2017-00124 - PHD Reasoning with Link Keys

Level of qualifications required : Graduate degree or equivalent

Fonction : PhD Position

About the research centre or Inria department

Grenoble Rhône-Alpes Research Center groups together a few less than 800 people in 35 research teams and 9 research support departments.

Staff is localized on 5 campuses in Grenoble and Lyon, in close collaboration with labs, research and higher education institutions in Grenoble and Lyon, but also with the economic players in these areas.

Present in the fields of software, high-performance computing, Internet of things, image and data, but also simulation in oceanography and biology, it participates at the best level of international scientific achievements and collaborations in both Europe and the rest of the world.

Context

Working environment: The candidate will join the mOeX team, a team of the Laboratoire d’Informatique de Grenoble (LIG), and a joint research team between the Université Grenoble Alpes (UGA) and Inria. The student will also collaborate closely with the Laboratoire d’Informatique Avancée de Saint-Denis (LIASD).

Work place: The position will be held at Inria Grenoble Rhône-Alpes (Montbonnot-Saint-Martin, France). Grenoble is the capital of the Alps in France with excellent train connections to Lyon (1h), Geneva (2h) and Paris (3h). Grenoble is a top skiing destination in Europe. The candidate is also expected to travel to Paris often during the PhD.

Funding: This PhD will be funded by the ELKER ANR French project (grant number: ANR-17-CE23-0007-01). The partners of this project are LIG, Inria and Université Paris 8. The selected candidate is expected to be very involved in this project.

Doctoral school: Ecole doctorale Mathématiques, Sciences et technologies de l’information, Informatique - MSTII.

Advisors: Manuel Atencia (UGA & Inria) and Chan Le Duc (LIASD).

Salary: the gross annual salary is 25,200 euros and the net annual salary 20,160 euros.

File: Provide CV, motivation letter, master’s dissertation and transcript of marks. Recommendation letters are a plus.

Assignment

Context of the PhD

The purpose of the semantic web is to take advantage of formalised knowledge at the scale of the worldwide web. This has led to the release of a vast quantity of data, continuously growing, expressed in semantic web formalisms (RDF) generally called linked data [5]. Part of the added value of linked data lies in the links identifying the same entity in different datasets as it allows for making inferences across datasets. They may identify, for example, the same books and articles in different
bibliographical data sources. So finding the manifestation of the same entity across several datasets is a crucial task of linked data [4].

One way of identifying entities is to use link keys, which generalise keys in relational databases to the case of two different RDF datasets [2]. An example of a link key is:

\[
\{\langle \text{auteur, creator}\rangle, \langle \text{titre, title}\rangle\} \text{ linkkey } \langle \text{Livre, Book}\rangle
\]

stating that whenever an instance of the class Livre has the same values for properties auteur and titre as an instance of class Book has for properties creator and title, then they denote the same entity.

Link keys can be used for data interlinking — they specify the properties and classes to compare for discovering links — but, interestingly, they can also be treated as logical axioms and can thus be combined with other kinds of knowledge, like ontologies, to support logical reasoning. Reasoning with link keys may be helpful to deduce further links, either directly or indirectly by deducing new link keys to use for interlinking.

**Goal**

The goal of this PhD is to study reasoning procedures for link keys. Two main directions are expected to be followed. First, and in line with the works on reasoning with ontologies written in OWL, mainly based on tableau methods for description logics (DLs) [6, 3], the selected candidate will extend these methods for reasoning with link keys. Then, the student will study our previous work on rule-based methods for data interlinking [1] and adapt them to be able to reason with different kinds of link keys. The optimal joint use of these two approaches for data interlinking will be the ultimate objective of the PhD. The proposed methods will be implemented to come up with computationally efficient link key reasoners.

**References**


**Main activities**

**Expected results:**

- Formalisation of different kinds of link keys.
- Identification of the most expressive combinations of DLs with link keys that remain decidable and proof of undecidability beyond them. Design of efficient reasoning algorithms for the identified combinations.
- Implementation of the algorithms obtained from the research results. Building different benchmarks allowing to evaluate the correctness and scalability of the implemented reasoners.
- Rule-based methods for data interlinking using any kind of link keys.
- Evaluation of link key reasoning for data interlinking using real-world RDF datasets.

**Skills**
- Good knowledge of formal languages such as description logics and first-order logic.
- Good knowledge of semantic web formalisms and technologies such as RDF, OWL languages and OWL reasoners.
- Design and programming skills in Java or C++.

Benefits package

- Subsidised catering service
- Partially-reimbursed public transport
- Social security
- Paid leave
- Flexible working hours
- Sports facilities