

# **Transactions on Computational Collective Intelligence XXI**

## **Special Issue on Keyword Search and Big Data**

### **Preface**

Keyword-based search has become the de facto standard for information discovery on the Web. This is mainly due to its simplicity, when compared with more formal query languages such as SQL or SPARQL. Although these are more efficient and effective, they require specialized knowledge usually held by computer scientists and engineers. However, as the Web evolves and new structured databases are brought on-line, as semantics, social, and contextual aspects associated with the data become more relevant, the challenges to the feasibility of keyword-based search increase. But these have to be overcome, if the latent socio economic impact of the new data sources is to be unlocked by end-users. For keyword-based search to be possible in this new reality, various problems have to be addressed. Furthermore, these are multifaceted in nature and the solutions require contributions from many disciplines. This special issue provides a contribution to this research space.

In the first paper, entitled “Keyword-Based Search over Databases: A Roadmap for a Reference Architecture Paired with an Evaluation Framework,” by Sonia Bergamaschi, Nicola Ferro, Francesco Guerra, and Gianmaria Silvello, the authors argue the need for a reference architecture for keyword search in databases, as well as a companion evaluation framework. The goal is to focus on the end-user goals, both by considering compelling use cases and by bringing together diverse techniques to fully address them. In the quest for a more effective keyword-based search, Enrico Sartori, Yannis Velegrakis, and Francesco Guerra, the authors of the second paper, “Entity-Based Keyword Search in Web Documents,” propose a way to increase the expressive power of a source document by representing it as a set of structures that describe relationships among the entities mentioned in the text. The third paper, “Evaluation of Keyword Search in Affective Multimedia Databases,” discusses a case where keyword-based search is used beyond the realms of simple text documents. In this case, the authors – Marko Horvat, Marin Vuković, and Željka Car – present an experimental evaluation of multi-keyword search in affective multimedia databases, where the emotional responses to the pieces of content are stored with a description of their semantics using keywords from unsupervised glossaries, expected emotion elicitation potential, and other contextual information. Another field of application of keyword-based search is addressed in the fourth paper – “Data-Driven Discovery of Attribute Dictionaries” – where Fei Chiang, Periklis Andritsos, and Renee J. Miller present techniques to extract attribute values from textual product descriptions of on-line retailers. Automated techniques such as this are increasingly important to ensure effective searches by customers, as the number of products available at these

outlets increase. Moving on to social media, the next paper addresses the challenge of efficiently retrieving and understanding messages. In “Subject-Related Message Filtering in Social Media Through Context-Enriched Language Models,” Alexandre Davis and Adriano Veloso exploit context in the analysis of social media messages using computational linguistics techniques, with interesting results. The importance of context is again emphasized in the sixth paper – “Improving Open Information Extraction for Semantic Web Tasks” – by Cheikh Kacfar Emani, Catarina Ferreira da Silva, Bruno Fiès, and Parisa Ghodous, who take Open Information Extraction, which aims at automatically identifying all the possible assertions within a sentence, and show how it can be improved when working on a given domain of knowledge. Last but not least, in “Searching Web 2.0 Data Through Entity-Based Aggregation,” Ekaterini Ioannou and Yannis Velegrakis draw on suggestions from the Linked Data movement to design an approach for more efficient and effective integration of Social and Semantic Web data using entities.

The papers in this special issue have been accepted after, at least, two iterations of a double-blind peer-review process. I would like to express my sincere thanks to all the authors and the Reviewers of the Special Issue for their contributions. A special word of appreciation to the Editor-in-Chief, Ngoc Thanh Nguyen, for embracing this project, and, finally my gratitude to Bernadetta Maleszka, Assistant Editor of TCCI, for all the help throughout the review process and in preparing the final documents for the issue.

January 2016

Paulo Rupino da Cunha

Transactions on Computational Collective Intelligence  
XXI

Special Issue on Keyword Search and Big Data

Nguyen, N.T.; Kowalczyk, R.; Rupino da Cunha, P. (Eds.)

2016, IX, 175 p. 60 illus. in color., Softcover

ISBN: 978-3-662-49520-9