

## Internship subject proposition

A simplified view of the Web of Data consists in seeing it as a collection of RDF bases that can be consulted on the Web, each containing its own specificities. The merging of these different bases into one uses mechanisms that allow naming and managing the bases expressed as graphs and expressing relations between bases. This requires the obligation to express meta-information on the graph bases in order to keep track of the origin of the information in the bases. This is what led the W3C community to introduce the concept of graphs named <https://www.w3.org/2009/07/NamedGraph.html> . In this project we propose to study the concept of named graphs in order to facilitate the management, the visualization and the evolution of the global graph and its parts in order, to a certain extent, to optimize the connections.

It is within this framework that a research project named KYOUKAN was defined with Pr T Amagasa of the University of Tsukuba. This project highlights the strategic partnership between the University of Tsukuba and the UGA. In a publication proposed in the International [Journal of Web Information Systems Volume 17 Issue 6](#), Prof. Amagasa and his team proposed a schema structure that allows users to interactively explore the relationships between entities in knowledge bases (KBs). This new relationship between entities allows an optimization of the connections between entities through a clustering algorithm CANDECOMP /PARAFAC (CP), which produces clusters. The application chosen to test this algorithm and to understand how to better structure the graphs is the Trajectories application <https://trajectories.univ-grenoble-alpes.fr/> whose graph is available through the Trajectories software Catalogue: <http://lig-tdcge.imag.fr/steamer/trajectories/public/> .

During the course of the TER the trainee will have to understand :

1. Understand the RDF data structure and the CANDECOMP/PARAFAC (CP) algorithm with the publication.
  - a. This step will be concluded by a short presentation in visio between UGA and UT.
  - b. Comprendre les structures de données et adapter les données de Trajectories pour concevoir un data set coherent.
2. To set up a collaborative work environment based on gitlab between our two laboratories LIG-UGA and CS Department at the University of Tsukuba
  - a. UGA engineer will organise the gitlab of the common project (Kyoukan space)
  - b. Une discussion aura lieu pour voir quelles modifications nous devons faire sur le code et le projet gitlab sera structurer en ce sens
  - c. Predicate from different sources, additional information from different predicates, add predicate in data set.
3. Code documentation
  - a. Code is written in Python and need to documented by the student
  - b. A set of elementary tests (unit test) need by done.
4. Branches structuration for the experimentations
5. Replay Dbpedia experiment from [the publication](#) (1st first branch of the gitlab)
  - a. Reproduce the original experiment
  - b. This step will be concluded by a short presentation in visio between UGA and UT.
6. Understand the RDF database resulting from the Trajectories project,

7. Understand the structure of the relationships and the structure of the namespaces for the different data producers
8. Experiments with the CANDECOMP/PARAFAC (CP) algorithm on the Trajectories graph.
  - a. This step will be concluded by a short presentation in visio between UGA and UT.
  - b. After each experiment some improvements can be done at each iteration of the research process.
  - c. Decompose the RDF knowledge graph in a set of named graphs.
  - d. One final step of these experiments could be to make some optimisation of named graphs with CP algorithm

The different tools and techniques implemented in these two projects and design a work plan to prepare Trajectories for the use of CANDECOMP/PARAFAC (CP). No major results will be requested for the TER. The complete experiments will be done in the framework of a master 2 topic. This step will be concluded by a short presentation in visio between UGA and UT.