Anonymization of personal stories

Lars Ahrenberg

Niklas Blomstrand

Department of Computer and Information Science

Linköping university

Marinette Fogde

Andreas Nilsson

The Museum of Work

Norrköping

An important activity of many Swedish museums and archives is the collection of personal stories that form an important part of the Swedish cultural heritage. They reflect their times in many ways: historically, linguistically and culturally. They are therefore of interest to researchers with different specializations such as ethnologists, historians, and linguists.

The Museum of Work (Arbetets museum) in Norrköping has in its archive more than 2600 personal stories from individuals about work and everyday life. The archive of the museum holds documentations of working life, depictions of life in industrialized societies, on housework in the 1950s and 1960s, and professional memories of miners, social workers, retail employees, and many other professions. Efforts to collect stories are ongoing and take place in various socially relevant areas.

The main target group are scientists, but a selection of stories are presented in anthologies, in exhibitions and on the museum's website. In these cases the informants have given their consent to publication. However, in other cases the informants have not signed a personal consent form regulating the use of their personal stories. Therefore, in accordance with the Personal Data Act, the museum has not made them public.

The most appropriate channel to make the personal stories public would be the Internet. This would make them searchable in a much more flexible manner than has been possible so far. To support Internet publication we are planning a project aimed at design and implementation of a system for document anonymization of personal stories. Language technology tools such as part-of-speech taggers, named entity recognizers and coreference recognizers will be critical components of the system. In addition, there will be functions for text transformation and metadata annotation.

There are an increasing number of proposals for the design of document anonymization systems (Vico and Calegari, 2015). They must obviously be interactive and support a natural work flow with points of approval by appropriate domain experts for different stages of the process. Ultimately, it is the museum that guarantees that publication is in accordance with the Personal Data Act.

At the workshop we will report the work that has been done so far and our plans for the design and implementation of the system.

Reference

Horacio Vico and Daniel Calegari, 2015. Software Architecture for Document Anonymization. Electronic Notes in Theoretical Computer Science 314 (2015) 83-100.