

Workshop on Resources and Evaluation for Identity Matching, Entity Resolution and Entity Management

Call for Papers

Structured repositories of data about people are being created through information extraction from unstructured text as well as from sources that may themselves be structured documents such as passports or customer transactions. Problems arise in managing these structured repositories and integrating information from diverse sources. For example, newly added information must be consistent with existing information, must avoid duplication, and must be associated with an existing entity when that is appropriate. Researchers have addressed these problems in different contexts with goals such as name or record matching, identity resolution, and entity disambiguation. In this workshop, researchers with different perspectives will focus on the development of resources, algorithms and evaluation methodologies to improve the technology for managing structured repositories of identity data.

Evaluation measures for tasks that integrate person information are especially challenging. Whereas it is generally accepted that entity extraction systems can be evaluated using MUC scoring metrics, the case is less clear for “follow-on” technologies. Even a seemingly simple task such as matching person names in a database context is deceptively complex, and although measures like precision and recall have been used to evaluate **name** and **record** matching, there are methodological issues to resolve before we can refer to a “standard” evaluation methodology for this task. Moreover, it is much less clear how to effectively evaluate **identity** matching, resolution, and management systems, or even what it means to perform an effective identity match, particularly in the context of data containing identity attributes of varying quality and in which we have varying degrees of confidence.

We solicit papers that address the following areas:

1. Position papers which:
 - Discuss metrics for evaluation of the above-mentioned technologies
 - Discuss resources that can be brought to bear on these tasks
 - Descriptions of the use cases for these technologies and/or discussion of which evaluation methods are appropriate for different use cases
 - Discuss approaches for dealing with uncertainty and assigning confidence values, particularly with respect to manual ground truth annotation, system output and evaluation metrics
2. Papers discussing state-of-the-art systems and resources for performing or evaluating name and record matching, identity resolution, and identity management. Suggested topics include, but are not limited to:
 - Research in name / record matching
 - Research in entity disambiguation / resolution / deconfliction
 - Descriptions of systems or best practices for entity management
 - Evaluation of record matching, entity disambiguation, or entity management in different use cases
 - Resources for identity matching, identity resolution, or identity management which can contribute to improvements in system performance
 - Resources for evaluation of these technologies

Although we are distributing a call for papers to be presented, our vision is to organize an interactive and dynamic event rather than a “mini-conference.” Presentations will serve to introduce topics for discussion and to shape the day. However, a significant portion of time will be devoted to interactive exercises, brainstorming, and other “work” activities. The workshop will be organized such that all attendees (including the organizers) come out better informed about the problems associated with managing structured data and with initial ideas on a methodology for evaluating techniques designed to solve these problems. In particular, we hope to have made progress toward an understanding of contextualized and principled evaluation of identity matching, resolution, and management systems, and to connect a group of researchers interested in collaborating on this research

Organizing Committee:

Keith J. Miller (The MITRE Corporation)
Mark Arehart (The MITRE Corporation)
Sherri Condon (The MITRE Corporation)
Jason Duncan (U.S. Department of Defense)
Louise Guthrie (University of Sheffield)
Richard Lutz (The MITRE Corporation)
Massimo Poesio (Universita' di Trento)

Important Dates:

Deadline for Paper Submission: Friday, 8 February 2008
Notification to Authors: Friday, 29 February 2008
Submission of Final Version: Friday, 14 March 2008
Workshop at LREC 2008: Saturday, 31 May 2008

Papers should be submitted in MS Word, PDF, or text (ASCII) format to Keith J. Miller at keith@mitre.org, with a copy to Mark Arehart at marehart@mitre.org. Please include the text “LREC Workshop Submission” in your subject line.