SUJET DE THESE / PhD SUBJECT

TITRE / TITLE: Generic Business Information System, customizable with rich multimedia services concerning data indexation, storage, enrichment, security and presentation

ABSTRACT:
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.

Mots clés (Keywords):

CONDITIONS D’EXERCICE / WORKING CONDITIONS

Laboratoire : LIUPPA Site web : https://liuppa.univ-pau.fr/

PhD n°1
Directeur de thèse (PhD Director): Richard Chbeir
Co-Directeur de thèse (PhD co-Director):
En collaboration avec (In Collaboration with – if any)

PhD n°2
Directeur de thèse (PhD Director): Christian Sallaberry
Co-Directeur de thèse (PhD co-Director): Sébastien Laborie
En collaboration avec (In Collaboration with – if any)

Lieu (Place) : Univ. Pau & Pays Adour located within Parc de Montaury – Anglet

Date début (start): November 2018 Durée (duration): 3 ans (years)

Employeur (employer): Université de Pau et des Pays de l’Adour (UPPA)

 Salaire mensuel brut (monthly salary before taxes): 1880 €

SAVOIR-FAIRE DU LABORATOIRE / HOST LABORATORY PROFILE

The LIUPPA is the UPPA computer science research laboratory. It addresses digital sciences in the broadest sense. Its research project focuses on the requirements and issues of a digital society in which the Internet (Internet of Things, social networks, etc.) and networks as a whole are increasingly impacting our daily activities. This affects IT systems, which are complex due to:

- Heterogeneous, multimedia and highly decentralized (cloud) data (in larger and larger volumes).
- Distributed and outsourced data processing.
• Different users (machines, humans) whose profiles and requirements differ and evolve.

In this context, even the interactions with and among these users need to be redefined to achieve better communication and more efficient cooperation.

The solutions put forward by the LIUPPA fall within the scope of two complementary areas:

• Information sciences and technologies based on information, knowledge and web data processing (data, documents, etc.).

• Software engineering centered on model, service and software architecture engineering.

This includes cross-functional issues related to security, image/signal processing and visualization, distributed systems, interaction and adaptation. The figure below summarizes this work.

---

**MISSION - ACTIVITES PRINCIPALES / MISSION – PRINCIPAL ACTIVITIES**

**I. Le contexte scientifique Scientific Context**

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.

**II. Les objectifs Objectives**

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.

**III. Plan de travail Work plan**

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). Consequently, as done in the Information Retrieval research domain, optimized indices would be defined in the project.

2. Data access:
   - **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.
   - **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:
   - **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.
   - **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.
   - **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.

IV. Références bibliographiques (Literature References)


COMPETENCES REQUISES / REQUIRED COMPETENCES

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.

CRITÈRES D’ÉVALUATION DE LA CANDIDATURE / CRITERIA USED TO SELECT CANDIDATE

Selection process steps:
- Establishment of the selection committee.
- Evaluation of the applicants’ cv
- Interview with the selected candidates and ranking.

Criteria used in selection of the candidate:
- The candidate's motivation, scientific maturity and curiosity.
- Candidate's knowledge.
- Candidate's marks and rankings in M1 and M2.
- English proficiency

CONSTITUTION DU DOSSIER DE CANDIDATURE, DATE LIMITE DE DEPOT / REQUIRED DOSSIER, DATE

Send an e-mail with your candidature containing:

- CV
- Cover letter detailing candidate's motivations
- Candidate's MSc marks and ranking
- Any letters of recommendation
- Contact details (for 2 referees)

DATE LIMITE DE DEPOT DU DOSSIER (limiting date): 15th of June 2018 but consideration of candidates will continue until the position is filled

CONTACTS

e-mail : Richard.Chbeir@univ-pau.fr, Sebastien.Laborie@univ-pau.fr and Christian.Sallaberry@univ-pau.fr