

The Construction of an Intelligent Campus Electronic Documents Retrieval System

魏召欣、陳鴻文

E-mail: 9222623@mail.dyu.edu.tw

ABSTRACT

Due to the large amount of electronic documents in enterprise, text retrieval systems become necessary. The study will utilize the technique of full-text document retrieval, and try to automatically analyze the contents of electronic documents from the viewpoints of five attributes of words such as 「how」, 「what」, 「when」, 「where」 and 「who」. With a domain knowledge database, the proposed system can make the meaning of users' queries more complete and clear by taking synonyms and five attributes of words into consideration. Under the scheme, the relevant information of documents will be automatically abstracted and listed as the retrieval reports of the system. An interface of simple question-and-answer is thus provided for users to quickly find out the unknown attribute within a query through documents. The proposed document retrieval system is expected to promote the efficiency of information communication and enterprise knowledge management.

Keywords : text retrieval, retrieval system, knowledge management

Table of Contents

第一章 緒論	1	第一節 研究背景與動機	1	第二節 研究目的	4	第三節 研究範圍與限制	5	第四節 研究內容與研究流程	6
第二章 文獻探討	8	第一節 搜尋引擎	8	第二節 文件檢索系統的探討	10	第三節 檢索詞彙的相關研究	13	第四節 判斷文句的重要性	15
第五節 中文斷詞	16	第六節 語意網	17	第三章 文件檢索方法與系統設計	19	第一節 公文文件特性的觀察與探討	19	第二節 資訊處理的方法	23
第三節 系統設計	31	第四節 預期成果	41	第四章 系統實作與效能	42	第一節 系統開發工具與環境	42	第二節 系統人機介面之設計	42
第三節 公文檢索系統實驗	44	第四節 實驗結果與分析	56	第五章 結論	60	第一節 結論	60	第二節 後續研究建議	61
參考文獻	64	附錄A	67	附錄B	68				

REFERENCES

1. 卜小蝶(民89), 台灣網路使用者檢索行為探析, 大學圖書館第四卷第二期, 頁23-37.
2. 王秀卿(民90), 網路使用與資訊尋求行為之文獻探討, 大學圖書館第五卷第一期, 頁144-162.
3. 吳文峰(民91), 中文郵件分類器之設計及實作, 逢甲大學資訊工程學系碩士班碩士論文.
4. 李建興等(民91), 基於Ontology之中文文件自動摘要技術之研究, 第三屆產業資訊管理學術暨新興科技實務研討會論文集(下集), 頁870-876.
5. 林隆祺(民89), 運用字詞位置的文件索引技術初探, 台灣大學資訊管理學研究所碩士論文.
6. 施建新(民89), 網路資訊學習系統之中文全文探勘工具, 中央大學資訊工程研究所碩士論文.
7. 唐大任(民91), 中文斷詞器之研究, 交通大學電信工程學系碩士班碩士論文.
8. 翁鴻加(民90), 多文件摘要一些新技術及評估模型之建立, 台灣大學資訊工程學研究所碩士論文.
9. 高虹(民91), 電腦也能看懂語意網, 科學人雜誌第六期, 頁47-56.
10. 高嘉祺(民89), 線上圖文購物引擎-以手機應用為例, 大葉大學資訊管理所碩士論文.
11. 陳佳君(民87), 檢索詞彙來源探析, 大學圖書館第二卷第三期.
12. 曾志軒(民90), 中文結構化文件之語意索引, 交通大學資訊科學研究所碩士論文.
13. 黃雲龍(民87), 中文全文資訊檢索研究架構與重要議題探討, 大學圖書館第二卷第三期.
14. 葉政輝(民91), 以語料為基礎的中文專有名詞分類之研究, 交通大學資訊科學研究所碩士論文.
15. 葉鎮源(民91), 文件自動化摘要方法之研究及其在中文文件的應用, 交通大學資訊科學研究所碩士論文.
16. Asuncion Gomez-Perez and Oscar Corcho (2002), "Ontology Languages for the Semantic Web", IEEE INTELLIGENT SYSTEMS, pp:54-60.
17. J. Bosak and T. Bray (1999), "XML and the Second-Generation Web", Scientific American.
18. Natalya F. Noy, Deborah L. McHuinness (2001), "Ontology Development 101: A guide to Creating Your First Ontology", Stanford University.
19. Nicholas J. Belkin, W. Bruce Croft (1992), "Information filtering and information retrieval: two sides of the same coin?", Communications of the ACM, Vol.35, NO.12., pp:29-38.
20. Saracevic T., Kantor P. (1988), "A Study of Information Seeking and Retrieving. III. Searchers, Searches and Overlap", Journal Of The American Society For Information Science, Vol. 39, Iss. 3, pp:214.
21. Tim Berners-Lee(1999), "Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor", Mark Harper San Francisco.
22. Yihong G., Xin L. (2001), "Generic text summarization using relevance measure and latent semantic analysis", Annual international ACM SIGIR conference on Research and development in information retrieval, New Orleans, Louisiana, United States, pp:19-25.
23. <http://www.keenage.com>
24. <http://www.w3c.org/>
25. <http://www.w3.org/2001/sw>
26. <http://www.SemanticWeb.org/knowmarkup.html>
27. <http://www.cs.umd.edu/projects/SOHE/faq.html>
28. <http://www.daml.org/>
29. <http://survey.yam.com/survey2000/>
- 30.

