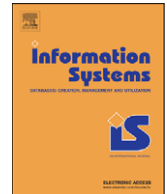




Contents lists available at ScienceDirect

Information Systems

journal homepage: www.elsevier.com/locate/infosys

Guest Editorial

Information systems special issue on ACM CIKM 2007

We are pleased to bring you this special issue of Information Systems based upon the 16th International Conference on Information and Knowledge Management (CIKM 2007), which was held from November 6 to 9 of 2007 in Lisbon, Portugal. CIKM is a venue that covers three distinct fields: databases (DB), information retrieval (IR) and knowledge management (KM), allowing leading researchers and practitioners to show their research and interchange ideas. The conference itself consisted of both research and industrial paper sessions, invited talks by leading specialists, and several workshops.

This special issue includes seven papers selected among the best ones by the Program Committee and that were considered representative of the types and diversity of research presented at the conference. These papers were selected from the 86 full papers accepted from 512 submissions. Hence, this sample is less than 9% of all accepted papers or 2% of all submitted papers. Each one of these papers is an extended and revised version of the conference original paper and went through a rigorous reviewing process, before being accepted for inclusion in this special issue.

There are two papers from the DB track. The first paper, titled “A Strategy for Allowing Meaningful and Comparable Scores in Approximate Matching”, by Dorneles et al. addresses the definition of threshold values in approximate data matching in terms of the estimated precision that is expected from the matching process. The proposed method maps the scores returned by similarity functions into adjusted scores. Experiments using single attributes show that adjusted scores can be used instead of the original similarity scores. Moreover, experiments using records show that existing strategies for combining similarity scores may be enhanced when adjusted scores are used. The second paper, titled “Self-Correcting Queries for XML”, by Cohen and Brodianskiy explores query correction in XML databases. Due to the heterogeneous nature of XML documents, and their intricate hierarchical structures, user queries over XML are often unsatisfiable. Hence, this paper presents a general framework for automatically correcting queries over XML, by returning

queries that are similar to the user's query, but are satisfiable.

The IR track is represented by three papers. The first paper, titled “Implementing and Evaluating Phrasal Query Suggestions for Proximity Search”, by Feuer, Savev and Aslam describes and evaluates a unified approach to phrasal query suggestions in the context of a high-precision search engine. They describe methods for generating longer and shorter suggestions and present algorithms for ranking the suggestions. Specifically, they introduce the problem of counting proximal sub phrases for specialization and the problem of counting unordered super phrases for generalization. In the second paper, titled “Using Semantic Components to Search for Domain-Specific Documents: An Evaluation from the System Perspective and the User Perspective”, Price et al. seek to leverage an expert user's knowledge about how information is organized in a domain and how information is presented in typical documents within a particular domain-specific collection, to effectively and efficiently meet the expert's targeted information needs. For that they developed the semantic components model to describe important semantic content within documents, including document classes, showing how that can be used to improve an IR system. Finally, in the paper titled “Exploring Models for Semantic Category Verification”, Roussinov and Turetken explore completely automated on-the-fly verification of a membership in any arbitrary category which has not been expected a priori. Their approach does not rely on any manually codified knowledge (such as WordNet or Wikipedia) but instead capitalizes on the diversity of topics and word usage on the World Wide Web, thus can be considered “knowledge-light” and complementary to the “knowledge-intensive” approaches. They have created a quantitative verification model and tested it empirically in a fact seeking engine on the well-known TREC conference test questions improving the answer accuracy.

Finally, the KM track includes two papers centered around issues on ontologies for knowledge management. In the paper entitled “Requirements-oriented Methodol-

ogy for Evaluating Ontologies”, Yu, Thom, and Tam introduce the ROMEO methodology. They provide requirements and mappings to evaluation measures for use in ontology evaluation. They demonstrate the use of these measures for evaluating ontologies for use in browsing Wikipedia. In the second paper, titled “Metric-based Stochastic Conceptual Clustering for Ontologies”, Fanizzi, d’Amato, and Esposito present a framework for stochastic conceptual clustering. They have shown that this can be applied in knowledge base settings such as those found on the Semantic Web. One use of this work is to find interesting groupings of annotated resources, such as the annotations enabled by semantic web languages. The stochastic clustering algorithm is an extension of distance-based clustering. They also present extensions aimed at producing hierarchical clustering and also at detecting drift of existing concepts or the rise of novel concepts.

So we hope that you enjoy reading these selected papers!

Guest Editorial

Ricardo Baeza-Yates *

*Universidad de Chile, Center for Web Research, Blanco
Encalada 2120, Yahoo! Research, Spain & Chile
E-mail address: rbaeza@acm.org*

Alberto H.F. Laender

Federal University of Minas Gerais, Brazil (DB Co-chair)

Deborah L. McGuinness

Rensselaer Polytechnic Institute, USA (KM Co-chair)

* Corresponding author. Tel.: 56 2 689 5531; fax: +56 2 689 2736.