2 Ph.D. positions at LIUPPA
(University Pau & Pays Adour)

Summary

We are looking for **two highly-motivated Ph.D. students** to propose a new generic Business Information System that is customizable with rich multimedia services concerning data indexation, storage, enrichment, security and presentation for several domains, and especially Energy and Environment applications. In the LIUPPA lab, the two selected candidates will develop research activities related to **Data Modeling and Reasoning, Information Retrieval, Semantic Web** (e.g., Linked Data and ontologies)**, and Privacy Protection**.

Ph.D. context and objectives

The ever-increasing need for extracting, managing, storing, publishing and retrieving data from heterogeneous contents has become a major concern in modern large-scale projects. This is particularly observed in many application domains where several actors (with different expertise, roles, preferences and rights) exchange a great amount of information at any stage of a project. As an example, in the construction industry, actors (i.e., owners, consultants and contractors) exchange contracts, technical specifications, administrative forms, technical drawings and on-site photos throughout the different stages of a construction process. The interchanged documents, originated from different sources, do not usually have a common standard structure. Also, they show heterogeneity in their formats (e.g., pdf, docx, xlsx, jpeg, etc.), contents (e.g., architecture, electrical, mechanical, structure, etc.), media types (e.g., image, text, etc.) and versions.

Consequently, there is a crucial need to help any project actors: (1) **to integrate, to store, to model and to index large collections of data**, especially multimedia data; (2) **to access data** according to specific needs and preferences with a user-friendly system; and (3) **to offer added-value services,** such as data enrichments and/or data security. We would like to **propose a personalized Business Information System for various companies or organizations**.

Offering such a generic and tunable system to any companies or organizations raises many research challenges that we have listed and detailed hereafter.

1. **Data integration, modeling, storing and indexing:**

   o **Multimedia Cloud:** the system should consider any types of contents (multimedia documents, sensors data…) and any specific metadata models. This information would be incorporated and stored in a Cloud-based infrastructure. Hence, users would have access to such a data at anytime, anywhere and anyhow. The cloud would also be used to execute (on-demand or not) specific indexing processing/services to complete the metadata.

- o **Data representation and optimized indices:** In recent research publications (Charbel et al., INFORSID 2017[1] and DEXA 2017[2]), we have proposed an ontology based-model that combines existing metadata standards (e.g., DC, TEI, EXIF, MPEG-7) and introduces new components augmenting the capabilities of these standards. This model offers a pluggable layer that makes it adaptable to different domain-specific knowledge. To validate our model, we have conducted experiments with Nobatek (www.nobatek.com), a French technological resource center which is involved in the sustainable construction domain. More precisely, its main role is to ensure the compliance of a construction project with the environmental standards and quality performance. The RDF and OWL W3C standards have been used to implement our model, however the instances based on these languages may be verbose in syntax and may contain redundancies, hence slowing down query treatments and inferences (Regina et al., ER 2015[3]). Consequently, as done in the Information Retrieval research domain, optimized indices would be defined in the project.

2. **Data access:**

- o **User needs and preferences:** the system should consider various kind of users with different expertise levels and roles. Consequently, the system has to handle various vocabularies used by different communities and to align these terms if needed for adapting the results presentation.

- o **Query representation and optimized processing:** Many standard query languages (e.g., SQL and SPARQL) can be used to retrieve information. However, users that are not experts in databases or semantic web technologies cannot specify such kind of technical queries. Hence, the system may guide the user interactively to find his/her desired information. In this project, specific actions will be devoted to user-friendly query specification and query rewriting.

3. **Added value services:**

- o *Enrichments:* the system should connect its internal resources to external data, such as Open Data and/or Linked Open Data. These connections could contextualize data, i.e., present more details about a concept or some related contents.

- o *Anonymization and security:* the system may contain sensitive data and/or data with specific rights (legacy aspects). The system should protect these data or anonymize them automatically if needed.

- o *Matching, results presentation and navigation:* the system should offer a user-friendly interface to end-users to present his/her results and let him navigate through them by considering all the handled data and connections. It may also consider data preferences and security.

---

[1] Nathalie Charbel, Christian Sallaberry, Sébastien Laborie, Gilbert Tekli, Richard Chbeir: LinkedMDR: un modèle sémantique de représentation de corpus de documents multimédia. In *Actes du 35ème congrés INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID 2017).*

[2] Nathalie Charbel, Christian Sallaberry, Sébastien Laborie, Gilbert Tekli, Richard Chbeir: LinkedMDR: A Collective Knowledge Representation of a Heterogeneous Document Corpus. In *Proc. of the 28th International Conference on Database and Expert Systems Applications (DEXA 2017)*. Springer LNCS.

[3] Regina Ticona-Herrera, Joe Tekli, Richard Chbeir, Sébastien Laborie, Irvin Dongo, Renato Guzman: Toward RDF Normalization. In *Proc. of the 34th Int. Conf. on Conceptual Modeling (ER 2015)*, pp. 261-275, Springer LNCS.

Candidate specification

We are looking for two excellent students with a Master degree in Computer Science. The candidate should have:

- Experience in Data Modeling, Information Retrieval and the Semantic Web (e.g., Linked Data and ontologies);
- Strong programming skills in Java and Web programming languages;
- Excellent writing and communication skills in English;
- A good ability to communicate with others;
- The ability to work autonomously.

Good Theoretical background would be a plus. Experience in Multimedia Cloud and Big Data will be a plus too. And French language knowledge would appreciated.

How to apply

The applicants will be required to send a detailed up-to-date CV, Master degree transcripts, two references (one of which must be academic) and a covering letter by email to: Richard.Chbeir@univ-pau.fr, Sebastien.Laborie@univ-pau.fr and Christian.Sallaberry@univ-pau.fr.

The deadline to apply is the 15th of June 2018 but consideration of candidates will continue until the position is filled.

Start date and salary

The start date is by **November 2018** for both Ph.D students.
The two Ph.D. will be funded by a grant from the University of Pau, and especially E2S (Energy and Environment Solutions) with a gross salary of **1880** euros/month (take-home pay 1500€/month).

Location and Environment

- Location : Univ. Pau & Pays Adour located within Parc de Montaury – Anglet.
- Environment : The PhD students will integrate the T2I group of LIUPPA lab (http://liuppa.univ-pau.fr). The project is conducted in partnership with Bertin Technologies (https://bertin.fr) and Hupi (https://www.hupi.fr).