

## Master Internship (and associated PhD):

### Timeline Visualization of document collections and extracted knowledge, for data journalism

#### Internship Summary

The Master internship, and associated PhD topic, is in the area of data analysis and visualization. The work will be conducted in Paris France, within the ILDA Inria team and the University Paris-Sud. If the internship progresses well, there is a PhD position funded associated to it.

#### Context and Description of larger project:

One of the challenges data journalists face today, is how to organize and understand large document collections, such as existing news articles and media libraries, to help them infer new knowledge. Their task often involves collaboratively inspecting and analyzing these heterogeneous information sources. Our multidisciplinary project stands at the crossroad of multiple research fields— visualization, content analysis, data management, knowledge representation, and of course data journalism. It is a joint project between multiple Inria teams, journalist organizations and news agencies (le Monde, Ouest France, AFP).

The goal of the internship (and later the PhD) is to explore different ways to visualize and interact with knowledge extracted automatically from very large heterogeneous document collections. This extracted knowledge will be in the form of a multi-attribute graph of extracted entities and relationships between them. The interactive visualizations developed in the internship and later PhD thesis, will include approaches to collaboratively analyze and organize content, and to interactively update and validate the extracted knowledge. Approaches may include the use of new collaborative platforms, such as high-resolution wall displays and tabletops. User evaluation sessions with journalists will be part of the work, in order to validate the visualization designs.

The student will be working with visualization researchers, but also researchers and engineers in text and data analysis, that will provide the tools for the automatic analysis and extraction of the entities and their relationships.

**Internship:** As a first step, the internship will look at proposing ways to visualize (and interact) with timelines that show how particular pieces of knowledge have evolved over time. E.g., the timeline of positions that one or more people have held in different organisms and if they co-inside, all the events associated with a geographical locations over a time period, etc. As a second step, it will look at how to indicate possible contradictory information within the timeline. The proposed designs will be implemented and evaluated (if needed with journalists). The student can develop the prototypes either on desktop or on a collaborative platform (digital wall / table).

### **Requirements and Applications:**

A BSc in computer science or closely related field is required. Prior focus on visualization and experience in programming (Java and/or JS) are also required. User-study experiences, and French knowledge are a plus.

Applications must include a detailed CV/resume, information on educational background, a brief description of practical work and research experience in computer science, and a short statement of motivation and goals.

Applications should be sent by **Jan 15<sup>th</sup>** 2018, but the position may be filled before that. Start time is estimated March/April 2018.

### **Contact:**

Anastasia Bezerianos

Dep. of Informatics Univ. Paris-Sud & Inria ILDA team

email: [anab@lri.fr](mailto:anab@lri.fr)

url: <http://www.lri.fr/~anab>

Team url: <http://ilda.saclay.inria.fr/>