UNIVERSITI TEKNOLOGI MARA

DEVELOPMENT OF DURIAN ONTOLOGY FROM UNSTRUCTURED TEXT AND EXTERNAL KNOWLEDGE SOURCE

KHAIRUL NURMAZIANNA BINTI ISMAIL

Thesis submitted in fulfillment
of the requirements for the degree of
Master of Science

Faculty of Computer and Mathematical Sciences

November 2014

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any other degree or qualification.

I, hereby, acknowledge that I have been supplied with the Academic Rules and Regulations for Post Graduate, Universiti Teknologi MARA, regulating the conduct of my study and research.

Name of Student   :   Khairul Nurmazianna Binti Ismail
Student ID No.   :   2011592573
Programme   :   Master of Science in Information Technology
                          and Quantitative Sciences
Faculty   :   Faculty of Computer and Mathematical Sciences
Thesis Title   :   Development of Durian Ontology from
                          Unstructured Text and External Knowledge Source

Signature of Student :   ....
Date   :   November 2014

ABSTRACT

Semantic web aims to improve current web retrieval ability by implementing ontology. Ontology can be developed using manually and automatically ontology development. However, ontology development method to suit domain ontology constrains are lacking. 40 HTML files related to durian plantation are collected. Durian plantation is chosen as domain focused because durian has high potential in the market. From the files, 512 words are categorized as relevant and 538 words are not relevant to durian plantation. The relevant words are used in this research for ontology development process. Firstly, manual ontology development called durian ontology development or DuriO Methodology is constructed by integrating Methontology and Ontology Development 101 methodology. DuriO Methodology has five phases namely, determine scope, webpages analysis, reuse knowledge model, implementation and validation phases. Each phase has specific activities. Techniques to determine the scope of ontology is by construction competency questions. In the webpages analysis phase, it has two activities namely HTML files selection and terms enumeration. In the third phase, external knowledge sources such as knowledge model from Agropedia is embedded as the basic element of ontology development. Then, the basic element is use ontology editor for implementation phase. The last phase is to validate the ontology by using competency question and Agrovoc class alignment. From the construction of DuriO, 21 durian characteristics are produced in Webpage analysis phase. Furthermore in reuse knowledge model phase, it produces initial and domain expert verified ontology element. In ontology validation phase, it shows the validation results using competency questions and Agrovoc concept alignment. Secondly, the research progress of constructing automatic ontology development is to ensure DuriO is reliable. Automatic ontology development use ontology learning tool. It has three phases that include workflow view, corpus view and Probabilistic Ontology Model (POM) view. In the first phase, the ontology element and algorithm techniques should be selected. While in the second phase, the HTML files are uploaded for processing. The final phase, the ontology in Web Ontology Language (OWL) is display. Thirdly, ontology elements from both ontology methods are extracted by using RDF Query Language. Ontology element extraction of manual ontology development consists of three processes namely ontology element selection, query construction and ontology element extracted. Meanwhile in ontology element extraction of automatic ontology development, started with upload the OWL file and the next process is the three processes that mention before. Ontology element extraction of manual ontology development yields 116 classes, 97 subclasses, 50 properties and 126 instances. Meanwhile, ontology element extraction of automatic ontology development produce 3186 classes, 104 subclasses, 97 properties and 2569 instances. Finally this research shows that manual ontology development DuriO produces completeness domain ontology upon its development scope and less comprehensiveness domain ontology than automatic ontology.

ACKNOWLEDGEMENT

TABLE OF CONTENTS