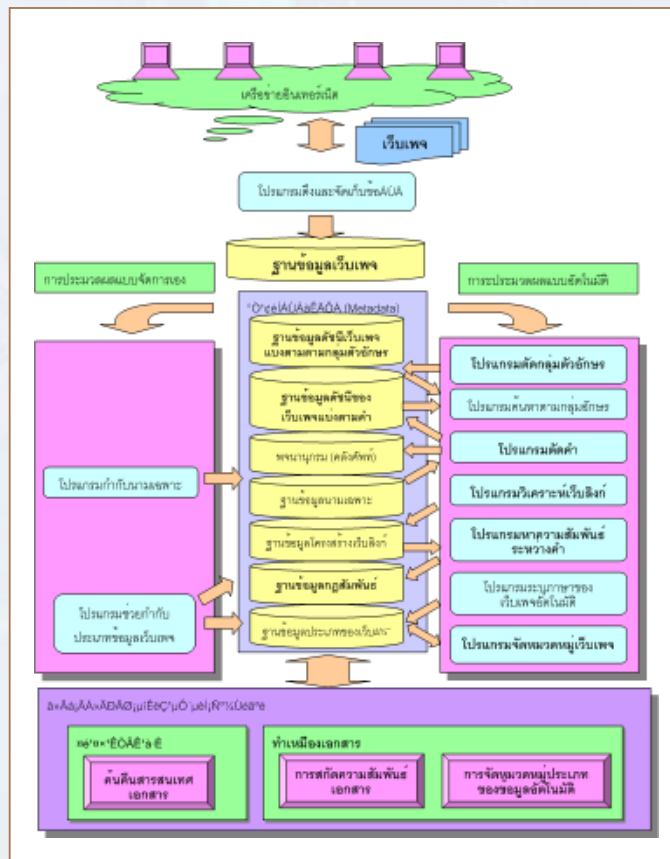


โครงการวิจัยพัฒนาศูนย์ทรัพยากรสำหรับการจัดการข้อมูลขนาดใหญ่บนเครือข่ายอินเทอร์เน็ต การสืบค้นข้อมูลและการขุดหาข้อมูล

Research and Development of Resources for Processing Very Large-Scaled Information on the Internet/Information Retrieval and Data Mining



สรุป

โครงการวิจัยพัฒนาศูนย์ทรัพยากรสำหรับการจัดการข้อมูลขนาดใหญ่บนเครือข่ายอินเทอร์เน็ต ได้ดำเนินการต่อเนื่อง 3 ปี โดยได้กำกับหมวดหมู่ของเอกสารได้มากกว่า 60,000 เอกสาร รวมทั้งนามเฉพาะภาษาไทยจำนวน 8,000 คำ และได้พัฒนาโปรแกรมช่วยในการจัดการข้อมูลบนเครือข่ายอินเทอร์เน็ตทั้งภาษาไทยและภาษาอังกฤษ ได้พัฒนาระบบค้นคืนสารสนเทศที่ใช้กับภาษาไทย ระบบทำเหมืองข้อมูลเพื่อจัดหมวดหมู่เอกสารและระบุภาษาของเอกสาร และพัฒนาระบบหาความสัมพันธ์ระหว่างเอกสาร ผลงานของโครงการนี้ได้รับการยอมรับและตีพิมพ์ในวารสารต่างประเทศจำนวน 8 ผลงาน และอื่นๆ อีก 18 ผลงาน และพัฒนานักวิจัยออกไปจำนวน 6 คน ทางโครงการได้จัดทำเว็บไซต์เพื่อเผยแพร่ข้อมูลโครงการที่ (URL) kind.siiit.tu.ac.th/irdm ซึ่งรวมไปถึงการเผยแพร่เนื้อหาวิชาการที่ได้จากงานวิจัยนี้

เอกสารอ้างอิง

- Lertnattee, Verayuth and Theeramunkong, Thanaruk, *Multi-dimensional Text Classification for Drug Information*, IEEE Transactions on Information Technology in Biomedicine, Vol. 8, No. 3, pp. 306-312, IEEE Computer Society Press, September 2004
- Lertnattee, Verayuth and Theeramunkong, Thanaruk, *Effect of Term Distributions on Centroid-based Text Categorization*, Information Sciences, Vol. 158, No. 1, pp. 89-115, Elsevier Science, January 2004
- Theeramunkong, Thanaruk, *Applying Passage in Web Text Mining*, International Journal of Intelligent Systems, Vol. 19, No. 1-2, pp. 149-158, Wiley Periodicals, January 2004
- Sripaew Kritsada and Theeramunkong, Thanaruk, *Fast Algorithms for Mining Generalized Frequent Patterns of Generalized Association Rules*, IEICE Transaction on Information and Systems, Vol. E87-D, No. 3, pp. 761-770, The Institute of Electronics, Information and Communication Engineers (IEICE), March 2004
- Theeramunkong, Thanaruk and Tanthemsong, Thanassan, *Pattern-Based Features vs. Statistical-Based Features in Decision Trees for Word Segmentation*, IEICE Transaction on Information and Systems, Vol. E87-D, No. 5, pp. 1254-1260, The Institute of Electronics, Information and Communication Engineers (IEICE), May 2004
- Theeramunkong, Thanaruk and Tanthemsong, Thanassan, *Pattern-Based Features vs. Statistical-Based Features in Decision Trees for Word Segmentation*, IEICE Transaction on Information and Systems, Vol. E87-D, No. 5, pp. 1254-1260, The Institute of Electronics, Information and Communication Engineers (IEICE), May 2004
- Nuanwan Soonthornphaisal, Boonserm Kijskiruk *Iterative cross-training: An algorithm for learning from unlabeled Web pages*, Int. J. Intell. Syst. 19(1-2): 131-147 (2004)
- Nuanwan Soonthornphaisal, Boonserm Kijskiruk *Iterative cross-training: An algorithm for web page categorization*, Intell. Data Anal. 7(3): 233-253 (2003)
- Lertnattee, Verayuth and Theeramunkong, Thanaruk, *Analysis of Inverse Class Frequency in Centroid-based Text Classification*, In Proceedings of International Symposium on Communication and Information Technology (ISCIT-2004), 26-29 October 2004, Sapporo, Japan.