Paper Review Form

Reviewer Name: ShaSha(Amy) Liu

Paper Name: Database and Information-Retrieval Methods for Knowledge Discovery

Section I. Overview

A. Reader Interest

1. Which category describes this manuscript?
   ___Practice/Application/Case Study/Experience Report
   ___Research/Technology
   _x_Survey/Tutorial/How-To

B. Content

1. Please explain how this manuscript advances this field of research and/or contributes something new to the literature.
   This paper gives a survey about how to integrate database and information-retrieval methods together in order to discover knowledge. Database and information-retrieval are separate fields in computer science and both have different views towards knowledge representation, query and ranking, etc. However in recent years, with the explosion of digital information, some researchers try to combine two methods and build a comprehensive knowledge base from both the web page and structured literature. The authors identified the motivations for bringing IR concepts to DB and vice versa, and give a summary on the existing and emerging applications which extract information from web, organize them into a knowledge base and discover relationship among them.

C. Presentation

1. Does the introduction state the objectives of the manuscript in terms that encourage the reader to read on?
   _x_Yes
   ___Could be improved
   ___No

2. How would you rate the organization of the manuscript? Is it focused? Is the length appropriate for the topic?
   _x_Satisfactory
   ___Could be improved
   ___Poor

3. Please rate and comment on the readability of this manuscript.
   _x_Easy to read
   ___Readable - but requires some effort to understand
Section II. Evaluation

Please rate the manuscript. Explain your choice.
___Award Quality
___Excellent
_x_Good
___Fair
___Poor

Section III. Detailed Comments (provide your thoughts/criticism about the ideas in the paper; not only summarize the paper but have a critical look here)

Database and Information-Retrieval emphasized different aspects of information management. Database aims at data storage, data consistency, transaction and recovery, while IR focuses on how to understand the meaning of text, statistical language models and user interface. It would be interesting to see these two methods can go side-by-side and help build a better way to represent knowledge and facilitate query.

Additional Comments:
1. Provide one aspect that you liked the most in this paper.
The paper provides several query examples that are hard to answer by today's search engine. Google and other keyword-based search engine cannot process two unrelated concepts in one single query, which brings more and more concerns. The paper points out these problems and shows us there are ongoing research efforts that are trying to address them.

2. Provide one aspect that you disliked the most in this paper.
The paper gave an outline of what knowledge base might look at, which is part of YAGO's typed entity-relationship graph. However the capability of YAGO's graph approach, namely RDF or Owl-Lite, is limited in its expressiveness and conceptualization of knowledge, thus it is not sufficient to capture the entity and relationship in the web. Furthermore, the authors do not mention LoD (Linked open Data) Cloud, which includes more than 70 large datasets contributed by experts belonging to diverse communities and aims at building a knowledge base and enable user to navigate among those separate dataset as easy as they browse web pages. YAGO might be among one of the possibilities but it is not the only one.

Section IV. Discussion Points (provide at least 3 discussion topics/questions related to ideas/techniques described in the paper; these will be used for discussions in the class)
1. how to obtain information from web
2. how to represent information from web and form expressive knowledge remains a big problem
3. how to query knowledge