# HERB ONTOLOGY: MATURITY-BASED ANALYSIS OF LIGHTWEIGHT ONTOLOGY ON HERB USAGES

NOOR HIDAYAH ZAKARIA

A thesis submitted in fulfilment of the

requirements for the award of the degree of

Masters in Science (Computer Science)

Faculty of Computing

Universiti Teknologi Malaysia

OKTOBER 2013

## ACKNOWLEDGEMENT

"In the name of Allah, the most Gracious and the most Merciful"

This thesis would have not been completed without the help and splendid support from many individuals and teams. Firstly, my sincere gratitude goes to the backbone of my research, my supervisors, Dr. Rohayanti Hassan and Dr. Muhamad Razib Othman for their excellent supervision, knowledge, belief, patience and interest in the work and for pushing me farther than I thought I could go. To my beloved parents, thank you for always being there and never fail to give me words of encouragement. Your endless support is making me for who I am today. To my dearest colleagues and research-mates, I am thankful for the friendship, supportive comments and ideas in reviewing each other's works and also fun times throughout last one and half year of this study. To my precious close friends and loved ones, thank you for helping me surviving all the stress and not letting me giving up. To everyone who has consistently giving support and advice directly or the other way round, including the team of Software Engineering Research Group (SERG) and GATES IT Solutions Sdn. Bhd., I refer my appreciation. My greatest thanks should also credit to all lecturers at the Faculty of Computing, Universiti Teknologi Malaysia (UTM) for their understanding and support. Last but not least, I appreciate the financial support from the GATES IT Solutions Sdn. Bhd. under GATES Scholar Foundation (GSF) and Malaysian Ministry of Higher Education under MyMaster funds.

# ABSTRACT

Ontology serves as a basis for denominating objects in a certain domain. A lightweight ontology is built using classes, instances and relationships and does not include any axiomatic definitions such as the ones found in heavyweight ontology. However, the lightweight ontology needs to be matured in order to detail the concepts and relationship that occur in a domain. To date, there is no suitable ontology design that exists in the herb domain and a design that is measured due to the heterogeneity of the ontology structures. In this study, a lightweight ontology specializing in herbal domain known as Herb Ontology (HO) is developed to explore the complete use of herbs based on their profiles. It began with the design of an informal domain modelling followed by an informal HO design that would manipulate the Unified Modelling Language (UML) notations to highlight the functionality, services and procedural strategies. In conjunction with that, eleven ontology metrics covering three maturity principles namely: reuse, extend and evolve are presented in this study. The principles are measured in both class-level and ontology-level so that different aspects of the ontology designs can be evaluated and would aid in controlling the development process of HO. Besides that, the HO design was compared with other types of ontology such as COIN, Gene Ontology and OntoCAPE. It was found that HO has Inheritance Richness = 0.99301 with the potential to be reused and a denser network ontology (Edge Node Ratio = 1.84) indicating the possibility of HO being extended and evolved. The results have proven that this proposed HO design has compiled herb usage for use by conventional and modern herbalists.

# ABSTRAK

Ontologi ialah perkongsian perbendaharaan kata tentang perkara umum sesuatu domain. Ontologi mempunyai pelbagai darjah ekspresi. Ontologi tidak kompleks dibina menggunakan kelas, atribut dan hubungan tanpa takrifan aksiom seperti yang terdapat dalam ontologi kompleks. Sehingga hari ini tiada lagi reka bentuk ontologi yang wujud dalam lapangan herba serta yang boleh diukur kerana kepelbagaian struktur ontologi. Dalam kajian ini ontologi tidak kompleks dalam domain herba yang dikenali sebagai Ontologi Herba (HO) telah dibangunkan untuk meneroka penggunaan herba lengkap berdasarkan profil herba. HO dimulakan dengan pemodelan tidak formal domain diikuti dengan reka bentuk tidak formal HO yang memanipulasikan notasi "Unified Modelling Language" (UML) untuk mengetengahkan fungsi, servis dan strategi prosedur. Sehubungan itu, sebelas metrik ontologi yang merangkumi tiga prinsip kematangan, iaitu: menggunakan semula, melanjutkan dan berkembang dibentangkan dalam kajian ini. Prinsip-prinsip ini diukur dalam kedua-dua peringkat, iaitu tahap kelas dan tahap ontologi, supaya aspek yang berbeza dalam reka bentuk ontologi boleh dinilai dan akan membantu pengawalan proses pembangunan HO. Kemudiannya, reka bentuk HO dibandingkan dengan ontologi yang lain seperti COIN, Ontologi Gen dan OntoCAPE, dan didapati bahawa HO mempunyai Kekayaan Warisan = 0.99301 dengan potensi untuk digunakan semula dan rangkaian ontologi yang padat (Nisbah Nod Pinggir = 1.84). Ini menunjukkan kemungkinan untuk HO dikembangkan dan berevolusi. Keputusan ini membuktikan cadangan reka bentuk HO telah menyusun penggunaan herba untuk kegunaan ahli herba konvensional dan moden.

# TABLE OF CONTENTS


| CHAPTER | TITLE | PAGE |
|---|---|---|
| | **DECLARATION** | ii |
| | **ACKNOWLEDGEMENT** | iii |
| | **ABSTRACT** | iv |
| | **ABSTRAK** | v |
| | **TABLE OF CONTENTS** | ix |
| | **LIST OF TABLES** | x |
| | **LIST OF FIGURES** | xii |
| | **LIST OF ABBREVIATIONS** | xiii |
| 1 | **INTRODUCTION** | |
| | 1.1 Background | 1 |
| | 1.2 Challenges in Designing Lightweight Ontology | 3 |
| | 1.3 Current Method in Designing Lightweight Ontology | 4 |
| | 1.4 Problem Statement | 6 |
| | 1.5 Objectives of the Study | 7 |
| | 1.6 Scope of the Study | 7 |
| | 1.6 Significance of the Study | 9 |
| | 1.7 Organization of the Thesis | 10 |

**2 LITERATURE REVIEW**

2.1 Introduction 11
2.2 Herb Domain 12
2.3 Lightweight Ontology 15
2.4 Informal Domain Modelling 18
2.5 Informal Ontology Design Method 21
2.6 Ontology Maturity Analysis 27
2.7 Trends and Directions 29
2.8 Summary 31

**3 RESEARCH METHODOLOGY**

3.1 Introduction 32
3.2 The Framework of the Study 33
3.3 Data Sources and Preparation 37
3.4 Instrumentation and Result Analysis 39
3.4.1 Hardware and Software Requirements 39
3.4.2 Application and Analysis 40s
3.4.3 Evaluation Metrics 40
3.5 Summary 41

**4 INFORMAL DOMAIN MODELLING**

4.1 Introduction 43
4.2 Taxonomy of Herb Domain 44
4.3 Thesaurus and General Properties of Herb 48
4.4 Reuse of Resources Collection 50
4.5 Features of Herb 54
4.6 Summary 57

**5 INFORMAL DESIGN OF HERB ONTOLOGY (HO)**

5.1 Introduction 58
5.2 Methodology of HO 62
5.3 Informal Specification of HO Structure 66
5.4 HO Applications 70
5.4.1 HO Functionality 71
5.4.1.1 Include and Extend Dependencies 74
5.4.1.2 Inheritance between Actors 75
5.4.2 HO Services 75
5.4.3 HO Procedural Strategies 78
5.5 Discussions on Maturity of HO in Terms of Extension, Reuse and Evolve 80
5.6 Summary 83

**6 MATURITY ANALYSIS OF HERB ONTOLOGY**

6.1 Introduction 84
6.2 The Herb Ontology 87
6.2.1 Scope 87
6.2.2 Structure 88
6.2.3 Content 90
6.3 Ontology Maturity Metrics 92
6.3.1 Reuse Metric 93
6.3.2 Extend Metric 94
6.3.3 Evolve Metric 97
6.4 Result and Analysis 99
6.4.1 Data Sources in Evaluation Phase 99
6.4.2 Relating Metrics with Ontology-Level Evaluation 100
6.4.3 Relating Metrics with Class-Level Evaluation 101

| | | | |
|---|---|---|---|
| | | 6.4.4 Relating Metrics with Reuse Maturity Principle | 103 |
| | | 6.4.5 Relating Metrics with Extend Maturity Principle | 104 |
| | | 6.4.6 Relating Metrics with Evolve Maturity Principle | 105 |
| | 6.5 | Summary | 106 |
| **7** | **CONCLUSION** | | |
| | 7.1 | Concluding Remarks | 107 |
| | 7.2 | Contributions | 109 |
| | 7.3 | Future Works | 111 |
| | 7.4 | Summary | 111 |
| | **REFERENCES** | | 112-117 |

# LIST OF TABLES


| TABLE NO. | TITLE | PAGE |
|---|---|---|
| 2.1 | Comparison between lightweight and heavyweight ontology | 18 |
| 2.2 | Advantages of domain model | 20 |
| 2.3 | Examples of ontology level formality | 22 |
| 2.4 | Ontology levels of generality | 26 |
| 2.5 | Summary of related works in ontology metrics | 30 |
| 3.1 | Heterogeneous herbs data resources | 37 |
| 3.2 | List of ontologies | 38 |
| 3.3 | Summary of ontology metrics | 41 |
| 4.1 | Example of herb general properties | 49 |
| 4.2 | Example reusing herb resources | 53 |
| 4.3 | Example of herb edibility, medicinal and other features | 56 |
| 5.1 | Ontology maturity levels | 61 |
| 5.2 | The structure development of HO | 68 |
| 5.3 | The use case description table | 72 |
| 5.4 | The nodes descriptions | 77 |
| 5.5 | Ontology based on the complexity of its structure | 81 |
| 5.6 | Examples of ontology classification | 81 |
| 5.7 | Examples of how ontologies are extend, reuse and evolved | 82 |
| 6.1 | The ontology complexity structure | 85 |
| 6.2 | Summary of ontology maturity metrics | 93 |
| 6.3 | List of ontologies | 99 |
| 6.4 | Ontology-level results | 101 |

## LIST OF FIGURES


| FIGURE NO. | TITLE | PAGE |
|---|---|---|
| 2.1 | Content structure of Chapter 2 | 11 |
| 2.2 | Classification of plant | 12 |
| 2.3 | Classification of herb usage | 15 |
| 2.4 | Concept hierarchy (taxonomy). Source: http://www.ai-one.com/tag/ lightweight-ontology/ | 17 |
| 2.5 | Concept hierarchy with additional relationships. Source: http://www.ai-one.com/tag/ lightweight-ontology/ | 17 |
| 2.6 | Ontology formality-complexity classification. Source: Lassila and McGuinness (2001) | 23 |
| 2.7 | Ontology levels of generality | 26 |
| 3.1 | Overview of the research framework | 34 |
| 4.1 | Plant classification | 44 |
| 4.2 | Taxonomic hierarchy of ginseng | 45 |
| 4.3 | Taxonomic hierarchy of common ginkgo | 46 |
| 4.4 | Pre-structure of herb | 47 |
| 4.5 | Example of resources collection | 52 |
| 4.6 | Example of herb features description on medicinal | 54 |
| 4.7 | Another example of herb features description on medicinal | 54 |
| 4.8 | Example of herb features description on edibility | 55 |
| 5.1 | The ontological formality-complexity graph | 60 |
| 5.2 | Ontology classification | 60 |
| 5.3 | HO methodology | 63 |
| 5.4 | Sources in HO | 65 |

| | | |
|---|---|---|
| 5.5 | The structure development of HO | 67 |
| 5.6 | The example of interterm relations in HO | 70 |
| 5.7 | Use case representation of the functionality involved in HO | 73 |
| 5.8 | Examples of mapping the processes occurred in HO with use cases | 73 |
| 5.9 | Example of inheritance between actors in HO | 75 |
| 5.10 | Services offered in HO | 76 |
| 5.11 | The procedural for data access and creating enquiries; **(b)** The procedural for manual curations; **(c)** The procedural for improving HO keeping record and check missing database. **(d)** The procedural for triggering HO weekly updates | 79 |
| 6.1 | HO specification | 88 |
| 6.2 | HO structure | 89 |
| 6.3 | HO data sources | 91 |

# LIST OF ABBREVIATIONS

| | |
|---|---|
| CCO | - Cell-Cycle Ontology |
| CLEPE | - Conceptual Level Programming Environment |
| CO-EDE | - Collaborative Open Ontology Development Environment |
| CPU | - Central Processing Unit |
| EU | - European Union |
| FMA | - Foundational Model of Anatomy |
| FOAF | - Friend of a Friend |
| GO | - Gene Ontology |
| GOC | - Gene Ontology Consortium |
| HO | - Herb Ontology |
| ISWC | - International Semantic Web Series |
| ITIS | - Integrated Taxonomic Information System |
| MGI | - Mouse Genome Informatics |
| OBO | - Open Biomedical Ontologies |
| ODP | - Ontology Design Pattern |
| OPPL | - Ontology PreProcessor Language |
| OWL | - Web Ontology Language |
| PFAF | - Plants for a Future |
| PO | - Plant Ontology |
| RDF | - Resource Description Framework |
| RDF-S | - RDF-Schema |
| SGD | - Saccharomyces Genome Database |
| SNOMED-CT | - Systematized Nomenclature of Medicine Clinical Terms |
| TAIR | - The Arabidopsis Information Resources |
| TCM | - Traditional Chinese Medicine |

| | |
|---|---|
| UML | - Unified Modeling Language |
| WHO | - World Health Organization |
| W3C | - World Wide Web Consortium |
| XML | - Extensible Markup Language |

# CHAPTER 1

# INTRODUCTION

## 1.1 Background

Herb has evolved from an alternative source of effective means in medicine (Tesch, 2002; Ali *et al.*, 2008; Kaefer *et al*., 2008), cosmetics (Antignac *et al*., 2011) and culinary (Hayaloglu *et al.*, 2011; Mielnik *et al.*, 2008) into a more mainstream method. Although Western medical practices seem to have questioned or even denied the efficacy of many traditional herbal remedies, traditional plants undoubtedly continue to play a key role in the well-being of indigenous communities (Darko, 2009). Therefore, despite the dramatic advances of conventional medicine, it is clear that a vast amount of herb usages continue to possess a high level of significance in many social settings. The dramatic increment of interest in the usage of herb these days are due to the critical scientific analysis and quality control of their therapeutic potential and safety. Today, there are thousands of herb information resources created by a wide range of information providers including herbalists, government agencies, charitable organizations, and non-profitable agencies who publish herbs data in one form or another. The abundance of herb information leads to the overflow of available resources; hence it is essential to treat them as an asset rather than a problem. In recent years, the proliferation in the usage of traditional herb has prompted researchers and regulators around the world to focus their attention on how to regulate this group of product and bringing them to the mainstream of research area.

The fundamental issue to be addressed in organizing herbs knowledge concerns the diverse and broadest definition of herbs. Variations in definition of herb are due to the cross-cultural differences that exist years ago. This study attempts to bridge knowledge across countries and species in herb domain.

Ontology is now in widespread use as a means in representing domain knowledge. An example in plant domain is Plant Ontology (PO: http://www.plantontology.org/). The goal of Plant Ontology is to produce a dynamic, controlled vocabulary to describe plant structure and developmental stages. It collaborates with several model plant genome databases which are The Arabidopsis Information Resources (TAIR: Rhee *et al.*, 2003), Gramene (Ware *et al.*, 2002) and MaizeGDB (Lawrence *et al.*, 2004) to enable comparative plant genomics research. In herbs domain, Traditional Chinese Medicine (TCM) has been actively researched and there are an enormous amount of websites in detailing amount of TCM resources. However, most of the databases are either inaccessible or highly restricted for information sharing (Chen, 2011). They solved the problem by introducing TCM Database@Taiwan (http://tcm.cmu.edu.tw/) which facilitates the virtual screening process in the experiment design for the TCM lead drug discovery. This database is actually the source to derive novel pharmaceutical compound and is currently the largest non-commercial TCM database available for download.

Ontologies have different degrees of expressiveness. The heavyweight ontology is built using classes, instances, relationships and including axiomatic definitions, in which lightweight ontology is lacking. However, an initial lightweight ontology is needed to learn and acquire hints on what concepts to be considered in final model in explored domain. This is very helpful as in the beginning, it is often that only partial of relevant domain concepts are known. Dublin Core (http://dublincore.org/) is recognized as the simplest ontology due to the simplicity of its internal structure, which consists of terms that stipulate the meta-data for documents. The COIN Ontology (Zhu and Madnick, 2006) is also classified as a lightweight ontology, but unlike WordNet and the Yahoo! Dictionary, this ontology includes high level concepts and uses formal ontology language. The famous full-fledged ontologies include Gene Ontology (Ashburner *et al*., 2000), Plant Ontology (Jaiswal *et al*., 2005) and OntoCAPE (Morbach *et al*., 2008).

One of the most important roles of lightweight ontology is that it provides a vocabulary of terms and some specifications of their meaning. This includes definitions and an indication of how concepts are inter-related. This would collectively impose a structure on the domain and constrain the possible interpretations of terms. The following sections in this chapter discuss the motivation and challenges involved in representing knowledge in a lightweight ontology. Then, the current methods in designing lightweight ontology will be presented. This is followed by the problems to be solved in this study. Research goal, objectives, scopes and significance ensue thereafter. The chapter ends with thesis organization.

## 1.2 Challenges in Designing Lightweight Ontology

Although there are overflow of herb knowledge providers, yet the broad definition of herb across the countries contributes to the heterogeneity of herb information. As a consequence, the herb usages are often overlapped from one herb database to another (e.g. same species of herb may have different usages in different countries considering some herbs are known for their different cultural background), and a common name for particular herb species is inaccurately interchanged by the layman. In turn, the first challenge belongs to the heterogeneous herb information which leads to inconsistent quality of information.

In order to produce an evolving, extending and reusable lightweight ontology in herb domain, the second challenge must be tackled, which is pertaining to the informal design of lightweight ontology. The ontology design released by researchers especially in herb domain is relatively very small in number. Although there are several works that involve ontology in life science field, it seems that there is no suitable ontology that exists in herb domain that can be reused or extended. For instance, Plant Ontology (Plant Ontology Consortium, 2003) presents controlled vocabularies that reflect the biology of plant structures and developmental stages. It collaborates with several model plant genome databases which are The Arabidopsis Information Resources (TAIR: Rhee *et al.*, 2003), Gramene (Dare *et al.,* 2002) and MaizeGDB (Lawrence *et al.,* 2004) to enable comparative plant genomics research. Hence, Plant Ontology is very species-specific which does not suit the herb definition in this study. Nonetheless, several methodologies for ontology building have been proposed and can be used as guidelines in designing lightweight ontology (Noy and McGuinness, 2001; Fernandez *et al*.; 1997; Uschold and Gruninger, 1996). Performing technical analysis to explore these concepts requires a careful analysis and is very time consuming. This will eventually leads to load of work in defining the terms, relationship and herb species annotation in this study.

The third challenge lies in evaluating the lightweight ontology design. The maturity evaluation can help ontology developers and maintainers with a better understanding of the current status of ontology, therefore allowing them to enhance their evaluation on its design and have a better control on its development process. In software engineering domain, metrics play an important role in designing, developing and maintaining software defects for future maintenance problems (Binstock and Andrew, 2010; Lincke *et al*., 2008). Intently, the concepts of software metrics are being used in measuring the maturity of ontology designs. Nonetheless, the problem with ontology metrics is that ontologies are heterogeneous in their structure, objectives and level of formality. Hence, this leads to the study of combinations of ontology metrics from several researchers, to find the suitable metrics that fit the design of lightweight ontology in this study.

## 1.3 Current Methods in Designing Lightweight Ontology

Generally, there are many ways to design lightweight ontology, depending on the category of ontology. Ontology design can be classified into four categories: formality, internal complexity, generality and pattern.

(a) Formality: The ontologies range from informal representations which can be automatically or semi-automatically derived from user classifications (e.g. the structure of folders in a file system) and web directories (Yahoo!, http://www.yahoo.com/ and Google, http://www.google.com/), to progressively more formal representations like enumerative classification schemes, Dewey Decimal Classification (http://www.oclc.org/dewey/) and the Library of Congress Classification (http://www.loc.gov/). This is followed by more strictly defined but still informal structures, such as thesauri and taxonomies, AGROVOC (http://aims.fao.org), faceted classification schemes, Colon Classification (Ranganathan, 2006), and, eventually, formal ontologies which are expressed into a logic formal language and represented using formal specifications such as Description Logic (DL) or Web Ontology Language (OWL).

(b) Internal complexity: Lassila and McGuinness (2001) described ontology complexity continuum ranges from lightweight ontology, which is typically defined as a hierarchical or taxonomy-like structure, to a full-fledged ontology as more relationships are captured. The complexities of ontological structures are linearly correlated with the level of formality.

(c) Generality: According to Gruber (2008), the classification of ontologies with respect to its generality starts with top-level ontologies (Dublin Core, WordNet, and Yahoo! Dictionary), domain ontologies (Gene Ontology and OntoCAPE), task ontologies (Nunes *et al.*, 2009) and lastly application ontologies (Shaw *et al.*, 2008).

(d) Pattern: Aranguren (2008) described ontology design pattern as a reusable solution to common recurrent object-oriented design problems. He used this technique to support the migration of Open Biomedical Ontologies (OBO) language to OWL DL and the creation of OWL DL ontology can be done with ease. This will produce more maintainable and expressive ontologies where more complex queries can be done and the biological knowledge is represented with higher fidelity.

## 1.4 Problem Statement

The problem in representing knowledge specifically in herb domain is described as follows:

> "Given broad definition of herb across the countries which contributed to the heterogeneity of herb information, the challenge is to design a lightweight ontology to get the fundamental concepts of herb domain which eventually, is expected to be reused, extended and evolved following the characteristics of heavyweight ontology. In addition, the design must be able to be measured to overcome heterogeneous aspects of ontology design."

Based on the above challenges, some factors need to be addressed by the possible solution. The first factor is related to the overflow of sources, which results in the information overlapping in herb domain. Besides, the heterogeneity of data also contributes to the false description and irrelevant answers to the users. It is observed that herb has the broad definition across countries and species, hence it is important to have a domain modelling that captures the terms and concepts existed in herb domain. Thus, this study aims to have the informal domain modelling that will describe the basic component according to herbalists and plant researchers through their websites and databases.

The second factor is relating to the informal design of lightweight ontology in herb domain. Currently, it seems that there is no suitable ontology design that exists in herb domain that can be reused or extended. The existing ontologies are either referring to herb in particular countries or the pharmacology of certain herb species. In contrast, this study aims to design an ontology of herb domain with non-species-specific across taxa. This study targets to support long-term ontological development as it moves forward on the ontological complexity continuum. The design would be able to cater the progress of its maturity which takes in form of extension, reuse and evolution.

The third factor is related to measuring the ontology complexity which is formed by various combinations of dimensional characteristics. Evaluation by a single metric would not be able to cover the overall insights of ontologies explored. Thus, this study aims to have a different set of ontology metrics to gain better results in interpreting the ontology insights in terms of ontology maturity. The results of this study would point out the complexity of ontology and their relation with maturity principles in extending, evolving and addressing the reusability issue in lightweight ontologies design.

## 1.5 Objectives of the Study

The goal of this study is to represent knowledge of herb domain with features that can be reused, extended and evolved using lightweight ontology. In order to reach this goal, several objectives have to be achieved:

(a) To investigate the related herb terms and relationships in order to design the informal domain modelling of HO.
(b) To design the lightweight HO by implementing informal domain modelling in (a).
(c) To evaluate the lightweight HO by using ontology metrics that covers class-level and ontology-level in order to meet with the ontology maturity principles.

## 1.6 Scope of the Study

The “herb terms” or “terms” in this study refers to the class and instances of herb. In this study, the data sources are catalogued into four major categories: (a) personal repositories; (b) government regulators repositories; (c) charitable repositories; and (c) non-profitable repositories. The personal repositories are obtained from CookBook Herbalism (http://earthnotes.tripod.com); and Herb Health

Guide (http://www.herb-health-guide.com). The government regulators repositories are obtained from two different sources which are PLANT Database (http://plants.usda.gov) and Integrated Taxonomic Information System (ITIS: http://itis.gov). On the other hand, the charitable repositories are contributed by The Herb Society (http://www.herbsociety.org.uk/) and Plants for a Future database (PFAF: http://www.pfaf.org). Lastly, the non-profitable repositories are given by Complementary and Alternative Healing (http://alternativehealing.org/) and American Botanical Council and Holistic Healing Webpage (http://www.holisticmed.com/www/ herbalism.html herbalism.html).

Unlike Plant Ontology, which combines several developed ontologies which are Gramene, MaizeGDB, and The Arabidopsis Information Resources to describe anatomy and morphology of flowering plants in their growth stage and developmental stage, HO is not species-specific. As herb can be defined in its broadest definition, hence, there is no complete structure of herbs in plant taxonomy. Therefore, it is reliable to deliberately design HO to be species-neutral. This includes terms in HO that are applicable to angiosperm and gymnosperm, woody and non-woody herb. Thus, HO covers any herbs from any species and niches. However, HO focuses on the divergence of herb usages querying from the botanical or common names of herbs. Hence, HO represents common concept that covers usages across herb species.

HO will be evaluated using eleven metrics by several authors that are being collected. These metrics would then be divided into three categories that would contribute to the maturity principles. They are: (i) reuse; (ii) extend; and (iii) evolve categories. In order for the lightweight ontologies to mature, they need to be on the same level or better than the established full-fledged ontologies. Therefore, the famous full-fledged ontologies, Gene Ontology and OntoCAPE are being used as comparisons to the lightweight ontologies (e.g. HO and COIN ontology).

## 1.7 Significance of the Study

In this study, lightweight ontology will be used to represent the herb domain. The reason of using lightweight ontology is to give a fundamental to the relevant concepts occurred in herb domain. The lightweight ontology could give a formal representation to a set of concepts within a domain and the relationships between those concepts. The design of lightweight ontology is required to solve the problems of terms heterogeneity in herb domain, specifically in herb profile and usage of herbs.

Therefore, this study comes out with the lightweight ontology across taxon which is called HO. Even though this lightweight ontology is simple and involves only a few relationships, it could surprisingly be a powerful tool for domain researchers when meticulously done. This ontology could help in the description of herbs' common name and usages which require uniform terminology that describes properties of certain herb species. Moreover, it would also facilitate in cross species comparative studies and comparison of herbs taxonomy and chemical composition found in certain species as well as herbs useful properties. The herb information from HO could also aid in drug-herb and food-drug interaction studies that are being rapidly conducted by researchers (Chen *et al*., 2011; Yoshikawa and Konagaya, 2006; Dragland *et al*., 2003). Besides, the chemical components in HO could help in providing detailed approach in order to address the complexity in biomedical domain, by combining herb function with modern pharmaceutics and biomedicine (Yu, 2008; Abel and Busia, 2005).

Apart from analysing evaluation results in general ways, indication of their complexity and its relations in maturing ontologies design to the lightweight ontologies especially in HO is done at the end of this study. The proposed set of metrics are aimed towards the improvement of lightweight ontology specifically HO.

## 1.8 Organization of the Thesis

This thesis is organized into seven chapters. A brief description on each chapter is as follows:

(a) Chapter 1 defines the challenges, problems, current methods, objectives, scopes and significance of the study.

(b) Chapter 2 reviews the main subjects of interest, which are the herb domain, lightweight ontology, informal domain modelling, ontology design methodology, ontology maturity analysis. The last section of this chapter will present the trend and tendencies related to this study.

(c) Chapter 3 begins with a brief review of the proposed ontology development framework, followed by detailed descriptions of hardware and software requirements, data sources, testing and analysis procedures and performance measurement used.

(d) Chapter 4 gives a brief overview on the basic component described by herbalists and plant researchers through their websites and databases. This includes explanations on the taxonomy of herb domain, thesaurus and general properties of HO, the reuse of resources collection and features of HO.

(e) Chapter 5 lays out the informal design of HO. The methodological framework of HO, informal HO specification and designs and applications of HO maturity, which includes the definition of extension, reuse and evolution, will be depicted in this chapter.

(f) Chapter 6 proposes on the maturity analysis of HO. This chapter provides a short overview of HO, background of ontology metrics, description of datasets, the proposed metrics and their relations to the extension, reuse and evolution of ontologies, analysis of the metrics towards experimented ontologies and discussion on the impact of proposed metrics towards ontology maturing process especially HO.

(g) In Chapter 7, the conclusion of the study and the achieved results to date are presented. The contributions and future works of the study will also be described.

# REFERENCES


Abel, C., Busia, L. (2005). An Exploratory Ethnobotanical Study of the Practice of Herbal Medicine by the Akan People of Ghana (Ethnobotanical Study: Herbal Medicine in Ghana). *Alternative Medicine Review*. 10: 112-122.

Ali, S. S., Kesoju, N., Luthra, A. (2008). Indian Medicinal Herbs as Sources of Antioxidant. *Food Research International*. 41(1): 1–15.

Antignac, A., Nohynek, G. J., Re, T., Clouseau, J., Toutain, H. (2011). Safety of Botanical Ingredients in Personal Care Products/Cosmetics. *Food and Chemical Toxicology*. 49(2): 324–341.

Aranguren, M. E., Antezana, E., Kuiper, M., Stevens, R. (2008). Ontology Design Patterns for Bio-Ontologies: A Case Study on the Cell Cycle Ontology. Proceedings of the 10-th Bio-Ontologies Special Interest Group Workshop 2007. 20 July 2007. Vienna, Austria: BMC Bioinformatics.

Ashburner, M., Ball, C. A., Blake, J. A. et al. (2000). Gene Ontology: Tool for the Unification of Biology. *Natural Genetics*. 25: 25-29.

Ball, C. A., Dolinski, K., Dwight, S. S., *et al*. (2000). Integrating Functional Genomic Information into the Saccharomyces Genome Database. *Nucleic Acids Research*. 28 (1): 77-80.

Binstock, A. (2010). Integration Watch: Using Metrics Effectively. *SD Times*. BZ Media. Retrieved December 19, 2012, from http://www.sdtimes. com/link/34157.

Birkedal, L., Mogelberg, R. E., Petersen, R. L. (2007). Domain-theoretical models of parametric polymorphism. *Journal Theoretical Computer Science*. 388 (1-3): 152-172.

Blake, J. A., Eppig, J.T., Richardson, J. E., *et al*. (2000). The Mouse Genome Database (MGD): Expanding Genetic and Genomic Resources for the Laboratory Mouse. *Nucleic Acids Research*. 28 (1): 108-111.

Brickell, C. (2008). *RHS A-Z Encyclopaedia of Garden Plants*. United Kingdom: Dorling Kindersley.

Bruijn, J. D. (2003). Using Ontologies: Enabling Knowledge Sharing and Reuse on the Semantic Web. DERI Technical Report. Digital Enterprise Research Institute (DERI), University Road Galway, Ireland.

Chen, C. Y-C. (2011). TCM Database@Taiwan: The World's Largest Traditional Chinese Medicine Database for Drug Screening *In Silico*. *PLoS ONE*. 6(1): e15939.

Corho, O., Lopez, E. F., Perez, A.G. (2003). Methodologies, Tools and Languages for Building Ontologies. Where Is Their Meeting Point? *Data & Knowledge Engineering*. 46 (1): 41-64.

Crosby, M. A., Goodman, J. L., Strelets, V. B., *et al*. (2007). FlyBase: Genomes by the Dozen. *Nucleic Acids Research*. 35: 486-491.

Darko, I. N. (2009). *Ghanaian Indigenous Health Practices: The Use of Herbs*. Master Thesis. University of Toronto, Canada.

Denaux, R., Dolbear, C., Hart, G., Dimitrova, V., Cohn, A. G. (2011). Supporting Domain Experts to Construct Conceptual Ontologies: A Holistic Approach. *Web Semantics: Science, Services and Agents on the World Wide Web*. 9 (2): 113-127.

Ding, Y., Fensel, D. (2001). Ontology Library Systems: The Key to Successful Ontology Re-Use. *First Semantic Web Working Symposium (SWWS01).* California, USA.

Dragland, S., Senoo, H., Wake, K., Holte, K., Blomhoff, R. (2003). Several Culinary And Medicinal Herbs Are Important Sources of Dietary Antioxidants. *Journal of Nutrition*. 133: 1286-1290.

Ehrlich, S. D. (2011, February 10). Herbal Medicine. *Avera*. Retrieved January 31, 2013, from http://averaorg.adam.com

Flouris, G., Plexousakis, D., Antoniou, G. (2006). Evolving Ontology Evolution. *Proceedings of the 32nd International Conference on Current Trends in Theory and Practice of Computer Science (SOFSEM 06).* Berlin, Heidelberg. ACM Digital Library: 14-26.

Fellbaum, C. (2006). Encyclopedia of Language & Linguistics (Second Edition). USA: Princeton University.

Fernandez, M., Gomez-Perez, A., Juristo, N. (1997). METHONTOLOGY: From Ontological Art Towards Ontological Engineering. *Proceedings of the AAAI97 Spring Symposium*. Stanford, USA, 33-40.

Gangemi, A., Catenacci, C., Ciaramita, M., Lehmann, J. (2006). Modelling Ontology Evaluation and Validation. *Proceedings of the 3rd European Semantic Web Conference (ESWC'06)*. Budva, Montenegro, 140–154.

Garcia-Penalvo, F. J., Colomo-Palacios, R., Garcia, J., Theron, R. (2012). Towards An Ontology Modeling Tool. A Validation in Software Engineering Scenarios. *Expert Systems with Applications*. 39 (13): 11468-11478.

Gruber, T. (2009). Ontology. In Liu, L., and Ozsu, M. T. (Ed). Encyclopedia of Database Systems (pp. 1963-1965). Springer-Verlag.

Gruniger, M., Fox, M. S. (1995). Methodology for the Design and Evaluation of Ontologies. *Proceedings of IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing*. Montreal, Canada.

Guarino, N. (1998). Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. *Summer School on Information Extraction*. 14-19 July. Frascati, Italy: Springer-Verlag.

Hayaloglu, A. A., Farkye, N.Y. (2011). Cheese with Added Herbs Spices and Condiments. New York, United States: Academic Press.

Hill, D. P., Blake, J.A., Richardson, J. E., Ringwald, M. (2002). Extension and Integration of the Gene Ontology (GO): Combining GO Vocabularies with External Vocabularies. *Genome Research*. 1982-1991.

Jaiswal, P., Avraham, S., Ilic, K., *et al.* (2005). Plant Ontology (PO): A Controlled Vocabulary of Plant Structures and Growth Stage. *Comparative and Functional Genomics.* (6): 388–397.

Kim, I-W., Lee, K-H. (2009). Model-Driven Approach for Describing Semantic Web Services: From UML to OWL-S. *IEEE Transactions on Systems, Man, And Cybernetics.* 39 (6): 637-646.

Klein, M., Fensel, D. (2001). Ontology Versioning on the Semantic Web. *First Semantic Web Working Symposium (SWWS01).* California, USA.

Law, K. S., Wong, C-S., Mobley, W. H. (1998). Toward A Taxonomy of Multidimensional Constructs. *The Academy of Management Review*. 23 (4): 741–755.

Lawrence, C. J., Dong, Q., Polacco, M. L., *et al*. (2004). MaizeGDB: The Community Database for Maize Genetics and Genomics. *Nucleic Acids Res*. 32: D393-397.

Lesley, J. (1997). *Herbs: The Visual guide to more than 700 herb species from around the world*. New York: Dorling Kindersley.

Lincke, R., Lundberg, J., Lowe, W. (2008). Comparing Software Metrics Tools. *International Symposium on Software Testing and Analysis*. 20-24 July, Seattle, Washington, USA, 131-142.

Lionelli, S., Diehl, A. D., Christie, K. R., Harris, M. A., Lomax, J. (2011). How the Gene Ontology Evolves. *BMC Bioinformatics*. 12(325): 1471-2105.

MacKenzie, R. E., Rakel, B. (2006). *Complementary and Alternative Medicine for Older Adults: A Guide to Holistic Approaches to Healthy Aging*. New York: Springer Publishing Company.

Mamat A., Rahman A. A. (2009). Designing a Conceptual Model for Herbal Research Domain using Ontology Technique. *Ninth International Conference on Intelligent Systems Design and Applications*. 13 Nov-2 Dec. Pisa, Italy: IEEE, 1167-1172.

McLeod, G. (2009, January 29). Brief Introduction to Domain Modelling. *Inspired*, Retrieved January 31, 2013, from http://www.slideshare.net

Mielnik, M. B., Sem, S., Egelandsdal, B., Skrede, G. (2008). By–Products from Herbs Essential Oil Production as Ingredient in Marinade for Turkey Thighs. *LWT–Food Science and Technology*. 41(1): 93-100.

Miller, G. A. (1995). WordNet: A Lexical Database for English. *Communications of the ACM*. (11): 39-41.

Morbach, J., Yang, A., Marquardt, W. (2008). OntoCAPE-A Large Ontology for Chemical Process Engineering. *Engineering Applications of Artificial Intelligenc*e. 20: 147-161.

Nicola, A. D., Missikoff, M., Navigli, R. (2009). A Software Engineering Approach to Ontology Building. *Information Systems*. 34 (2): 258-275.

Noy, N. F. and McGuinness, D. L. (2001). *Ontology Development 101: a Guide to Creating your First Ontology*. Technical Report, Stanford Knowledge Systems Laboratory, Standford, USA.

Nunes, V. T., Santoro, F. M., Borges, M. R. (2009). A Context-Based Model for Knowledge Management Embodied In Work Processes. *Information Sciences*. 179 (15): 2538-2554.

Obrst, L. (2003). Ontologies for Semantically Interoperable Systems. *Proceedings of the Twelfth International Conference on Information and Knowledge Management*. New York, USA, 366-369.

Oldfield, P. (2002). Domain Modelling. Appropriate Process Movement. Retrieved January 31, 2013, from http://www.aptprocess.com.

Olivier, C. (2009). Improving the Data Quality of Relational Databases using OBDA and OWL2QL. *W3C Web Ontology Language (OWL) - Experiences and Directions Workshop (OWLED '09)*. Virginia, USA.

Owen, D. J. (2002). *The Herbal Internet Companion: Herbs and Herbal Medicine Online*. New York, United States: Haworth Herbal Press.

Qi, X., Cungen, C., Baoyan, L., *et al*. (2010). Establishment of the TCM Meta Conceptual Model Based on Domain Ontology. *World Science and Technology*. 11 (4): 621–625.

Qiu, J. (2007). China Plans to Modernize Traditional Medicine. *Nature*. 446 (7136): 590-591.

Ranganathan, S. R. (2006). *Colon Classification* (6$^{th}$ ed). Ess Ess Publications.

Rhee, S. Y., Beavis, W., Berardini, T. Z., *et al*. (2003). The Arabidopsis Information Resource (TAIR): A Model Organism Database Providing a Centralized Curated Gateway to Arabidopsis Biology, Research Materials and Community. *Nucleic Acid Res*. 31: 224-228.

Robles, K., Fraga, A., Morato, J., Llorens, J. (2012). Towards an Ontology-Based Retrieval of UML Class Diagrams. *Information and Software Technology*. 54 (1): 72–86.

Shang, A., Huwiler, K., Nartey, L., et al. (2007). Placebo-Controlled Trials of Chinese Herbal Medicine and Conventional Medicine Comparative Study. *International Journal of Epidemiology*. 36 (5): 1086–1092.

Shaw, M., Detwiler, L. T., Brinkley, J. F., Suciu, D. (2008). Generating Application Ontologies from Reference Ontologies. *AMIA Annual Symposium Proceeding*. PubMed: 672-676.

Smith, B., Ashburner, M., Rosse, C., *et al*. (2007). The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration. *Nature Biotechnology*. (25): 1251-1255.

Strasunskas, D., Hakkarainen, S. E. (2012). Domain Model-Driven Software Engineering: A Method for Discovery of Dependency Links. *Journal Information and Software Technology*. 54 (11): 1239-1249.

Tartir, S., Arpinar, I. B., Moore, M., et al. (2005). Ontoqa: Metric-based ontology quality analysis. *IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources*. Houston, Texas.

Tesch, B. J. (2002). Herbs Commonly Used by Women: An Evidence–Based Review. *Disease-a-Month.* 48 (10): 671–696.

Uschold, M. and Gruninger, M. (1996). ONTOLOGIES: Principles, Method and Application. *Knowledge Engineering Review*. 11(2): 93-155.

Uschold, M. and Grüninger, M. (2004). Ontologies and Semantics for Seamless Connectivity. *SIGMOD Record*. 33: 58-64.

Valaski, J., Malucelli, A., Reinehr, S. (2012). Ontologies Application in Organizational Learning: A Literature Review. *Expert Systems with Applications.* 39: 7555-7561.

Vrandecic, D., Sure, Y. (2007). How to Design Better Ontology Metrics. *ESWC '07: Proceedings of the 4th European conference on The Semantic Web*. Innsbruck, Austria: Springer-Verlag, 311–325.

Ware, D. H., Jaiswal, P., Ni, J., *et al*. (2002). Gramene: A Tool for Grass Genomics. *Plant Physiol*. 130: 1606-1613.

Weyuker, E. J. (1998). Evaluating Software Complexity Measures. *IEEE Transaction Software Engineering*. 14 (9): 1357-1365.

Yao, H., Orme, A. M., Etzkorn, L. (2005). Cohesion Metrics for Ontology Design And Application. *Journal of Computer Science*. 1 (1): 107–113.

Yoshikawa, S., Satou, K., Konagaya, A. (2004). Drug Interaction Ontology (DIO) for Inferences of Possible Drug-drug Interactions. *Medinfo2004: Proceedings of the 11th World Congress on Medical Informatics*. Amsterdam: IOS Press, 454-458.

Zhang, L., Xie, D. (2002). Comments on the Applicability of Weyuker Property 9 to Object-Oriented Structural Inheritance Complexity Metrics. *IEEE Transactions on Software Engineering*. 28 (5): 526–527.

Zhang, H., Li, Y-F., Tan, H. B. K. (2010). Measuring Design Complexity of Semantic Web Ontologies. *Journal of Systems and Software*. 83 (5): 803-814.

Zhang, H., Zhang, X., Gu, M. (2007). Predicting Defective Software Components from Code Complexity Measures. *PRDC '07: Proceedings of the 13th Pacific Rim International Symposium on Dependable Computing*. Washington, USA: IEEE Computer Society, 93–96.

Zhu, H., Madnick, S. (2006). A Lightweight Ontology Approach to Scalable Interoperability. Working Paper CISL# 2006-06. Composite Information Systems Laboratory (CISL), Massachusetts Institute of Technology.