

# Seventh International Workshop for Information Integration on the Web (IIWeb'09)

Welcome to the Workshop on Information Integration on the Web (IIWeb'09), organized in conjunction with IJCAI-07, at Pasadena, California, USA on 11th July, 2009. The theme of the workshop is "Data Issues in Services and Service Integration".

## Information Integration on the Web

The explosive growth of the World Wide Web has amassed a huge number of information sources on the Internet with unprecedented potential for information access. Information originating from different sources needs to be integrated in order to gain full advantage of having that information available. Information integration techniques enable the interaction between users and data sources through a centralized access point and uniform query interfaces that give users the illusion of querying a homogeneous system.

Along with the growth of the World Wide Web, there has been phenomenal growth of the Internet as the medium to deliver services. Services in this context span a wide range: business processes (e.g., like call center) of organizations being outsourced, to applications made available as per Service Oriented Architecture which can be readily composed by 3<sup>rd</sup> party, to infrastructure services where IT resources is made available on-demand (e.g., cloud computing). Services are driven by information and increasingly, both of them are distributed. The data challenges in Services that a service provider faces stem from the fact that although service and information needs of different customer engagements are broadly similar, yet they are presently recorded and managed differently. The economies of scale and reuse will come from organizing information in some uniform way, yet there are customer sensitivities about data (e.g., privacy and regulation) that must be tackled. The magnitude of data involved could also be enormous. For the customers, the data challenges stem from their need to bring efficiency to operations, yet control their differentiation from their competition and meet all the regulatory requirements. We can clearly see the connections to problems of interest to the IIWeb community.

## Aim of the Workshop

This workshop, *seventh in the IIWeb series*, is focused on making the research in information integration on the web more relevant to the challenges in services. The workshop will continue its traditional themes of interest, namely *integration architectures, information extraction, web object extraction, record linkage, named entity extraction, source meta-data learning, query execution and optimization*. However, we will give special emphasis to how this can be applied to services and service integration. After information is extracted and integrated, how should it be organized? How can we build or exploit *models, taxonomies or ontologies; trust models for information on the web, service specification in multi-model environment, searching for services, impact of data changes on composition and runtime adaptation, scale-up of methods to web levels*.

The workshop will provide a platform to discuss research directions, share experience and insights from both academia and industry. The anticipated outcome of the workshop is to assess the state of the art in the area, as well as to identify critical next steps to pursue in this topic. As

information integration and services is interdisciplinary in nature, its researchers have spanned the related areas of data mining, machine learning, databases, information retrieval, semantic Web, Web services, and others.

There are 6 papers selected for full presentation and 4 for short presentation. We hope you enjoy the workshop.

**Co-Chairs:**

Biplav Srivastava  
IBM T.J. Watson Research Center, Hawthorne, USA  
Email: [sbiplav@in.ibm.com](mailto:sbiplav@in.ibm.com), Ph: +1-914-784-7894

Ullas Nambiar  
IBM India Research Lab, New Delhi, India,  
Email: [ubnambiar@in.ibm.com](mailto:ubnambiar@in.ibm.com), Ph: +91-11-4129-2231

Craig Knoblock  
ISI, University of Southern California, USA.  
Email: [knoblock@isi.edu](mailto:knoblock@isi.edu), Ph: +1-310- 448-8786

**Steering Committee:**

- Subbarao Kambhampati, Arizona State University, USA
- Rakesh Mohan, IBM Research, USA

**Program Committee:**

- Avigdor Gal, Technion - Israel Institute of Technology
- Biplav Srivastava, IBM Watson Research Lab
- Chen Li, University of California, Irvine
- Craig Knoblock, Univ of Southern California
- Eran Toch, Carnegie-Mellon University
- Felix Naumann, Hasso Plattner Institut, Potsdam
- Gautam Das, University of Texas, Arlington
- Kevin Chang, University of Illinois at Urbana-Champaign
- L Venkat Subramaniam, IBM India Research Lab
- Lise Getoor, University of Maryland, College Park
- Louiqa Raschid, University of Maryland, College Park
- Michael Sheng, Univ of Adelaide, Australia
- Mong Li Lee, National University of Singapore
- Prasan Roy, Aster Data Systems
- Prashant Doshi, University of Georgia
- Rakesh Mohan, IBM Watson Research Lab
- Richard Goodwin, IBM Watson Research Lab
- Subbarao Kambhampati, Arizona State University
- Thomas Lee, The Wharton School, Univ of Pennsylvania
- Ullas Nambiar, IBM India Research Lab
- Vikram Pudi, IIIT Hyderabad
- Weiyi Meng, Binghamton University, NY
- Yuan Chi Chang, IBM Watson Research Lab
- Zaiqing Nie, Microsoft Research Asia, China

## History of the Workshop

IIWeb-09 will follow the tradition established by the following six workshops, each having a different focus theme each time and held at premier venues across related communities.

- **Intelligent Information Integration (III 99)**, held at **IJCAI 1999** **Co-chairs:** D. Fensel, C.A. Knoblock, N. Kushmerick and M.C. Rousset  
URL: <http://www.sigmod.org/dblp/db/conf/ijcai/ijcai99iii.html>
- **AI and Information Integration**, held at **AAAI 1998** **Co-chairs:** Craig Knoblock and Alon Levy  
URL: <http://www.aaai.org/Library/Workshops/ws98-14.php>
- **IIWeb 2003**, held at **IJCAI 2003**, received 40 submissions (33 accepted), attracted 45 attendees. **Co-chairs:** Craig Knoblock and Subbarao Kambhampati,  
URL: <http://www.isi.edu/info-agents/workshops/ijcai03/iiweb.html>
- **IIWeb 2004**, held at **VLDB 2004**, received 45 submissions (25 accepted), attracted 45 attendees. **Co-chairs:** Hasan Davulcu and Nick Kushmerick.  
URL: <http://cips.eas.asu.edu/iiweb.htm>
- **IIWeb 2006**, held at **WWW 2006**. **Co-chairs:** Kevin Chang and Avigdor Gal  
URL: <http://iiweb2006.cs.uiuc.edu/>
- **IIWeb 2007**, held at **AAAI 2007**, received 24 submissions, (12 accepted), attracted 30 attendees. **Co-chairs:** Ullas Nambiar and Zaiqing Nie  
URL: <http://research.microsoft.com/users/znie/IIWeb07/>

# Table of Content

## Full Papers:

1. Discovering and Learning Semantic Models of Online Sources for Information Integration;  
Jos'e Luis Ambite, Bora Gazen, Craig A. Knoblock, Kristina Lerman, Thomas Russ;  
Contact: knoblock@isi.edu
2. Towards Scalable Information Integration with Instance Coreferences;  
Abir Qasem, Dimitre A. Dimitrov, Jeff Heflin;  
Contact: abir.qasem@gmail.com
3. Unstructured information integration through data-driven similarity discovery;  
Rema Ananthanarayanan, Sreeram Balakrishnan, Berthold Reinwald, Yuen Yee;  
Contact: arema@in.ibm.com
4. Combining Multiple Query Interface Matchers Using Dempster-Shafer Theory of Evidence;  
Jun Hong, Zhongtian He and David A. Bell;  
Contact: j.hong@qub.ac.uk
5. Cross document person name disambiguation using entity profiles;  
Harish Srinivasan, John Chen, Rohini Srihari;  
Contact: hsrinivasan@janyainc.com
6. Challenges in Moving from Documents to Information Web for Services;  
Rakesh Mohan, Biplav Srivastava, Pietro Mazzoleni, Richard Goodwin;  
Contact: sbiplav@in.ibm.com

## Short Papers:

7. A Flexible Machine Learning Framework with User Interaction for Ontology Matching;  
Hoai-Viet To, Ryutaro Ichise, and Hoai-Bac Le;  
Contact: thviet@fit.hcmuns.edu.vn
8. Flexible query formulation for federated search;  
Matthew Michelson, Sofus A. Macskassy, and Steven N. Minton;  
Contact: mmichelson@fetch.com
9. On Developing Service-Oriented Web Applications;  
Sabah S. Al-Fedaghi;  
Contact: sabah@alfedaghi.com
10. Composition of Services that Share an Infinite-State Blackboard;  
Fabio Patrizi and Giuseppe De Giacomo;  
Contact: fabio.patrizi@dis.uniroma1.it