Preface

OVERVIEW

Research into information retrieval has strong impact on both the social and economic world (Debons, 1971). Information Retrieval – IR involves a wide range of applications including social science, science, and engineering (Lu, 2005; Wang & Lu, 2009). The improvement in both methodology and technology in IR is tremendously demanding. The areas like scientific experiments, engineering productions, business processes, et cetera, could generate huge amount of data daily, hourly, or per second, which could lead to a huge problem in efficient search and retrieval or secure storage, for the scientists, engineers, doctors, librarian, managers, et cetera in the real world. In recent years, a number of emerging technologies significantly contribute to the IR, together with fundamental IR theories and concepts, which are adopted into new tools or systems in target applications, such as XML technology, content based image retrieval, and business processes management.

XML technology can be counted as one of key contributors to the modern IR technology because it operates based on meaning not like other computing languages, which are based on notation or human-unreadable binary code. Using XML, people can instruct computer to search information semantically within language itself. However, XML is not perfect. Typical examples are XML files look like a large amount of text, which may cause problem to reduce retrieval efficiency, waste storage space with redundant data, and lower the extent of security, particularly data integrity. Thus, research into XML data compression is of important in both users and researchers. Also, XML is strong in data representation and exchanging in database management systems. The native XML database or industrial supported XML documents are popular in the contemporary computing in both applications and development environments (Lu & Arabnia, 2009). Relational database – RDB - is simple and easy to manipulate database management systems, and still very popular for a large number of database users among the database markets. Thus, automatic mapping XML into RDB system could be useful to enable XML DB system can be deployed by RDB users who has no XML background.

Image retrieval is another important area in IR as image uses a special way to deliver the information to human beings. Image information is normally stored in a huge repository and retrieved when needed. Applications are involved in a wide range of areas, such as medical diagnoses, personal collections, space agencies, and geographic information from general to specific systems. Content Based Image Retrieval – CBIR - is one of emerging technologies in image retrieval areas (Pein & Lu, 2010). A typical application of CBIR is multimedia publishing and design. Query languages and learning algorithms extract new features based on low level information could be beneficial from semantic search and satisfactory of both machine and human understanding.

Traditional IR was not often extended into business process models. In fact, information fusion in the business world cannot be ignored. Life cycle for Business Process Management – BPM, marketing analysis, process design, sales reports, et cetera, may easily produce billions of data sets, documents, or even images in today's digital world. Design, implementation, and optimization of BPM models are crucial for the success of organizational business. Therefore, BPM becomes an important component to join the areas of IR research.

THE TARGET AUDIENCE

Immediate audiences for this book are from the area of information retrieval around the world. The book targets at the readers who are interested in the latest theories, methods, technologies and tools for IR in interdisciplinary and multidisciplinary research and applications; researchers who are working in higher education, industrial companies, and professional bodies can be also benefited from the book; professors, lecturers, and teachers from a wide range of subject areas can be benefited from the book if they are interested in IR. The book can be an inspiration for research initiatives and reading materials for educationists and students, and for a library collection for this fast developing subject area with emerging cutting age technologies.

THE IMPORTANCE OF EACH OF THE SECTIONS

This book is organized in 5 sections. Section 1 has 5 chapters. Section 2 has 6 chapters. Section 3 has 7 chapters. Section 4 has 6 chapters. Section 5 has 6 chapters. The importance of each part will be introduced as follows.

Section 1 introduces research in searchability and manipulation for XML data stream files. One current approach is to use keyword to query requested information. The following areas in query semantics, computing algorithms, the ranking of the results, view-based keyword search, and the comparison between data-centric and text-centric approaches are discussed within context. Another interesting approach is to build up XML filtering systems, which aim to provide fast, on-the-fly matching of XML-encoded data to large numbers of query specifications containing constraints on both structure and content. Approaches use event-based parsing and Finite State Machines (FSMs) to provide the basis for highly scalable structure-oriented XML filtering systems. Finally, optimization technique is shown to improve the processing efficiency.

Section 2 researches into a very common issue in XML world, i.e. dealing with large amount of XML documents generated by legend systems. This part focuses mainly on introducing compression of large documents and querying information without decompression through a system called *XCVQ-QP* that uses the path-dictionary, which contains all the elements and attributes names, to specify the relevant documents from thousands of XML documents. It is claimed that the system is able to retrieve information from unspecified document(s). To be against existing XML query processors that require the users to pre-specify the required documents to retrieve information, *XCVQ-QP* has the ability to retrieve information from one or more than one XML document without the need to specify exactly which document could contain the required information.

Section 3 presents a research outcome which can automatically map XML documents into relational database. The research aims to make people's lives easier, especially in the reasons for which they are using database without XML background. The system developed is called MAXDOR, which has achieved the following advantages:

- **High Flexibility of Updating:** MAXDOR approach performed updating processes of inserting new tokens in any location in the document and at any level of relevance to the candidate element (i.e. parent, child, left-sibling and right-sibling), updating token name and value at constant cost of execution time since there is no need to relabeling following tokens IDs or overwrite tokens paths.
- **Stability:** The approach worked fine in both directions; mapping and rebuilding for large documents: *"Auction"* document with 600MB size and 9244050 tokens can be processed without trouble.

Section 4 introduces a set of research work based on the emerging technology – Content Based Image Retrieval (CBIR). Investigation is against traditional machine learning approaches. The study has been evaluated by a number of case studies based on a developed query language. The achievements show that developed query language is good enough to deal with general retrieval tasks, e.g. mapping image similarity, merging multiple sub-queries, querying features semantically, through a set of advanced algorisms.

Section 5 presents a set of collaborative approaches in business process in information retrieval. The studies focus on the building up business model, developing advanced algorisms, implementing the theories and concepts into advanced service oriented architecture systems. The key methods used in the investigation are Petri-net for analysis of Business Process Modeling Methods; Pi-calculus for advanced branching and synchronization patterns description; rule based approach for process modeling; finally WS-CDL (Web Service Choreography Description Language) for describing multiple party how to work with together to accomplish a work in the context of SOA.

CONCLUSION

In conclusion, this book presents a picture for the latest concepts and the state of the art technology in the information retrieval and its related technologies and approaches. In particular, it addresses the following.

1. The contributions of key technologies to IR, i.e. XML, database, image retrieval, and business modelling process.
2. The contributions of key application areas to IR, i.e. data technology, document engineering, and business sectors.
3. New challenges in the research into retrieval efficiency, accuracy, and correctness for a wide range of disciplines.

REFERENCES

Debons, A. (1971). Command and control: Technology and social impact. *Advances in Computers*, *11*, 319–390. doi:10.1016/S0065-2458(08)60634-8

Lu, Z. (2005). A survey of XML applications on science and technology. *International Journal of Software Engineering and Knowledge Engineering*, *15*(1), 1–33. doi:10.1142/S0218194005001902

Lu, Z., & Arabnia, H. R. (2009). An introduction to contemporary computing with XML in science & technology . In *Natural language engineering & machine learning: Self adaptive systems pervasive computing*. Athens, GA: Elliott & Fitzpatrick.

Pein, R. P., & Lu, J. (2010). Multi-feature query language for image classification. *Procedia Computer Science*, *1*(1), 2539–2547. doi:10.1016/j.procs.2010.04.287

Wang, L., & Lu, Z. (2009). *Efficient spatial patterns mining*. USA: Elliot & Fitzpatrick, Inc.