Master's Thesis proposal

General Information

Master's Thesis Title: **Learning to Match**

Orientation:
☐ professional
☒ research

M.Sc. Th. Advisor's Dept. & University: LSI, UPC

M.Sc. Th. Advisor: Xavier Carreras, Ariadna Quattoni

M.Sc. Th. Advisor e-mail: carreras@lsi.upc.edu, aquattoni@lsi.upc.edu

Observations: This project has possible funding for a full PhD.

Student's Name:
(if already known)

M.Sc. Thesis Description

Main issues / Brief Description:

This project is about learning to match people that offer or request things online, using textual descriptions of their offers or needs.

Detailed Description:

Imagine an online website of classified advertisements, where people can offer or request things. For example, in the job domain employers offer jobs requiring certain skills, while people look for jobs for which they are qualified. There are many other applications that fall into this framework of people looking for something (who we call "seekers") and people offering something (who we call "providers"). To make this application appealing to users, the system should allow them to describe their offers or requests in free text, using everyday's natural language. At the same time, users should be able to find matches quickly.

The goal of this thesis project is to develop automatic methods to match seekers with providers. We will use Natural Language Processing (NLP) techniques to automatically extract relevant information from advertisements in free text of seekers and providers. We will use Machine Learning methods in order to evaluate the compatibility between seekers and providers, relying on the information extracted with NLP tools.

There are two main challenges in this thesis. The first is to induce the semantics behind advertisements, that is, what concepts and relations are mentioned in advertisements of a particular domain. Our aim is to discover such concepts and relations automatically, from large collections of data. The second challenge is to exploit the semantics of advertisements in order to match seekers and providers, and to ponderate which features are more important for a seeker when choosing a provider, and vice-versa. Again, we aim at learning these matching functions from collections of data.

The project will combine Natural Language Processing and Machine Learning methods. We will exploit methods for learning ranking functions, together with methods for inducing the semantics of a target application.

References

[A10] Shivani Agarwal. Ranking Methods in Machine Learning. A Tutorial Introduction. SIAM conference on Data Mining, 2010.

[CGCR10] James Clarke, Dan Goldwasser, Ming-Wei Chang and Dan Roth. 2010. Driving Semantic Parsing from the World's Response. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL-2010), pages xx–yy. Uppsala, Sweden.

[DS09] Dipanjan Das and Noah A. Smith. Paraphrase identification as probabilistic quasi-synchronous recognition. In Proceedings of ACL, 2009

[BCZB09] S.R.K. Branavan, Harr Chen, Luke Zettlemoyer and Regina Barzilay. Reinforcement Learning for Mapping Instructions to Actions. Proceedings of ACL, 2009.

[CCK08] Xavier Carreras, Michael Collins and Terry Koo. TAG, Dynamic Programming, and the Perceptron for Efficient, Feature-rich Parsing. In Proceedings of CoNLL, 2008.