PhD proposal : Artificial Intelligence as a pedagogical tool: Knowledge extraction and decision support for the analysis of learners' trajectories in Computer Science

In 2018, the University of New Caledonia set up an innovative research and teaching centre containing high-tech infrastructure and equipment for university training and research. This institutional dynamic accompanies and encourages a profound transformation of teaching practices in the digital age. While digital tools are nowadays a valuable asset for the transmission and evaluation of knowledge, their use in the processes of *acquisition* of this knowledge by the learner remains very limited. Indeed, these processes are part of the expertise and pedagogical know-how of each teacher whom it would be utopian to want to replace with a machine. On the other hand, the real challenge in terms of pedagogical innovation consists in designing new digital tools that place the teacher at the heart of the system and that facilitate the adaptation and individualization of *his* pedagogy, particularly in situations of teaching in large groups or even at a distance.

The thesis takes place in the field of Artificial Intelligence and will contribute to the development of fundamental methodologies for knowledge extraction and decision support applied to digital humanities for education. This will involve exploiting learners' productions in the field of computer programming (scripts). These data are by nature very complex because of their temporality/sequentiality, their potential volume (Big Data) or their polymorphic nature (multiple representations). This thesis will seek to remove several major scientific obstacles to their analysis with a view to their exploitation for the purpose of assisting pedagogical "piloting".

A first work will necessarily focus on the representations of learners' productions by the construction or even the automatic extraction of descriptors from different levels of analysis (syntactic, semantic, structural, etc.). In particular, deep-learning techniques may be considered to build a semantic representation space (embedding) on these data. The second step will be to define and exploit learners' "trajectories" (production sequences) by reconsidering certain well-chosen techniques for analyzing classical or even symbolic data in a particularly multi-represented context and in a user-centered process. Particular attention will be paid to methodologies for clustering analysis, detection of atypical trajectories (outliers), projection for visualization (dimensionality reduction) or characterization of trajectories oriented towards "skills and know-how" (supervised learning).

Keywords : Artificial Intelligence, Data Mining, Deep Learning, Big Data, Digital Humanities for Education.

Profile : Master level students in Computer Sciences
Desired skills:
- Data analysis
- Machine Learning: supervised and unsupervised approaches
- Data Mining for complex data : temporal, multiview, etc.
- Interactive learning
- Interest in educational sciences

<u>**Laboratory**</u> : Institut des Sciences Exactes et Appliquées (ISEA) - Nouméa, New Caledonia

<u>**Contacts**</u> :   Guillaume Cleuziou (guillaume.cleuziou@unc.nc)
         Frédéric Flouvat (frederic.flouvat@unc.nc)