IFN712 Research Project proposal

|  |  |
| --- | --- |
| Cluster name | Data Science/ computer Science 2 |
| Relevant MIT Major (e.g. Data Science, Computer Science,Software Development, User Experience, Cyber security and networks, Business Process Management, Business Analysis, Executive IT; Library and Information Practice) | Data Science  Computer Science |
| Supervisors/Academics | Aspro Yue Xu |
| Contact emails | yue.xu@qut.edu.au |
| Tutor’s name/email | Dakshi Kapugama Geeganage - kapugama@qut.edu.au |
| Project title | Filtering related news stories based on user’s interest by applying topic modelling techniques |
| Brief description of your research topic and aims (100-200 words) | A large volume of text content is available over the word wide web. Social media, digitized libraries and news portals contain an enormous amount of text contents, and it is challenging to filter the required contents based on the user’s interest. Similarly, a user may have several topic interests like politics, economy, sports. It is important to semantically understand the user’s multiple topic interests and retrieve the relevant news. This project aims to retrieve the related news based on the users’ topic interests based on text mining techniques.  The project opens an opportunity for the students to acquire knowledge and skills related to text mining such as topic modelling, semantic extraction by reviewing the state of the are systems. Finally, students need to present an effective algorithm to categorise and retrieve the relevant news stories related to the topics of user’s interest based on his/her previous news articles. This approach can be applied in social media text mining and will be able to retrieve the relevant tweets related to the topics of user’s interest based on his/her tweets. For the experiment purposes RCV1 dataset need to be used. |
| Research problems and gaps | Users have multiple topic interests and it is challenging to filter the contents according to multiple topic interests. |
| Specific research questions (2-4) | How to extract the user interest based on previous news?  How to apply a semantic based approach to filter user’s interest?  How to generate a model to filter the user's topics of interest? |
| Key references (4 to 8) | [1] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation”. Journal of Machine Learning Research. 2003.  [2]D. Kapugama Geeganage, Y. Xu and Y. Li, "Semantic-based topic representation using frequent semantic patterns", Knowledge-Based Systems, vol. 216, p. 106808, 2021. Available: 10.1016/j.knosys.2021.106808.  [3] Y. Gao, Y. Xu and Y. Li, "Pattern-based Topics for Document Modelling in Information Filtering", IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 6, pp. 1629-1642, 2015. Available: 10.1109/tkde.2014.2384497. |
| Expected new knowledge and outputs to generate from the research project | Topic modeling , Information Filtering |
| Required studies skills, knowledge, and speciality (if any). For example, specify programming language/skill level required, or required proficiency in data analytics. | Programming skills  Analytical skills |
| Number of preferred students | 5 |
| Industry supervisor/sponsor (if any) |  |
| Conditions of offer (such as IP ownership) |  |
| Students/emails (tutor to add in Week 1) |  |
|  |
|  |
|  |
|  |
|  |
|  |

Digit libruary

People may be interested in more than one topic

How do we filter the interest

How do we group the related topics

Filter other document related to the domains

Clustering, classification…

RCV1 dataset

Topic modelling

How do we apply pattern

Filtering tech.

Libruary for finding research paper:

Science Direct

ACM

TKDD

TKDE

Knowledgebase systems

ICDE

AAAI

WWW

WISE

KDD

CIKM

PAKDD

Scimago

Keywords:

Information filtering

Extracting tweets

Hashtag

Topic modelling:

mallet

LDA

Genism

Scikit learn