

CO7201 MSc Individual Project

Computer Science Department



**A Web-Based Recommender System using Semantic Web
and Linked Data for Suggesting Supervisors
to Postgraduate Research Students**

Student: **TRUNG THANH DANG**

CFS username: **TTD2**

DISSERTATION

Supervisor: **Dr. Monika Solanki**

Second Marker: **Prof. Thomas Erlebach**

Date of submission: 13/09/2013

Number of words: 11822

DECLARATION

All sentences or passages quoted in this report, or computer code of any form whatsoever used and/or submitted at any stages, which are taken from other people's work have been specially acknowledged by clear citation of the source, specifying author, work, date and page(s). Any part of my own written work, or software coding, which is substantially based upon other people's work, is duly accompanied by clear citation of the source, specifying author, work, date and page(s). I understand that failure to do these amounts to plagiarism and will be considered grounds for failure in this module and the degree examination as a whole.

Name: Trung Thanh Dang

Signed:

Date:

Abstract

One obstacle of postgraduate research students is finding and selecting a suitable supervisor to their research areas. A content-based recommender system promises to help these students reducing time and difficulties of the tasks. However, traditional approaches of content-based recommender systems have the over-specification and new user issues. In this project, an ontology-based recommender system suggesting supervisor for postgraduate research students is developed; it makes use the Semantic Web and Linked-Data technologies for overcoming such limitations. This recommender system consists of three main components including Automated Data Curation Tool, Profile Learner, and Recommendation Engine components. The components allow the system to retrieve data from Linked-Data sources and stores them as the background data, input student preferences and match them to the background data to suggest recommendation lists to the students. Along with the components, the Postgraduate Education ontology is also designed to represent all relevant concepts and relationships relating to the decision making process of the system. Besides, a triple store is employed in storing the ontology data in order to make use the reasoning feature as well as improving stability and performance of the system. The recommendation algorithms are designed to exploit semantic characteristics of the Postgraduate Education ontology. The accuracy of the recommendation algorithms is evaluated with a data set mainly coming from Linked Data sources, and the evaluation results shows that it highly satisfies the requirements.

Acknowledgements

I would like to special thanks to Dr. Monika Solanki for valuable guidance and advice as well as lots of time on discussions and reviewing my work. I am very grateful to Prof. Thomas Elerbach for his critical questions about my methodology and guidance for the shortest path algorithm.

Contents

Abstract.....	3
Acknowledgements.....	4
Contents	5
List of Figures.....	8
List of Tables	9
List of Acronyms / Abbreviations.....	10
Chapter 1. Introduction.....	11
1.1 Motivations.....	11
1.2 Aims and objectives	11
1.3 Contributions.....	12
Chapter 2. Background and Related Work.....	14
2.1 Supervisor Selection in Postgraduate Research Education.....	14
2.1.1 Education and Research in Postgraduate Education.....	14
2.1.2 Supervisor Selection	14
2.2 Semantic Web and Linked-Data	16
2.2.1 Semantic Web.....	16
2.2.2 Ontology and Ontology Engineering.....	18
2.2.3 Linked-Data	18
2.3 Content-Based Recommender Systems	20
2.3.1 Overview.....	20
2.3.2 System Architecture.....	20
2.3.3 Item and User Profile Representation.....	21
2.3.4 Recommendation Algorithms.....	22
2.3.5 Other Approaches to Improve Content-based Recommender Systems.....	22
Chapter 3. Ontology Development and Linked-Data Deployment.....	24
3.1 Competency Questions.....	24
3.2 Supervisor Suggestion Criteria	24
3.3 Postgraduate Education Ontology.....	25

3.3.1	Reuse Existing Ontologies	25
3.3.2	Concepts and properties	26
3.4	Deploying Ontology Data	28
3.5	Important SPARQL Queries	29
Chapter 4.	System Architecture	32
4.1	Overview	32
4.1.1	System Architecture	33
4.1.2	Data Access Package	34
4.2	Profile Learner Component	35
4.2.1	Use Cases	35
4.2.2	Component Structure	35
4.2.3	Main Functions	37
4.3	Recommendation Engine component.....	38
4.3.1	Use Cases	38
4.3.2	Component Structure	39
4.3.3	Recommendation Process	40
4.4	Automated Data Curation Tool	40
4.4.1	Overview	40
4.4.2	Data Curation Process	41
4.4.3	Structure of the Tool	43
Chapter 5.	Recommendation Algorithms	44
5.1	Reduction-Based Approach Algorithm	44
5.2	Jaccard Recommendation Algorithm	45
5.3	Ranked Recommendation Algorithm	47
5.4	Graph-Based Recommendation Algorithm	49
5.4.1	Graph-Based Approach	49
5.4.2	Algorithm	51
Chapter 6.	Evaluation	54
6.1	Evaluation Process	54

6.2	Datasets	55
6.2.1	Lecturers Data Set.....	55
6.2.2	Student Profiles Data Set	55
6.3	Results Analysis	56
Chapter 7.	Conclusions and Future Work	58
7.1	Conclusions	58
7.1.1	Results.....	58
7.1.2	Limitations	59
7.2	Future Work	60
References	61
Appendixes	63
Appendix 1.	List of Linked-Data Sources Used by the Supervisor Recommender System	63
Appendix 2.	Postgraduation Education Ontology.....	64
Appendix 3.	Snapshots of the Supervisor Recommender System’s User Interface	78

List of Figures

Figure 2-1: A layer approach to the Semantic Web [12].....	17
Figure 2-2: Linked Data Publishing Options and Workflows [16].....	19
Figure 2-3: High level architecture of a Content-based Recommender [1].....	21
Figure 3-1: The ontological model for the Supervisor Recommendation System.....	27
Figure 4-1: Connectivity model of the Supervisor Recommender System	32
Figure 4-2: Architecture of the Supervisor Recommender System	33
Figure 4-3: Packages of the Supervisor Recommender System	34
Figure 4-4: Classes in the co7201.data package	35
Figure 4-5: Use cases of the Profile Learner component.....	36
Figure 4-6: Classes used by the Profile Learner component	36
Figure 4-7: Use cases of the Recommendation Engine component	38
Figure 4-8: Classes used by the Recommendation Engine component	39
Figure 4-9: Use cases of the Automated Data Curation Tool.....	41
Figure 4-10: The data curation process.....	42
Figure 4-11: Classes used by the Automated Data Curation Tool.....	43
Figure 6-1: Precision at N of GraphBased, Ranked, and Jaccard recommendation algorithms	57
Figure 6-2: Recall at N of GraphBased, Ranked, and Jaccard recommendation algorithms	57

List of Tables

Table 1: Importance Weight of Factors [9].....	15
Table 2: Criteria for selection of a thesis supervisor [10].....	16
Table 3: Competency questions	24
Table 4: Criteria for recommending supervisors	25
Table 5: List of name spaces and prefixes used in the Lecturer and Student Profile datasets.....	29

List of Acronyms / Abbreviations

ACM	Association for Computing Machinery
API	Application Programing Interface
FOAF	Friend of a friend
HTML	HyperText Markup Language
OWL	Web Ontology Language
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
RS	Recommender system
SKOS	Simple Knowledge Organization System
SPARQL	SPARQL Protocol and RDF Query Language
SW	Semantic Web
URI	Uniform Resource Identifier
XML	eXtensible Markup Language