

Working Towards Ontology Generation from Context of Listening to Presentations

Yuki Matsuoka, Ikki Ohmukai, and Hideaki Takeda
The Graduate University for Advanced Studies, National Institute of Informatics
{m-yuki, i2k, takeda}@nii.ac.jp

Abstract. Many studies have been conducted concerning ontology generation from the tags of social tagging services and some tagging services target academic literature. We think that it is unsuitable to search for technical information using the generated ontology from these tags, because they are often taken directly from the contents. In order to obtain contextual information on the papers, we focused on the notes taken when users listen to a presentation. We implemented and applied a system where the participants can input two kinds of memos on Web pages for presentations at conferences. As a result of this operation, we obtained words related to the content of the papers and to the background knowledge and research details.

1 Introduction

Social tagging services like del.io.us have recently become very popular. There are even social tagging services for scholarly information like CiteULike and BibSonomy. Some studies have been made on ontology generation based on social taggings [1][2][3]. This means that tags are used for a semantic understanding of the contents. Tags are useful information, but they are often taken directly from the contents themselves. Users sometimes need extra information from the contents to better understand professional documents, such as technical papers.

We propose a new method that gathers useful words from the memos written about presented papers. These memos are written by people who have listened to the presentations at a conference. The presenters may use some important words related to the content of their papers even if they are not written in the paper. Moreover, users may hit on words related to the content of the literature by using their background knowledge when they listen to the presentation. We implemented and applied the system, which we call “memoQ” [4], with which participants can input two kinds of memos on the Web pages for a presentation at the conference.

2 Operation of the system

The memoQ system anonymously publishes the memos that include a “?” as *question memos*, and keeps other memos private as *personal memos*. We offered the

system as a part of the support system for the 21st annual conference for the Japanese Society for Artificial Intelligence 2007, which was held from June 20th to 22nd, 2007. We targeted Web pages that included information on the papers presented at the conference. After the operation, there were 61 participants who input question or personal memos. We obtained 2556 memos (1849 question memos and 707 personal memos). There were 2836 nouns in the question memos, of which 1383 of them were written in the literatures (48.8%). Also, there were 1968 nouns in the personal memos, of which 1189 of them were written in the literatures (60.4%). The personal memos had a higher ratio of nouns written in the literatures, because the participants tended to input their personal memos as conference notes.

We investigated the question and personal memos input in the session for the “semantic Web” at the conference. The nouns in the personal memos that were not written in the papers include words like “OWL” and “BLAST”. “OWL” was a word that the presenter used to describe ontology. In addition, “BLAST” was a kind of bioinformatics Web service that the presenter spoke of. On the other hand, the nouns in the question memos that were not written in the papers included words like “primitive task” and “maintenance”. The words “primitive task” were used by a questioner about the task resolution of Web services, and the word “maintenance” was used by a questioner about the maintenance of ontology. In this way, the words related to the content of the papers were found as personal memos and the words concerning the background knowledge and research details were found as question memos. In other sessions, a similar tendency was seen.

3 Conclusion and Future work

We propose gathering useful words from two kinds of memos input into a system when users listen to presentations at academic conferences. According to the system, the words related to the content of the literature were included in the personal memos and the words about the research techniques were included in the question memos. We found that the types of words input into the memos were different between the memos, based on question use and personal use. Our future work is to find the semantic relationship between the words in these memos.

References

1. Mika, P., Ontologies are us: A Unified Model of Social Networks and Semantics, in: Proceedings of the 4th International Semantic Web Conference (ISWC 2005), 2005.
2. Xian Wu, Lei Zhang, and Yong Yu: Exploring social annotations for the semantic web, in: Proceedings of 15th International World Wide Web Conference (WWW2006), 2006
3. L. Specia and E. Motta: Integrating Folksonomies with the Semantic Web, in: Proceedings of the European Semantic Web Conference (ESWC 2007), 2007.
4. Yuki Matsuoka, Ikki Ohmukai, and Hideaki Takeda, Collaborative Memo-based System in the Academic Conference, IADIS International Conference WWW/Internet 2007 (ICWI2007),2007 (to appear)