

Preface

Welcome to the Sixth International Workshop on Information Integration on the Web (IIWeb'07), organized in conjunction with AAAI-07, at Vancouver, British Columbia, Canada on 23 July, 2007.

Information Integration on the Web

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. Most integration solutions have assumed structured sources with the heterogeneity being introduced by the variety in source schemas and data models. But a large portion of the web consists of pages that contain information presented as unstructured text, such as blogs, wikis, reviews, and so on. Therefore, integration systems that can match entities/objects across both structured and unstructured sources are the need of the hour. Recent research in web object extraction, record linkage and named entity recognition have generated some initial solutions. However, many challenges remain in developing such a system.

Aim of the Workshop

This workshop, sixth in the IIWeb series, was proposed to bring together researchers looking to effectively integrate information from both structured and unstructured sources of information. The anticipated outcome of the workshop was to assess the state of the art in the area, as well as to identify critical next steps to pursue in this topic. Accordingly we asked submissions in topics such as: automatic wrapper induction, schema matching, Web entity extraction and search, Record linkage and object consolidation, DB and IR integration, applications & experiences.

The call attracted 24 submissions out of which 12 were selected for full presentation and 6 for short presentation. The accepted submissions have been classified into four broad topics:

- Novel Integration Architectures and Web Applications
- Online Information Integration
- Name Disambiguation and Entity Resolution
- Semantic Data Integration and Schema Matching

We are very indebted to all program committee members who helped us organize the workshop and reviewed the papers very carefully. We would also like to thank all the authors who submitted their papers to the workshop; they provided us with an excellent workshop program. We would also like to thank the AAAI staff for helping us with logistics for the workshop. A very special thanks to our sponsors — AAAI, Microsoft Research, Yahoo Research, and QL2 Software, Inc., for supporting the workshop.

We hope you enjoy the workshop.

– Ullas Nambiar (*IBM India Research Lab, New Delhi, India*)
– Zaiqing Nie (*Microsoft Research Asia, Beijing, P.R. China*)