Master Thesis title / MAL izenburua
Deep Learning Question Answering systems exploration

Proposer(s) / Proposatzailea(k): Aitor García Pablos, Vicomtech

Contact / Kontaktua: agarciap@vicomtech.org

Description / Deskribapena

Deep Learning models have dramatically boosted the capabilities and performance of Question Answering Systems. Almost each month a new model or system is proposed, beating the previous state-of-the-art in some way. There are plenty of different Question Answering Systems: extractive (which select pieces of existing text as the answer), extractive but disjoint (capable of selecting non-contiguous pieces of text as the answer), abstractive/generative (which uses natural language generation to create its own answer), open domain (which are general an capable of answer from a large volume of knowledge), knowledge-based (which make use of semi-structured or fully-structured information, such as knowledge graphs to craft their answers). They also have different capabilities like answering yes/no questions, deciding when there is not suitable answer for a given question, or making some simple reasoning or arithmetic operations based of text, etc.

The project consists of getting familiar with the area of Question-Answering, exploring the State-of-the-Art of existing systems and model architectures, make an inventory of existing Question Answering datasets (e.g. SQUAD dataset, but there are many more), sorting them by difficulty, question types, language, domain, etc., and implement/evaluate some of them to make a comparison and, if possible, propose ideas or improvements for future work.

Goals / Helburuak

The overall objective is to get familiar with how the modern, neural-based, Question-Answering models work:
- which is the current State-of-the-Art
- learn to implement/train/evaluate a Deep Learning based Question-Answering model
- learn how to use the trained models to showcase them (for example, making a Q/A demonstrator)
- all this knowledge and work will be documented in a report.

Requirements / Betebeharrak
- Basic knowledge of Natural Language Processing and Python programming language
- Interest on Deep Learning, Neural Language Models and Question-Answering systems

Tasks and plan / Atazak eta plana
1. Make and exploratory analysis of the State-of-the-Art (become familiar with the are, the technologies)
    a. Elaborate a report to guide the rest of the steps
2. Find or create a manually labelled Question-Answering dataset for evaluation purposes
    a. In Spanish and/or Basque, and general or specific domain (to be decided)
3. Pick and implement a question-answering system based on State-of-the-Art architectures and evaluate it
    a. The objective is to train and integrate one or more Question-Answering models into a working system that can be evaluated, compared, and showcased