Search Computing

Challenges and Directions
Preface

Who are the strongest European competitors on software ideas? Who is the best doctor to cure insomnia in a nearby hospital? Where can I attend an interesting conference in my field close to a sunny beach? This information is available on the Web, but no software system can accept such queries nor compute the answer. At most, users can identify sub-problems that can be addressed by specific search engines and interact with each of them serially, but then they have the responsibility of building global answers by manually composing results. Search computing is a new multi-disciplinary discipline which will provide the abstractions, methods, tools and computing systems required to express these queries and to build their answer.

The emerging paradigm of software services has so far been neutral to search. Search computing is an evolution of service computing focused on building the answers of complex queries by interacting with a constellation of cooperating search services, using ranking as the dominant factor for service composition. New language and description paradigms are required for interconnecting services and for expressing queries. Semantic domain knowledge helps enrich terminological knowledge about objects being searched. New protocols help capture ranking preferences and their refinement; new interfaces present complex results with simple visual descriptions. Ranking is relative to individuals and context and therefore reflects personal and social contributions. Financial and legal implications of search computing must be understood and mastered. In summary, search computing is a multi-disciplinary effort which requires adding to sound software principles contributions from other sciences such as knowledge representation, human–computer interfaces, psychology, sociology, economics and legal sciences.

The Search Computing (Seco) Project is funded by the European Research Council (ERC), responding to the 2008 Call for “IDEAS Advanced Grants,” a program dedicated to the support of investigation-driven frontier research. SeCo started on November 1, 2008 and will last until October 31, 2013 (see www.searchcomputing.eu). This book describes the outcome of the first SeCo “Workshop on Search Computing Challenges and Directions,” held in Como during June 17–19, 2009.

The book is divided into three parts. The first part presents visions of the current evolution in search, which is becoming more and more task-oriented and is now starting to use ontological knowledge in order to manage complex queries; these visions are marking the new trends in search.

The second part provides some background and related technologies. These can be considered as parallel fields of research, useful both for setting the theoretical premises for search computing and for providing a technological framework for building search computing systems and applications.
The third part dwells on the *technological problems and issues* which arise when dealing with search computing as a new search paradigm. It provides a unified view of the results of search computing as achieved exactly one year after its starting date.

The book is the result of a collective effort of all the project participants and has been reviewed with the help of the project’s advisory board members and of several other experts. We thank all of them for their effort.

January 2010

Stefano Ceri

Marco Brambilla
Organization

Reviewers

Luciano Baresi  Politecnico di Milano
Marco Brambilla  Politecnico di Milano
Fabio Casati  Università di Trento
Tiziana Catarci  Università di Roma “La sapienza”
Stefano Ceri  Politecnico di Milano
Georg Gottlob  Oxford University
Ihab Ilyas  University of Waterloo
Ioana Manolescu  INRIA and Université de Paris Sud
Giansalvatore Mecca  Università della Basilicata
Roberto Verganti  Politecnico di Milano
Gerhard Weikum  Max-Planck Institute for Informatics

Sponsoring Institutions

The Search Computing (Seco) Project is funded by the European Research Council (ERC), responding to the 2008 Call for “IDEAS Advanced Grants”, a program dedicated to the support of investigation-driven frontier research. SeCo started on November 1st, 2008 and will last until October 31, 2013.
# Table of Contents

## Part I: Visions

Chapter 1: Search Computing ........................................... 3  
*Stefano Ceri*

Chapter 2: Next Generation Web Search ............................ 11  
*Ricardo Baeza-Yates and Prabhakar Raghavan*

Chapter 3: Search for Knowledge ................................. 24  
*Gerhard Weikum*

## Part II: Technology Watch for Search Computing

Chapter 4: The Search Engine Industry ............................. 45  
*Tommaso Buganza and Emanuele Della Valle*

Chapter 5: From Mashup Technologies to Universal Integration: Search Computing the Imperative Way ................................. 72  
*Florian Daniel, Stefano Soi, and Fabio Casati*

Chapter 6: Web Data Extraction for Service Creation ............ 94  
*Robert Baumgartner, Alessandro Campi, Georg Gottlob, and Marcus Herzog*

Chapter 7: Dataspaces ................................................. 114  
*Cornelia Hedeler, Khalid Belhajjame, Norman W. Paton, Alessandro Campi, Alvaro A.A. Fernandes, and Suzanne M. Embury*

Chapter 8: Multimedia and Multimodal Information Retrieval .... 135  
*Alessandro Bozzon and Piero Fraternali*

## Part III: Issues in Search Computing

Chapter 9: Service Marts .............................................. 163  
*Alessandro Campi, Stefano Ceri, Georg Gottlob, Andrea Maesani, and Stefania Ronchi*

Chapter 10: Join Methods and Query Optimization ............... 188  
*Daniele Braga, Stefano Ceri, and Michael Grossniklaus*

Chapter 11: Rank-Join Algorithms for Search Computing ............ 211  
*Ihab F. Ilyas, Davide Martinenghi, and Marco Tagliasacchi*
Chapter 12: Panta Rhei: Flexible Execution Engine for Search Computing Queries .............................................. 225
    Daniele Braga, Stefano Ceri, Francesco Corcoglioniti, and Michael Grossniklaus

Chapter 13: Liquid Queries and Liquid Results in Search Computing . . . 244
    Alessandro Bozzon, Marco Brambilla, Stefano Ceri, Piero Fraternali, and Ioana Manolescu

Chapter 14: Building Search Computing Applications ................................. 268
    Alessandro Bozzon, Marco Brambilla, Stefano Ceri, Francesco Corcoglioniti, and Nicola Gatti

Chapter 15: Search Computing and the Life Sciences .............................. 291
    Marco Masseroli, Norman W. Paton, and Irena Spasić

Appendixes

Appendix A: Search Computing Dictionary ........................................ 307

Author Index ..................................................................................... 321