Second International Workshop on Searching and Integrating New Web Data Sources (VLDS 2012)

Co-located with VLDB (Very Large DataBases) 2012
`http://vlds.search-computing.org`

Marco Brambilla, Stefano Ceri
Politecnico di Milano, DEI
Milano, Italy
firstname.lastname@polimi.it

Tim Furche, Georg Gottlob
Dept. of Computer Science, Oxford University
Oxford, UK
firstname.lastname@cs.ox.ac.uk

1. CONTEXT

Recent years witnessed an exponential growth of data providers available on the Web. These providers offer a plethora of different ways of accessing their data sources, spanning from APIs over proprietary query languages (such as Yahoo! Query Language, YQL) to endpoints accessible through standard query languages (e.g., SPARQL). At the same time, data is increasingly being labeled, tagged, and linked with existing data, partially due to social networking applications. These data sources expose their data as semi-structured information and an increasing number also provide the information in the *linked data cloud*, with URI-based references between resources. Linked Open Data (LOD) emerges as a best practice for exposing, sharing, and connecting pieces of data, information, and knowledge.

This is a major change of paradigm. On one side, this augments the power of search methods which access and query information with respect to the old-fashioned page based Web paradigm. On the other side, though, this challenges the current information retrieval, data integration, and Web search practices to comply with the new shape and capabilities of new Web data sources. Searching for data upon such new, Web-enabled data sources has the potential of reshaping the scenario of current Web applications, going beyond the capabilities of conventional search engines in solving search problems, but it also presents new technical challenges, for search as well as for surfacing techniques. Current web pages all to often stick with the old-fashioned page-based Web paradigm. Therefore, methods for turning such web pages into search services or other forms of knowledge are very much necessary for search services to be universally useful.

2. GOAL

This years' VLDS workshop gathers, as in previous years, leading researchers and practitioners in the diverse fields related to data integration, deep web search, and the construction of knowledge bases from the web with the purpose of discussing innovative strategies for combining search facilities with integration aspects for Web data sources. The workshop represents a unique venue for discussing all the aspects related to the surfacing, publication, and orchestration of services over new Web data sources, the most suitable paradigms to improve the user experience in context, as well as the application scenarios which may better benefit of these new technologies.

VLDS'12. Istanbul, August 31st, 2012.

Solving these problems requires new solutions on the intersection of data integration, multi-domain search, deep web extraction, and information extraction. In this edition, a particular focus is the construction of search services and knowledge bases from unstructured web data and the deep web.

3. TOPICS OF INTEREST

The topics of interest for this workshop include:

Methods and tools for Search Services, including:
- Modeling and Exposing search functionalities as services
- Deploying and Using search services
- Languages and platforms for composing search services
- Best practices and methodologies for designing and composing search services
- Mashup platforms and practices applied to search

Methods and tools for deep web information access:
- Exploitation of public APIs for search (e.g., Google APIs, Yahoo Query Language YQL)
- Implementation issues of ranking, ordering, and chunking in queries on data sources
- Use of query languages (including SQL, SPARQL, XQuery) for deep web data sources
- Mashup platforms and practices for deep web data

Methods and Tools for domain-specific search, including:
- Algorithms and tools for domain-specific or purpose-specific search
- Best practices and methodologies for domain or purpose-specific search

Methods and Tools for Open Linked Data, including:
- Algorithms and tools for search and exploration over linked and semantically-enriched data
- Methods for preparing and labeling data to support search applications

User experience of search
- User interfaces for search, including purpose- or domain-specific services
- Information exploration and exploratory search over Web structured, semi-structured and unstructured data
- Continuous, incremental and push-based search

Applications of search
- Warehousing and integration of searchable data
- Enterprise search applications
- Social search
- Web recommender systems

Benchmarks for search applications on integrated data

4. PROGRAM COMMITTEE

We wish to thank the PC members that contributed to the success of the workshop by carefully reviewing the submitted papers and providing the authors with useful suggestions for improving the papers:

Robert Baumgartner	Lixto Software GmbH
Michael Benedikt	Oxford University
Florian Daniel	University of Trento
Anish Das Sarma	Google Research
Arjen de Vries	CWI
Sergio Flesca	DEIS - University of Calabria
Alejandro Jaimes	Yahoo! Research
Arnd Christian König	Microsoft Research
Jens Lehmann	Universität Leipzig
Ioana Manolescu	INRIA Saclay–Île-de-France and LRI, Université Paris Sud-11
Hamid Motahari	HP Labs
Neoklis Polyzotis	University of California Santa Cruz
David Robertson	University of Edinburgh
Mike Rosner	University of Malta
Sebastian Schaffert	Salzburg Research Forschungsgesellschaft
Klara Weiand	University of Munich
Gerhard Weikum	KPI
Clement Yu	University of Illinois at Chicago

We also wish to thank the additional reviewers that kindly helped to PC to select the best papers for VLDS 2012.

5. WORKSHOP PROGRAM

The workshop received about 15 submissions, of which only about 50% (7 papers) have been accepted. For the workshop, the papers are divided into three sessions on "Web Knowledge Bases", "Deep Web", and "Wrappers". These papers are joined by two invited keynotes by Gerhard Weikum (Max-Planck-Institut, Germany) and Raghu Ramakrishnan (Microsoft).

Invited Speakers

Gerhard Weikum
> Semantic Search: from Names and Phrases to Entities and Relations

Raghu Ramakrishnan
> The Future of Information Discovery and Search: Content Optimization, Interactivity, Semantics, and Social Networks

Web Knowledge Bases

Marilena Oita, Antoine Amarilli and Pierre Senellart
> Cross-Fertilizing Deep Web Analysis and Ontology Enrichment

Jianfeng Si, Qing Li, Tieyun Qian and Xiaotie Deng
> Hierarchical Clustering on HDP Topics to build a Semantic Tree from Text

Ndapandula Nakashole, Mauro Sozio, Fabian Suchanek and Martin Theobald
> Query-Time Reasoning in Uncertain RDF Knowledge Bases with Soft and Hard Rules

Deep Web

Meghyn Bienvenu, Daniel Deutch, Davide Martinenghi, Pierre Senellart and Fabian Suchanek
> Dealing with the Deep Web and all its Quirks

Feng Niu, Ce Zhang, Christopher Re and Jude Shavlik
> DeepDive: Web-scale Knowledge-base Construction using Statistical Learning and Inference

Wrappers

Tim Furche, Giovanni Grasso, Christian Schallhart, Andrew Sellers and Antonino Rullo
> Think before you Act! Minimising Action Execution in Wrappers

Rolando Creo, Valter Crescenzi, Disheng Qiu and Paolo Merialdo
> Minimizing the Costs of the Training Data for Learning Web Wrappers

6. ACKNOWLEDGEMENTS