

Lecture Notes in Artificial Intelligence 3847

Edited by J. G. Carbonell and J. Siekmann

Subseries of Lecture Notes in Computer Science

Klaus P. Jantke Aran Lunzer
Nicolas Spyratos Yuzuru Tanaka (Eds.)

Federation over the Web

International Workshop
Dagstuhl Castle, Germany, May 1-6, 2005
Revised Selected Papers

Series Editors

Jaime G. Carbonell, Carnegie Mellon University, Pittsburgh, PA, USA
Jörg Siekmann, University of Saarland, Saarbrücken, Germany

Volume Editors

Klaus P. Jantke
FIT Leipzig, Forschungsinstitut für InformationsTechnologien
Postfach 30 11 66, 04251 Leipzig, Germany
E-mail: jantke@meme.hokudai.ac.jp

Aran Lunzer
Yuzuru Tanaka
Hokkaido University, Meme Media Laboratory
North 13 West 8, Sapporo 060-8628, Japan
E-mail: {aran,tanaka}@meme.hokudai.ac.jp

Nicolas Spyratos
Université Paris-Sud, Laboratoire de Recherche en Informatique
LRI-Bât. 490, 91405 Orsay Cedex, France
E-mail: spyratos@lri.fr

Library of Congress Control Number: 2005938389

CR Subject Classification (1998): I.2, H.2.8, H.3, H.4, J.1

ISSN 0302-9743
ISBN-10 3-540-31018-5 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-31018-1 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11605126 06/3142 5 4 3 2 1 0

Preface

The lives of people all around the world, especially in industrialized nations, continue to be changed by the presence and growth of the Internet. Its influence is felt at scales ranging from private lifestyles to national economies, boosting the pace at which modern information and communication technologies influence personal choices along with business processes and scientific endeavors.

In addition to its billions of HTML pages, the Web can now be seen as an open repository of computing resources. These resources provide access to computational services as well as data repositories, through a rapidly growing variety of Web applications and Web services.

However, people's usage of all these resources barely scratches the surface of the possibilities that such richness should offer. One simple reason is that, given the variety of information available and the rate at which it is being extended, it is difficult to keep up with the range of resources relevant to one's interests. Another reason is that resources are offered in a bewildering variety of formats and styles, so that many resources effectively stand in isolation.

This is reminiscent of the challenge of enterprise application integration, familiar to every large organization be it in commerce, academia or government. The challenge arises because of the accumulation of information and communication systems over decades, typically without the technical provision or political will to make them work together. Thus the exchange of data among those systems is difficult and expensive, and the potential synergetic effects of combining them are never realized.

Motivation for overcoming this challenge with respect to the Web is found in many domains. In academia there is a recognized need for interdisciplinary and international availability, distribution and exchange of intellectual resources, which include both static information (such as publications and other research results) and the tools that support research activities. Similar pressure derives from the development and deployment of pervasive computing, which extends the types of resources that are present on the Web to encompass embedded devices and mobile computing resources, communicating over wireless networks. In such domains, just as for enterprise application integration, the rich variety of resources and the boundless human creativity applied in developing new solutions conspire to increase the number of compatibility barriers.

To what extent can this challenge be addressed by standardization? While the development and broad adoption of standards are crucial to the advance of information and communication technologies, any attempt to find a general solution to resource incompatibility through global standardization is doomed to fail because of the diversity of resources and the pace at which they change. That said, given an appropriately narrowed target scope, it is reasonable for providers to agree on a shared middle ground that will increase the mutual

compatibility of their resources. This is the key to the approach known as *mediation*. Mediation, at least as usually understood in the area of databases, requires cooperation among providers in specifying (a) a well-defined community of cooperating sources, and (b) a common schema (called the mediator schema) to which the sources address their queries and/or provide answers.

In contrast to mediation, the study of *federation* involves working to bridge the differences between resources without such a predefined common ground. The process of resource federation in general involves selecting the resources that are to be combined, discovering the relationships that will allow them to work together, establishing the necessary connections, then driving the assembly in a coordinated way to achieve certain desired goals. The science of federation requires new theoretical foundations and enabling technologies for analysis of syntactic and/or semantic interrelations among resources, matching of service requesters and providers, and reliable and secure establishment and coordination of their execution.

Federation over the Web has attracted the attention of researchers aiming to support interdisciplinary and cross-border reuse and interoperation of heterogeneous intellectual resources, for example in support of scientific research, simulation, and digital libraries. Federation over enterprise intranets, also based on Web technologies, is being pursued as a way to bring large numbers of legacy application systems into cooperation with each other.

Existing work on federation can be divided broadly into programmatic and interactive approaches. Programmatic approaches are based on standardization at the level of communication protocols and languages for discovering compatibilities between resources, including the run-time matching of requesters and providers; for the Web, such federation tends to be based on Web-service technologies. On the other hand, interactive federation places in the hands of the user all responsibility for judging which resources are suitable for connection, then establishing and coordinating such connections. A simple example of interactive federation on the Web is the use of visual operations to connect result elements of one Web application, found at predictable locations within its HTML results, to input fields within an HTML form of another application.

From 1 to 6 May 2005 we held a workshop at Dagstuhl Castle to discuss advances in this area, drawing together active researchers from several institutes in Japan and Europe. The workshop focused on theoretical foundations and enabling technologies for federation of resources offered over the Web or within pervasive computing environments. We invited the participants to present and discuss work falling under any of the following topics:

- Knowledge look-up and matching
- Knowledge search and clustering
- Knowledge ontology and mediation
- Interoperation of Web-based resources
- Knowledge extraction and Web wrappers
- Computational models for knowledge federation

Based on the 18 workshop presentations, we went through a process of consultation, reviewing and editing to arrive at the 12 papers in this book. As shown in the table of contents, these papers touch on most of the above topics.

Future research and development of Web-based federation stands to influence how humans use intellectual resources in local and global networks. In combination with the rise of ubiquitous computing, introducing new forms of mobile computing devices and smart objects, computer systems will increasingly form location-based, on-demand federations. By the meeting and cooperation of these systems, humans will be dynamically connected with other humans and with a greater variety of systems and services. One can envisage future workshops on resource federation including contributions from the humanities, including sociology and psychology.

Let the present volume set the stage for such exciting developments.

October 2005

Klaus P. Jantke
Aran Lunzer
Nicolas Spyratos
Yuzuru Tanaka

Table of Contents

Knowledge Look-Up and Matching

Text Mining Using Markov Chains of Variable Length <i>Björn Hoffmeister, Thomas Zeugmann</i>	1
Faster Pattern Matching Algorithm for Arc-Annotated Sequences <i>Takuya Kida</i>	25
VSOP (Valued-Sum-of-Products) Calculator for Knowledge Processing Based on Zero-Suppressed BDDs <i>Shin-ichi Minato</i>	40

Knowledge Search and Clustering

A Method for Pinpoint Clustering of Web Pages with Pseudo-Clique Search <i>Makoto Haraguchi, Yoshiaki Okubo</i>	59
Specific-Purpose Web Searches on the Basis of Structure and Contents <i>Mineichi Kudo, Atsuyoshi Nakamura</i>	79
Graph Clustering Based on Structural Similarity of Fragments <i>Tetsuya Yoshida, Ryosuke Shoda, Hiroshi Motoda</i>	97

Knowledge Mediation

Connecting Keywords Through Pointer Paths over the Web <i>Mina Akaishi, Nicolas Spyratos, Koichi Hori, Yuzuru Tanaka</i>	115
Querying with Preferences in a Digital Library <i>Nicolas Spyratos, Vassilis Christophides</i>	130

Interoperation of Web-Based Resources

An Enhanced Spreadsheet Supporting Calculation-Structure Variants, and Its Application to Web-Based Processing <i>Aran Lunzer, Kasper Hornbæk</i>	143
---	-----

Knowledge Federation over the Web Based on Meme Media Technologies
Yuzuru Tanaka 159

Knowledge Evolution

Towards Understanding Meme Media Knowledge Evolution
Roland Kaschek, Klaus P. Jantke, István-Tibor Nébel 183

Mechanisms of Knowledge Evolution for Web Information Extraction
Carsten Müller 202

Author Index 215