



Post-Doctoral job offer

Automatic classification of business documents

The LabCom IDEAS calls for applications for a post-doctoral position in computer science, in the field of machine learning, for the classification of business documents.

Duration: 12 months (with possibilities of renewal)

Desired hiring date: September 1st, 2020 (may be modified according to the sanitary situation)

Take-home salary: 2100 € / month (French public healthcare coverage included)

Workplace: LabCom IDEAS, on the premises of the L3i laboratory in La Rochelle, France

Specialities: Computer Science / Machine Learning / Classification / Image Analysis / Automatic Language Processing

Description of the LabCom:

The work carried out by the candidate will be part of the LabCom IDEAS. IDEAS is co-funded by the French National Research Agency (ANR) and the Région Nouvelle-Aquitaine, and brings together the Yooz company and the L3i laboratory. The IDEAS objective is to imagine, invent, design, develop, optimise and train the best algorithms for processing automatically business documents. The goal is to offer services, based on artificial intelligence, that are capable of automatically analysing and understanding various types of business documents.

The post-doc fellow will be based in the LabCom, located in the premises of the Laboratory L3i, in La Rochelle, France.

The L3i laboratory, created in 1993 at La Rochelle University brings together researchers in Computer Science and Signal Processing from different faculties. The L3i brings together the skills of its researchers in order to address the issues of digital content enhancement from a systemic perspective. This relies, in particular, on a cross exploitation of interactive applications, content indexing and knowledge representation. The laboratory is structured around three scientific themes (Knowledge Engineering, Content Analysis and Management, Interactivity and Dynamic Systems), centred on the common goal of interactive and intelligent management of digital content.

Yooz, Labcom's industrial partner, is a provider of a Cloud service for automating purchasing and payment processes. Yooz integrates Artificial Intelligence technologies to automate the processes and document processing involved in these processes. The yooz service is used daily by nearly 3000 users.

The research and development carried out within the LabCom focuses on 3 main areas:

- Document classification
- Document search
- Document fraud detection

Job description:

The work of the post-doc fellow will fall within the framework of the area "Document Classification". The aim is to design innovative approaches for classifying documents (according to their nature: invoice, quotation, bank account details, ...) in multi-channel incoming document flows and to create, from these approaches, a prototype.

There are many scientific bottlenecks arising from this applicative context, mainly in the field of machine learning:

- document classes are generally very unbalanced in existing corpuses. Indeed, some classes are very well represented in the learning base (with many training documents), while others are much less represented (if present at all). As a result, the approaches developed so far offer very uneven rates of accuracy between classes.
- the intra-class variability is very large (sometimes even greater than the inter-class variability). An example of this is the fact that two documents from the same company but from different classes (*e.g.* an invoice and a quotation) may be closer than two invoices from different companies in the representation space, both visually and in terms of textual content.

This post-doctoral work will be based on a detailed state of the art of existing approaches, to identify their limits and propose innovative approaches that will help to overcome the bottlenecks mentioned above. To solve these problems, we plan to use machine learning techniques that, based on existing image and/or text content classification techniques, allow to:

- take both modalities (image and text) into account jointly for document classification (multi-modality), in order to improve accuracy for most classes
- learn a class from very few (or even 0) examples (zero-shot learning), possibly on the fly, in the document flow
- effectively implement a rejection strategy when the document to be classified is too "far" from existing classes, or when the ambiguity between classes is too great (with thresholds that will be set automatically or semi-automatically, depending on the corpus).

If the post-doc fellow would like to acquire/reinforce his/her experience of working in a private company, we could arrange short collaborative working stays in the premises of Yooz company, at Aimargues (Mediterranean coast).

Candidate Profile:

The candidate, who holds a Ph.D. in the fields of computer science, computer engineering, signal processing, or applied mathematics, must have a significant research experience in at least two of the following areas:

- Machine learning / classification
- Image analysis
- Pattern recognition

Moreover, knowledge or experience of Automatic Language Processing would be appreciated.

The candidate's skills will include:

- Mastering one or more programming languages (Java, Python, C/C++...)
- Very good teamwork skills, having knowledge or experience of Agile methods would be a plus (as the work will be carried out both in conjunction with researchers from the L3i laboratory and the R&D department of the Yooz company).
- Good scientific writing skills, and fluency in writing and speaking English.

To apply:

Candidates for this position should send a CV and a cover letter (names and reference details would be appreciated) to:

- [muriel.visani \[at\] univ-lr.fr](mailto:muriel.visani@univ-lr.fr)
- [nicolas.sidere \[at\] univ-lr.fr](mailto:nicolas.sidere@univ-lr.fr)
- [Vincent.PoulaindAndecy \[at\] getyooz.com](mailto:Vincent.PoulaindAndecy@getyooz.com)
- [Aurelie.Joseph \[at\] getyooz.com](mailto:Aurelie.Joseph@getyooz.com)

Applications will be considered as they arise, so there is no strict deadline for applying, even though we would prefer to have selected the best applicant by mid-July.