

# Exploratory Search in Databases and the Web

Georgia Koutrika  
HP Labs, Palo Alto  
koutrika@hp.com

Laks V.S. Lakshmanan  
Department of Computer  
Science, University of British  
Columbia  
laks@cs.ubc.ca

Mirel Riedewald  
College of Computer and  
Information Science,  
Northeastern University  
mirek@ccs.neu.edu

Kostas Stefanidis  
ICS-FORTH, Heraklion  
kstef@ics.forth.gr

## Introduction

The traditional way a user interacts with a database system is through queries. Structured query languages, such as SQL for relational data, XQuery for XML, and SPARQL for RDF data, allow users to submit queries that may precisely capture their information needs, but users need to be familiar with the underlying ontology and data structure and of course the query language itself. Moreover, users need to some extent be familiar with the content of the database and have a clear understanding of their information needs. These requirements stand as the weaknesses of this interaction mode. As data stored in databases grows in unprecedented rates and becomes accessible to diverse and less technically oriented audience, new forms of data exploration and interaction become increasingly more attractive.

The World Wide Web represents the largest and arguably the most complex repository of content. Users seek information on the web through two predominant modes: by browsing or by searching. In the first mode, the interaction between the user and the data repository is driven directly by the user's interpretation of their information need and their information foraging constraints. In the latter mode, a search engine typically mediates the user-data interactions and the process starts with the user entering query-terms that act as surrogates for the user information goals. Free-text queries allow end-users a simple way to express their information needs independently from the underlying data model and structure, as well as from a specific query language. Given a query, the most common strategy has been to present the results as a ranked list. Users have to subsequently peruse the list to satisfy their information needs through browsing the links and/or by issuing further queries.

However, the information in the web gets rapidly diversified both in terms of its complexity as well as in terms of the media through which the information is encoded, spanning from large amounts of unstructured and semi-structured data to semantically rich available knowledge. Increasing de-

mands for sophisticated discovery capabilities are now being imposed by numerous applications in various domains such as social media, healthcare, telecommunication, e-commerce and web analytics, business intelligence, and cyber-security. Yet, many of these data are hidden behind barriers of language constraints, data heterogeneity, ambiguity, and the lack of proper query interfaces.

Furthermore, the complexity and heterogeneity of the information implies that the associated semantics is often user-dependent and emergent. Individual aspects like age, gender, profession or experience are often not taken into account, for example the difference in searching between children and adults. In addition, most common systems still assume that the user has a static information need, which remains unchanged during the seeking process. Hence, they are strongly optimized for lookup searches, expecting that the user is only interested in facts and not in complex problem solving.

Consequently, there is a need to develop novel paradigms for exploratory user-data interactions that emphasize user context and interactivity with the goal of facilitating exploration, interpretation, retrieval, and assimilation of information. A huge number of applications need an exploratory form of querying. Ranked retrieval techniques for relational databases, XML, RDF and graph databases, text and multimedia databases, scientific and statistical databases, social networks and many others, is a first step towards this direction. Recently, several new aspects for exploratory search, such as preferences, diversity, novelty and surprise, are gaining increasing importance. From a different perspective, recommendation applications tend to anticipate user needs by automatically suggesting the information which is most appropriate to the users and their current context. Also, a new line of research in the area of exploratory search is fueled by the growth of online social interactions within social networks and web communities. Many useful facts about entities (e.g. people, locations, organizations, products) and their relationships can be found in a multitude of semi-structured and structured data sources such as Wikipedia<sup>1</sup>, Linked Data cloud<sup>2</sup>, Freebase<sup>3</sup>, and many others. Therefore, novel discovery methods are required to provide highly expressive discovery capabilities over large amounts of entity-relationship data, which are yet intuitive for end-users.

(c) 2014, Copyright is with the authors. Published in the Workshop Proceedings of the EDBT/ICDT 2014 Joint Conference (March 28, 2014, Athens, Greece) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0.

<sup>1</sup><http://wikipedia.org>

<sup>2</sup><http://linkeddata.org>

<sup>3</sup><http://freebase.com>

## The ExploreDB Workshop

The purpose of the ExploreDB workshop is to bring together researchers and practitioners that approach data exploration from different angles, ranging from data management, information retrieval to data visualization and human computer interaction, in order to study the emerging needs and objectives for data exploration as well as the challenges and problems that need to be tackled.

In this first workshop instance, we have put together a program, comprising a keynote talk, six research papers, and a panel, that examines data exploration from the standpoints of data visualization, information retrieval, web search, data mining, and database queries. We are grateful to all the authors who submitted papers. We would also like to thank Daniel Keim for accepting to be the keynote speaker, and our reviewers who did their best in delivering thorough reviews on time.

## The ExploreDB Organization

### The workshop co-chairs

Georgia Koutrika  
*HP Labs, Palo Alto, USA*

Laks V.S. Lakshmanan  
*Department of Computer Science,  
University of British Columbia, Canada*

Mirek Riedewald  
*College of Computer and Information Science,  
Northeastern University, USA*

Kostas Stefanidis  
*Institute of Computer Science, FORTH, Hellas*

### Program committee

Ira Assent, Aarhus University, Denmark

Wolf-Tilo Balke, TU Braunschweig, Germany

Carlos Castillo, Qatar Computing Research Institute, Qatar

Nicola Ferro, University of Padua, Italy

Minos Garofalakis, Technical University of Crete, Greece

Melanie Herschel, University of Paris South, France

H. V. Jagadish, University of Michigan, Ann Arbor, USA

Haridimos Kondyakis, ICS-FORTH, Greece

Mohamed Mokbel, University of Minnesota, Minneapolis, USA

Kjetil Kjetil Nørnvåg, Norwegian University of Science and Technology, Norway

Evaggelia Pitoura, University of Ioannina, Greece

Julia Stoyanovich, Drexel University, Philadelphia, USA

Letizia Tanca, Politecnico di Milano, Italy

Martin Theobald, University of Antwerp, Belgium

Panayiotis Tsaparas, University of Ioannina, Greece

Yannis Tzitzikas, University of Crete and ICS-FORTH, Greece

Jun Yang, Duke University, Durham, USA

Demetris Zeinalipour, University of Cyprus, Cyprus