Title: Implementation of a Question Generation System to promote Critical Readings in STEM Education

Proposers: Itziar Aldabe, Oier Lopez de Lacalle, Montse Maritxalar (IXA http://ixa.eus)

Contact: Itziar Aldabe (itziar.aldabe@ehu.eus)

Description: Setting the foundations for the generation of pedagogically relevant questions and for the representation of the domain specific knowledge. Building a Question Generation (QG) system that generates deep question relying on the most advanced NLP methods. Perform a series of evaluations to measure the appropriateness of the approach.

Goals
The main goal of the project is to generate content-dependent deep questions to support critical reading in the Biology Domain. First, a domain specific knowledge-base will be generated by automatically extending existing resources. After that, a QG system will be developed that, given a text and relying on distant supervision, will identify the relevant pieces (e.g. formalizing it as a graph matching problem) and link them to the knowledge-base. Some templates will be manually described to make the QG system able to generate questions about the information from the text.

The student will focus on identifying the relevant pieces of information occurring in text that would make able to generate the questions. As this step is critical for the main goal, latest information extraction techniques need to be developed, this includes deep learning based models. More specifically, the student has to design and implement a Deep Neural Relation Extraction model, in which architectures like convolutional network, recurrent neural modules (e.g. LSTM) and attention mechanism will be combined to make successful the textual inference.

The student will apply deep learning in order to build a Deep Learning model able to extract relations between entities occurring in a document. The key objectives are the following:
Analysis of the state of the art techniques for deep learning based information
extraction systems.
Design of a deep learning architecture that finds relations between entities.
Implementation and evaluation of the model.

Requirements
English. Machine learning. Good programming skills, basic math skills.

Although it is not a requirement, taking the course “Seminar on language technologies. Deep
Learning” (see below) will allow the student to accomplish more ambitious goals. Contact us for
further details.

The dissertation can be written in Basque, English or Spanish.

Framework
Python, Tensorflow (but this is not a hard constraint, the student may use other framework like
Pytorch)

Tasks and plan
Dec-Jan: Study literature
Feb: Attend course “Seminar on language technologies. Deep Learning” (see below)
Mar-May: Development and experiments
June: Write down and presentation

References:
Ming, L., Calvo, R., &Rus, V. (2014) Automatic Generation and Ranking of Questions for
Dialogue &Discourse, 3(2), 75 – 99
Y. Peng, J. Qi, and Y. Yuan. Modality-specific cross-modal similarity measurement with

Seminar on Language Technologies. Deep learning. (LAP18)
Deep Learning neural network models have been successfully applied to natural language
processing. These models are able to infer a continuous representation for words and sentences,
instead of using hand-engineered features as in other machine learning approaches. The seminar
will introduce the main deep learning models used in natural language processing, allowing the
students to gain hands-on understanding and implementation of them in Tensorflow.

Topics
Introduction to machine learning and NLP with Tensorflow
Deep learning
Word embeddings
Language modeling and recurrent neural networks
Convolutional neural networks
Attention mechanisms

Prerequisite. Basic programming experience, a university-level course in computer science and experience in Python. Basic math skills (algebra or pre-calculus) are also needed.