

Deep Learning and Natural Language Processing for Expert Finding from Web Search Results

PhD position

January 2020

Description

The EXPERTRACK project (*Deep Learning and Natural Language Processing for Expert Tracking from Web Search Results*) funded by the Empirical Foundations of Linguistics ([EFL](#)) laboratory of excellence and the ECOS Nord fund is looking for a PhD candidate to study and develop machine learning and natural language processing methods for web search query results in order to find experts from a certain scientific discipline, a country of origin or a person's name by mining the web.

Most of current expert finding approaches work on highly organized data sources, like scientific publications or the semantic web [Flores et al., 2012]. We propose an original approach based on information extraction from a noisier data source: web search query results. Deep Reinforcement Learning (DRL) has shown promising results on information extraction from web search results [Narasimhan et al., 2016]. However, the association of web search results to experts, and the extraction of meaningful semantic information (affiliation institution, city, year, scientific discipline) from the resulting dataset is challenging both from a scientific and a technological perspective. As web search querying is expensive from an algorithmic point of view, reinforcement learning is intended to optimize the query formulation process and result exploration policy with neural network based algorithms such as Deep-Q network [Mnih et al., 2015] [Alizadeh et al., 2020].

The goal of the PhD will be to provide an end to end deep learning method for expert finding, while integrating state of the art query generation and information extraction strategies to the method. This approach will be evaluated on an existing dataset of international experts coming from developing countries and having obtained a scientific degree abroad. An ongoing expert finding system is currently being developed, and the PhD candidate is expected to participate in this development.

While most of the PhD will be carried out at the Northern Paris Computer Science Lab ([LIPN](#)) at [Sorbonne Paris Nord University](#) (Villetaneuse campus), the PhD student will perform yearly scientific of one or two months at UNAM, funded by an ongoing ECOS Nord project in association with Mexico.

Required Skills

- Sound background in computer science
- Familiar with deep learning and information extraction and Natural Language Processing
- Excellent programming skills (Python, and some Pythorch or Tensorflow experience)

- Autonomy, creativity, curiosity
- Master's degree in Computer Science or Natural Language Processing
- Full proficiency (spoken, written) in English
- Any French and Spanish skills would be appreciated

Contact

Send motivational cover letter and CV before April 15th to:

Jorge Garcia Flores / jgflores@lipn.univ-paris13.fr
Pegah Alizadeh / pegah.alizadeh@devinci.fr
Iván Vladimir Meza / ivanvladimir@turing.iimas.unam.mx

References

- Alizadeh, P., Flores, J. J. G., and Meza Ruiz, I. V. (2020). Apprentissage par renforcement pour la recherche d'experts sur le web. In *Extraction et Gestion des Connaissances (EGC)*, pages 261–268.
- Flores, J. J. G., Zweigenbaum, P., Yue, Z., and Turner, W. (2012). Tracking researcher mobility on the web using snippet semantic analysis. In Isahara, H. and Kanzaki, K., editors, *Advances in Natural Language Processing*, pages 180–191, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., and Riedmiller, M. A. (2015). Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602.
- Narasimhan, K., Yala, A., and Barzilay, R. (2016). Improving information extraction by acquiring external evidence with reinforcement learning. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2355–2365.