

---

## PhD position – Thesis offer (M/F)

(Reference: DS\_PP\_PhD\_KBC\_042018)

---

<b>Research topics</b>	<b>Data management, Data quality</b>
<b>Department</b>	Data science
<b>Parution date</b>	April 2018
<b>Start date</b>	First semester of 2018
<b>Duration</b>	Duration of the thesis
<b>Description</b>	<p>A Knowledge Base System (KBS) is a large-scale integration of information built on two main components. The first is the knowledge base (KB), which is the repository of data. It is composed of triples in the form of (subject, predicate, object), where the entities (e.g., <i>The Mona Lisa painting</i> and <i>Leonardo Da Vinci</i>) are the subject/object and their relationship is an instance of a predicate (e.g., <i>createdBy</i>). The second component is the language used to express knowledge representation. This is made of a logical formalism for expressing facts and rules, and a reasoning engine that uses it. By understanding the data semantics, KBSs enable analytics such as querying with structured languages (e.g. SQL/SPARQL), and a wide array of learning and reasoning tasks. However, creating and maintaining a KBS is difficult. Domain specialists guarantee <b>quality</b>, but, because of the large and increasing number of documents, cannot do it manually. Automatic algorithms <b>scale</b> over large number of documents, but cannot provide high quality in the results, because of process complexity. Creating and curating high quality KBS cannot be done manually at scale and new automatic methods designed around the users need to be developed. We envision new hybrid systems that use the machine to scale over large datasets, and the human experts to teach the algorithms how to solve the hard cases. This vision involves challenges from multiple fields, including data integration, information extraction, data mining, natural language understanding, and probabilistic reasoning.</p> <p>This PhD program will address the following topics:</p> <ol style="list-style-type: none"><li>1. Combine and extend techniques to involve users in the extraction, cleaning and integration of structured data from text resources.</li><li>2. Develop methods to extract intentional knowledge from the data. This includes refining the schema, the types (classes), the predicates among types, and the rules that apply over the data (such as business regulations). The goal is to develop profiling and mining methods that discover metadata from the combination of evidence from the data, existing structured information, and the users.</li><li>3. Given the natural evolution of a KBS over time (new data source, changes in schemas, changes in extractors), study how to effectively monitor changes and identify what to expose to the users to keep high quality in the KBS.</li><li>4. Investigate and evaluate querying and reasoning tools to support real use cases provided by an industrial partner.</li></ol> <p>The objectives will be implemented through a number of work-packages and project-wide use cases that will also serve as ways to measure success in reaching expected impacts.</p>
<b>Requirements</b>	<ul style="list-style-type: none"><li>• Education Level / Degree: MSc (with excellent grades)</li><li>• Field / specialty: Computer Science</li><li>• Technologies: Data Management / Machine Learning / Excellent programming skills</li><li>• Languages : English [German is a plus]</li><li>• Strong mathematical and algorithmic background</li><li>• Solid motivation and personal preparation</li></ul>



## What we offer

- Strong international research environment with supervision from experienced faculty.
- Opportunities to collaborate and for research stays with renowned collaborators worldwide.
- Well-paid PhD position, in an area known for its beauty and the mild weather.
- English used for research and communication, but opportunity to learn French language will be provided free of cost.

## Application

The position is available immediately, The cutoff date for the application is May 15, 2018, but the application evaluation will start immediately, so early applications are encouraged. Interested individuals should submit the following documents (in English):

- I-Curriculum Vitae, including your current contact address, transcript of certificates and grades (with a list of university courses taken), and your previous publications (if any)
- II-Motivation letter including research and education perspectives
- III-Contact information for 2 referees at your current and/or previous affiliations

**Applications should be submitted by e-mail to [secretariat@eurecom.fr](mailto:secretariat@eurecom.fr) with the reference: DS\_PP\_PhD\_KBC\_032018**

**Postal address** CS 50193 - 06904 Sophia Antipolis, France

**Contact** [secretariat@eurecom.fr](mailto:secretariat@eurecom.fr)

*EURECOM is a French graduate school and a research center in digital sciences based in the international science park of Sophia Antipolis, which brings together renowned universities such as Télécom ParisTech, Aalto University (Helsinki), Politecnico di Torino, Technische Universität München (TUM), Norwegian University of Science and Technology (NTNU), Chalmers University (Sweden) and Czech Technical University in Prague (CTU). The Principality of Monaco is a new institutional member. The Institut Mines-Télécom is EURECOM's founding member.*

*EURECOM benefits from a strong interaction with the industry through its specific administrative structure: Economic Interest Group (kind of consortium), which brings together international companies such as: Orange, BMW Group Research & Technology, Symantec, Monaco Telecom, SAP, IABG.*

*EURECOM deploys its expertise around three major fields: Digital Security, Data Science and Communication Systems. EURECOM is particularly active in research in its areas of excellence while also training a large number of doctoral candidates. Its contractual research is recognized across Europe and contributes largely to its budget.*

*Thanks to its strong ties set up with the industry, EURECOM was awarded the "Institut Carnot" label jointly with the Institut Telecom right from 2006. The Carnot Label was designed to develop and professionalize cooperative research. It encourages the realization of research projects in public research centers that work together with socioeconomic actors, especially companies.*