

## Text selection tool for scanned handwritten documents

Duration: 6 months

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

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

Project: [GeneaLire](#)

### Description

This internship is part of a larger project which aims at designing helping tools for transcribing ancient documents. This tool will genuinely combine interactive and automatic methods. Indeed, automatic methods such as machine learning 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.

We are interested in designing, developing and evaluating an interactive text selection tool for scanned handwritten documents. Classical selection tools such as free forms and various *magic wands* are not adapted. Our approach is a selection brush with 4 degrees of freedom: x-y position, brightness threshold and selection size. Our first investigations are promising [1], but we still have to evaluate it with a controlled experiment.

The mapping of 4 degrees of freedom is complicated with a keyboard + mouse/touchpad settings. We will investigate a pen + touch setting on a tablet [2].

This work will consist in:

- Studying pixel selection tools, in particular for text manuscripts.
- Defining a design space for pixel selection techniques.
- Evaluating the current technique.
- Adapt the technique for pen + touch interaction on a tablet.

Depending on the progress of the project, these results may be submitted as a research paper.

### 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. It is not required to speak French.

### 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](#), [Sylvain Malacria](#) et [Stéphane Huot](#).

### Bibliography

[1] Alawoe, E., Pietrzak, T., Huot, S. *Outil de sélection de texte manuscrit sur des documents numérisés*. 2018. Symposium international francophone sur l'Écrit et le document (SIFED 2018).

[2] Hinckley, K. et al. *Pen + touch = new tools*. 2010. Proc. UIST '10, 27–36.