

Report on the Fourth International Workshop on Web Information and Data Management (WIDM'2002)

Ee-Peng Lim

Department of Systems Engineering and Engineering Management
Chinese University of Hong Kong, Shatin, Hong Kong
aseplim@ntu.edu.sg

Roger Chiang Hsiang-Li

College of Business, University of Cincinnati
Cincinnati, Ohio 45221, USA
Roger.Chiang@uc.edu

1 Introduction

The Fourth International Workshop on Web Information and Data Management (WIDM'2002) was held at the SAIC Headquarters, McLean in Virginia on November 8, 2002. The workshop was held in conjunction with the International Conference on Information and Knowledge Management (CIKM'2002).

WIDM2002, compared to WIDM2001, has expanded its program to include 12 full paper and 6 short paper presentations. The workshop received 53 paper submissions from 21 countries, an increase of 50% over that of WIDM2001. The authors of these papers are from 9 countries. The twelve full papers were divided into 3 sessions: **Advanced XML Technologies and Applications**, **Web Mining**, and **Web Services and Performance Evaluation**. All short papers were presented in together in a separate **Short Papers** session.

2 Workshop Overview

The workshop included a keynote address on **Improving Access to Scientific Literature** by **Steve Lawrence** from NEC Research Institute. As a key researcher and developer behind the **CiteSeer** Scientific Literature Digital Library project, Steve gave a comprehensive overview of the design, implementation and operation of the CiteSeer system. His interesting talk was immediately followed by several questions from the workshop audience on the extraction and indexing techniques used in CiteSeer and its future directions.

The first session on *Advanced XML Technologies and Applications* consists of 4 paper presentations. The paper, **Consistently Updating XML Documents using Incremental Constraint Check Queries** by **Kane, Su and Rundensteiner**, describes a query rewriting approach to ensure safeness for updates on XML documents. The second paper, **Hybrid Applications over XML: Integrating the Declarative and Navigational Approaches** by **Manghi, Simeoni, Lievens and Connor**, proposes a quasi-statically typed language for XML to allow XML data to be associated with different structures and algebras. **Zhang, Pielech and Rundensteiner**, in the paper entitled **Honey, I Shrunk the XQuery! - An XML Algebra Optimization Approach**, describes a new way to optimize XQuery using query algebra. The paper, **A Mapping Schema and Interface for XML** by **Amer-Yahia and Srivastava**, addresses the problem of mapping XML data to relational schemas.

The second session entitled *Web Mining* began with the paper, **Using Latent Semantic Analysis (LSA) to Find Different Names for the Same Entity in Free Text** by **Oates, Bhat and Shanbhag** describes a new technique based on LSA to identify names that correspond to the same entities. The paper also reports promising results using the proposed technique. The second paper, **A Framework for Web Table Mining** by **Yang and Luk** investigate different ways to extract web table information. In the paper, **Effects of Maximum Flow Algorithm on Identifying Web Community**, **Imafuji and Kitsuregawa** reports the results of using maximum flow algorithm to determine web communities. The last paper, **Ranking User's Relevance to a Topic through Link Analysis on Web Logs** by **Wang, Zheng, Li, Liu and Ma**, proposes the use of a modified link analysis method to determine user relevance to a given topic.

In the third session on *Web Services and Performance Evaluation*, the first paper entitled **Preserving Privacy in Web Services** by **Rezgui, Ouzzani, Bouguettaya** and **Medjahed** proposes a web service infrastructure that aims to protect privacy in an e-government context. The second paper, **Indexing Web Access-Logs for Pattern Queries** by **Nanopoulos, Manolopoulos, Zakrzewicz** and **Morzy** describes a new indexing technique for web access logs. The next paper **Collecting Hidden Web Pages for Data Extraction** by **Lage, Silva, Golgher** and **Laender** investigates new ways to extract information from hidden web sites. The last paper, **A Fine-Grained Replacement Strategy for XML Query Cache** by **Chen, Wang, Cash, Ryder, Hobbs** and **Rundensteiner**, proposes a new cache replacement strategy for XML data.

In the final short paper session, **Ginsburg** first presented the paper entitled **The Catacomb Project: Building a User-Centered Web Portal the Conversational Way**. The next paper, **An Algebraic Approach for Incremental Maintenance of Materialized XQuery**, by **EL-Sayed, Wang, Ding**, and **Rundensteiner** deals with the incremental update problem of materialized XQuery views. The third paper by **Liu, Ng, Li** and **Lim** proposes **A Visual Tool for Building Logical Data Models of Websites**. This is followed by the paper on **Web Classification Using Support Vector Machine** by **Sun, Lim** and **Ng**. The paper describes the feature engineering issues for web classification. The paper, **Concept Extraction and Association from Cancer Literature** by **Fu, Bauer, Mostafa, Palakal** and **Mukhopadhyay**, attempts to relate different concepts found in biomedical literature. The last paper, **Design and Performance Evaluation of a Web-based Multi-Tier Federated System for a Catalogue of Life** by **Xu, Jones , Gray, Fiddian, White** and **Bisby**, presents a distributed database architecture design for accessing biological data.

3 Conclusions

At the end of WIDM'2002, a happy hour session was organized for the workshop participants at a nearby restaurant. About 9 people turned up at the happy hour to enjoy the finger food and beers. Some photographs have been posted at <http://www.se.cuhk.edu.hk/~eplim/widm2002>.

At the time of writing this report, the next WIDM workshop has been finalized and will be held together with CIKM'2003 in New Orleans, Louisiana (<http://www.cikm.org/2003>). A new workshop co-chair, Alberto Laender, has been recruited, and the workshop call-for-paper announcement will be out soon. In WIDM2003, a two-day workshop program will be introduced to encourage more paper submissions and participation.