



APPLICATIONS OF SOCIAL DATA MINING TO LEARNING ANALYTICS

Kais Dai, Supervised by Dr. Ana Fernández Vilas and Dr. Rebeca P. Díaz Redondo

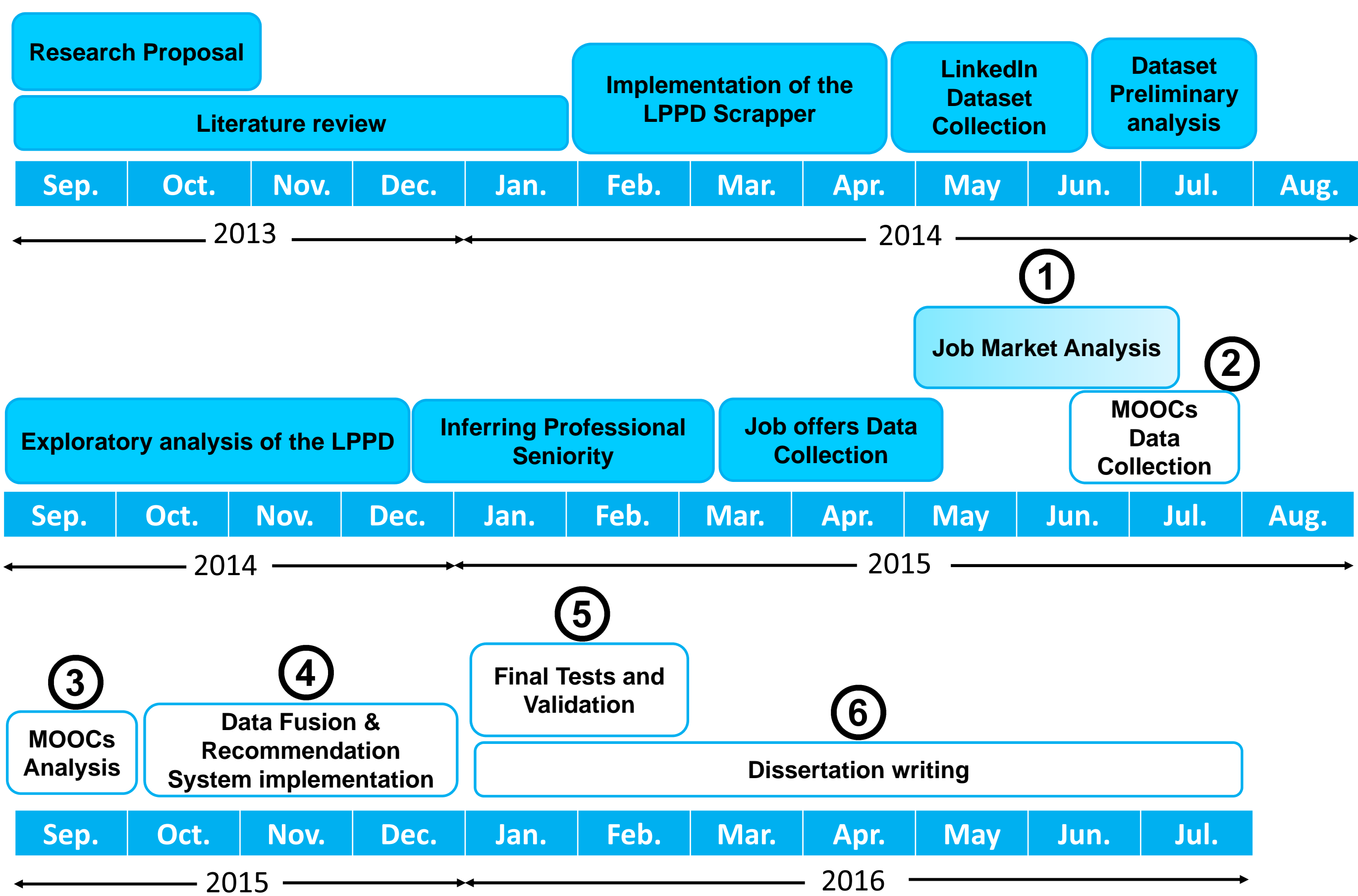
Affiliation: AtlantTIC Research Centre, Department of Telematics Engineering (University of Vigo)



Motivation of the work

- The emergence of *Massive Open Online Courses (MOOCs)* as a new learning alternative for lifelong learners highlights the need of a more personalized learning experience especially regarding the offers diversity and abundance.
- Furthermore, the dynamics of the job market inherits new interactions' models among the different stakeholders (recruitment companies, job seekers, etc.). So, using Online Social Networks (OSNs) as a playground for these interactions became more than an evidence. In this sense, collaborative data sources such as OSNs (LinkedIn, Twitter, etc.) must be considered in the *Learning Analytics* context in order to get more featured view and address these challenges.
- Our work is motivated by the need of providing the learning stakeholders the right information in the right moment by optimizing the learning environment and furnishing decision support at early steps of the learning process.

Research Plan



Next Year Planning

- Job Ads analysis using domain knowledge based classification and cross-classification for validating our approach (in progress).
- Collection of data about online learning resources in general (Courses, lifelong learners, Q&A, etc.) and from MOOCs in particular (Coursera, edX, Udemy, etc.).
- Analysis and characterisation of the MOOCs offers: MOOCs' classification (expertise level requirements, duration, etc.) but also the use of both text mining and clustering techniques to discover the trending MOOCs' topics.
- Data fusion by merging data and results from the different sources: LinkedIn profiles characterization, Twitter and LinkedIn job offers and E-learning resources' data (LMS, MOOCs). We will also establish the recommendation strategy: Which MOOC and When it would be recommended?
- Implementation and validation of the MOOCs recommender system.
- Dissertation writing and Thesis Defense.

Within the next academic year, we plan to attend two international conferences/workshops (submission of a conference paper by the end of July 2015) and submit in at least two international journals (indexed in JCR).

Thesis Objectives

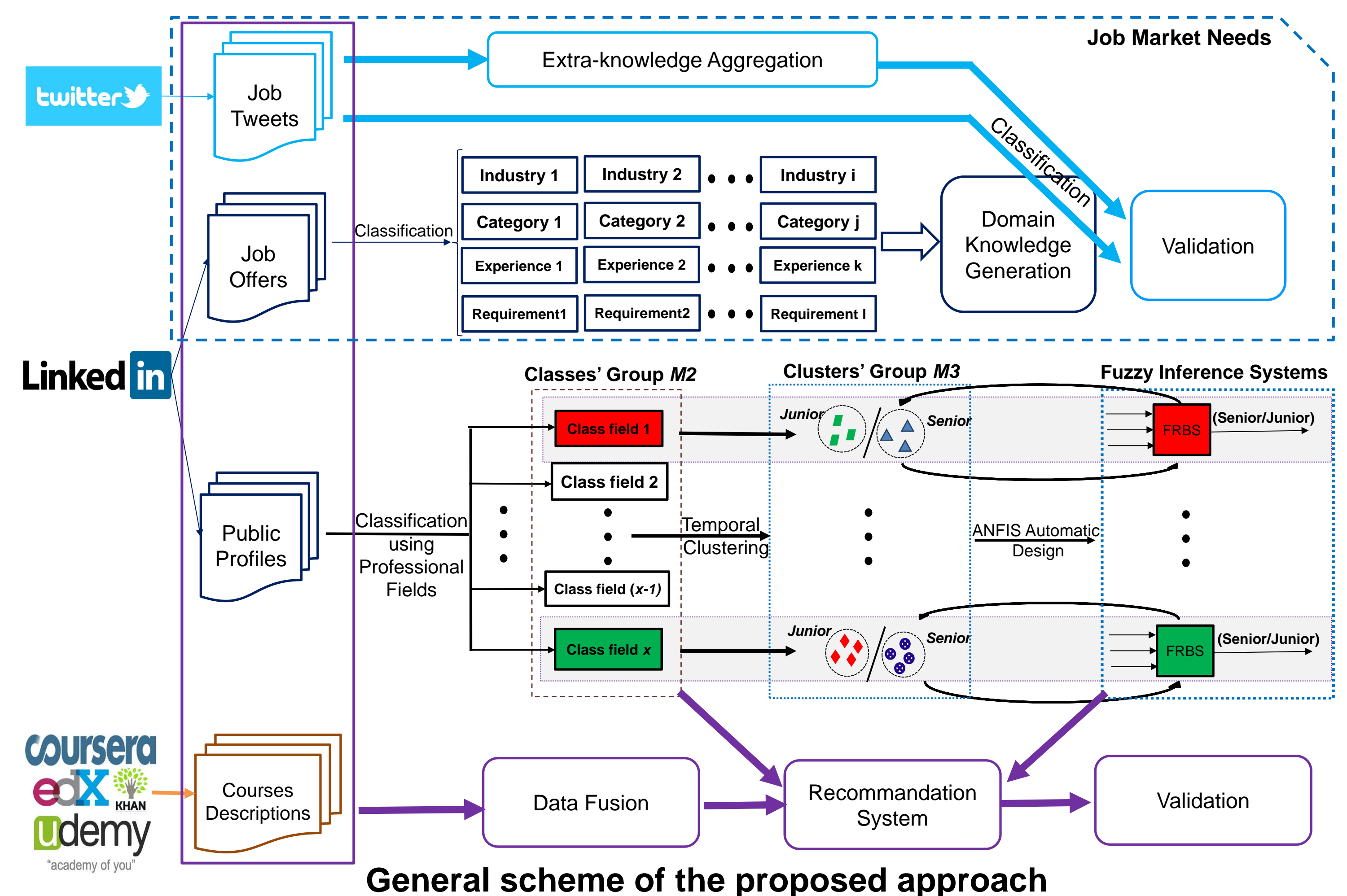
The main objective of the thesis is to take advantage of collaborative data sources (OSNs) in order to provide a personalized recommendation's system for lifelong learners (MOOCs users) and enhance their career path according to an objective driven approach guided by the job market needs.

In this context, we are applying social mining techniques to merge and process a big amount of data (regarding previous learners' records and curriculums but also their social media fingerprints) to discover and characterize OSNs' users, to analyze and detect the job market needs and MOOCs offers. So, the idea is to merge all these data sources and be able to recommend MOOCs and warn both learners and online courses providers (E-learning platforms, MOOCs) to react properly.

Finally, our objectives can be summarized in answering the following questions:

- How to characterize and discover group of profiles in OSNs' (educational and professional background)?
- How to detect the job market needs from an OSN perspective?
- How to characterize web 2.0 learning materials?
- How would be a personalized MOOCs' recommendation system? (Which MOOC and when it would be appropriate to perform these recommendations?)

Results & Discussions



- Scraping over 5 million public profiles of LinkedIn and applying Clustering techniques for their characterization (results were published in [1])
- Extended Analysis of the LinkedIn Public Profiles Dataset (LPPD):
 - ✓ A novel classification method according to the professional fields was proposed and performed over the LPPD
 - ✓ Clustering the obtained profiles' groups using temporal parameters
 - ✓ A novel professional seniority inference system was designed and tested over our dataset
- Job offers data were collected (LinkedIn and Twitter) and pre-processed using text mining techniques.

References

- Kais Dai, Celia González Nespereira, Ana Fernández Vilas and Rebeca P. Díaz Redondo: "Scraping and Clustering Techniques for the Characterization of LinkedIn Profiles", International Conference on Information Technology Convergence and Services (ITCS'2015), Zurich (Switzerland), 2-3 January 2015.
- Kais Dai, Ana Fernández Vilas and Rebeca P. Díaz Redondo: "Inferring professional Seniority using Public LinkedIn Profiles", PLOS ONE journal (Submitted in May 2015).
- Mathieu Bastian, Matthew Hayes, William Vaughan, Sam Shah, Peter Skomoroch, Hyungjin Kim, Sal Uryasev, and Christopher Lloyd. 2014. "LinkedIn skills: large-scale topic extraction and inference", In Proceedings of the 8th ACM Conference on Recommender systems (RecSys '14). ACM, New York, NY, USA, 1-8.
- S. B. Aher and L. M. R. J. Lobo: "Combination of machine learning algorithms for recommendation of courses in E-Learning System based on historical data", Knowledge-Based Syst., vol. 51, pp. 1-14, Oct. 2013.

