

中文摘要

在人類經濟邁入以知識為主軸的知識經濟之際，為了提升產業競爭力，世界各國不斷努力於產業升級與轉型，而企業競爭的優勢便在於其知識的品質，重視創意領先和科技研發。對企業而言，專利資訊不只為人類的智慧寶庫，也是研究開發人員重要的參考資料。企業所重視的是如何在茫茫大海般的專利文件中，將專利資料轉換為企業所需之有效資訊與情報。另外，由於專利資訊亦揭露專利侵權的警訊，智權管理人員藉由專利資訊隨時監控競爭對手的專利核准動向，以降低企業因侵權所應付出的龐大成本。另一方面，企業亦可進行專利部署，以專利作為擴散武器，增加市場佔有率或是進行策略性授權、交互授權、專利聯盟、技術轉移等。在本研究中，提出了一個以本體論為基之智慧型專利文件分類系統。本研究方法論的步驟如下：首先，本研究利用解析 Web Ontology Language(OWL) 文件，來取得領域知識的本體論；接著，藉由 Term Frequency - Inverse Document Frequency (TF-IDF) 為基之技術來擷取出專利文件中重要的關鍵詞彙，並以擷取出的關鍵詞彙為基礎，計算該關鍵詞彙所隱含本體論概念的機率。再者，將本體論與類神經網路結合，運用分類文件中關鍵詞彙出現的頻率與隱含本體論概念的機率及本體論關係的計算來進行專利文件的自動分類。此外，本研究還包含了專利文件的搜尋模組，來加強分類後文件的分析與使用。而本研究還提出修正回饋的機制，藉由更新詞彙機率及類神經網路的學習過程來增進分類的準確率。最後，本研究以化學機械研磨 (CMP) 領域和無線射頻識別 (RFID) 領域的專利文件為案例來測試自動分類系統之成效。

關鍵字：本體論、關鍵詞彙、文件分類、類神經網路、TF-IDF

Abstract

In order to stimulate novel ideas and avoid patent infringement during new product development, R&D engineers need to obtain existing patent information related to the development domain accurately and in a timely matter. Further, patent documents if we organized and categorized, can provide IP managers with a clean view of the state-of-the-art technologies in an efficient and effective way. Equipped with IP knowledge, companies can set R&D directions and develop patent portfolio and territory strategy to stay competitive in the global market place. This thesis proposes a patent categorization methodology by using Artificial Neural Network (ANN) to classify patent documents based on pre-constructed ontology. The proposed methodology not only recognizes Web Ontology Language (OWL) created by protégé but also acquire probabilities which key phrases belong to the specific concepts in domain ontology. The procedure of the proposed methodology, firstly, extracts key phrases from documents based on Term Frequency - Inverse Document Frequency (TF-IDF) method, and then summarizes a probability matrix between key phrases and concepts to calculate the probability that a specific key phrase contains a certain concept. Because combining frequencies and probabilities of key phrases, this study can get better representative input values for ANN model. In addition, this research provides document searching module by selecting key phrases and setting weights to execute IP document analysis. Finally, this research uses patents of Chemical Mechanical Polishing (CMP) and Radio Frequency Identification (RFID) domains as case examples to illustrate and demonstrate the proposed methodology at work with superior results.

Keywords: Ontology, Key Phrases, Document Categorization, Neural Network, TF-IDF