Ontology-Driven Automatic Entity Disambiguation in Unstructured Text

by

JOSEPH EDWARD HASSELL IV

(Under the Direction of I. Budak Arpinar)

ABSTRACT

Precisely identifying entities in web documents is essential for document indexing, web search and data integration. Entity disambiguation is the challenge of determining the correct entity out of various candidate entities. Our novel method utilizes background knowledge in the form of a populated ontology. Additionally, it does not rely on the existence of any structure in a document or the appearance of data items that can provide strong evidence, such as e-mail addresses, for disambiguating authors for example. Originality of our method is demonstrated in the way it uses different relationships in a document as well as in the ontology to provide clues in determining the correct entity. We demonstrate the applicability of our method by disambiguating authors in a collection of DBWorld posts using a large scale, real-world ontology extracted from the DBLP. The precision and recall measurements provide encouraging results.

INDEX WORDS: Entity disambiguation, ontology, semantic web.

Ontology-Driven Automatic Entity Disambiguation in Unstructured Text

by

JOSEPH EDWARD HASSELL IV

B.S., Columbus State University, 2003

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillments of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2006

© 2006

Joseph Edward Hassell IV

All Rights Reserved

Ontology-Driven Automatic Entity Disambiguation in Unstructured Text

by

JOSEPH EDWARD HASSELL IV

Major Professor: I. Budak Arpinar

Committee: John Miller
Amit Sheth

Electronic Version Approved:

Maurren Grasso
Dean of the Graduate School
The University of Georgia
August 2006

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

1. Introduction

A significant problem with the World Wide Web today is that there is no explicit semantic information about the data and objects being presented in the web pages. Most of the content encoded in HTML format serves its purpose of describing the presentation of the information to be displayed to human users. HTML lacks the ability to semantically express or indicate that specific pieces of content refer to real-world named entities or concepts. For instance, if “George Bush” is mentioned on a web page, there is no way for a computer to identify which “George Bush” the document is referring to or even if “George Bush” is the name of a person.

The Semantic Web aims at solving this problem by providing an underlying mechanism to add semantic metadata to any content, such as web pages. However, an issue that the Semantic Web currently faces is that there is not enough semantically annotated web content available. One aspect of adding semantic metadata is that of stating an explicit relationship from each appearance of named entities within a document to some identifier or reference to the entity itself. The architecture of the Semantic Web relies upon URIs [Berners-Lee, 2005] for this purpose. An example of this would be the entity “UGA” pointing to http://www.uga.edu and “George Bush” pointing to a URL of his official web page at the White House. However, the most benefit can be obtained by referring to actual entities of an ontology where such entities would be related to concepts and/or other entities. The problem that arises is that of entity disambiguation, which is concerned with determining the right entity within a

document out of various possibilities due to same syntactical name match. For example, “A. Joshi” is ambiguous due to various real-world entities (i.e. computer scientists) having the same name.

Entity disambiguation is an important research area within Computer Science. The more information that is gathered and merged, the more important it is for this information to accurately reflect the objects they are referring to. It is a challenge in part due to the difficulty of exploiting, or lack of background knowledge about the entities involved. If a human is asked to determine the correct entities mentioned within that document, s/he would have to rely upon some background knowledge accumulated over time from other documents, experiences, etc. The research problem that we are addressing is how to exploit background knowledge for entity disambiguation, which is quite complicated particularly when the only available information is an initial and last name of a person. In fact, this type of information is already available on the World Wide Web in databases, ontologies or other forms of knowledge bases. Our method utilizes background knowledge stored in the form of an ontology to pinpoint, with high accuracy, the correct object in the ontology that a document refers to. Consider a web page with a “Call for Papers” announcement where various researchers are listed as part of the Program Committee. The name of each of them can be linked to their respective homepage or other known identifiers maintained elsewhere, such as the DBLP bibliography server. Our approach for entity disambiguation is targeted at solving this type of problem, as opposed to entity disambiguation in databases which aims at determining similarity of attributes from different database schemas to be merged and identifying which record instances refer to the same entity (i.e., [Dey, 2002]).

The contributions of our work are two-fold:

1. A novel method to disambiguate entities within unstructured text by using clues in the text and exploiting metadata from an ontology, and
2. An implementation of our method that uses a very large, real-world ontology to demonstrate effective entity disambiguation in the domain of Computer Science researchers.

According to our knowledge, our method is the first work of its type to exploit an ontology and use relations within this ontology to recognize entities without relying on structure of the document. We show that our method can determine the correct entities mentioned in a document with high accuracy by comparing to a manually created and disambiguated dataset.

2. Background

In this chapter, we will discuss background information for our system. In section 2.1, we present the ontology and describe its uses in today's applications. Section 2.2 describes the vision of the Semantic Web and presents some scenarios where the Semantic Web would be greatly beneficial to a user. In section 2.3, we introduce RDF [RDF, 2006] and explain its significance in the representation of an ontology. Section 2.4 explains why we chose to use the Sesame repository as opposed to other leading ontology repositories.

2.1. The Ontology

An ontology, in computer science terms, is created as an enabling technology for the sharing and manipulation of information. It is a body of information that contains conceptualizations of real world objects, concepts and other entities representing some domain [Gruber, 1993]. One of its key benefits is the ability to explicitly express relationships between each entity within the ontology. Figure 1 illustrates an example ontology where each blue object is an entity. The connections between the entities represent relationships which can be named.

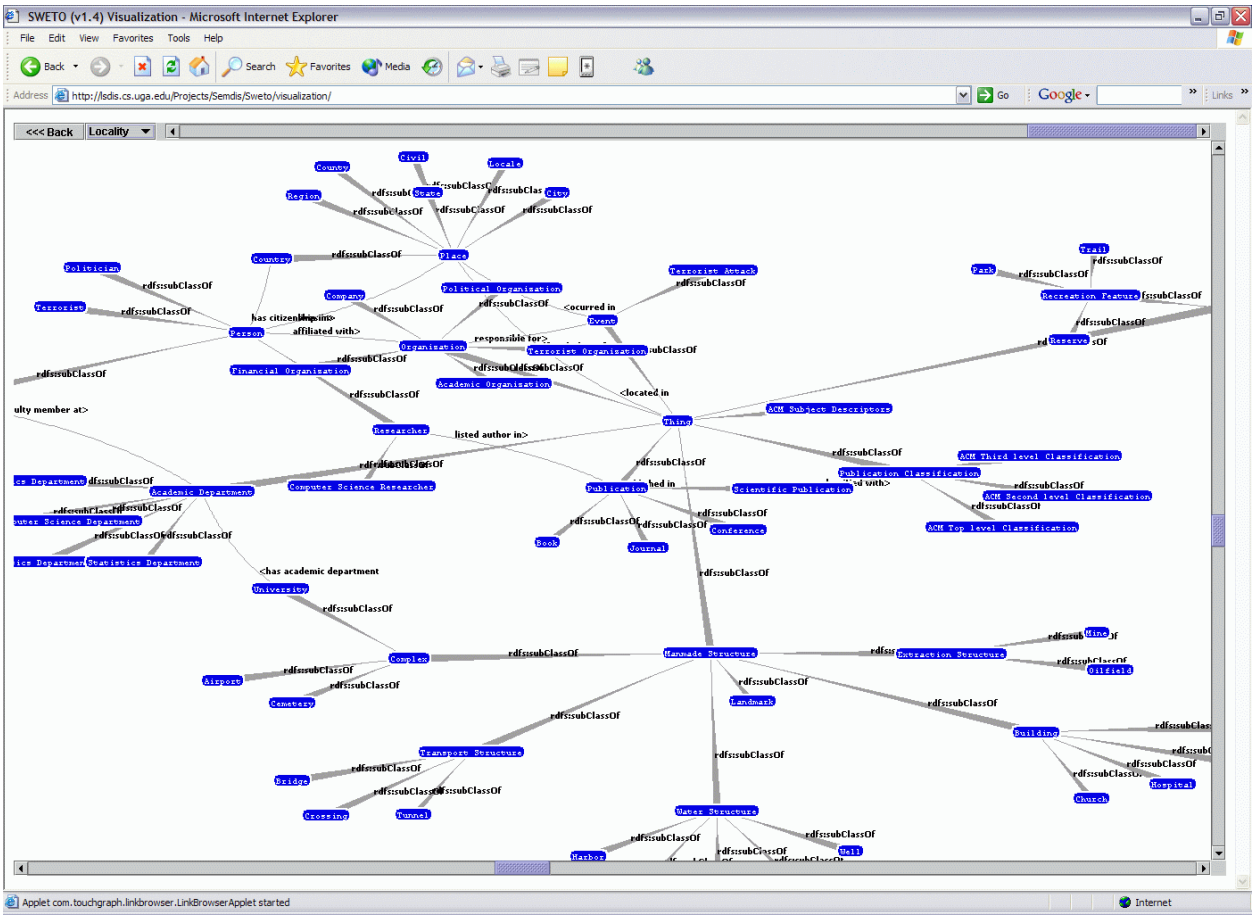

Figure 1: Example of an Ontology

The use of ontologies is not specific to the computer science research domain. They are heavily used in areas such as life sciences, chemistry, health care and national security. [Sheth, 2005] explains the applicability of these methods used in today's market of information and technology. Many of these ontologies are updated on a regular basis and are publicly available via the World Wide Web. For example, the TAP knowledge base is a public ontology maintained by Stanford University [Guha, 2003]. It is a shallow but broad knowledge base that consists of the lexical and taxonomic information of a wide range of popular objects. The main goal of the TAP ontology is to help bootstrap (provide simple data to help start the movement) the Semantic Web with a wide variety

of general information. SWETO is an ontology created by the LSDIS Lab at the University of Georgia [Aleman, 2004]. It is a heavily maintained ontology containing millions of real world facts and is designed to be a testing tool for large-scale algorithms including the discovery of semantic associations. As of April 2006, this ontology consists of 811,819 entities and 1,545,320 relationships between these entities. GlyCo is an ontology also created by the LSDIS Lab and is a focused domain ontology designed for the glycochemistry domain [Sahoo, 2006]. It contains knowledge of concepts and relationships that are crucial in biosynthesis and glycosylation processes of complex carbohydrates and proteins. As of April 2006, the GlyCo ontology consists of 573 classes and 113 types of named relationships.

2.2. The Semantic Web

The World Wide Web is a huge corpus of information that humans can read and understand easily. Imagine the possibilities if a computer could also read and process everything that the World Wide Web had to offer. This would greatly increase the usefulness and productivity of computers in everyday life. The vision of the Semantic Web is to achieve this goal and aim at making the World Wide Web readable and processable to machines by using a universal medium for data exchange that allows the possibility for computers to process the meaning of information. By doing this, information can be automatically exchanged, understood and processed by computers with minimal human intervention.

In *"*Semantic Web*"* [Berners-Lee, 2001], the author depicts a few scenarios where the Semantic Web comes into play in everyday life. In these scenarios, a

“computer agent”, which is a theoretical personal device much like a palm pilot, is able to access the Internet and lookup information automatically by searching out appropriate information and performing complex task, ultimately returning detailed information back to the user’s agent. The information returned to the user is the result of the agent accessing the Internet, finding data that has been posted on the Semantic Web and processing it automatically without any human intervention. The agent is able to process and understand this information on the Semantic Web because it is available in machine understandable format. Thus, the agent displays the relevant information as results to the user.

Another very interesting example that demonstrates the wide range of possibilities of the Semantic Web is Haystack [Quan, 2004]. Haystack is a Semantic Web browser that illuminates the user's control over information being displayed. The user has direct control over what information is to be displayed as well as the layout of this information. Haystack's goal is to have, “an end user application that automatically locates metadata and assembles point-and-click interfaces from a combination of relevant information, ontological specifications, and presentation knowledge, all described in RDF and retrieved dynamically from the Semantic Web.” Haystack is aimed at newcomers to the Semantic Web and presents them with a way to discover and explore this technology.

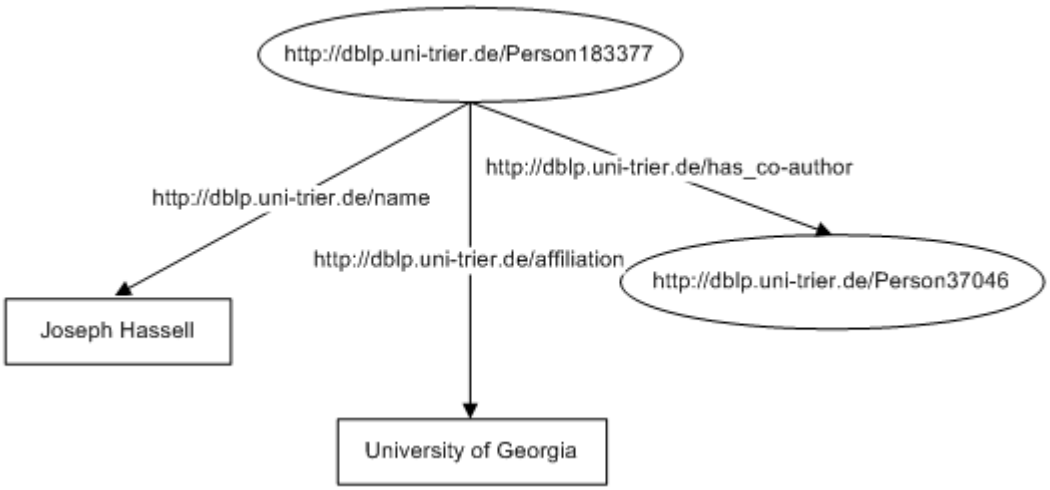

Figure 2: Example of a Simple RDF Graph

2.3. RDF

To represent an ontology using a computer, a standard format must be used that is human understandable as well as machine processable. This is why the Resource Description Framework [RDF, 2006], also known as RDF, was created. In RDF everything is represented as a triple in the form of subject, predicate and object. This closely resembles a graph with vertices and edges. The entities within the RDF coincide with the vertices of a graph, while the relationships within the RDF coincide with the edges of a graph. Each entity within the RDF may have special attributes associated with it, such as label, date, size, etc. Some RDF graphs take advantage of a taxonomy by incorporating this representation into their structure while still having the possibility of relationships between any entity in the RDF. Relationships in the RDF model are first-class objects that can be created without the constraint of being a sub-object. RDF data can be serialized in XML syntax so it inherits the standards set by its overseer, the World Wide Web Consortium. It is a centerpiece of the Semantic Web because of its

ability to describe different resources that can be found on the World Wide Web. Many other fields are making heavy use of RDF because of its flexibility with the representation of abstract concepts and its universal applicability.

The reason we are interested in RDF instead of a traditional relational database is because of the ability to explicitly represent relationships between the entities. These relationships can be named, which is an advantage of RDF over a relational database. Although this can be done in a relational database scheme, RDF is tailored toward this type of representation and naturally works with our needs of explicitly representing our named relationships between entities. If we were to use a relational database, each relationship expressed in the ontology would be represented in its own table and would have a complicated and confusing configuration of primary and foreign keys. Using a method of this nature would be extremely inefficient compared to using RDF because of the large number of join operations that would be needed when using data from multiple tables.

2.4. RDF Repository

An RDF repository is much like a relational database system except that it acts as an ontology and keeps track of all attributes and relationships between the objects. The repository is in charge of maintaining the information and providing ways to access it via a query language. When making the decision of what system to use as a RDF repository, there are several options to consider such as Jena [Carroll, 2004], Sesame [Broekstra, 2002] and Brahms [Janik, 2005]. We chose to use Sesame because of its ability to store large amounts of information by not being dependant on memory storage

alone. Although Brahms is designed to run extremely efficient on very large knowledge bases, its current version does not support updates to the repository.

Sesame has three modes in which it stores its information so that it can cater to many different scenarios:

- In-memory – Stores everything in RAM.
- Native – Stores everything in proprietary files on disk.
- Database – Uses a database as a backbone to store information.

Although memory mode would be faster, the large size of our DBLP dataset (described later in Section 3.1) warrants use of the native mode repository because it is typically too large to fit into memory and using the database option has proved to be too slow in update operations. The performance of native mode is somewhere in the middle of that of database and in-memory modes.

3. Dataset

In this chapter, we discuss the datasets that we have chosen to use in our evaluations of our method. In Section 3.1, we discuss the DBLP dataset [Ley, 2002] and explain how we converted XML data set into a RDF repository for use with our system. Section 3.2 explains the DBWorld [DBWorld, 2006] website that we chose to populate our corpus of documents with. Our dataset consist of two parts: an ontology, which consist of the converted DBLP information, and a corpus of documents that we evaluate our system with, which is the collection of DBWorld posts. We chose the DBLP dataset because it is a rich source of information in the Computer Science domain and DBWorld because it contains documents which include names of people which typically exist in DBLP, which often require entity disambiguation of authors.

3.1. DBLP

Our goal is to demonstrate real-world applicability of our approach; therefore, we use a real-world dataset. We chose to use the DBLP dataset which has been around since the 1980's. This is a web site that contains bibliographic information for computer science researchers, journals and proceedings. Currently, it indexes more than 725,000 articles and contains several thousand links to home pages of computer scientists. The DBLP dataset is used everyday by researchers around the world. Conveniently, the site provides two XML files that contain information stored in its servers. One of the files contains objects such as authors, proceedings and journals.

The other file contains lists of papers usually organized by tracks or sessions of the conference or workshop where they were presented. We have taken the former and converted it into RDF. The resulting RDF is very large, approximately one and a half gigabytes. It contains 3,079,414 entities and 447,121 of these are authors from around the world. Below is a table that breaks down the main instances per class.

Table 1: Instances of Classes in the DBLP Ontology

Authors	447,121
Journal Articles	262,562
Articles in Proceedings	445,530

The conversion to RDF was designed to create entities out of peoples' names, instead of just treating the names as literal values that are a part of the metadata of a publication. For this reason, we did not make use of other available RDF-converted data of DBLP, such as http://www.semanticweb.org/library/#dblp. Additionally, the data in RDF is enriched by adding relationships to affiliations (i.e., universities) and research topics for researchers. For further details, see http://lsdis.cs.uga.edu/~aleman/data/. The quality of the data in DBLP is continually improving yet there are still cases of different peoples' names appearing as the same person. Dealing with this situation is outside the scope of this paper.

3.1.1. Creation of Ontology from DBLP

We use a SAX parser to process the XML file and feed its information into a Sesame repository. We chose to use the SAX parser because of its unique ability to process an

XML file line by line. This is a distinct advantage, compared to other methods such as a DOM parser, because of the large size of our dataset. Using a DOM parser would require far too much memory in this situation. Because of the structure of the DBLP information, two passes were necessary for our conversion to take place. The first pass seeks out all authors who have specific information in the DBLP dataset and stores this information into the ontology while making a mapping of name to URI. Conveniently, some authors with the same name have been labeled with a unique number assigned and appended to them for separation. We take advantage of this by creating two or more separate entities within the ontology. The second pass processes the rest of the XML file, which consists of papers, journals, etc., and populates the ontology with this information.

During the conversion from XML to RDF, we keep track of certain aspects of the DBLP dataset. Mainly, we explicitly create relationships between authors, such as co-authors. Due to the possibility that some documents we will be searching will not contain all of the correct international characters, we have chosen to put in some alternate spellings for the authors' names that will eliminate false negatives due to an accented letter, etc. For example, if our system comes across the name "María del Carmen Bañuls", our method will store an alternate spelling of this as "Maria del Carmen Banuls".

In addition to explicitly adding relationships to the repository, we take advantage of external information regarding authors' affiliations and areas of interest. In "Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection" [Aleman, 2006], the authors created a dataset containing this

information which we have access to. So, as we add authors to the repository, we also check them against this dataset to see if any external information can be included in the form of affiliations or areas of interest of the authors.

Once all of this information has been stored into the Sesame repository, we export this information to RDF format. Thus, the information of DBLP in XML format is successfully converted into an ontology (following the RDF model) whose domain is Computer Science bibliography. This ontology is an ideal input to our system because it contains valuable information that will help in the disambiguation process.

3.2. DBWorld

DBWorld is a mailing list of information for mainly upcoming conferences related to the databases field. Although it does contain some random post about open positions etc., we are only interested in postings about conferences, workshops, and symposiums.

We have created a HTML scraper that visits the DBWorld web site and downloads only the posts that we are interested in. To be precise, it only downloads posts that contain “Call for Papers”, “Call for Participation” or “CFP” in the subject. Our system disambiguates the people listed in these postings and provides a URI to the corresponding entity in the ontology.

A DBWorld post typically contains an introduction, topics of interest, important dates and a list of committee members. The general layout of the DBWorld post is rarely consistent in terms of its structure. For example, sometimes the participants of a conference are listed with their school or company affiliation and sometimes they are listed with the name of their country.

4. Approach

In our approach, different relationships in the ontology provide clues on determining the correct entity out of various possible matches. We argue that rich semantic metadata representations allow a variety of ways to describe a resource. We characterize several relationship types that we identified and explain how they contribute towards the disambiguation process. As mentioned, we use the scenario of disambiguating researchers' names in DBWorld postings. However, we believe that the following relationship types are applicable to other scenarios, such as disambiguating actor names in movie reviews.

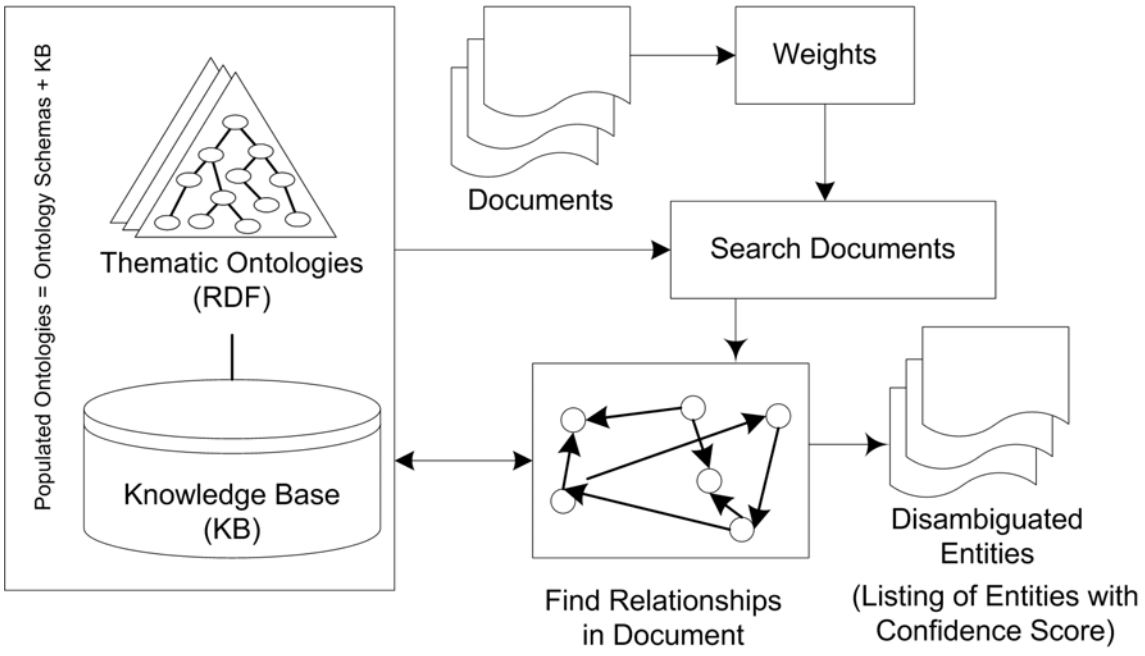

Figure 3: Overview of Our System

4.1. Entity Names

An ontology contains a variety of concepts and instance data. The first step of our approach is specifying which entities from a populated ontology are to be spotted in text and later disambiguated. To do this, it is necessary to indicate which literal property is the one that contains the 'name' of entities to be spotted. In most cases, such a literal property would be 'rdfs:label.' However, in some cases, additional 'name' properties may need to be listed, such as aliases and alternate names. Additionally, a different ontology may have its own way of representing the name for each entity.

4.2. Text-proximity Relationships

Various relationships contain metadata that can be expected to be in 'text-proximity' of the entity to be disambiguated. For example, affiliation data commonly appears near researchers' names in DBWorld posts. Thus, when the known affiliation (from the ontology) appears 'near' an entity, there is an increased likelihood that this entity is the correct entity that the text refers to. This 'nearness' is measured by the number of character spaces between two objects. Figure 4 illustrates an example where the affiliation "Stanford University" appears next to the entity of interest, "Michael Kassoff", whose known affiliation is "Stanford University" according to the populated DBLP ontology. We acknowledge the fact that the up to date status of an ontology can have an impact on the quality of disambiguation results yet measuring the degree of such impact is outside the scope of this paper.

Figure 4: Snippet from a DBWorld Post

4.3. Text Co-occurrence Relationships

Text co-occurrence relationships are similar to text-proximity relationships with the exception that 'proximity' is not relevant. For example, the intuition of using affiliation data is applicable as long as it appears 'near' an entity, but it would not be relevant if it appears somewhere else in the text because it could be the affiliation of a different entity (or referring to something else). Text co-occurrence relationships are intended to specify data items that when appearing in the same document, provide clues regarding the correct entity being referred to in the text. For example, in DBWorld posts, the listed 'topics' fit the idea of text co-occurrence relationships. Figure 5 shows a portion of the same document in Figure 4, where "Web mining" and "Semantic Web" are spotted and are both areas of interest that match research topics related to "Michael Kassoff." Thus, by specifying the text co-occurrence relationship, specific metadata contained in the ontology helps disambiguate the correct person, depending on the topics mentioned in the text.

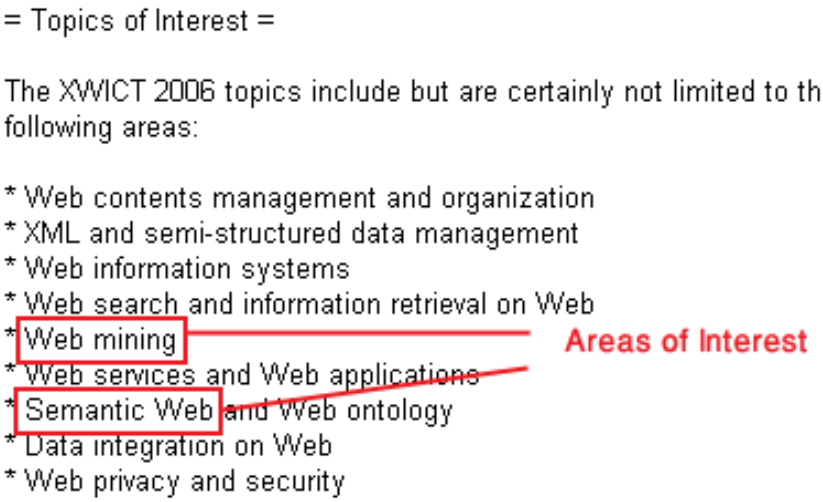

Figure 5: Snippet from the Same DBWorld Post in Figure 4

It is important to mention that this co-occurrence relationship is applicable only on well-focused content. That is, if a document contains multiple DBWorld postings (which is highly unlikely), then their content could bring 'noise' and negatively impact the results of the disambiguation process. In such cases, it may be necessary to perform a text-segmentation process [Embley, 1999, Zhang, 2006] to separate and deal with specific subparts of a document.

4.4. Popular Entities

The intuition behind using popular entities is to specify relationships, which have more frequent occurrences, that will bias the right entity to be the most popular entity. For example, "A. Joshi" is a name that matches up to 20 entities in DBLP but only a couple of authors with that name have published more than 70 papers. Arguably, the most prolific of the "A. Joshi" names would be more likely to appear in texts such as DBWorld posts. Indeed, the rationale for this measure fits very well in some scenarios. For

example, most researchers that are listed in program committees of conferences typically have a relatively higher number of publications compared to researchers who do not appear in any program committees. This aspect for entity-disambiguation should be used with care, depending on the domain.

4.5. Semantic Relationships

Semantic relationships are intended to consider relationships that go beyond metadata which consists of literal values, such as syntactical matching of peoples' names [Bilenko, 2003]. For example, researchers are related to other researchers by means of their collaboration network. Researchers are also closely related to their co-authors and other authors through complex relationships. In DBWorld posts, it is common that listed people have relationships among themselves within a list of accepted papers and/or program committee members of a conference. Thus, the semantic relationship helps with determining the correct entity being referred in the text.

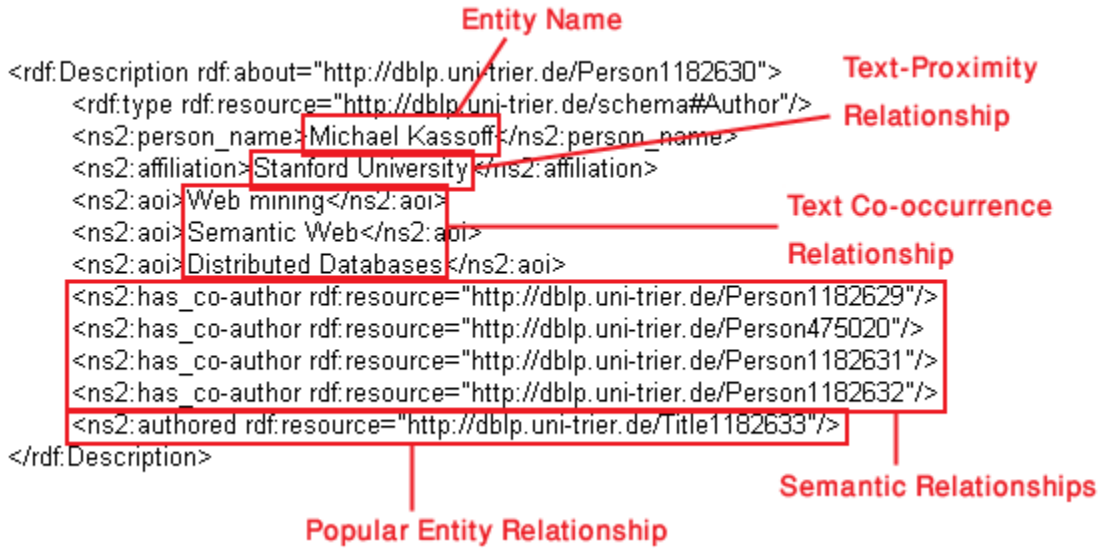

Figure 6: Sample RDF Object

In Figure 6, we present a part of our DBLP RDF file, which is an input to the system for entity disambiguation. In this example, the entity's name is "Michael Kassoff" who is affiliated with "Stanford University" and has authored one paper. The author has three areas of interest and is related to four other authors via the semantic relationships (i.e. has_co-author).

5. Algorithm

In this section, we describe our method for disambiguating entities in unstructured text. Figure 7 explains the steps of our method using pseudocode. The general idea is to spot entity names in text and then assign each potential match a confidence score. The confidence score for each ambiguous entity is adjusted based on whether existing information of the entity from the ontology matches accordingly to the relationship types found in the ontology as explained in the previous section. Throughout this paper, we will use *cf* to represent the initial confidence score, *acf* to represent the initial, abbreviated confidence score, *pr* to represent proximity score, *co* to represent text co-occurrence score, *sr* to represent the semantic relationship score and *pe* to represent the popular entity score. These variables are adjustable to capture the relative importance of each factor in the disambiguation process.

```
Algorithm Disambiguation( ) {
  for (each entity in ontology) {
    if (entity found in document) {
      create 'candidate entity'
      Cs  for 'candidate entity' ← cf / (entities in ontology)
    }
  }
  for (each 'candidate entity') {
    search for 'candidate entity's text proximity relationship
    if (text proximity relationship found near 'candidate entity'){
      Cs for 'candidate entity' ← Cs for 'candidate entity' + pr
    }
    search for 'candidate entity's text co-occurrence relationship
    if (text co-occurrence relationship found) {
      Cs for 'candidate entity' ← Cs for 'candidate entity' + co
    }
    if (ten or more popular entity relationships exist){
    {
       Cs for 'candidate entity' ← Cs for 'candidate entity' + pe
    }
  }
  iterate ← false
  while (!iterate) {
    iterate ← true
    for (each 'candidate entity') {
      search  for semantic relationships in the ontology to other 'candidate entities'
      for (each relation found that has not been seen AND
           target entity Cs is above 'threshold') {
        Cs for 'candidate entity' ← Cs for 'candidate entity' + sr
        mark relation as seen
        if ('candidate entity' score has risen above 'threshold') {
          iterate ← false
}}}}}
```

Figure 7: Algorithm Pseudocode

5.1. Spotting Entity Names

The first step of our algorithm consists of searching within a document for the names of the entities to be disambiguated (see Section 4.1). The system only looks for entity names within the ontology. Each entity name found in the document is a potential match

for one or more entities in the populated ontology. Each of the entities of the ontology that matches a name becomes a candidate entity that will help disambiguate the spotted entity in the text. A confidence score is initially assigned to each candidate entity depending on how many of them match the same name. The formula for assigning this confidence score (c_s) is as follows.

$$e_s = \frac{cf}{\text{Number of entities with the same label}} \quad (1)$$

Other methods for spotting names find anything that looks like a name, such as two words having their first letter capitalized. We did not choose these types of techniques to avoid spotting irrelevant information, which would have had to be filtered out later or result in low performance.

In addition to spotting based on name, our method also looks for abbreviated names, such as “A. Joshi”. These types of entities get a c_S that is initialized differently to reflect the fact that many more entities from the ontology can syntactically match to the same name. This aspect is not included in the pseudocode in Figure 7 for the sake of simplicity. The formula for assigning this confidence score is as follows.

$$e_s = \frac{acf}{\text{Number of related entities in the ontology}} \quad (2)$$

This consideration for abbreviated names is a feature that can be turned on or off. We found that it is suitable for use with peoples' names yet we did not explore further considerations such as canonical names (i.e., Tim and Timothy) and other techniques for name matching [Han, 2004, Torvik, 2005].

5.2. Spotting Literal Values of Text-proximity Relationships

The second step of our algorithm consists of spotting literal values based on text-proximity relationships (see Section 4.2). In order to narrow down the search for such literals, only the candidate entities found in the previous step are considered when determining literal values of text-proximity relationships to be spotted. By checking the ontology, it is then possible to determine whether a candidate entity appears near one of the spotted literal values based on text-proximity relationships, such as a known affiliation of a person appearing within a predefined window of the person's name. We argue that this type of evidence is a strong indication that it might be the right entity. Hence, the confidence-score of an entity is increased substantially. Figure 4 shows an example where the affiliation is a highly relevant hint toward the disambiguation of the candidate entity "Michael Kassoff."

5.3. Spotting Literal Values of Text Co-occurrence Relationships

This step consists of spotting literal values based on text co-occurrence relationships (see Section 4.3). For every candidate entity, if one of its text co-occurrence relationship values is found within the document, its confidence score is increased. In our DBLP dataset, this step finds literal values appearing in the document based on the relationship 'aoi' which contains areas of interest of a researcher. For example, in Figure 5 "Web mining" and "Semantic Web" are spotted as areas of interest that match those of candidate entities spotted already. Thus, any candidate entity having such areas of interest receives an increase on its disambiguation C_S.

5.4. Using Popular Entities

The degree of popularity among the candidate entities is considered to adjust the C_S of candidate entities (see Section 4.4). The intention is to slightly increase the C_S for those entities that, according to the ontology, have many relationships that were predefined (e.g. authored). In the scenario of DBWorld posts, this step increases the score of candidate entities that have many publications slightly, as indicated in the ontology. We acknowledge that this step may not be applicable in disambiguating other types of entities. However, we found that it is a useful tie-breaker for candidate entities that have the same C_S.

5.5. Using Semantic Relationships

This step goes beyond just using literal values as evidence for disambiguating entities. The intuition is to use relationships to create a propagation or network effect that can increase the C_S of candidate entities based on *semantic* relationships (see Section 4.5). In the scenario of disambiguating researchers in DBWorld posts, this step considers whether the candidate entities have co-authorship relationships and increases the C_S for the ones that do. Such C_S adjustments can only be done fairly by starting with the candidate entities having the highest score so far. Each candidate entity with a high score is analyzed through its semantic relationships in the ontology to increase the score of other candidate entities whenever they are connected. On the other hand, it may not be necessary to perform this analysis on candidate entities with very low C_S. To deal with this issue, our algorithm uses a threshold C_S which can be customized. Additionally, the process of adjusting C_S is repeated if at least one candidate entity's C_S increases over such threshold. Any such entity could then help boost the C_S of

remaining candidate entities with low scores until no more adjustments to c_S take place. Thus, this step is iterative and always converges.

```
<entity>
      <uri>http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/k/Kassoff:Michael.html</uri>
      <entityName>Michael Kassoff</entityName>
      <confidence>90</confidence>
      <charOffset>5688, 5703</charOffset>
</entity>

<entity>
      <uri>http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Schroeder:Michael.html</uri>
      <entityName>Michael Schroeder</entityName>
      <confidence>100</confidence>
      <charOffset>16241, 16259</charOffset>
</entity>

<entity>
      <uri>http://www.informatik.uni-trier.de/~ley/db/indices/a-tree/s/Schroeder_0002:Michael.html</uri>
      <entityName>Michael Schroeder</entityName>
      <confidence>45</confidence>
      <charOffset>16241, 16259</charOffset>
</entity>
```

Figure 8: Sample Output

5.6. Output

As shown in Figure 8, we have chosen to output our results in XML format because of its universally accepted syntax. For each entity found in the document and the ontology, we output its URI, name, confidence score and character offset. The URI of each entity represents the DBLP web page containing information regarding it. The name is the literal found in the documents and the character offset is the location of the entity within the document. This information is output in a generic form which can be used in many scenarios such as annotating documents with Microformats (microformats.org) or RDFa (w3.org/TR/xhtml-rdfa-primer/), populating knowledge bases, etc.

5.7. Data structures

Because we are using the native-mode repository in Sesame due to the large size of our DBLP dataset, our queries to the repository are slower than expected. To compensate for this, we have created several different data structures using Java objects, such as hashmaps where the entity name is used as the key and the URI is used as the value. By creating these data structures, we are able to perform the most common queries that access memory instead of having to access files on disk. This drastically decreases the amount of queries issued to the repository, which in turn improves performance of our method.

6. Evaluation

We chose to evaluate our method for entity disambiguation using a gold standard which we created manually and refer to as the disambiguated dataset. This dataset consists of 20 documents from DBWorld and is available at http://arches.uga.edu/~jhassell/data/testData.html. For the purpose of having a representative dataset, the documents were chosen by first picking a random DBWorld announcement and the next 19 documents as they were posted in chronological order. Each document is processed manually by inspecting peoples' names. For each name, we add a link to its corresponding DBLP web page which we use in the ontology as the URI that uniquely identifies a researcher. Ideally, every DBWorld post would have a gold standard representation but this does not exist because it is extremely time consuming to create. By creating this disambiguated dataset, it is possible to evaluate our method's results and measure precision and recall. To define precision and recall, we use a set *A* as the set of unique names identified using the disambiguated dataset and a set *B* as the set of entities found by our method. The intersection of these sets represents the set of entities correctly identified by our method. We measure precision as the proportion of correctly identified entities with regard to *B*.

$$Precision = \frac{sizeof(A \cap B)}{sizeof(B)} \quad (3)$$

We measure recall as the proportion of correctly disambiguated entities with regard to *A*.

$$\mathcal{Recall} = \frac{sizeof(A \cap B)}{sizeof(A)} \quad (4)$$

Our method computes the C_S of candidate entities using weights for the different disambiguation aspects in Section 5. These weights are part of the input settings that allow fine tuning depending on the domain and importance of available relationships in a given ontology much like [Anyanwu, 2005] uses a ranking system to distinguish important relationships. We adjusted the settings so that an entity's affiliation and relations (co-authorship) to other researchers are considered far more valuable than the areas of interest of the researcher. Table 2 lists the assignments that produced the most accurate results when running our test data:

Table 2: Input Values

Description	**Variable**	**Value**
charOffset		50
Text proximity relationships	*pr*	50
Text co-occurrence relationships	*co*	10
Popular entity score	*pe*	10
Semantic relationship	*sr*	20
Initial confidence score	*cf*	90
Initial abbreviated confidence score	*acf*	70
Threshold	*threshold*	90

Within our disambiguated set, we were able to find 758 entities that have representations in our ontology. In the 20 documents of our disambiguated set, only 17 entities were not represented in the DBLP ontology. These mainly consisted of local organizers and researchers listed in cross-disciplinary conferences. When comparing the results of our method with the disambiguated set, our method was able to find 620 entities. Only 18 of these were incorrectly disambiguated. Using formula (3), we calculated the precision to be 97.1 percent. To measure recall, we used formula (4) and calculated the recall to be 79.6 percent. Table 3 is a summary of our results.

Table 3: Precision and Recall

Correct Disambiguation	**Found Entities**	**Total Entities**	**Precision**	**Recall**
602	620	758	97.1%	79.4%

Figure 9 illustrates the precision and recall evaluation on a per document basis. The document numbers coincide with our disambiguated set available at http://arches.uga.edu/~jhassell/data/testData.html. The precision is quite accurate in most cases and the recall promising but varies from document to document.

There are several situations where our method did not disambiguate the correct entity. This was mostly due to the ontology which, although largely populated, does not have complete coverage. For example, some of the authors within the ontology have a single relationship to a paper while some authors have a variety of relationships to papers, other authors, affiliation, etc. Because of this, it was not possible to precisely disambiguate some entities. Another error that is common is the situation where we find

an entity's name that matches a portion of the name of another entity. We provide some safeguards against this as long as both of the candidate entities exist in the ontology, but the algorithm still misses in a few cases.

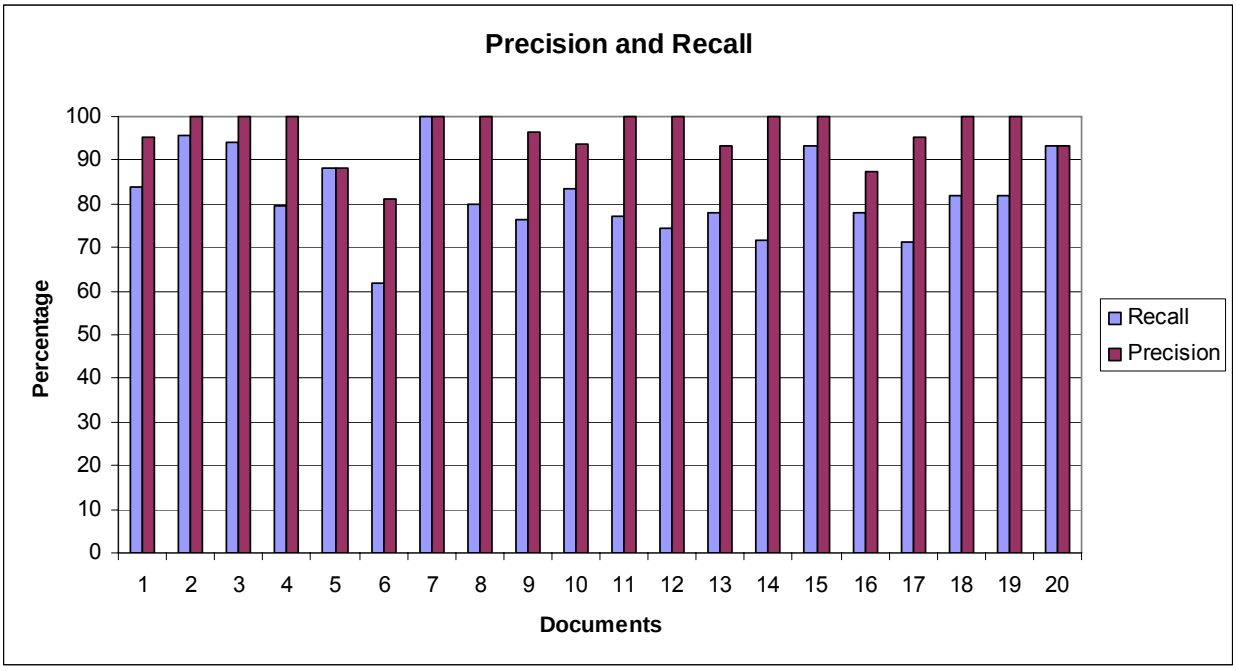

Figure 9:. Measures of Precision and Recall in a Per-Document Basis

Figure 10 is an example of a DBWorld post where the entities found within the document are highlighted. The bold entities represent names of people found in the DBLP ontology. The underlined entities represent an affiliation found next to an entity that it is affiliated with. The italic entities represent an abbreviated entity found in the document that represents an object within the ontology with a full name. For more examples of DBWorld documents and our system's output in regard to these documents, refer to the appendices of this paper.

S E C O N D C A L L F O R P A P E R S

5th ACM International Workshop on Data Engineering
for Wireless and Mobile Access (MobiDE'06)

Sunday, June 25, 2006, Chicago, Illinois, USA
(co-located with ACM SIGMOD/PODS 2006)

http://db.cs.pitt.edu/mobide06

AIMS & TOPICS OF INTEREST

Being the fifth in a series of successful workshops, MobiDE'06 aims to act as a bridge between the data management, wireless networking, and mobile computing communities.

The previous MobiDE workshops took place in Seattle, WA, in conjunction with MobiCom 1999; in Santa Barbara, CA, together with SIGMOD 2001; in San Diego, CA, together with MobiCom 2003; and in Baltimore, MD, in conjunction with SIGMOD 2005.

MobiDE'06 will serve as a forum where researchers and technologists discuss the state-of-the-art, present their contributions, and set future directions in the general area of data management for mobile and wireless access, including sensor networks. The specific focus of MobiDE'06 is on novel mobile applications and services including games, entertainment, and infotainment.

The topics of interest related to mobile and wireless data engineering include, but are not limited to, the following.

* ad-hoc networked databases
* consistency maintenance and management
* context-aware data access and query processing
* data caching, replication and view materialization
* data publication modes: push, broadcast, and multicast
* data server models and architectures
* database issues for moving objects: storage, indexing, etc.
* location-based and context-aware services
* m-commerce
* mobile agent models and languages
* mobile database security
* mobile databases in scientific, medical, and engineering applications
* mobile peer-to-peer applications and services
* mobile transaction models and management
* mobile web services
* mobility awareness and adaptability
* mobile services: in transportation, agriculture, industries, etc.
* pervasive computing
* prototype design of mobile data management systems
* quality of service for mobile data management
* sensor network data management
* small footprint data management
* transaction migration, recovery and commit processing
* wireless multimedia systems
* wireless web

As in the past, the workshop will be organized in a manner that fosters interaction and exchange of ideas among the participants. In addition to paper presentations, time will be allocated for open discussion forums, informal discussions or panels.

IMPORTANT DATES

Abstract Submissions:	Mon, Mar 27, 2006 (4:00 p.m. EST)
Regular Paper Submissions:	Mon, Apr 3, 2006 (4:00 p.m. EST)
Notification of acceptance:	Mon, May 1, 2006
Camera-ready version due:	Mon, May 15, 2006
Workshop date:	Sun, Jun 25, 2006

SUBMISSION INSTRUCTIONS AND PROCEEDINGS

In addition to regular, scientific papers reporting on original research results, vision and work-in-progress papers that have the potential to stimulate debate on existing solutions or identify novel challenges are especially encouraged. Proposals for panels on emerging or controversial topics are also especially welcome.

Submissions will be handled electronically. Research paper submissions should be formatted in the ACM proceedings format and must not exceed 8 pages in that format. Each paper will be assigned for reviewing to at least three members of the program committee.

Accepted papers will be presented at the workshop and will appear in the workshop proceedings, which will be published by ACM. Electronic versions of the papers will be included in the ACM DL and DiSC'07.
Detailed submission information will be posted on the workshop web site, db.cs.pitt.edu/mobide06.

ORGANIZING COMMITTEE

Workshop Chairs:

Vijay Kumar
School of Computing and Engineering
University of Missouri-Kansas City
kumarv @ umkc.edu

Alexandros Labrinidis
Department of Computer Science
University of Pittsburgh
labrinid @ cs.pitt.edu

Program Chairs:

Panos K. Chrysanthis
Department of Computer Science
University of Pittsburgh
panos @ cs.pitt.edu

Christian S. Jensen
Department of Computer Science
Aalborg University
csj @ cs.aau.dk

Publicity Chair:

Magdalena Balazinska
Dept of Computer Science and Engineering
University of Washington
magda @ cs.washington.edu

STEERING COMMITTEE

Sujata Banerjee	HP Labs
Ugur Cetintemel	Brown University
Mitch Cherniack	Brandeis University
Panos K. Chrysanthis	University of Pittsburgh
Alexandros Labrinidis	University of Pittsburgh
Evaggelia Pitoura	University of Ioannina

PROGRAM COMMITTEE

Amr El Abbadi	University of California, Santa Barbara, USA
Walid Aref	Purdue University, USA
Magdalena Balazinska	University of Washington, USA
Sujata Banerjee	HP Labs, USA
Christian Becker	University of Stuttgart, Germany
Ugur Cetintemel	Brown University, USA
Nigel Davies	Lancaster University, UK
Maggie Dunham	Southern Methodist University, USA
Takahiro Hara	Osaka University, Japan
Ravi Jain	Google, USA
Vana Kalogeraki	University of California, Riverside, USA
Dimitris Katsaros	University of Thessaloniki, Greece
Dik Lun Lee	Hong Kong University of Science and Technology, Hong Kong
Mong Li Lee	National University of Singapore, Singapore
Wolfgang Lehner	Technische Universitat Dresden, Germany
Ling Liu	Georgia Institute of Technology, USA
Sanjay Madria	University of Missouri, Kansas City, USA
Pedro Jose Marron	University of Stuttgart, Germany
Daniela Nicklas	University of Stuttgart, Germany
Maria Papadopouli	University of North Carolina at Chapel Hill, USA
Evi Pitoura	University of Ioannina, Greece
Claudia Lucia Roncancio	Lab. LSR-IMAG/University of Grenoble, France
George Samaras	University of Cyprus, Cyprus
Joerg Sander	University of Alberta, Canada
Bernhard Seeger	University of Marburg, Germany
Jianwen Su	University of California, Santa Barbara, USA
Yufei Tao	City University of Hong Kong, Hong Kong
Vassilis Tsotras	University of California, Riverside, USA
Anthony K H Tung	National University of Singapore, Singapore
Wai Gen Yee	Illinois Institute of Technology, Chicago, USA
Vladimir Zadoroshny	University of Pittsburgh, USA
Arkady Zaslavsky	Monash University, Australia
Donghui Zhang	Northeastern University, USA

Figure 10: Sample DBWorld Post with Entities Highlighted

7. Related Work

Research on the problem of entity disambiguation has taken place using a variety of techniques which mostly work on structured parts of a document. The applicability of disambiguating peoples' names is evident when finding citations within documents. Han et al provides an assessment of several techniques used to disambiguate citations within a document. These methods use string similarity techniques and do not necessarily address various candidate entities that have the same name.

Our method differs from other approaches by a few important features. First, our method performs well on unstructured text. Second, by exploiting background knowledge in the form of a populated ontology, the process of spotting entities within the text is more focused and reduces the need for string similarity computations. Third, our method does not require any training data, as all of the data that is necessary for disambiguation is straightforward and provided in the ontology. Last but not least, our method exploits the capability provided by relationships among entities in the ontology to go beyond techniques traditionally based on syntactical matches.

The iterative step in our work is similar in spirit to a recent work on entity reconciliation [Dong, 2005]. In such an approach, the results of disambiguated entities are propagated to other ambiguous entities, which could then be reconciled based on recently reconciled entities. This method is intended to be part of a Personal Information Management system that works with a user's desktop environment to facilitate access and querying of his/her email address book, personal word documents, spreadsheets,

etc. However, this method processes data that in general, has a predictable structure such as fields that contain known types of data (i.e., emails, dates and peoples' names) whereas our method deals with more unstructured data. In our approach, we do not make such assumptions about the structure of text containing named entities to be disambiguated. This is a key difference as the characteristics of the data to be disambiguated pose different challenges. Our method uses an ontology and runs on un-structured text, an approach that theirs does not consider.

Citation matching is a related problem aiming at deciding the right citation referring to a publication [Giles, 1998]. In our work, we do not assume the existence of citation information such as publication venue and date. However, we believe that our method is a significant step to the Identity Uncertainty problem [Pasula, 2002] by automatically determining unique identifiers for peoples' names with respect to a populated ontology.

KIM is an application that aims to be an automatic ontology population system that runs over text documents to provide content for the Semantic Web [Popov, 2003]. The KIM platform has many components that are unrelated to our work but within these components, there is an entity recognition portion. KIM disambiguates entities within a document by using a natural language processor and then attempts to index these entities. The evaluation of the KIM system is done by comparing the results to human-annotated corpora, much like our method of evaluation.

Finally, our approach is different to that of disambiguating word senses [Basili, 1997, Gomes, 2003, Navigli, 2005]. Instead, our focus is to disambiguate named entities such as peoples' names, which has recently gained attention for its applicability

in Social Networks [Bekkerman, 2005]. Thus, instead of exploiting homonymy, synonymy, etc., our method works on relationships that real-world entities have, such as affiliation of a researcher and his/her topics.

8. Conclusion

We proposed a new ontology-driven solution to the entity disambiguation problem in unstructured text. In particular, our method uses relationships between entities in the ontology to go beyond traditional syntactic-based disambiguation techniques. The output of our method consists of a list of spotted entity names, each with an entity disambiguation CS. We demonstrated the effectiveness of our approach through evaluations against a manually disambiguated document set containing over 700 entities. This evaluation was performed over DBWorld announcements and using an ontology created from DBLP (consisting of over three million entities). The results of this evaluation lead us to claim that our method has successfully demonstrated its applicability to scenarios involving real-world data. Currently, we are evaluating our method against another similar tool called the Semagix Freedom Toolkit [Sheth, 2002] which is used to disambiguate entities. To the best of our knowledge, this work is among the first that successfully use relationships within a large, populated ontology for identifying entities in text without relying on the structure of the text.

In future work, we plan to integrate the results of entity disambiguation into a more robust platform, such as UIMA [Ferrucci, 2004]. In addition, the results of entity-disambiguation can be presented as part of the document using initiatives such as Microformats (microformats.org) and RDFa (w3.org/TR/xhtml-rdfa-primer/). Another positive addition would be that of automatic relationship ranking which would automatically adjust the network relationship weights based on their relevancy to the

situation, as suggested in [Halaschek, 2004] and/or considering multi-hop relationships as those described in [Anyanwu, 2003].

9. References

Anyanwu, K., Maduko, A., Sheth, A.: SemRank: Ranking Complex Relationship Search Results on the Semantic Web. *Proceedings of the 14th International World Wide Web Conference*, Japan (2005)

Anyanwu, K., Sheth, A.: ρ-Queries: Enabling Querying for Semantic Associations on the Semantic Web. *12th International World Wide Web Conference*, Budapest, Hungary (2003)

Aleman-Meza, B., Halaschek, C., Sheth, A., Arpinar, B., and Sannapareddy, G.: SWETO: Large-Scale Semantic Web Test-bed. *Proc. of the 16th International Conference on Software Engineering & Knowledge Engineering: Workshop on Ontology in Action,* (2004)

Aleman-Meza, B., Nagarajan, M., Ramakrishnan, C., Ding, L., Kolari, P., Sheth, A.P., Arpinar, I.B., Joshi, A. and Finin, T.: Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. *15th International World Wide Web Conference*, Edinburgh, Scotland (2006)

Basili, R., Rocca, M. D., Pazienza, M. T.: Contextual Word Sense Tuning and Disambiguation. *Applied Artificial Intelligence,* 11(3) (1997) 235-262

Bekkerman, R., McCallum, A.: Disambiguating Web Appearances of People in a Social Network. *14th International World Wide Web Conference* (WWW 2005), Chiba, Japan, (2005) 463-470

Berners-Lee, T., Fielding, R., Masinter, L.: Uniform Resource Identifier (URI): Generic Syntax. *RFC 3986, IETF, (*2005)

Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. *Scientific American*, (May, 2001)

Bilenko, M., Mooney, R., Cohen, W., Ravikumar, P., Fienberg, S.: Adaptive Name Matching in Information Integration. *IEEE Intelligent Systems,* 18(5) (Sept/Oct, 2003) 16-23

Broekstra, J., Kampman, A., Harmelen, F.: Sesame: A generic architecture for storing and querying RDF and RDF schema. *The Semantic Web – ISWC* (2002)

Carroll, J., Dickinson, I., Dollin, C., Reynolds, D., Seaborne, A., Wilkinson, K.: Jena: Implementing the Semantic Web Recommendations. *13th World Wide Web Conference* (2004)

DBWorld. http://www.cs.wisc.edu/dbworld/ April 9, 2006.

Dey, D., Sarkar, S., De, P.: A Distance-Based Approach to Entity Reconciliation in Heterogeneous Databases. *IEEE Transactions on Knowledge and Data Engineering*, 14(3) (May 2002) 567-582

Dong, X. L., Halevy, A., Madhaven, J.: Reference Reconciliation in Complex Information Spaces. *Proc. of SIGMOD*, Baltimore, MD. (2005)

Embley, D. W., Jiang, Y. S., Ng, Y.: Record-Boundary Discovery in Web Documents. *Proc. of SIGMOD*, Philadelphia, Pennsylvania (1999) 467-478

Ferrucci, D., Lally, A.: UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Natural Language Engineering*, 10(3-4) (2004) 327-348

Giles, C.L., Bollacker, K.D., Lawrence, S.: CiteSeer: An Automatic Citation Indexing System. *Proc. of the 3rd ACM International Conference on Digital Libraries*, Pittsburgh, PA. (June 23-26, 1998) 89-98

Gomes, P., Pereira, F. C., Paiva, P., Seco, N., Carreiro, P., Ferreira, J. L., Bento, C.: Noun Sense Disambiguation with WordNet for Software Design Retrieval. Proc. of the 16th Conference of the Canadian Society for Computational Studies of Intelligence (AI 2003), Halifax, Canada (June 11-13, 2003) 537-543

Gruber, T.: A Translation Approach to Portable Ontologies. *Knowledge Acquisition, 5(2)* (1993)

Guha, R., McCool, R.: Tap: A semantic web platform. *Network Computing* (2003)

Halaschek, C., Aleman-Meza, B., Arpinar, B., Sheth, A.: Discovering and Ranking Semantic Associations over a Large RDF Metabase. *30th International Conference on Very Large Data Bases*, Toronto, Canada (2004)

Han, H., Giles, L., Zha, H., Li, C., Tsioutsiouliklis, K.: Two Supervised Learning Approaches for Name Disambiguation in Author Citations. *Proc. ACM/IEEE Joint Conf on Digital Libraries*, Tucson, Arizona (2004)

Hassell, J., Aleman-Meza, B., Arpinar, B.: Ontology-Driven Automatic Entity Disambiguation in Unstructured Text. Under review in *The Semantic Web – ISWC*, Athens, Georgia (2006)

Janik, M., Kochut, K.: BRAHMS: A WorkBench RDF Store and High Performance Memory System for Semantic Association Discovery. *Fourth International Semantic Web Conference*, Galway, Ireland (2005)

Ley, M.: The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives. *Proc. of the 9th International Symposium on String Processing and Information Retrieval*, Lisbon, Portugal (Sept. 2002) 1-10

Navigli, R., Velardi, P.: Structural Semantic Interconnections: A Knowledge-Based Approach to Word Sense Disambiguation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(7) (2005) 1075-1086

Pasula, H., Marthi, B., Milch, B., Russell, S., Shpitser, I.: Identity Uncertainty and Citation Matching, Neural Information Processing Systems. Vancouver, British Columbia (2002) 1401-1408

Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D., Goranov, M.: KIM - Semantic Annotation Platform. *Proc. of the 2nd Intl. Semantic Web Conf*, Sanibel Island, Florida (2003)

Quan, D., D. Karger. How to Make a Semantic Web Browser. *Proc. of 13th International World Wide Web Conference* (2004)

RDF Primer. http://www.cc.gatech.edu/gvu/usersurveys/survey1997-10/ March 20, 2006.

Sahoo, S., Thomas, C., Sheth, A., York, W., Tartir, S.: Knowledge Modeling and its Application in Life Sciences: A Tale of two Ontologies. *15th Int. World Wide Web Conferenc,* (2006)

Sheth, A.: Enterprise Applications of Semantic Web: The Sweet Spot of Risk and Compliance. *IFIP International Conference on Industrial Applications of Semantic Web*, Jyväskylä, Finland (2005)

Sheth, A., Bertram, C., Avant, D., Hammond, B., Kochut, K., Warke, Y.: Managing semantic content for the Web. *IEEE Internet Computing*, 6(4) (2002) 80-87

Torvik, V., Weeber, M., Swanson, D., Smalheiser, N.: A Probabilistic Similarity Metric for Medline Records: A Model for Author Name Disambiguation. *Journal of the American Society for Information Science and Technology*, 56(2) (2005) 40-158

Zhang, W., Torvik, V., Smalheiser, N., Yu, C.: Segmentation of Publication Records of Authors from the Web. *Proceedings of the 22nd IEEE International Conference on Data Engineering* (2006)

Appendix 1.1: Document 1

Last Call for Papers --- DEADLINE APPROACHES

* * * HADIS 2006 * * *

Second International Workshop on
High Availability of Distributed Systems

in conjunction with DEXA 2006

5 September 2006, Krakow, Poland

http://www.iti.upv.es/madis/hadis06/cfp.html

Deadline for submissions: 13 March 2006.

Scope

High availability (HA) and wide-area distribution of mission-critical IT services are becoming indispensable. HA is receiving rapidly increasing attention, since many web-based services need to be reachable and reliable at any place and any time. System downtimes resulting in a lack of availability are undesirable or even intolerable, due to ergonomic, economical and security reasons. HA has established itself as a benchmark for user acceptance, market viability, system dependability and trustworthiness of distributed systems.

In general, HA refers to the continuous availability and seamless recovery of resources in a computer system, particularly in the wake of component failures. HA is concerned with avoiding single points of failure. This can be achieved in a variety of ways, ranging from vendor-specific support, redundant software and hardware, to solutions that provide distribution and consistent replication of data and processes, reliable group communication, membership management, quorum subsystems, concurrency control, and various other forms of middleware support.

HADIS 2006 will provide a forum for discussing state-of-the-art, novel and on-going research and development in HA. We are calling for papers reporting original research results, work in progress, industrial experience, or case studies. Submitted contributions should highlight aspects of HA in topic areas including, but not limited to:

* Distributed Databases and Information Systems

* Protocols for Communication, Replication and Recovery

* Failure Resilience, Fault Tolerance and Failover

* Replicated Data and Processes

* Dependability and Reliability

* Local and Wide Area Clustering

* Middleware Architectures

* P2P and GRID Computing

* Load Sharing and Load Balancing

* Software and Hardware Redundancy

* High Performance Computing

* Transparency of Distribution

* Distributed Storage Systems

* HA and Aspect Orientation

* HA and Formal Methods

* HA and Integrity

* HA and Security

* HA and Standardisation

* HA and Trust

* HA and Ubiquity

* HA in Wireless, Mobile and Nomadic Networks

* HA for Service Architectures and Applications

Authors are asked to submit an an electronic version (preferably .pdf) of original unpublished work by 3/03/2006. Full papers must currently be neither revised nor published elsewhere, contain an abstract of up to 200 words, be written in English and not exceed 5000 words. An electronic submission site is operational. Submissions can also be sent directly to the programme chair. We recommend to use the format of the final versions of accepted papers. Accepted papers will be published by IEEE in a proceedings volume of the DEXA 2006 workshops.

Important Dates

* Paper Submission: 13 March 2006
* Notification: 14 April 2006
* Camera Ready: 15 May 2006
* Workshop: 5 Sept. 2006

Organisers

 * Stephane Bressan, National Univ. Singapore (co-chair)
 * Hendrik Decker, Instituto Tecnol. de Informatica, Valencia, Spain (chair)
 * Francesc D. Munyoz, Univ. Politec. Valencia, Spain (programme chair)

Programme Committee

 * Lorenzo Alvisi, Univ. Texas Austin, U.S.A.
 * Roberto Baldoni, Univ. Roma La Sapienza, Italy
 * Maria del Carmen Banuls, Max Planck Inst. Quantenoptik, Germany
 * Stephane Bressan, National Univ. Singapore
 * Jesus Carretero, Univ. Carlos III, Madrid, Spain
 * Vicent Cholvi, Univ. Jaume I, Castellon, Spain
 * Hendrik Decker, Inst. Tecnologico de Informatica, Spain
 * Pascal Felber, Univ. Neuchatel, Switzerland
 * Christof Fetzer, Tech. Univ. Dresden, Germany
 * Pablo Galdamez, Univ. Politec. Valencia, Spain
 * Karl Goeschka, Tech. Univ. Vienna, Austria
 * Hamidah Ibrahim, Univ. Putra Malaysia, Malaysia
 * Bettina Kemme, McGill Univ. Montreal, Canada
 * Marc-Olivier Killijian, CNRS Toulouse, France
 * Mikel Larrea, Euskal Herriko Univ., Spain
 * Pietro Manzoni, Univ. Politec. Valencia, Spain
 * Jose R. Gonzalez de Mendivil, Univ. Publica Navarra, Spain
 * Anirban Mondal Univ. Tokyo, Japan
 * Francesc D. Munyoz, Univ. Politec. Valencia, Spain
 * Simin Nadjm-Tehrani, Linkoping Univ., Sweden
 * Marta Patino-Martinez, Univ. Politec. Madrid, Spain
 * Rasmus Pedersen, Copenhagen Business School, Denmark
 * Fernando Pedone, Univ. Lugano, Switzerland
 * Luigi Romano, Univ. Napoli, Italy
 * Juan-Carlos Ruiz-Garcia, Univ. Politec. Valencia, Spain
 * Jesus Villadangos, Univ. Publica Navarra, Spain

For submissions, please see instructions on
http://www.iti.upv.es/madis/hadis06/cfp.html
and
http://www.iti.upv.es/madis/hadis06/index.html
for further information.

We apologize for multiple postings.

Appendix 1.2: Document 1 Results

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Mondal:Anirban.html</uri>
      <entityName>Anirban Mondal</entityName>
      <confidence>90</confidence>
      <charOffset>4465, 4480</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kemme:Bettina.html</uri>
      <entityName>Bettina Kemme</entityName>
      <confidence>105</confidence>
      <charOffset>4208, 4222</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Fetzer:Christof.html</uri>
      <entityName>Christof Fetzer</entityName>
      <confidence>115</confidence>
      <charOffset>4009, 4025</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Pedone:Fernando.html</uri>
      <entityName>Fernando Pedone</entityName>
      <confidence>125</confidence>
      <charOffset>4718, 4734</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/i/Ibrahim:Hamidah.html</uri>
      <entityName>Hamidah Ibrahim</entityName>
      <confidence>95</confidence>
      <charOffset>4157, 4173</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Decker:Hendrik.html</uri>
      <entityName>Hendrik Decker</entityName>
      <confidence>125</confidence>
      <charOffset>3413, 3428</charOffset>
      <charOffset>3902, 3917</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/P=eacute=rez:Jes=uacute=s_Carretero.html</uri>
      <entityName>Jes&uacute;s Carretero</entityName>
      <entityName>Jesus Carretero</entityName>
      <entityName>Jes&uacute;s Carretero P&eacute;rez</entityName>
      <entityName>Jesus Carretero Perez</entityName>
      <confidence>105</confidence>
      <charOffset>3798, 3814</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/r/Ruiz-Garcia:Juan-
Carlos.html</uri>
      <entityName>Juan-Carlos Ruiz-Garcia</entityName>
      <confidence>100</confidence>
      <charOffset>4803, 4827</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Alvisi:Lorenzo.html</uri>
      <entityName>Lorenzo Alvisi</entityName>
      <confidence>105</confidence>
      <charOffset>3587, 3602</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Romano:Luigi.html</uri>
      <entityName>Luigi Romano</entityName>
      <confidence>95</confidence>
      <charOffset>4765, 4778</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/k/Killijian:Marc-
Olivier.html</uri>
      <entityName>Marc-Olivier Killijian</entityName>
      <confidence>105</confidence>
      <charOffset>4256, 4279</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Ba=ntilde=uls:Mar=iacute=a_del_Carmen.html</uri>
      <entityName>Mar&iacute;a del Carmen Ba&ntilde;uls</entityName>
      <entityName>Maria del Carmen Banuls</entityName>
      <confidence>110</confidence>
      <charOffset>3684, 3708</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/p/Pati=ntilde=o-
Mart=iacute=nez:Marta.html</uri>
      <entityName>Marta Patino-Martinez</entityName>
      <entityName>Marta Pati&ntilde;o-Mart&iacute;nez</entityName>
```

```
      <confidence>70</confidence>
      <charOffset>4606, 4628</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/p/Pati=ntilde=o-
Martinez:Marta.html</uri>
      <entityName>Marta Pati&ntilde;o-Martinez</entityName>
      <entityName>Marta Patino-Martinez</entityName>
      <confidence>45</confidence>
      <charOffset>4606, 4628</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Larrea:Mikel.html</uri>
      <entityName>Mikel Larrea</entityName>
      <confidence>95</confidence>
      <charOffset>4306, 4319</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Gald=aacute=mez:Pablo.html</uri>
      <entityName>Pablo Gald&aacute;mez</entityName>
      <entityName>Pablo Galdamez</entityName>
      <confidence>115</confidence>
      <charOffset>4059, 4074</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Felber:Pascal.html</uri>
      <entityName>Pascal Felber</entityName>
      <confidence>135</confidence>
      <charOffset>3961, 3975</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Manzoni:Pietro.html</uri>
      <entityName>Pietro Manzoni</entityName>
      <confidence>105</confidence>
      <charOffset>4352, 4367</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Gonz=aacute=lez:R=.html</uri>
      <entityName>R. Gonz&aacute;lez</entityName>
      <entityName>R. Gonzalez</entityName>
      <confidence>45</confidence>
      <charOffset>4408, 4420</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Gonzalez:R=.html</uri>
      <entityName>R. Gonzalez</entityName>
      <confidence>45</confidence>
      <charOffset>4408, 4420</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Baldoni:Roberto.html</uri>
      <entityName>Roberto Baldoni</entityName>
      <confidence>105</confidence>
      <charOffset>3633, 3649</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/n/Nadjm-
Tehrani:Simin.html</uri>
      <entityName>Simin Nadjm-Tehrani</entityName>
      <confidence>105</confidence>
      <charOffset>4558, 4578</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bressan:St=eacute=phane.html</uri>
      <entityName>St&eacute;phane Bressan</entityName>
      <entityName>Stephane Bressan</entityName>
      <confidence>105</confidence>
      <charOffset>3356, 3373</charOffset>
      <charOffset>3752, 3769</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Cholvi:Vicent.html</uri>
      <entityName>Vicent Cholvi</entityName>
      <confidence>105</confidence>
      <charOffset>3851, 3865</charOffset>
</entity>
```

Appendix 2.1: Document 2

[Apologies if you receive multiple copies of this message.]

S E C O N D C A L L F O R P A P E R S

5th ACM International Workshop on Data Engineering
for Wireless and Mobile Access (MobiDE'06)

Sunday, June 25, 2006, Chicago, Illinois, USA
(co-located with ACM SIGMOD/PODS 2006)

http://db.cs.pitt.edu/mobide06

AIMS & TOPICS OF INTEREST

Being the fifth in a series of successful workshops, MobiDE'06 aims to act as a bridge between the data management, wireless networking, and mobile computing communities.

The previous MobiDE workshops took place in Seattle, WA, in conjunction with MobiCom 1999; in Santa Barbara, CA, together with SIGMOD 2001; in San Diego, CA, together with MobiCom 2003; and in Baltimore, MD, in conjunction with SIGMOD 2005.

MobiDE'06 will serve as a forum where researchers and technologists discuss the state-of-the-art, present their contributions, and set future directions in the general area of data management for mobile and wireless access, including sensor networks. The specific focus of MobiDE'06 is on novel mobile applications and services including games, entertainment, and infotainment.

The topics of interest related to mobile and wireless data engineering include, but are not limited to, the following.

* ad-hoc networked databases
* consistency maintenance and management
* context-aware data access and query processing
* data caching, replication and view materialization
* data publication modes: push, broadcast, and multicast
* data server models and architectures
* database issues for moving objects: storage, indexing, etc.
* location-based and context-aware services
* m-commerce
* mobile agent models and languages

* mobile database security
* mobile databases in scientific, medical, and engineering applications
* mobile peer-to-peer applications and services
* mobile transaction models and management
* mobile web services
* mobility awareness and adaptability
* mobile services: in transportation, agriculture, industries, etc.
* pervasive computing
* prototype design of mobile data management systems
* quality of service for mobile data management
* sensor network data management
* small footprint data management
* transaction migration, recovery and commit processing
* wireless multimedia systems
* wireless web

As in the past, the workshop will be organized in a manner that fosters interaction and exchange of ideas among the participants. In addition to paper presentations, time will be allocated for open discussion forums, informal discussions or panels.

IMPORTANT DATES

Abstract Submissions:	Mon, Mar 27, 2006 (4:00 p.m. EST)
Regular Paper Submissions:	Mon, Apr 3, 2006 (4:00 p.m. EST)
Notification of acceptance:	Mon, May 1, 2006
Camera-ready version due:	Mon, May 15, 2006
Workshop date:	Sun, Jun 25, 2006

SUBMISSION INSTRUCTIONS AND PROCEEDINGS

In addition to regular, scientific papers reporting on original research results, vision and work-in-progress papers that have the potential to stimulate debate on existing solutions or identify novel challenges are especially encouraged. Proposals for panels on emerging or controversial topics are also especially welcome.

Submissions will be handled electronically. Research paper submissions should be formatted in the ACM proceedings format and must not exceed 8 pages in that format. Each paper will be assigned for reviewing to at least three members of the program committee.

Accepted papers will be presented at the workshop and will appear in the workshop proceedings, which will be published by ACM. Electronic versions of the papers will be included in the ACM DL and DiSC'07.

Detailed submission information will be posted on the workshop web site, db.cs.pitt.edu/mobide06.

ORGANIZING COMMITTEE

Workshop Chairs:

Vijay Kumar Alexandros Labrinidis

School of Computing and Engineering
University of Missouri-Kansas City
kumarv @ umkc.edu

Department of Computer Science
University of Pittsburgh
labrinid @ cs.pitt.edu

Program Chairs:

Panos K. Chrysanthis
Department of Computer Science
University of Pittsburgh
panos @ cs.pitt.edu

Christian S. Jensen
Department of Computer Science
Aalborg University
csj @ cs.aau.dk

Publicity Chair:

Magdalena Balazinska
Dept of Computer Science and Engineering
University of Washington
magda @ cs.washington.edu

STEERING COMMITTEE

Sujata Banerjee	HP Labs
Ugur Cetintemel	Brown University
Mitch Cherniack	Brandeis University
Panos K. Chrysanthis	University of Pittsburgh
Alexandros Labrinidis	University of Pittsburgh
Evaggelia Pitoura	University of Ioannina

PROGRAM COMMITTEE

Amr El Abbadi	University of California, Santa Barbara, USA
Walid Aref	Purdue University, USA
Magdalena Balazinska	University of Washington, USA
Sujata Banerjee	HP Labs, USA
Christian Becker	University of Stuttgart, Germany
Ugur Cetintemel	Brown University, USA
Nigel Davies	Lancaster University, UK
Maggie Dunham	Southern Methodist University, USA
Takahiro Hara	Osaka University, Japan
Ravi Jain	Google, USA
Vana Kalogeraki	University of California, Riverside, USA
Dimitris Katsaros	University of Thessaloniki, Greece
Dik Lun Lee	Hong Kong University of Science and Technology, Hong Kong
Mong Li Lee	National University of Singapore, Singapore
Wolfgang Lehner	Technische Universitat Dresden, Germany
Ling Liu	Georgia Institute of Technology, USA
Sanjay Madria	University of Missouri, Kansas City, USA
Pedro Jose Marron	University of Stuttgart, Germany
Daniela Nicklas	University of Stuttgart, Germany
Maria Papadopouli	University of North Carolina at Chapel Hill, USA
Evi Pitoura	University of Ioannina, Greece
Claudia Lucia Roncancio	Lab. LSR-IMAG/University of Grenoble, France
George Samaras	University of Cyprus, Cyprus
Joerg Sander	University of Alberta, Canada
Bernhard Seeger	University of Marburg, Germany

Jianwen Su	University of California, Santa Barbara, USA
Yufei Tao	City University of Hong Kong, Hong Kong
Vassilis Tsotras	University of California, Riverside, USA
Anthony K H Tung	National University of Singapore, Singapore
Wai Gen Yee	Illinois Institute of Technology, Chicago, USA
Vladimir Zadoroshny	University of Pittsburgh, USA
Arkady Zaslavsky	Monash University, Australia
Donghui Zhang	Northeastern University, USA

Appendix 2.2: Document 2 Results

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Labrinidis:Alexandros.html</uri>
      <entityName>Alexandros Labrinidis</entityName>
      <confidence>135</confidence>
      <charOffset>4238, 4260</charOffset>
      <charOffset>5080, 5102</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Abbadi:Amr_El.html</uri>
      <entityName>Amr El Abbadi</entityName>
      <confidence>115</confidence>
      <charOffset>5201, 5215</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/z/Zaslavsky:Arkady.html</uri>
      <entityName>Arkady Zaslavsky</entityName>
      <confidence>90</confidence>
      <charOffset>7124, 7141</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Seeger:Bernhard.html</uri>
      <entityName>Bernhard Seeger</entityName>
      <confidence>135</confidence>
      <charOffset>6657, 6673</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Becker:Christian.html</uri>
      <entityName>Christian Becker</entityName>
      <confidence>125</confidence>
      <charOffset>5420, 5437</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/j/Jensen:Christian_S=.html</uri>
      <entityName>Christian S. Jensen</entityName>
      <confidence>115</confidence>
      <charOffset>4525, 4545</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/n/Nicklas:Daniela.html</uri>
      <entityName>Daniela Nicklas</entityName>
      <confidence>105</confidence>
      <charOffset>6278, 6294</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Lee:Dik_Lun.html</uri>
      <entityName>Dik Lun Lee</entityName>
      <confidence>115</confidence>
      <charOffset>5863, 5875</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/z/Zhang:Donghui.html</uri>
      <entityName>Donghui Zhang</entityName>
      <confidence>125</confidence>
      <charOffset>7180, 7194</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Pitoura:Evaggelia.html</uri>
      <entityName>Evaggelia Pitoura</entityName>
      <confidence>135</confidence>
      <charOffset>5131, 5149</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Samaras:George.html</uri>
      <entityName>George Samaras</entityName>
      <confidence>125</confidence>
      <charOffset>6544, 6559</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/s/Su:Jianwen.html</uri>
      <entityName>Jianwen Su</entityName>
      <confidence>165</confidence>
      <charOffset>6715, 6726</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/l/Liu:Ling.html</uri>
      <entityName>Ling Liu</entityName>
      <confidence>105</confidence>
      <charOffset>6086, 6095</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Balazinska:Magdalena.html</uri>
```

```
      <entityName>Magdalena Balazinska</entityName>
      <confidence>115</confidence>
      <charOffset>4759, 4780</charOffset>
      <charOffset>5323, 5344</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Papadopouli:Maria.html</uri>
      <entityName>Maria Papadopouli</entityName>
      <confidence>90</confidence>
      <charOffset>6338, 6356</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Cherniack:Mitch.html</uri>
      <entityName>Mitch Cherniack</entityName>
      <confidence>155</confidence>
      <charOffset>4983, 4999</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Davies:Nigel.html</uri>
      <entityName>Nigel Davies</entityName>
      <confidence>105</confidence>
      <charOffset>5529, 5542</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Chrysanthis:Panos_K=.html</uri>
      <entityName>Panos K. Chrysanthis</entityName>
      <confidence>165</confidence>
      <charOffset>4486, 4507</charOffset>
      <charOffset>5029, 5050</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Marr=oacute=n:Pedro_Jos=eacute=.html</uri>
      <entityName>Pedro Jos&eacute; Marr&oacute;n</entityName>
      <entityName>Pedro Jose Marron</entityName>
      <confidence>105</confidence>
      <charOffset>6218, 6236</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/j/Jain:Ravi.html</uri>
      <entityName>Ravi Jain</entityName>
      <confidence>105</confidence>
      <charOffset>5694, 5704</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Madria:Sanjay.html</uri>
      <entityName>Sanjay Madria</entityName>
      <confidence>90</confidence>
      <charOffset>6150, 6164</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Banerjee:Sujata.html</uri>
      <entityName>Sujata Banerjee</entityName>
      <confidence>145</confidence>
      <charOffset>4906, 4922</charOffset>
      <charOffset>5380, 5396</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/h/Hara:Takahiro.html</uri>
      <entityName>Takahiro Hara</entityName>
      <confidence>105</confidence>
      <charOffset>5643, 5657</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/=/=Ccedil=etintemel:Ugur.html</uri>
      <entityName>Ugur &Ccedil;etintemel</entityName>
      <entityName>Ugur Cetintemel</entityName>
      <confidence>125</confidence>
      <charOffset>4940, 4956</charOffset>
      <charOffset>5480, 5496</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kalogeraki:Vana.html</uri>
      <entityName>Vana Kalogeraki</entityName>
      <confidence>105</confidence>
      <charOffset>5733, 5749</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/t/Tsotras:Vassilis_J=.html</uri>
      <entityName>Vassilis J. Tsotras</entityName>
      <confidence>115</confidence>
      <charOffset>6853, 6871</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kumar:Vijay.html</uri>
      <entityName>Vijay Kumar</entityName>
      <confidence>155</confidence>
      <charOffset>4199, 4211</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/y/Yee:Wai_Gen.html</uri>
      <entityName>Wai Gen Yee</entityName>
      <confidence>95</confidence>
      <charOffset>6993, 7005</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Aref:Walid_G=.html</uri>
      <entityName>Walid G. Aref</entityName>
      <confidence>135</confidence>
      <charOffset>5272, 5284</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Lehner:Wolfgang.html</uri>
      <entityName>Wolfgang Lehner</entityName>
      <confidence>105</confidence>
      <charOffset>6019, 6035</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/t/Tao:Yufei.html</uri>
      <entityName>Yufei Tao</entityName>
      <confidence>135</confidence>
      <charOffset>6787, 6797</charOffset>
</entity>
```

Appendix 3.1: Document 3

[We apologize for multiple copies]

CoMoGIS2006

3rd International Workshop on Conceptual Modeling for Geographic Information Systems
<A
HREF="http://lbdwww.epfl.ch/e/conferences/CoMoGIS2006/">http://lbdwww.epfl.ch/e/conferences/CoMoGIS2006/</A>

in conjunction with 25th International Conference on Conceptual Modeling (ER 2006)

November 6-9, 2006
Tucson, Arizona

The recent advances in remote sensing and GPS technologies have increased the production, collection and diffusion of geo-referenced data, thus requiring the rapid development and the wide deployment of various geographic information systems. With the popularity of World Wide Web and the diversity of GISs on the Internet, geographic information can now be available via personal devices anytime and anywhere. Nowadays GIS is emerging as a common information infrastructure, which penetrates into more and more aspects of our society, and converges with most areas in the IT scenario, such as office automation, workflow, digital libraries, Web searching and virtual reality. This has given rise to new challenges related to the development of conceptual models for GIS. Recently several new research approaches have been applied in the development of geo-spatial systems to accommodate new users requirements.

The workshop on Conceptual Modeling for GIS is intended to bring together researchers, developers, users, and practitioners carrying out research and development in geographic information systems and conceptual modeling, and fostering interdisciplinary discussions in all aspects of these two fields. The workshop provides a forum for original research contributions and practical experiences of conceptual modeling for GIS and will highlight future trends in this area. The workshop has been successively held in conjunction with the ER conference in 2004 and 2005.

We invite submissions that address theoretical, technical and practical issues of conceptual modeling for GIS. Suggested topics include, but are not limited to (as long as they are related to new research approaches in conceptual modeling and GIS):

-Spatial data modeling
-Spatio-temporal data modeling

-Conceptual and logical models for GIS
-GIS services modeling
-Ontologies for GIS applications
-Semantic issues in GIS
-Schema mapping and evolution
-Spatial data query and retrieval
-Geographical search engines
-Spatial information grid
-Digital geographical libraries
-Query languages and interfaces
-Spatial information integration
-Spatial information visualisation
-Spatial information personalisation
-Spatial data mining and data warehousing
-Location-based services
-Peer-to-peer computing for GIS
-Interoperability and standards
-Metadata management
-GIS Middleware architectures

The proceedings will be published by Springer-Verlag in the LNCS series.

Important dates

Paper abstracts: April 3, 2006
Full papers: April 10, 2006
Notification: June 14, 2006
Camera-ready papers: July 12, 2006

All paper proposals should be submitted in Springer format (details on the format at <A HREF="http://www.springer.de/comp/lncs/authors.html">http://www.springer.de/comp/lncs/authors.html</A>) via email to Christelle Vangenot (christelle.vangenot@epfl.ch) containing the title and abstract of your paper, authors' names, e-mail and post addresses, and phone and fax numbers. Attach your submission (as a MIME attachment) in PDF, Word, or portable postscript format to the same message. The suggested number of pages is 14, and the maximum number of pages is 16. Manuscripts not submitted in the LNCS style or having more than 16 pages will not be reviewed and thus automatically rejected. (The final, camera-ready version must not exceed 14 pages to avoid page charges.)

Workshop Chairs

Prof. Christophe Claramunt, Naval Academy Research Institute, France (claramunt@ecole-navale.fr)
Dr. Christelle Vangenot, EPFL - Ecole Polytechnique Federale de Lausanne, Switzerland (christelle.vangenot@epfl.ch)

Program Committee

Masatoshi Arikawa (University of Tokyo, Japan)
Natalia Andrienko (Fraunhofer Institute AIS, Germany)
Michela Bertolotto (University College, Dublin, Ireland)

Patrice Boursier (University of La Rochelle, France and Open University, Malaysia)
Elena Camossi (IMATI-CNR Genova, Italy)
James Carswell (Dublin Institute of Technology, Ireland)
Maria Luisa Damiani (University of Milano, Italy)
Thomas Devogele (Naval Academy Research Institute, France)
Max Egenhofer (University of Maine, USA)
Andrew Frank (Technical University of Vienna, Austria)
Bo Huang (University of Calgary, Canada)
Zhiyong Huang (National University of Singapore, Singapore)
Christian S. Jensen (Aalborg University, Denmark)
Ki-Joune Li (Pusan National University, South Korea)
Dieter Pfoser (CTI, Greece)
Martin Raubal (University of Munster, Germany)
Andrea Rodriguez (University of Concepcion, Chile)
Sylvie Servigne (INSA, France)
Kathleen Stewart Hornsby (University of Maine, USA)
George Taylor (University of Glamorgan, UK)
Nectaria Tryfona (Talent Information Systems, Greece)
Agnes Voisard, (Fraunhofer ISST and FU Berlin, Germany)
Nico van de Weghe (University of Gent, Belgium)
Nancy Wiegand (University of Wisconsin-Madison, USA)
Stephan Winter (University of Melbourne, Australia)
Ilya Zaslavsky (San Diego Supercomputer Centre, USA)
Esteban Zimanyi (Université Libre de Bruxelles, Belgium)

Appendix 3.2: Document 3 Results

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/v/Voisard:Agn=egrave=s.html</uri>
      <entityName>Agn&egrave;s Voisard</entityName>
      <entityName>Agnes Voisard</entityName>
      <confidence>115</confidence>
      <charOffset>5204, 5218</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Rodr=iacute=guez:Andrea.html</uri>
      <entityName>Andrea Rodr&iacute;guez</entityName>
      <entityName>Andrea Rodriguez</entityName>
      <confidence>90</confidence>
      <charOffset>4972, 4989</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Frank:Andrew_A=.html</uri>
      <entityName>Andrew A. Frank</entityName>
      <confidence>23</confidence>
      <charOffset>4637, 4651</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Frank:Andrew_O=.html</uri>
      <entityName>Andrew O. Frank</entityName>
      <confidence>23</confidence>
      <charOffset>4637, 4651</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Frank:Andrew_U=.html</uri>
      <entityName>Andrew U. Frank</entityName>
      <confidence>78</confidence>
      <charOffset>4637, 4651</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/h/Huang:Bo.html</uri>
      <entityName>Bo Huang</entityName>
      <confidence>105</confidence>
      <charOffset>4693, 4702</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/v/Vangenot:Christelle.html</uri>
      <entityName>Christelle Vangenot</entityName>
      <confidence>105</confidence>
      <charOffset>3286, 3306</charOffset>
      <charOffset>3994, 4014</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/j/Jensen:Christian_S=.html</uri>
      <entityName>Christian S. Jensen</entityName>
      <confidence>125</confidence>
      <charOffset>4794, 4814</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Claramunt:Christophe.html</uri>
      <entityName>Christophe Claramunt</entityName>
      <confidence>105</confidence>
      <charOffset>3899, 3920</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Pfoser:Dieter.html</uri>
      <entityName>Dieter Pfoser</entityName>
      <confidence>135</confidence>
      <charOffset>4897, 4911</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Camossi:Elena.html</uri>
      <entityName>Elena Camossi</entityName>
      <confidence>100</confidence>
      <charOffset>4391, 4405</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/z/Zim=aacute=nyi:Esteban.html</uri>
      <entityName>Esteban Zim&aacute;nyi</entityName>
      <entityName>Esteban Zimanyi</entityName>
      <confidence>105</confidence>
      <charOffset>5466, 5482</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/t/Taylor:George.html</uri>
      <entityName>George Taylor</entityName>
      <confidence>90</confidence>
      <charOffset>5106, 5120</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/z/Zaslavsky:Ilya.html</uri>
      <entityName>Ilya Zaslavsky</entityName>
      <confidence>90</confidence>
      <charOffset>5413, 5428</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Carswell:James_D=.html</uri>
      <entityName>James D. Carswell</entityName>
      <confidence>80</confidence>
      <charOffset>4430, 4446</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/l/Li:Ki-
Joune.html</uri>
      <entityName>Ki-Joune Li</entityName>
      <confidence>105</confidence>
      <charOffset>4844, 4856</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Damiani:Maria_Luisa.html</uri>
      <entityName>Maria Luisa Damiani</entityName>
      <entityName>Maria Damiani</entityName>
      <confidence>95</confidence>
      <charOffset>4488, 4508</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Raubal:Martin.html</uri>
      <entityName>Martin Raubal</entityName>
      <confidence>130</confidence>
      <charOffset>4925, 4939</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Arikawa:Masatoshi.html</uri>
      <entityName>Masatoshi Arikawa</entityName>
      <confidence>95</confidence>
      <charOffset>4150, 4168</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/e/Egenhofer:Max_J=.html</uri>
      <entityName>Max J. Egenhofer</entityName>
      <confidence>135</confidence>
      <charOffset>4596, 4611</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bertolotto:Michela.html</uri>
      <entityName>Michela Bertolotto</entityName>
      <confidence>125</confidence>
      <charOffset>4251, 4270</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/w/Wiegand:Nancy.html</uri>
      <entityName>Nancy Wiegand</entityName>
      <confidence>95</confidence>
      <charOffset>5308, 5322</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Andrienko:Natalia_V=.html</uri>
      <entityName>Natalia V. Andrienko</entityName>
      <confidence>75</confidence>
      <charOffset>4196, 4215</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/t/Tryfona:Nectaria.html</uri>
      <entityName>Nectaria Tryfona</entityName>
      <confidence>145</confidence>
      <charOffset>5150, 5167</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Boursier:Patrice.html</uri>
      <entityName>Patrice Boursier</entityName>
      <confidence>90</confidence>
      <charOffset>4308, 4325</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/w/Winter:Stephan.html</uri>
      <entityName>Stephan Winter</entityName>
      <confidence>105</confidence>
      <charOffset>5361, 5376</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Servigne:Sylvie.html</uri>
      <entityName>Sylvie Servigne</entityName>
      <confidence>100</confidence>
      <charOffset>5023, 5039</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Devogele:Thomas.html</uri>
      <entityName>Thomas Devogele</entityName>
      <confidence>100</confidence>
      <charOffset>4538, 4554</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/h/Huang:Zhiyong.html</uri>
      <entityName>Zhiyong Huang</entityName>
      <confidence>115</confidence>
      <charOffset>4734, 4748</charOffset>
</entity>
```

Appendix 4.1: Document 4

Fourth International XML Database Symposium (XSym 2006)
In Conjunction with VLDB 2006 Seoul, Korea
10-11 September 2006

The theme of the XML Database Symposium (XSym) is the convergence of database technology with XML technology. The goal of this symposium is to bring together academics, practitioners, users and vendors to discuss the use and synergy between the above-mentioned technologies. Many commercial systems built today are increasingly using these technologies together and it is important to understand the various research and practical issues. This symposium will provide the opportunity to debate new issues and directions for research and development work in the future.

Topics of Interest

* XML full-text search and ranking
* Approximate XML querying
* Query processing and optimization
* Indexing and access methods
* Access control and security
* Storage and compression
* Updates and integrity maintenance
* Concurrency control and recovery
* Performance evaluation
* Conceptual design: models and methodologies, expressiveness and usability
* Logical design: models and methodologies, expressiveness and usability
* Query languages, expressiveness and usability

Paper Submission:

Symposium submissions must generally be in electronic form using Portable Document Format (.pdf), PostScript (.ps) or WinWord (.doc). Papers should not be more than 15 pages in length. Papers should be formatted according to the Springer-Verlag Lecture Notes in Computer Science (LNCS) guidelines.

Paper submission site will be accessible soon from <A HREF="http://www.xsym.org/06/">http://www.xsym.org/06/</A>

Important Dates:

15 May 2006	Paper Submission Deadline
19 June 2006	Notification of Acceptance
9 July 2006	Camera Ready Copy
10-11 September 2006	Symposium

Organizing Committee:

General Chair
Zohra Bellahsene, LIRMM (France)

Local Chair
Kyuseok Shim, Seoul National University (Korea)

Program Committee Chairs
Sihem Amer-Yahia, AT&T Research (USA)
Jeffrey Xu Yu, Chinese University of Hong Kong (China)

Communications and Sponsorship
Ela Hunt, University of Zurich (Switzerland)

Proceedings
Rainer Unland, University of Duisburg-Essen (Germany)

Program Committee:

Ashraf Aboulnaga, University of Waterloo (Canada)
Bernd Amann, Universiti Paris 6, (France)
Denilson Barbosa, University of Calgary (Canada)
Omar Benjelloun, Stanford University (USA)
Veronique Benzaken, University Paris-Sud (France)
Philip Bernstein, Microsoft Research (USA)
Philip Bohannon, Bell Laboratories - Lucent Technologies (USA)
Jihad Boulos, American University of Beirut (Lebanon)
Stephane Bressan, National University of Singapore (Singapore)
Yi Chen, Arizona State University (USA)
Alex Dekhtayar, University of Kentucky (USA)
Alin Deutsch, University of California at San Diego (USA)
Yanlei Diao, University of Massachusetts at Amherst (USA)
Irini Fundulaki, University of Edinburgh (UK)
Minos Garofalakis, Intel Research (USA)
Giorgio Ghelli, Universita di Pisa (Italy)
Torsten Grust, Technical University of Munich (Germany)
Giovanna Guerrini, Universita di Genova (Italy)
Ihab Ilyas, University of Waterloo (Canada)
Zack Ives, University of Pennsylvania (USA)
Vanja Josifovski, Yahoo Research (USA)
Carl Christian Kanne, University of Mannheim (Germany)
Yaron Kanza, University of Toronto (Canada)
Raghav Kaushik, Microsoft Research (USA)
Laks Lakshmanan, University of British Columbia (Canada)
Dongwon Lee, Pennsylvania State University (USA)
Mong Li Lee, National University of Singapore (Singapore)
Qiong Luo, Hong Kong University of Science and Technology (China)
Murali Mani, Worcester Polytechnic Institute (USA)
Amelie Marian, Rutgers University (USA)
Peter McBrien, Imperial College - London (UK)
Tova Milo, Tel Aviv University (Israel)
Atsuyuki Morishima, University of Tsukuba (Japan)
Fatma Ozcan, IBM Almaden Research Center (USA)
Tamer Ozsu, University of Waterloo (Canada)
Tadeusz Pankowski, Poznan University of Technology (Poland)
Alkis Polyzotis, University of California at Santa Cruz (USA)
Philippe Pucheral, INRIA (France)

Prakash Ramanan, Wichita State University (USA)
Michael Rys, Microsoft (USA)
Monica Scannapieco, University of Roma "La Sapienza" (Italy)
Jayavel Shanmugasundaram, Cornell University (USA)
Jerome Simeon, IBM Research (USA)
Divesh Srivastava, AT&T Research (USA)
Martin Theobald, Max-Planck-Institut fur Informatik (Germany)
Vasilis Vassalos, Athens University of Economics and Business (Greece)
Stratis Viglas, University of Edinburgh (UK)
Yuqing Melanie Wu, Indiana University (USA)
Jun Yang, Duke University (USA)

Appendix 4.2: Document 4 Results

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Deutsch:Alin.html</uri>
      <entityName>Alin Deutsch</entityName>
      <confidence>105</confidence>
      <charOffset>2927, 2940</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Marian:Am=eacute=lie.html</uri>
      <entityName>Am&eacute;lie Marian</entityName>
      <entityName>Amelie Marian</entityName>
      <confidence>145</confidence>
      <charOffset>3824, 3838</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Aboulnaga:Ashraf.html</uri>
      <entityName>Ashraf Aboulnaga</entityName>
      <confidence>125</confidence>
      <charOffset>2385, 2402</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Morishima:Atsuyuki.html</uri>
      <entityName>Atsuyuki Morishima</entityName>
      <confidence>95</confidence>
      <charOffset>3950, 3969</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Amann:Bernd.html</uri>
      <entityName>Bernd Amann</entityName>
      <confidence>135</confidence>
      <charOffset>2435, 2447</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Barbosa:Denilson.html</uri>
      <entityName>Denilson Barbosa</entityName>
      <confidence>95</confidence>
      <charOffset>2477, 2494</charOffset>
</entity>
```

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Srivastava:Divesh.html</uri>
      <entityName>Divesh Srivastava</entityName>
      <confidence>165</confidence>
      <charOffset>4470, 4488</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Lee:Dongwon.html</uri>
      <entityName>Dongwon Lee</entityName>
      <confidence>130</confidence>
      <charOffset>3600, 3612</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/l/Lee_0002:Dongwon.html</uri>
      <entityName>Dongwon Lee</entityName>
      <confidence>45</confidence>
      <charOffset>3600, 3612</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/h/Hunt:Ela.html</uri>
      <entityName>Ela Hunt</entityName>
      <confidence>90</confidence>
      <charOffset>2252, 2261</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/o/Ozcan:Fatma.html</uri>
      <entityName>Fatma Ozcan</entityName>
      <confidence>105</confidence>
      <charOffset>4000, 4012</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Ghelli:Giorgio.html</uri>
      <entityName>Giorgio Ghelli</entityName>
      <confidence>105</confidence>
      <charOffset>3129, 3144</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Guerrini:Giovanna.html</uri>
      <entityName>Giovanna Guerrini</entityName>
      <confidence>105</confidence>
      <charOffset>3228, 3246</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Fundulaki:Irini.html</uri>
      <entityName>Irini Fundulaki</entityName>
      <confidence>135</confidence>
      <charOffset>3043, 3059</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Shanmugasundaram:Jayavel.html</uri>
      <entityName>Jayavel Shanmugasundaram</entityName>
      <confidence>205</confidence>
      <charOffset>4385, 4410</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/y/Yu:Jeffrey_Xu.html</uri>
      <entityName>Jeffrey Xu Yu</entityName>
      <confidence>105</confidence>
      <charOffset>2165, 2179</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Sim=eacute=on:J=eacute=r=ocirc=me.html</uri>
      <entityName>J&eacute;r&ocirc;me Sim&eacute;on</entityName>
      <entityName>Jerome Simeon</entityName>
      <confidence>135</confidence>
      <charOffset>4436, 4450</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Boulos:Jihad.html</uri>
      <entityName>Jihad Boulos</entityName>
      <confidence>95</confidence>
      <charOffset>2725, 2738</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/y/Yang:Jun.html</uri>
      <entityName>Jun Yang</entityName>
      <confidence>105</confidence>
      <charOffset>4731, 4740</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Shim:Kyuseok.html</uri>
      <entityName>Kyuseok Shim</entityName>
      <confidence>105</confidence>
      <charOffset>2053, 2066</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/t/Theobald:Martin.html</uri>
      <entityName>Martin Theobald</entityName>
      <confidence>95</confidence>
      <charOffset>4509, 4525</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Rys:Michael.html</uri>
      <entityName>Michael Rys</entityName>
      <confidence>105</confidence>
      <charOffset>4295, 4307</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Scannapieco:Monica.html</uri>
      <entityName>Monica Scannapieco</entityName>
      <confidence>105</confidence>
      <charOffset>4324, 4343</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Mani:Murali.html</uri>
      <entityName>Murali Mani</entityName>
      <confidence>105</confidence>
      <charOffset>3773, 3785</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Benjelloun:Omar.html</uri>
      <entityName>Omar Benjelloun</entityName>
      <confidence>125</confidence>
      <charOffset>2526, 2542</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/McBrien:Peter.html</uri>
      <entityName>Peter McBrien</entityName>
      <confidence>115</confidence>
      <charOffset>3864, 3878</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bohannon:Philip.html</uri>
      <entityName>Philip Bohannon</entityName>
      <confidence>115</confidence>
      <charOffset>2662, 2678</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Pucheral:Philippe.html</uri>
      <entityName>Philippe Pucheral</entityName>
      <confidence>105</confidence>
      <charOffset>4213, 4231</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Ramanan:Prakash.html</uri>
      <entityName>Prakash Ramanan</entityName>
      <confidence>90</confidence>
      <charOffset>4247, 4263</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/l/Luo:Qiong.html</uri>
      <entityName>Qiong Luo</entityName>
      <confidence>135</confidence>
      <charOffset>3707, 3717</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kaushik:Raghav.html</uri>
      <entityName>Raghav Kaushik</entityName>
      <confidence>105</confidence>
      <charOffset>3502, 3517</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/u/Unland:Rainer.html</uri>
      <entityName>Rainer Unland</entityName>
      <confidence>105</confidence>
      <charOffset>2310, 2324</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/a/Amer-
Yahia:Sihem.html</uri>
      <entityName>Sihem Amer-Yahia</entityName>
      <confidence>145</confidence>
      <charOffset>2127, 2144</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bressan:St=eacute=phane.html</uri>
      <entityName>St&eacute;phane Bressan</entityName>
      <entityName>Stephane Bressan</entityName>
      <confidence>105</confidence>
      <charOffset>2779, 2796</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/v/Viglas:Efstratios.html</uri>
      <entityName>Stratis Viglas</entityName>
      <entityName>Efstratios Viglas</entityName>
      <confidence>125</confidence>
      <charOffset>4642, 4657</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Pankowski:Tadeusz.html</uri>
      <entityName>Tadeusz Pankowski</entityName>
      <confidence>90</confidence>
      <charOffset>4091, 4109</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Grust:Torsten.html</uri>
      <entityName>Torsten Grust</entityName>
      <confidence>95</confidence>
      <charOffset>3172, 3186</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/m/Milo:Tova.html</uri>
      <entityName>Tova Milo</entityName>
      <confidence>165</confidence>
      <charOffset>3910, 3920</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/j/Josifovski:Vanja.html</uri>
      <entityName>Vanja Josifovski</entityName>
      <confidence>115</confidence>
      <charOffset>3364, 3381</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/v/Vassalos:Vasilis.html</uri>
      <entityName>Vasilis Vassalos</entityName>
      <confidence>105</confidence>
      <charOffset>4571, 4588</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Benzaken:V=eacute=ronique.html</uri>
      <entityName>V&eacute;ronique Benzaken</entityName>
      <entityName>Veronique Benzaken</entityName>
      <confidence>105</confidence>
      <charOffset>2569, 2588</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Diao:Yanlei.html</uri>
      <entityName>Yanlei Diao</entityName>
      <confidence>90</confidence>
      <charOffset>2985, 2997</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kanza:Yaron.html</uri>
      <entityName>Yaron Kanza</entityName>
      <confidence>95</confidence>
      <charOffset>3458, 3470</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/c/Chen:Yi.html</uri>
      <entityName>Yi Chen</entityName>
      <confidence>50</confidence>
      <charOffset>2842, 2850</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Chen_0002:Yi.html</uri>
      <entityName>Yi Chen</entityName>
      <confidence>45</confidence>
      <charOffset>2842, 2850</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/w/Wu:Yuqing_Melanie.html</uri>
      <entityName>Yuqing Melanie Wu</entityName>
      <confidence>90</confidence>
      <charOffset>4687, 4705</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bellahsene:Zohra.html</uri>
      <entityName>Zohra Bellahsene</entityName>
      <confidence>115</confidence>
      <charOffset>2007, 2024</charOffset>
</entity>
```

Appendix 5.1: Document 5

First Call for Papers

Advances in Semantics for Web services Workshop
(semantics4ws'06)
<A HREF="http://events.deri.at/semantics4ws2006/">http://events.deri.at/semantics4ws2006/</A>

at the Fourth International Conference on Business Process Management
(BPM 2006)
<A HREF="http://bpm2006.tuwien.ac.at/">http://bpm2006.tuwien.ac.at/</A>

Vienna, Austria, September 4, 2006

The theme of semantics4ws'06 is “Semantic Web Service in Business Processes”

GENERAL OVERVIEW

Web services have added a new level of functionality to the current Web by taking a first step towards seamless integration of distributed software components using Web standards. Nevertheless, current Web service technologies around SOAP, WSDL and UDDI operate at a syntactic level and, therefore, although they support interoperability (i.e. interoperability between the many diverse application development platforms that exist today) through common standards, they still require human interaction to a large extent. For example, the human programmer has to manually search for appropriate Web services in order to combine them in a useful manner, which limits scalability and greatly curtails the added economic value of envisioned with the advent of Web services.

Recent research (to which we refer to as Semantic Web Services - SWS), which draws on a variety of fields such as Semantic Web, knowledge representation, formal methods, software engineering, process modeling, workflow, and software agents, is gaining momentum, in particular in the context of Web services usage. Research in the mentioned fields can be exploited to automate Web services-related tasks, like discovery, selection, composition, mediation, monitoring, and invocation, thus enabling seamless interoperation between them while keeping human intervention to a minimum. Although several initiatives, like OWL-S, WSMO, WSDL-S, or IRS, have emerged in this area aiming at addressing the problem of semantics in Web services, many major challenges still need to be addressed and solved in this field.

In this context, this workshop aims to provide a forum in which to focus on selected core technical challenges for deployment of Semantic Web Services, and reach a better understanding of the relationships between commercial Web service standards, current SWS research efforts, and the ultimate requirements for full-scale deployment of these technologies. More specifically, this workshop

aims to tackle the research problems (as well as recent practical experiences) around methods, concepts, models, languages and technology that enable semantics in the context of Web services, as well as discussing recent advances in semantics for Web services. Of particular interest are the architectural, technical, and developmental foundations of SWS, and showing how they combine synergistically to enable service automation on the scale required by today's Internet-connected enterprises.

This proposed workshop aims to bring together researchers and industry practitioners (e.g. leading modelers, architects, system vendors, open-source projects, developers, and end-users) addressing many of these issues (including recent developments in tools and techniques, and real-world implementations of SWS applications), and promote and foster a greater understanding of how semantics can assist automation in Web services, thus helping people develop and manage services more efficiently and effectively.

TOPICS

The following indicates the general focus of the workshop. However, related contributions are welcome as well.

* case studies for (semantic) Web services
* OWL-S, WSMO, WSDL-S, IRS, SWSF-based systems and applications
* static and dynamic logics for Web services and related aspects
* ontologies for modeling (semantic) Web services
* formal languages for describing (semantic) Web services
* ontologies and languages for process modeling for (semantic) Web services
* ontological representation of quality of services (QoS), services level agreements (SLAs), and non-functional properties (NFPs) of Web services
* formal languages for QoS, SLAs, and NFPs
* reasoning tasks and their complexity in SWS
* formal methods and their applications in Web services
* validation and verification for Web services
* advertising, discovery, matchmaking, selection, and brokering of (semantic) Web services
* data/process/protocol mediation in (semantic) Web services
* composition, planning, and re-planning with (semantic) Web services
* execution and lifecycle management of (semantic) Web services
* monitoring, adaptability, and recovery strategies for (semantic) Web services
* policies for (semantic) Web services
* semantics in Web services contracts
* security and privacy for (semantic) Web services
* semantics for Grid services and e-Services
* architectures for (semantic) Web services deployment
* tools, middleware, and infrastructure for (semantic) Web services

WORKSHOP FORMAT AND ATTENDANCE

The program will occupy a full day, and will include presentations of papers selected from the full papers category (see 'submissions' below).

Please note that at least one author of each accepted submission must attend the workshop. The BPM 2006 conference formalities are applied for fees and respective organizational aspects. Submission of a paper is not required for attendance at the workshop. However, in the event

that the workshop cannot accommodate all who would like to participate, those who have submitted a paper (in any category) will be given priority for registration.

SUBMISSIONS

The workshop invites different types of contributions:

* Papers
* Demos
* Posters / Position papers

Papers:The papers should not exceed 12 pages and should have the Springer Lecture Notes of Computer Science (LNCS) layout.

Demos: Detailed description plus sufficient number of screenshots or a video of the demo are required. For paper-based submissions, please follow the Springer LNCS layout. Please note that at the workshop itself no technical support is provided except possibly Internet connection and power (to be confirmed).

Posters/Position papers: The posters/position papers should not exceed 5 pages and should have the Springer LNCS layout.

All contributions will be peer reviewed by a program committee that will incorporate well recognized experts in the area of semantic technologies and Web services.

All submissions should be formatted in Springer's LNCS style, should be submitted in electronic format using the link:
<A HREF="http://www.easychair.org/semantics4ws2006/">http://www.easychair.org/semantics4ws2006/</A>
.

All accepted full papers and all position papers of attendees will be published in the proceedings of the workshop. Workshop proceedings will be published with Springer LNCS and will be available at the workshop.

IMPORTANT DATES

Submissions: May 1, 2006
Acceptance: May 23, 2006
Final copy: June 7, 2006
Workshop day: September 4, 2006

ORGANIZING COMMITTEE

Steven Battle (Hewlett-Packard Labs, UK),
John Domingue (The Open University, UK),
David Martin (SRI International, USA)
Dumitru Roman (DERI Innsbruck, Austria)
Amit Sheth (University of Georgia, USA)

PROGRAM COMMITTEE (confirmed; to be extended)

- Rama Akkiraju, IBM, USA
- Abraham Bernstein, University of Zurich, Switzerland
- Carine Bournez, W3C, France
- Jorge Cardoso, University Mediera, Portugal
- Sanjay Chaudhary, DA-IICT, India
- Marin Dimitrov, Ontotext, Bulgaria
- Dieter Fensel, DERI, Austria
- Karthik Gomadam, University of Georgia, USA
- Michael Gruninger, University of Toronto, Canada
- Sung-Kook Han, Won Kwang University, South Korea
- Rick Hull, Lucent, USA
- Deepali Khushraj, Nokia, Finland
- Michael Kifer, State University of New York at Stony Brook, USA
- Michael Maximilien, IBM, USA
- Sheila McIlraith, University of Toronto, Canada
- Massimo Paolucci, DoCoMo Euro-Labs, Germany
- Tony Shan, Wachovia Bank, USA
- Stuart Williams, HP Bristol, UK

Appendix 5.2: Document 5 Results

```
<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Bernstein:Abraham.html</uri>
      <entityName>Abraham Bernstein</entityName>
      <confidence>145</confidence>
      <charOffset>7634, 7652</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/s/Sheth:Amit_P=.html</uri>
      <entityName>Amit P. Sheth</entityName>
      <confidence>135</confidence>
      <charOffset>7517, 7529</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/m/Martin:David.html</uri>
      <entityName>David Martin</entityName>
      <confidence>155</confidence>
      <charOffset>7440, 7453</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Khushraj:Deepali.html</uri>
      <entityName>Deepali Khushraj</entityName>
      <confidence>90</confidence>
      <charOffset>8041, 8058</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/f/Fensel:Dieter.html</uri>
      <entityName>Dieter Fensel</entityName>
      <confidence>145</confidence>
      <charOffset>7837, 7851</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/r/Roman:Dumitru.html</uri>
      <entityName>Dumitru Roman</entityName>
      <confidence>100</confidence>
      <charOffset>7478, 7492</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Domingue:John.html</uri>
      <entityName>John Domingue</entityName>
      <confidence>175</confidence>
      <charOffset>7399, 7413</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Cardoso:Jorge.html</uri>
      <entityName>Jorge Cardoso</entityName>
      <confidence>105</confidence>
      <charOffset>7719, 7733</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Gomadam:Karthik.html</uri>
      <entityName>Karthik Gomadam</entityName>
      <confidence>100</confidence>
      <charOffset>7868, 7884</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/d/Dimitrov:Marin.html</uri>
      <entityName>Marin Dimitrov</entityName>
      <confidence>90</confidence>
      <charOffset>7800, 7815</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/p/Paolucci:Massimo.html</uri>
      <entityName>Massimo Paolucci</entityName>
      <confidence>135</confidence>
      <charOffset>8223, 8240</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/g/Gr=uuml=ninger:Michael.html</uri>
      <entityName>Michael Gr&uuml;ninger</entityName>
      <entityName>Michael Gruninger</entityName>
      <confidence>105</confidence>
      <charOffset>7914, 7932</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/k/Kifer:Michael.html</uri>
      <entityName>Michael Kifer</entityName>
      <confidence>115</confidence>
      <charOffset>8076, 8090</charOffset>
</entity>

<entity>
```

```
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/a/Akkiraju:Rama.html</uri>
      <entityName>Rama Akkiraju</entityName>
      <confidence>95</confidence>
      <charOffset>7608, 7622</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/c/Chaudhary:Sanjay.html</uri>
      <entityName>Sanjay Chaudhary</entityName>
      <confidence>90</confidence>
      <charOffset>7765, 7782</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/b/Battle:Steven.html</uri>
      <entityName>Steven Battle</entityName>
      <confidence>90</confidence>
      <charOffset>7356, 7370</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-
tree/w/Williams:Stuart.html</uri>
      <entityName>Stuart Williams</entityName>
      <confidence>90</confidence>
      <charOffset>8301, 8317</charOffset>
</entity>

<entity>
      <uri>http://dblp.uni-trier.de/db/indices/a-tree/h/Han:Sung-
Kook.html</uri>
      <entityName>Sung-Kook Han</entityName>
      <confidence>100</confidence>
      <charOffset>7965, 7979</charOffset>
</entity>
```