INSA de Rennes recruits

a Post-doctoral researcher/engineer in computer science

Combination of skills in logic and statistics
For the automatic recognition of ancient hand-writing

Conditions

Establishment : **INSA de Rennes, IRISA laboratory**

Service : IRISA team of research **Intuidoc**

A 12-months contract to take as soon as possible

Context

Intuidoc project (https://www.irisa.fr/intuidoc): the IRISA team conducts one of its researches on the structure of old documents, hand-written or degraded (musical partitions, archives, newspapers, letters, electronic schema, …).

Doptim (https://www.doptim.eu) is a start-up created in 2017 and specialized in data science with two activities : consulting for local companies who want to get value from their data and R&D to develop software at state of the Art in machine learning and big data technologies.

The collaboration between Doptim and Intuidoc concern images of handwritten registers mainly studied by worldwide genealogists who try to reconstitute their rich family story.

Doptim develops a web platform, Geneafinder, (http://www.geneafinder.com) to facilitate the information searching in the millions of online images.

One technique is particularly considered : the possibility to cut register images into a collection ofbirth, death or unions acts to display rapidly such acts in readable format. Reader avoids time-consuming manipulation and continues on transcription in a simplified semantic context.

Missions

Intuidoc previous research led to the creation of the DMOS (Description and MOdification of the Segmentation) method. DMOS is a grammatical method to recognize parts of document images.

An extension of this method, DMOS-P, adds a supplementary dimension by taking into account several levels of perception of a same image, for example, considering different resolutions. The method is generic and can be applied to the recognition of any kind of document.

INSA RENNES
20, Avenue des Buttes de Cöesmes
CS 70839 - 35 708 Rennes Cedex 7
Tél.+ 33 (0)2 23 23 82 00 - Fax + 33 (0)2 23 23 83 96
www.insa-rennes.fr

The present project focuses on the analysis of documents written during the 17th and 18th centuries. A first prototype cuts images into acts by analyzing the position of text lines. It is not effective enough and we shall now consider additional visual indices, as output of statistical analysis or output of text recognition systems.

The objectives of this post-doc will be the following :

- Handle the different techniques used in the team to analyze documents
- Propose indices based on the statistical properties of the pages to improve the cutting
- Integrate together existing software blocks to improve the text recognition
- Apply the software and measure its performance of a set of images provided by Doptim
- Create a referential to measure the improvement of the performance at each software iteration
- Make the software able to cut an image on-the-fly (few milliseconds after a user loads an image for immediate recognition)
- Make the software flexible with different set of grammars to analyze images on-the-fly
- Document and harden the software to allow its integration in the cloud (multi-servers, robustness, operational statistics, remote debugging and updates, release control)

Main Skills

We are looking for a doctor having a thesis in the domain of document recognition or statistical analysis.

Skills in grammars and languages and/or logical programming are nice-to-have.

Environment

The 12-months contracts will be hosted in the IRISA laboratory premises, in Intuidoc team. It will be supported by Bertrand Coüasnon and Aurélie Lemaitre, associate professors.

Doptim will integrate results by iteration in an experimental project set up with local genealogists. The project is then built in agile mode with sustained deliveries and field feedbacks. Doptim will propose a validation platform, a technical support, especially on statistical algorithms and will organize weekly synchronization meetings.

For more information, you can contact via email :

bertrand.couasnon@irisa.fr, aurelie.lemaitre@irisa.fr sophie.tardivel@doptim.eu

INSA RENNES
20, Avenue des Buttes de Cöesmes
CS 70839 - 35 708 Rennes Cedex 7
Tél.+ 33 (0)2 23 23 82 00 - Fax + 33 (0)2 23 23 83 96
www.insa-rennes.fr