & Schweighofer, 1997; Schweighofer et al., 2001; Schweighofer, Winiwarter, & Merkl, 1995; Thompson, 2001).

Most previous research in legal informatics has focused on Western languages – seldom on oriental languages, such as Chinese, Japanese, or Korean. In this paper, we study the effects of applying legal informatics to legal documents written in one of the above oriental languages by using an unsupervised learning algorithm called the growing hierarchical self-organizing map (GHSOM). As this algorithm has a proven performance record in Western language legal knowledge systems, we apply it to the construction of Chinese legal maps.

According to Schweighofer et al. (2001), it is better to segment the topics in legal documents when presenting legal knowledge. We, therefore, selected a very important topic, namely, legal knowledge of Taiwan's securities and futures markets, as our test environment, because – according to statistics published by the ROC government<sup>1</sup> – the

\* Corresponding author. Tel.: +886 3 2118800x5410; fax: +886 2 27089989.

E-mail address: [jsyshiv@mail.cgu.edu.tw](mailto:jsyshiv@mail.cgu.edu.tw) (J.-Y. Shih).

<sup>1</sup> According to Major Indicators of the Securities and Futures Markets, Taiwan District, ROC, in 2005, the trading value of domestic individual investors is up to 74.6% of the total trading value, and 2.94 times larger than that of institutional investors.

majority of participants in these markets are individual investors. In contrast, institutional investors are the key players in the American and European markets. To protect individual investors, who are not as sophisticated as their institutional counterparts in gathering market information, Taiwan's securities and futures markets are highly regulated by several government agencies and the markets' self-regulating bodies, including the Securities and Futures Bureau (SFB), the Taiwan Securities Exchange (TSE), the ROC. Over-the-counter Securities Exchange (OTC), and Taiwan Futures Exchange (Taifex). However, as most investors are not familiar with the large number of laws, rules, and regulations promulgated by the competent authorities, it is hard for them to understand the relationships among the different types of legal information. As a result, they sometimes violate laws, and inadvertently commit crimes.

Currently, legal information retrieval systems used in Taiwan's markets emphasize key word search functions, but the search results are not presented in a meaningful way. As most market participants are not experienced in retrieving legal information, the retrieval procedure should be uncomplicated and user friendly so that legal knowledge about Taiwan's securities and futures markets can be accessed easily. Hence, the goal of this paper is to provide helpful and cost-effective legal guidance for market partic-

ipants, including institutional and individual investors, employees of public companies, and securities and futures service providers.

The remainder of this paper is organized as follows. In Section 2, we introduce related research about legal knowledge retrieval. In Section 3, we review the self-organizing map (SOM) related literature and point out some drawbacks of applying growing hierarchical SOM (GHSOM) to legal maps. We then describe how to construct legal maps for Taiwan's securities and futures markets in Section 4, and present the legal maps in Section 5. In Section 6, we evaluate GHSOM by a validity test and a task experiment. Finally, in Section 7, we present our conclusions.

## 2. Literature review of legal knowledge systems

In recent years, many researchers have focused on the area of legal knowledge presentation. Their methods can be categorized as either supervised indexing approaches, or unsupervised learning models. The former use thesauri, ontology, or other classifications for a simple form of conceptual search. Some researchers apply knowledge-based models composed from thesauri and rules, such as the FLEXICON project (Smith et al., 1995). Case-based reasoning (CBR) plays an important role in developing some legal expert systems. In the SPIRE project, it is applied

Table 1  
Summary of related research

Research	Methodology	Application
Flexicon project (Smith et al., 1995)	Knowledge base	A term extraction module that recognizes concepts, case citations, statute citations, and fact phrases resulting in a very helpful structured document profile. This profile is transformed into a weighted vector to enable a search for related documents or problems
SPIRE project (Daniels & Rissland, 1995)	CBR	Inference net presentations of important cases are used to search for similar documents
Salomon project (Moens, 1999)	CBR	Linguistic representations capable of abstracting legal documents automatically
Anandanpillai and Barta (1999)	CBR	The system, composed of the case knowledge base, pattern detection mechanism, case selection mechanism, adaptation mechanism and weight adjustment mechanism, is developed for housing discrimination law
ASHSD-III (Pal, 1999)	RBR & CBR	The goal of this research is to formulate a model by which one can find the appropriate precedents in an automated system. The ASHSD-III prototype has achieved the goal by using a model of case similarity assessment based on fuzzy proximity relations
BanXupport (Elhadi, 2000)	IR and CBR	The Statutes-Based Case Storage and Retrieval System (BanXupport) is a prototype system of case-based indexing and retrieval that is designed to represent and retrieve cases in bankruptcy law. It can assist lawyers in bankruptcy law and similar fields in doing their law research and reasoning using previously decided cases to solve new ones
KONTERM (Merkl et al., 1995)	SOM	Kobonen's self-organization feature map is used to represent 41 neutrally related text documents extracted from documents contained in the European Community law database, CELEX. The database prepares document weight vectors in the [0, 1] format, which represents every term (feature) in a document, whether or not such terms actually appear
KONTERM (Merkl & Schweighofer, 1997)	GSOM	The researchers present the results of a case study in legal document classification based on an experimental document archive comprised of 100 important treaties in public international law (5 MB of text). They use a layered architecture consisting of mutually independent unsupervised, neural networks. The length of each individual feature vector is equal to 1625 components
Schweighofer et al. (2001)	GHSOM & LabelSOM	Schweighofer et al. (2001) show the feasibility of using GHSOM and LabelSOM techniques in legal research by tests with text corpora in European case law. The results show the generalities and specialties of legal text corpora. The segmentation of a document improves the quality of labeling significantly. The next challenge will be a change from $tf \times idf$ vector representation to a modified vector representation that takes into account thesauri or ontologies considering learned properties of legal text corpora

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات