Annotation tool for scanned handwritten documents

Duration: 6 months

Team: Loki (Inria Lille – Nord Europe & CRISTAL)

Advisors: Thomas Pietrzak (thomas.pietrzak@univ-lille.fr) & Stéphane Huot (stephane.huot@inria.fr)

Project: GeneaLire

Description

This internship is part of a larger project which aims at designing a helping tool for transcribing ancient documents. This tool will genuinely combine interactive and automatic methods. Indeed, automatic methods are not sufficient, first of all because they require a hand-made knowledge database. Second, the user must keep the control on ambiguities management. Third, we would like users to gain skills, which will only be possible if the user has an active role.

Documents such as civic or church records follow typical structures, with typical pieces of information: places, names, dates, occupations, relationships… While transcribing a document, a researcher can identify the type of information of a word, without being able to understand the word itself. We would like users to be able to specify such information with an annotation tool. Annotation tools are used to link or add pieces of information to an existing document [1].

We propose to design, implement and evaluate a tag-based annotation tool. Tags have several advantages: they are non-hierarchical, and multiple tags can be attached to a single item. The system will feature different visualization techniques to highlight the structure of the document, and the progress of transcription. Thanks to these techniques, the researcher will be able to identify words with glossaries, or with machine learning techniques. These techniques will most likely be more efficient with knowledge about the context.

This work will consist in:

- Studying annotation tools, in particular tagging systems.
- Defining a design space for document annotation techniques.
- Designing, implementing and evaluating the current technique.

Candidate

The ideal candidate is a MSc student or equivalent with a major in computer science, and shows a great interest in HCI research. He must have experience or a strong interest in software development. Creativity, independence, team work as well as great communication skills are valuable advantages.

Working environment

The internship will take place in the Loki team in Lille, France, joint between Inria – Lille Nord Europe and the CRISTaL (UMR CNRS 9189) laboratory of the University of Lille. Supervisors: Thomas Pietrzak and Stéphane Huot.

Bibliography