Master in Artificial Intelligence (UPC-URV-UB)

Master's Thesis Proposal[1]

General Information

Title:	Intelligent Bibliography Discovery Tool
Expiry Date:	
Modality:	☒ technological project ☒ research work
Advisor/s:	Javier Béjar
Advisor's Dept. & Univ.:	LSI,UPC
Advisor/s e-mail:	bejar@lsi.upc.edu
Observations:	
Student's Name: (if already known)	

M.Sc. Thesis Description

Main issues / Brief Description [Mandatory]:

The goal is to develop a tool for intelligent research bibliography discovery. This projects involves recommendation systems, text analysis and ontology building.

1 **Each M.Sc. Th. Proposal should be in a separate file, named as follows: "MSc-Th-Proposal-2-or-3-title-first-words-Advisor/s-AcademicYear.pdf".**
For Example: "MSc-Th-Proposal-Syntactic-and-Semantic-LluisMarquez&JesusGimenez-1011.pdf"
The proposal could be elaborated with any text processor (Word, Openoffice, etc.), but **the file electronically delivered** to LSI Dept. Secretary (merce@lsi.upc.edu) **MUST BE a single PDF file**

One problem in research is to find papers related to an specific subject and to obtain a view of the state of the art. This can be done manually by intensive search in journals and proceeding and recollecting all their references.

There are some systems able to recommend bibliography by comparing the lists of bibliographic references other people has collected and the different words that have been used tag the references. The main problem is that this systems can not be focused in an specific area or set of keywords.

The idea of this project is to use different unstructured sources to help find bibliographic references using an initial set of references and a set of keywords. The goal is to be able to recommend new references by discovering them using specific search engines, citations and crossreference citations, keywords obtained from the references and authors connections.

The tasks to develop are the following:

* Study different mechanisms of recommender systems
* Study methodologies of keyword extraction from text
* Study of methodologies for graph subset discovery
* Study of methodologies for shallow ontologies building

References [Mandatory]:

Pazzani, M.J., Billsus, D.: Content-Based Recommendation Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)

Schafer, J.B., Frankowski, D., Herlocker, J., Sen, S.: Collaborative Filtering Recommender Systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) The Adaptive Web: Methods and Strategies of Web Personalization. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007).

Rajaraman, A, Ullman, Mining Massive Datasets http://infolab.stanford.edu/~ullman/mmds.html

T. Hastie and R. Tibshirani and J. Friedman The Elements of Statistical Learning: Data Mining, Inference and Prediction, Springer-Verlag, 2001.

CiteULike http://www.citeulike.org

Minimal Requirements & Previous Knowledge [Optional]:

Other comments [Optional]:

We recommend applicants to have minimal knowledge of machine learning and data mining although it is not a sine quan non condition.

Location and Date: Barcelona, March 8th 2011

To the Academic Commission of the Master in Artificial Intelligence (CAIMIA)