

Ontology Based Video Information Mining Tool

A.G. Gunawardena

08/10023

Dissertation submitted to the Faculty of Information Technology,
University of Moratuwa, Sri Lanka for the partial fulfilment of the
requirements of the Msc in
Information Technology

University of Moratuwa



105303

December 2012

004 "12"

004 (043)

105303

105303

Declaration

I declare that this thesis is my own work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education. Information derived from the published or unpublished work of others has been acknowledged in the text and a list of references is given.

Name of Student

Asiri Gunawardena

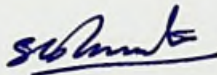

Signature of Student

Date: 07-05-2013

Supervised by

Name of Supervisor

Mr. Saminda Premaratne


Signature of Supervisor

Date: 13/05/2013

Acknowledgement

Foremost, I would like to express my sincere gratitude to my advisor Mr Saminda Premaratne for the continuous support of my Msc study and research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Msc study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. P. Wimalaratne, Dr. L. Ranathunga for their encouragement, insightful comments, and hard questions.

Last but not the least, I would like to thank my parents my brother and my wife for supporting me spiritually throughout my life.

Abstract

Finding information on World Wide Web or web is the most popular and interactive searching medium to retrieve information these days. Web is easy to access and there are lots of information available on the web. With growth of amount of the content or information, web has become increasingly difficult to find the data on the web quickly. The easiest way of searching on the web is search engines. Search engines returns thousands of results for small search keyword. It is difficult to filter out the required information from these search results.

Web content mining is one of the most popular and effective way of filtering data. Users will be able to retrieve their required information through this kind of filtering mechanism. Web content mining is specific to particular area. The filtered data can be saved locally in order to user later. This information can be stored in a database system to use as required. Integrating relational database with ontology is an active research area these days. Ontology will be a core part of the next generation web. Current web (web 2.0) mostly relies on keyword based search. The performance of the system depends on the matching the keyword with available data.

This study has focused on web content mining technologies and integrating ontology with relational database system to manipulate filtered data. The domain of the study is movies and related information. Movie related information is filtered using a web crawler and this information is saved in a database system after further cleaning. Web crawler is configured to get the movie related information from the web. Ontology is designed based on the movie related information. After designing the database and the ontology, these two layers should be mapped. This part is done using a data wrapper. This wrapper links ontology instances with database system.

After mapping is done there should be a proper way to retrieve the data from ontology. SPARQL is used to query the ontology and SPARQL will get the data from the database through ontology mapping. Finally user interface is designed to interact with the user. User can search through two interfaces called general search advanced search. Using general search user will be able to search a keyword and through advanced search user can search using multiple parameters. Users will be able to search based on their actual searching criteria's rather that searching a keyword.

Contents

	Page
Chapter 1 - Introduction.....	1
1.1. World Wide Web	1
1.2. Aim and Objectives.....	3
1.3. Solution in Brief.....	4
1.4. Structure of the thesis.....	5
Chapter 2 - Literature Review.....	6
2.1. Introduction	6
2.2. Background and Motivation.....	6
2.3. Existing Models.....	7
2.4. Internet Movie Database or IMDB.....	7
2.5. Drawbacks in Existing Models	8
2.6. Proposed solution	8
2.7. Summary	9
Chapter 3 - Technology Adapted	10
3.1. Introduction	10
3.2. Main Steps of the proposed solution.....	10
3.3. What is Web Crawler?	10
3.4. Pre-processing	11
3.5. DataMining.....	Error! Bookmark not defined.
3.6. What is Ontology?.....	12
3.7. Movie Ontology in Tabular Way	13
3.8. Ontology Design	13
3.9. Ontology Mapping	13
3.10. Information retrieval using SPARQL.....	14
3.11. Tools and technologies	14
3.12. Summary.....	14
Chapter 4 - Approach.....	15
4.1. Introduction	15
4.2. Technologies Adapted.....	15
4.3. Web Crawler	15
4.4. Data Warehouse	16

4.5.	Ontology.....	16
4.6.	Users.....	16
4.7.	Input	16
4.8.	Output.....	17
4.9.	Processes	17
4.10.	Advantages of Proposed Solution.....	17
4.11.	Summary.....	18
Chapter 5 - Analysis and Design		19
5.1.	Introduction	19
5.2.	System Architecture	19
5.3.	Requirements Analysis.....	20
5.4.	Design of Web Crawler.....	20
5.5.	Ontology Design	21
5.6.	Ontology Mapping	24
5.7.	SPARQL interface and User interface	24
5.8.	Summary	25
Chapter 6 - Implementation.....		26
6.1.	Introduction	26
6.2.	Implementation of Web Crawler.....	26
6.3.	Implementation of the Database.....	26
6.4.	Implementation of Ontology	27
6.5.	Implementation of Ontology Mapping.....	28
6.6.	Create New Mapping	29
6.7.	SPARQL Interface and User Interface.....	30
6.8.	Summary	32
Chapter 7 - Evaluation		33
7.1.	Introduction	33
7.2.	Evaluation & Testing of the proposed system	33
7.3.	Evaluation.....	33
7.4.	Final Evaluation of General Search	35
7.5.	Final Evaluation of Advanced Search.....	36
7.6.	User Based Evaluation	36
7.7.	Summary	38

Chapter 8 - Conclusion and Further Work.....39

8.1. Introduction 39

8.2. Whether Project Succeeded and Achieved Goal?..... 39

8.3. Problems Encountered and Limitations of the project 40

8.4. Future Work of the project 40

8.5. Summary 41

List of Figures and Tables

	Page
Figure 3.1:High Level Architecture of a Standard Web Crawler	11
Figure: 3.7Sample of Movie Ontology	13
Table 3.1 Relations of a Movie.....	13
Figure 5.1: Proposed System Architecture	19
Figure 5.1: Web Crawler Process Flow	21
Table 5.1 Classes and Data/Object Properties of Ontology.....	24
Figure 5.3: User Query Execution Flow	25
Figure 5.1: Proposed Ontology Design.....	28
Figure 6.2: Ontology Mapping With Movie Database Table	30
Figure 6.3: Advanced Search.....	32
Table 7.1 Evaluation of General Search	35
Figure 7.1: Evaluation Graph of Searching Module.....	37