

# A New Approach to Personal Network Search based on Information Extraction

Jie Tang, Mingcai Hong, Jing Zhang, Bangyong Liang, and Juanzi Li

Department of Computer Science and Technology, Tsinghua University  
10-201, East Main Building, Tsinghua University, Beijing, China, 100084  
{tangjie, hmc, zhangjing, lby, ljz}@keg.cs.tsinghua.edu.cn

**Abstract.** In this paper, we propose a new approach based on information extraction for constructing a personal network by discovering and identifying personal information from the web. We have developed a system called 'Personal Network Search'. The paper describes the architecture and main features of the system.

## 1. Introduction

Personal network search has become a popular research area in Semantic Web recently and many research efforts have been made so far. A person usually has different types of information: personal profile (including personal picture, homepage, position, affiliation, and publications), contact information (including address, email, telephone, and fax number), and social network information (including personal or professional relationships between persons). However, the information is often hidden in heterogeneous and distributed web pages.

Existing approaches usually construct the personal network manually [Mika, 2005]. In this paper, we try to address personal network search in a novel approach. We use the FOAF (Friend-Of-A-Friend) ontology in our system, which has been widely accepted and used for describing personal information and social network. The key point then is how to construct the personal network. Our proposal is to take a strategy of 'collecting-extracting-fusing'. We first collect documents 'related' to a person. By 'related', we mean documents that really describe the person rather than the keyword-based 'relevant' documents in the conventional search. Next, we employ information extraction technologies to extract the personal information from the collected documents [Tang, 2005]. After that, we integrate the extracted information from the documents of different sources. We use the extracted personal information and the publication information to create instances of the concepts "Person" and "Document" of FOAF. The created instances are stored in a local ontology base. In this way, we are able to construct a personal network automatically.

We have developed a system based on the proposed approach called 'Personal Network Search' (PNS shortly). Figure 1 (a) shows the processing flow of the system. In PNS, the user inputs a person name, and the system returns the information of the person. There are two types of searches: offline search and online search. Given a

person name, we first search in the constructed personal network. If the person can be found, the information of the person stored in the local ontology base will be displayed. (We call this type of search as offline search.) Otherwise, we perform the online search. We utilize Google API to get a list of relevant documents. Then a classification model is employed to identify whether or not a document in the list is really ‘related’ to the person. Typical related documents can be persons’ homepages, publication pages from DBLP, and persons’ introduction documents. Next, we extract personal information from the identified documents. We employ two classification models to detect the start position and the end position for each type of the personal information. And then we view the text between the start position and the end position as the target. As models, we use SVMs (Support Vector Machines). Features are defined in the SVM models respectively for each type of the information. We also extract person’s publications from DBLP. After that, we integrate together the personal profile, contact information, and publication information. We view coauthors in publications as the friends and finally construct the personal network.

Semantic search has been implemented based on the constructed personal network. There are four types of searches supported in the current system: person search, expert search, association search, and publication search. Figure 1(b) indicates a person search results. (In this example, the user types a person name, and he gets a detailed description of the person.)

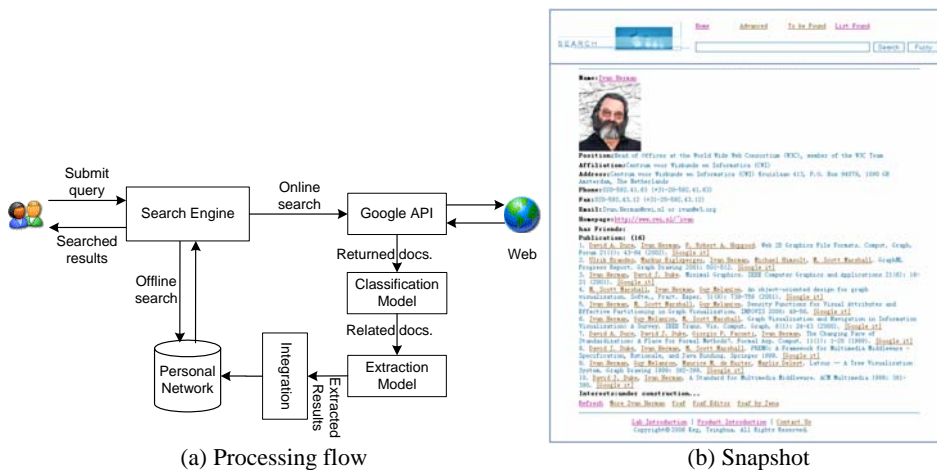


Fig. 1. Personal Network Search system

The demo will explain in detail the architecture and the main features of the system. We will also explain how the personal network is constructed and how the proposed approach is implemented in the Personal Network Search system.

## References

- [Mika, 2005] P. Mika. Flink: Semantic Web Technology for the Extraction and Analysis of Social Networks. *Journal of Web Semantics*. July, 2005
- [Tang, 2005] J. Tang, J. Li, H. Lu, B. Liang, and K. Wang. 2005a. iASA: learning to annotate the semantic web. *Journal on Data Semantic, IV*. Springer Press. pp. 110-145