FORMAL REPRESENTATION
OF ACUPUNCTURE:
CONCEPT, THEORY, AND LOGIC

A THESIS SUBMITTED TO THE UNIVERSITY OF SALFORD FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN THE FACULTY OF HEALTH AND SOCIAL CARE

July 2009

YONGSHENG GAO
SCHOOL OF HEALTH CARE PROFESSIONS
Contents

List of figures ................................................................................................................ vii
List of tables .................................................................................................................. viii
List of common acronyms .............................................................................................. ix
Acknowledgements ........................................................................................................ xi
Declaration ..................................................................................................................... xii
Copyright ....................................................................................................................... xiii
Dedication ....................................................................................................................... xiv
Abstract ......................................................................................................................... xv

Prolegomenon ................................................................................................................. 0

Chapter 1 ........................................................................................................................ 1
Introduction ..................................................................................................................... 1
  1.1 An introduction to the thesis ................................................................................... 1
  1.2 Research aim and objectives ................................................................................... 3
    1.2.1 Aim ................................................................................................................... 3
    1.2.2 Objectives ......................................................................................................... 4
  1.3 Contributions of this thesis ..................................................................................... 5
  1.4 Acupuncture — A short history and current practice ............................................... 5
    1.4.1 Subject areas of the study................................................................................. 5
    1.4.2 Brief history of acupuncture............................................................................. 8
    1.4.3 Current practices and regulations................................................................... 10
  1.5 Requirements for knowledge representation......................................................... 12
    1.5.1 Requirement for clinical acupuncture practice .............................................. 12
    1.5.2 Requirement for academic research ............................................................... 14
    1.5.3 Requirements for knowledge sharing ............................................................. 15
  1.6 A guide to the thesis ............................................................................................... 16
  1.7 Summary ............................................................................................................... 17

Chapter 2 ....................................................................................................................... 18
Research Background ..................................................................................................... 18
2.1 Introduction ........................................................................................................... 18
2.2 Ontology ................................................................................................................ 18
   2.2.1 Ontology – A philosophical perspective ....................................................... 18
   2.2.2 Ontology - The semiotic and linguistic perspective ....................................... 21
   2.2.3 Ontology – The informatics perspective ....................................................... 23
2.3 Principles and methodologies of ontology development ..................................... 27
   2.3.1 Principles ........................................................................................................ 27
   2.3.2 Methodologies .............................................................................................. 29
2.4 Systems, languages, and tools for ontology .......................................................... 34
   2.4.1 Systems for ontology ..................................................................................... 35
   2.4.2 Languages of ontology ................................................................................... 38
   2.4.3 Tools for ontology development .................................................................... 41
2.5 Ontologies in medicine ......................................................................................... 42
2.6 The challenges of acupuncture knowledge representation .................................... 46
   2.6.1 Paradigm heterogeneity .................................................................................. 46
   2.6.2 Language heterogeneity ................................................................................. 48
   2.6.3 Ontology heterogeneity .................................................................................. 49
   2.6.4 Content heterogeneity .................................................................................... 49
2.7 Summary ............................................................................................................... 50

Chapter 3 ...................................................................................................................... 52
Approach to the Research ........................................................................................... 52
3.1 Introduction ........................................................................................................... 52
3.2 Philosophical arguments of the research ............................................................... 52
   3.2.1 Philosophical ground ...................................................................................... 52
   3.2.2 Research assumptions .................................................................................... 54
3.3 Ontological arguments of research ....................................................................... 55
   3.3.1 The object of study – domain knowledge ....................................................... 55
   3.3.2 The characteristics of object – paradigms and disciplines ................................ 55
   3.3.3 The result of the study – ontology as knowledge representation ................. 57
3.4 Epistemological arguments of research ................................................................ 57
   3.4.1 Research decisions on system, language, knowledge representation formalism 57
3.4.2 Researcher engagement of development – the interpretation of domain knowledge ............................................................................................ 58
3.4.3 The interpretations of data from participants .......................................................... 59

3.5 Methodological arguments of research ........................................................................ 60
3.5.1 Quantitative vs. Qualitative ............................................................................ 60
3.5.2 Research methodologies for qualitative research ................................................. 61
3.5.3 The methods for data collection in evaluation ..................................................... 64

3.6 Research Pragmatics ............................................................................................. 65
3.6.1 Research participants ..................................................................................... 65
3.6.2 Research environment .................................................................................... 65
3.6.3 Research ethics ............................................................................................... 66

3.7 Summary .............................................................................................................. 66

Chapter 4 ...................................................................................................................... 68
Ontology Design and Development ............................................................................. 68
4.1 Introduction........................................................................................................... 68
4.2 The ontology development environment ................................................................... 69
4.2.1 Protégé ........................................................................................................... 69
4.2.2 OWL plug-in .................................................................................................. 71
4.2.3 Description logics reasoner ............................................................................ 73

4.3 Approach to the ontology development ................................................................... 74
4.3.1 Identify - stage: feasibility study and ontology kick off ........................................ 75
4.3.2 Categorise - stage: ontology kick off and refinement .......................................... 78
4.3.3 Test - stage: evaluation .................................................................................. 79

4.4 Flexible sets of distinctions approach ...................................................................... 80
4.4.1 Flexible sets of distinctions ............................................................................ 80
4.4.2 Experience of flexible sets of distinctions approach ........................................... 84
4.4.3 Flexibility and extensibility of model ............................................................... 89

4.5 Summary .............................................................................................................. 89

Chapter 5 ..................................................................................................................... 91
Acupuncture Ontology ................................................................................................. 91
5.1 Introduction......................................................................................................... 91

5.2 Acupuncture ontology ......................................................................................... 92
5.2.1 Scope ........................................................................................................... 92
5.2.2 Content ....................................................................................................... 92
## List of figures

<table>
<thead>
<tr>
<th>Figure number</th>
<th>Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Conceptual Realism</td>
<td>20</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>The meaning triangle</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>On-To-Knowledge methodology</td>
<td>32</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>Hierarchy of Sowa’s Top-Level Categories</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Architecture of a knowledge representation system based on Description Logics</td>
<td>38</td>
</tr>
<tr>
<td>Figure 2.6</td>
<td>Semantic Web Language cake</td>
<td>40</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Screen shot of Protégé OWL plug-in</td>
<td>71</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Two-way distinction</td>
<td>83</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>Three-way distinction</td>
<td>83</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Superclasses of LR3 after classification</td>
<td>103</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Stated (asserted) view of LR3 specification</td>
<td>105</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Inferred view of LR3 specification</td>
<td>105</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>Screen shot of annotation widget</td>
<td>108</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>CM perspective</td>
<td>122</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>OM perspective</td>
<td>123</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>CM and OM perspectives</td>
<td>124</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Screen shot of Sesame query interface</td>
<td>125</td>
</tr>
<tr>
<td>Figure 6.5</td>
<td>Screen shot of Transana</td>
<td>139</td>
</tr>
<tr>
<td>Figure 6.6</td>
<td>Transana database structure</td>
<td>140</td>
</tr>
</tbody>
</table>
## List of tables

<table>
<thead>
<tr>
<th>Table Number</th>
<th>Caption</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Two-way distinctions</td>
<td>33</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Classification of KR formalism according to the kind of primitives used</td>
<td>35</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>SNOMED CT hierarchies</td>
<td>44</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Contrasting paradigms</td>
<td>53</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Characteristics of practical and scientific knowledge</td>
<td>56</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>The differences between qualitative and quantitative research</td>
<td>61</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Key terms in Protégé</td>
<td>70</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Protégé representation of OWL expressions</td>
<td>72</td>
</tr>
<tr>
<td>Table 4.3</td>
<td>Description logics language expressivity supported by Protégé OWL plug-in</td>
<td>73</td>
</tr>
<tr>
<td>Table 4.4</td>
<td>Protégé system and two-way distinctions</td>
<td>85</td>
</tr>
<tr>
<td>Table 4.5</td>
<td>Protégé system and three-way distinctions</td>
<td>86</td>
</tr>
</tbody>
</table>
# List of common acronyms

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABox</td>
<td>Assertional part (Assertions on individuals)</td>
</tr>
<tr>
<td>AL</td>
<td>Attribute Language</td>
</tr>
<tr>
<td>ALC</td>
<td>Attribute Language with Complements</td>
</tr>
<tr>
<td>API</td>
<td>Application Programming Interface</td>
</tr>
<tr>
<td>AV</td>
<td>Audio Video</td>
</tr>
<tr>
<td>CAM</td>
<td>Complementary and Alternative Medicine</td>
</tr>
<tr>
<td>CM</td>
<td>Chinese Medicine</td>
</tr>
<tr>
<td>DAML</td>
<td>DARPA Agent Markup Language (DARPA – Defense Advanced Research Projects Agency)</td>
</tr>
<tr>
<td>DL</td>
<td>Description Logics</td>
</tr>
<tr>
<td>FaCT</td>
<td>Fast Classification of Terminologies</td>
</tr>
<tr>
<td>FRACTAL</td>
<td>Formal Representation of Acupuncture: Concept, Theory, And Logic</td>
</tr>
<tr>
<td>FMA</td>
<td>Foundational Model of Anatomy</td>
</tr>
<tr>
<td>GP</td>
<td>General Practitioner</td>
</tr>
<tr>
<td>GRAIL</td>
<td>GALEN Representation and Integration Language</td>
</tr>
<tr>
<td>HTTP</td>
<td>Hypertext Transfer Protocol</td>
</tr>
<tr>
<td>ICD</td>
<td>International Classification of Disease</td>
</tr>
<tr>
<td>KIF</td>
<td>Knowledge Interchange Format</td>
</tr>
<tr>
<td>KR</td>
<td>Knowledge Representation</td>
</tr>
<tr>
<td>nRQL</td>
<td>new Racer Query Language</td>
</tr>
<tr>
<td>OIL</td>
<td>Ontology Inference Layer</td>
</tr>
<tr>
<td>OKBC</td>
<td>Open Knowledge Base Connectivity</td>
</tr>
<tr>
<td>OM</td>
<td>Orthodox Medicine</td>
</tr>
<tr>
<td>OWL</td>
<td>Web Ontology Language</td>
</tr>
<tr>
<td>Term</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>OWL DL</td>
<td>Web Ontology Language Description Logics</td>
</tr>
<tr>
<td>OWL Full</td>
<td>Web Ontology Language Full</td>
</tr>
<tr>
<td>OWL Lite</td>
<td>Web Ontology Language Lite</td>
</tr>
<tr>
<td>OWL-QL</td>
<td>OWL query language</td>
</tr>
<tr>
<td>RACER</td>
<td>Renamed Abox and Concept Expression Reasoner</td>
</tr>
<tr>
<td>RDF</td>
<td>Resource Description Framework</td>
</tr>
<tr>
<td>RDF-S</td>
<td>Resource Description Framework Schema</td>
</tr>
<tr>
<td>RDQL</td>
<td>RDF Data Query Language</td>
</tr>
<tr>
<td>RMI</td>
<td>Remote Method Invocation</td>
</tr>
<tr>
<td>RQL</td>
<td>RDF Query Language</td>
</tr>
<tr>
<td>SAIQL</td>
<td>Schema And Instance Query Language</td>
</tr>
<tr>
<td>SeRQL</td>
<td>Sesame RDF Query Language</td>
</tr>
<tr>
<td>SHIRE</td>
<td>Salford Health Informatics Research Environment</td>
</tr>
<tr>
<td>SNOMED</td>
<td>Systematised Nomenclature of Medicine</td>
</tr>
<tr>
<td>SNOMED CT</td>
<td>Systematised Nomenclature of Medicine Clinical Terms</td>
</tr>
<tr>
<td>SNOMED RT</td>
<td>Systematised Nomenclature of Medicine Reference Terminology</td>
</tr>
<tr>
<td>SPARQL</td>
<td>SPARQL Protocol and RDF Query Language</td>
</tr>
<tr>
<td>SPARQL-DL</td>
<td>SPARQL Query for OWL-DL</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured Query Language</td>
</tr>
<tr>
<td>SquishQL</td>
<td>A simple SQL-ish RDF query language</td>
</tr>
<tr>
<td>TBox</td>
<td>Terminology part (assertions on concepts)</td>
</tr>
<tr>
<td>UMLS</td>
<td>Unified Medical Language System</td>
</tr>
<tr>
<td>URI</td>
<td>Uniform Resource Identifiers</td>
</tr>
<tr>
<td>W3C</td>
<td>World Wide Web Consortium</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>XML</td>
<td>eXtensible Markup Language</td>
</tr>
</tbody>
</table>
Acknowledgements

It has been a long journey for my PhD study and I have met some amazing people. Some of you cannot be named for ethical reasons.

The author would like to express greatest gratitude to Professor Stephen Kay, for his tremendous guidance and support as well as invaluable input and feedback throughout this programme of research.

He would like to also thank Dr. Nicholas Hardiker for his great input, and all members of the SHIRE (Salford Health Informatics Research Environment) at the University of Salford, for their assistance, support and friendship.

He would like to thank all experts who have been involved with the research. Your cooperation and help throughout this study is greatly appreciated.

The researcher makes use of open source software and would like to express appreciation for the important contributions of the software. In particular:

- Protégé and its OWL plug-in developed at Stanford University, is available from http://protege.stanford.edu/
- Sesame, open source project from Aduna, is available from www.openrdf.org
- Transana, qualitative analysis of video and audio data software, is developed at the University of Wisconsin-Madison Centre for Education Research. It is available from http://www.transana.org/
- NetBeans, java IDE from Sun, is available from www.netbeans.org
- Eclipse, java IDE, is available from www.eclipse.org

Lei – your love, patience and kindness make it all worthwhile. Thanks for reading the drafts of this thesis, for helpful discussion, and for giving me the support and courage to complete it.

I have experienced many changes in my life during the study. My little boy Kevin was born mid way through my study. When I first saw his face, I felt so peaceful. My little girl Kellie arrived before I started writing up my thesis. She is so energetic. They are the reason that I wish we can build a better world for them.
Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.
Copyright

Copyright in text of this thesis rests with the author. Copies (by any process) either in full, or of extracts, may be made only in accordance with instructions given by the author and lodged in the University of Salford Library. Details may be obtained from the Librarian. This page must form part of any such copies made. Further copies (by any process) of copies made in accordance with such instructions may not be made without the permission (in writing) of the author.

The ownership of any intellectual property rights which may be described in this thesis are vested in the University of Salford, subject to any prior agreement to the contrary, and may not be made available for use by third parties without the permission of the university, which will prescribe the terms and conditions of any such agreement.

Further information on the conditions under which disclosure and exploitation may take place is available from Professor Stephen Kay.
Dedication

For my parents, wife and children
Abstract

This research investigates the development of a single acupuncture ontology which can represent knowledge from two distinct perspectives, i.e., those of so-called orthodox or Western medicine and those of Chinese medicine. Acupuncture, initially a treatment from ancient Chinese medicine, has been widely adopted by increasing numbers of practitioners of orthodox medicine. However the differences between the two traditions make it difficult to exchange knowledge and to compare findings.

With a view to facilitate knowledge sharing, this thesis explores how heterogeneous types of acupuncture knowledge and information can be represented in a single ontology. The purpose of such an ontology is to provide a consistent representation of acupuncture across the different paradigms.

An ontology for acupuncture was conceptualised using the flexible sets of distinctions based on Sowa's distinction approach, and constructed using description logic technologies. To assess the contribution of the acupuncture ontology, evaluation was carried out by experts from two disciplines, i.e., from medical informatics and from health care, and with respect to three applications, i.e., ontology editing tool, ontology browser with choices of perspectives, and web ontology query application. The data collection methods comprised observation and interviews, coupled with video recording and screen capture. The video and transcripts were analysed and coded with software designed for supporting qualitative research.

The outcome of the evaluations concluded that the research had achieved its aim, i.e., to formally represent the concept, theory and logic of acupuncture from distinct paradigms of both orthodox and Chinese medicine in a single ontology. It succeeded in making implicit relationships in acupuncture knowledge explicit at concept level, instance level and meta level across the two paradigms.
**Prolegomenon**

The research presented in this thesis is conducted in the academic discipline of Health Informatics. This thesis considers the ontology as a means of knowledge representation in general and, in particular, describes the development of an acupuncture ontology which reflects knowledge from both perspectives of orthodox medicine and Chinese medicine. The support of a University of Salford bursary for this study is fully acknowledged.

Prior to this work, Yongsheng Gao worked as doctor in General Hospital of Air Force in Beijing, China. He managed implementation of the hospital information system before moving to the UK. He acquired certificate of Microsoft Certified Systems Engineer (MCSE). The passion and belief that informatics will play an important role in health care has led him to the study of health informatics.

During the study, he managed and developed several projects, such as Electronic Patient Record development for renal specialty, National renal registry submission programme at Salford Royal NHS foundation Trust. Currently, Yongsheng is a clinical terminology specialist working on SNOMED CT for Connecting for Health which provides national IT solutions for healthcare in the UK.
Chapter 1

Introduction

This opening chapter presents the aim, objectives and contributions of the thesis and then discusses the motivation and requirements for the research. The aim and objectives are revisited in the final chapter, where a discussion of how well they have been addressed is given. A guide to the thesis concludes this chapter.

1.1 An introduction to the thesis

The research is conducted in the emerging discipline Health informatics, described by Musen and van Bemmel (2003):

'Unlike computer science, for example, which typically places its emphasis on computation, our research community is dedicated to the study of information as a first-class object. It is modelling the data and knowledge required by our applications that requires unique skills, and where our academic research makes its distinguishing contribution. Ours is indeed the discipline that cares about the content. Our principal challenge in medical or health informatics is to understand better the structure of data, information, and knowledge, and to cast our scholarship in terms of appropriate models of these abstract entities.'

Although it is not a formal definition, essentially, it portrays some characteristics of health informatics. The data, information and knowledge are objects and contents. The task is to model these abstract entities into appropriate models for applications in healthcare.
There are heterogeneous types and sources of knowledge, information and data within the health care domain. Complementary and Alternative Medicine (CAM), recognised for its emerging role in healthcare (Ernst 2000), provides such an example with its wide spectrum of practices and philosophies.

Acupuncture, the best known and, relatively a much better researched area of Complementary and Alternative therapies, has been practised by increasing numbers of orthodox medicine practitioners. It has been an integral part of Traditional Chinese Medicine for some two and half millennia and is an essentially clinical empirical knowledge based medical system. The research and practice of acupuncture often raises conflicts with orthodox medicine, especially related to the Chinese medicine philosophies and its concepts. Arguments over whether substantive evidence exists relating to both the therapeutic process and its outcome (Gao and Kay 2005).

However, proponents argue that the power of close clinical observation extending over thousands of years offers insights which should not be simply ignored. They would argue that clinicians and researchers can draw parallels from Chinese medicine for systematic studies that promise therapeutic efficacy of modern medicine and health care (Ng and Liao 1997). With a view to facilitate knowledge sharing, this thesis has chosen acupuncture knowledge to explore how heterogeneous types and sources of knowledge and information can be represented in a single ontology to serve the purpose of both academic learning and research as well as have the potential for implementation in practice. The intended end users might be acupuncture practitioners and educators. The ontology can be used for diverse but related applications such as assisting education of new acupuncture practitioners or decision support in acupuncture practice.

In our study, we focused upon knowledge representation of both Chinese Medicine (CM) and Orthodox Medicine (OM) for the purpose of facilitating the sharing of knowledge.

Research and clinical practice of acupuncture have been carried out all over the world and some biological changes related to acupuncture treatments have been identified. However, mechanisms of acupuncture treatment are by and large still unclear. In the last decade, many researchers and practitioners have realized the importance of the internally coherent set of theoretical basis for the philosophies and concepts underpinning acupuncture. For the purpose of knowledge sharing, research,
and clinical practice, a common ground must be found as a starting point to ensure a clearer recognition of theoretical and philosophical differences in acupuncture.

Ontology was recognised as a key technology for knowledge sharing (Neches et al. 1991; Gruber 1993b; Gómez-Pérez and Benjamins 1999; Fensel 2001). The development of acupuncture ontology could help to build such common ground for a better understanding and a means of communicating and using acupuncture in research and clinical practice. More specifically, such an ontology will give an overview of relationships between OM and CM.

This chapter sets the scene of research. The research aim, objectives, and contributions of this thesis are introduced in this chapter. The motivation and requirements of conducting this study are explained. This chapter is organised as following:

Section 1.2 - introduces aim and objectives
Section 1.3 - explains the contributions of this thesis
Section 1.4 - describes the brief history and current status of acupuncture practice
Section 1.5 - addresses the necessity and importance for knowledge representation
Section 1.6 - provides a guide to the whole thesis
Section 1.7 - summary

1.2 Research aim and objectives

The aim and objectives described in this chapter were developed from careful considerations of the issues raised from the literature review. The aim and objectives are revisited in the final chapter, where a discussion of how well they have been addressed is given.

1.2.1 Aim

The specific aim of this doctoral research is:

To formally represent the concept, theory and logic of acupuncture from distinct paradigms of both orthodox and Chinese medicine in a single
ontology and thereby make implicit relationships explicit and accessible across the paradigms

We argue that there is a requirement for knowledge sharing between practitioners of orthodox and Chinese medicine. To satisfy this requirement, a better knowledge representation is needed that brings together the differing paradigms' perspectives. The ontology of acupuncture is the product of this research and it represents different views of knowledge and enables access to them via a unified specification.

1.2.2 Objectives

To meet the research aim, four primary objectives follow each with sub-objectives. The objectives are:

1. **To explore the significance of ontology as a means of knowledge representation for acupuncture in health care**
   a. To explore the characteristics of ontology in informatics
   b. To explore the development of ontologies in health care

2. **To investigate the ontology development methodology so as to represent different perspectives in a single ontology**
   a. To explore current approaches for ontology construction
   b. To investigate the flexible sets of distinctions approach
   c. To analyse system and domain by the flexible sets of distinctions approach

3. **To develop an ontology in a formal description logic language**
   a. To develop a single ontology which represents both paradigms
   b. To investigate how to maintain validity for cross-paradigms representation

4. **To evaluate the ontology for knowledge representation and knowledge sharing**
   a. To validate the ontology by programmatic and formative evaluation
   b. To evaluate ontology for knowledge representation for the two disciplines (informatics and health care)
   c. To evaluate ontology for knowledge sharing between the two paradigms (CM and OM)
Based on the aim and objectives, the next section highlights the specific contributions of this thesis to the academic field of Health Informatics as well as to practitioners in health informatics who involved with the design, development and implementation of ontology.

1.3 Contributions of this thesis

This thesis makes the following contributions:

Development of the first validated acupuncture ontology, composed of both orthodox and Chinese medicine knowledge representation of different paradigms of knowledge with relations being made explicit at the concept level, instance level and meta level. Contributed original work based on empirical evidence to the academic discipline of Health Informatics by hypothesising the flexible sets of distinctions approach for ontology construction and investigating its application in this ontology.

The findings may be of particular interest to those facing the challenges of knowledge representation for different perspectives.

1.4 Acupuncture – A short history and current practice

In this section, we introduce acupuncture in Chinese Medicine (CM) and Orthodox Medicine (OM) with a short history and an overview of current acupuncture practices.

1.4.1 Subject areas of the study

Acupuncture

Acupuncture as part of Chinese medicine is one of the oldest and most commonly used medical procedures in the world. Developed from the ancient era when needles were made from bamboo sticks, silver or gold, and then iron, its discursive development sees a rich history of the ancient principles of health care in a continuous process of critical thinking as well as extensive clinical observation and testing, with its theories being constantly refined and adapted (Kaptchuk 2000). The ‘holistic approach’ of CM employs a model of health, where a disease is seen in the close interaction between human, nature and society. This health model embraces the theory that the patient’s
constitution predisposes them to a given disease. This approach is different from what is traditional in western medicine; i.e., orthodox medicine’s focus relates to the specific agents as aetiology factors in disease (Dubos 1959).

Due to the diversity in acupuncture practice nowadays, e.g., the way of actual achieving the stimulation, the place where to apply such stimulation, etc., it is difficult to universally define acupuncture. Broadly speaking, acupuncture is a theory and practice of deploying equipment to stimulate at specific points on the human body in order to achieve therapeutic effect.

**Chinese medicine (CM)**

The Chinese pictograph ‘針灸’ for acupuncture, or ‘zhēn jiǔ’ in Chinese Pin-yin phonetic alphabet system, is pronounced approximately in English equivalent as ‘jen jeu’.

However, 针 is rather the Chinese for Acupuncture; 灸 is the Chinese for Moxibustion as they are commonly used in conjunction.

This ancient form of medicine is uniquely a complete and independent system covering health provision, prevention, promotion, treatment and rehabilitation, which has been documented for over two thousand years in literature records in China. Along with acupuncture and moxibustion as the original modalities, it also includes Chinese herbal medication, message and manipulation, Qigong and exercises, such as Tai Ji practice. The theory embodies concepts of Yin and Yang, Zang Fu (or internal organ frameworks), Qi, etc. For example, each of the internal organs, with terms such as heart, lung, liver, large intestine, represents a collection of body functions, which are associated with functions of certain systems known to knowledge of biomedicine. Clinical findings are grouped into Patterns of disharmonies to identify the underlying changes, often with interactions evaluated between different patterns as well as with a patient’s constitution.

Interestingly, although a later development than acupuncture, herbal medicine now has the dominant role in Chinese medicine. Whereas in the West, acupuncture has been the subject of intensive research to explore its scientific basis and to seek evidences of clinical efficacy with a plethora of literature enhancing its practice.

Historically around the end of nineteenth century, Chinese medicine as a whole suffered serious decline when medicine in the West established itself with developments in anti-bacterium, in pharmacology and surgery techniques. It was after 1950’s when
CM was lifted politically and strategically to provide health care where OM was too expensive or unavailable. Since then, but not without debate and controversy, however, CM has been gradually integrated into health care system pragmatically at the level of practice in China. Presently, CM is practiced as a specialty in hospitals in China, with practitioners trained and qualified in both CM and OM, and patients exercise choice to seek help from CM.

**Complementary and Alternative Medicine (CAM)**

The terms complementary and alternative, or non-conventional medicine are used interchangeably with traditional medicine in some countries. The concept seemingly covers disparate modalities under the board umbrella term of CAM and adds to the difficulty in defining what exactly CAM is. WHO (2000) defined CAM as:

> 'the knowledge, skills and practices based on the theories, beliefs and experiences indigenous to different cultures, whether explicable or not, used in the maintenance of health, as well as in the prevention, diagnosis, improvement or treatment of physical and mental illnesses.'

The above describes a set of characteristics of CAM, born out by other literatures principally consistent with the view, putting the key features as holism, vitalism, individualised care, self-healing and wellbeing maintenance while seeking and comprehensive treatment and management (Hirschkorn and Bourgeault 2006). However, what constitute CAM, as Ernst and his colleagues (2001) pointed out, varies greatly in the context of cultural differences and individual viewpoints.

Acupuncture distinctively embedded its classic theory in inquiring on into the human body and the mutual relationships with the internal and external environment so as to gain health improvement through the treatment and prevention of diseases (Kaptchuk 2000).

**Orthodox medicine (OM)**

Orthodox medicine refers to conventional or mainstream biomedicine in this thesis. It is also called the Western Medicine in China. Although Chinese medicine was the mainstream medicine for over two thousand years of Chinese history, the biomedicine became the dominate position in last century.
1.4.2 Brief history of acupuncture

1.4.2.1 Acupuncture in Chinese medicine

The long history of acupuncture in CM is attested by innumerable literature sources in China. They are unknown to wide audiences in the West because very little has been translated into other languages. The milestones in the development of acupuncture can be followed in the ancient literature wherein an increasing number of acupuncture points were recorded (Xiao and Mu 2000).

The very first historical record of acupuncture is rather an archaeological finding which dated as late 300BC to early 200 BC. The ‘Ma Wang Dui Script’ included text of Chinese philosophical and medical works written on silk. It contained the meridians with no points (White and Ernst 2004).

The earliest major source of acupuncture's theory is the Huang Di Nei Jing (Yellow Emperor's Inner Classic), written in 475-221 BC. There were 25 single points, 135 bilateral points. The unified and comprehensive theories of points and channels were presented in Nan Jing (Classic of Difficult Issues, 220 AD).

The Zhen Jiu Jia Yi Jing (Comprehensive Manual of Acupuncture and Moxibustion), by Huang-Pu Mi (256-262 AD) is the oldest existing classical text devoted entirely to acupuncture and moxibustion. There were 49 single points, 300 bilateral points, and total 349 points (Xiao and Mu 2000).

The Zhen Jiu Da Cheng (Great Compendium of Acupuncture and Moxibustion) by Yang Ji Zhou, was published in 1601. It synthesized many classical texts as well as unwritten traditions of practice, and became the most influential medical text for later generations in Asia and Europe. The book was recognised as the primary source translated into different languages before 20th century. There were 51 single points, 308 bilateral points, total is 359 points (Helms 1998). The Zhen Jiu Feng Yuan was published in 1817 which included 361 point for fourteen meridians that have been used till present (Xiao and Mu 2000).

The theoretical root of acupuncture practice is based on the foundational theories for CM, such as Yin Yang, Five Elements, etc.. Acupuncture theory employs the circulatory system of ‘channels’ or ‘meridians’ within the whole body which are interconnected, functioning like pathways for vital energy of ‘Qi’ (pronounced as ch’i) and nutrition (termed as Blood) to flow around body in a circadian rhythm.
Acupuncture points are where Qi rises to the body surface, thus capable of conveying internal pathologies to external presentation as well as receiving external stimuli to influence status of Qi and Blood. Insertion of needles into these points produces effects that correct the disharmonised channel flow, and thereby the state of health is improved.

Acupuncture has been used for various diseases on its own or combined with other therapeutic methods. Acupuncture analgesia attracted intensive attention in the 1970s. The subsequent research efforts produced evidence and suggested that acupuncture effects was linked to the central nervous system activities of endogenous opioid peptides and biogenic amines which attracted attentions from biomedicine (Kaptchuk 2002).

1.4.2.2 Acupuncture in orthodox medicine

Acupuncture was introduced to Europeans in the 17th century and the first record dates back to 1671 (Helms 1995, 1998; Baldry 2005). A number of articles and books appeared in the 19th century and they were outlined in the book from Birch and Felt (1999).

Clinical acupuncture has been developed within biomedical practices. The most influential impact on the development of 20th century European acupuncture was the work of George Soulié de Morant, who translated and introduced acupuncture to French and European medical community (Helms 1998). Acupuncture was generally accepted in France and Germany (Baldry 2005). In USA, a major step forward in the acceptance of acupuncture by Western medicine occurred with the founding of the American Academy of Medical Acupuncture in 1987. The AAMA represented the physician-based acupuncture society in North America (Helms 1998).

In the UK, an early movement of integration of acupuncture emerged in 19th century in clinical practice (Helms 1995; Bivins 2000). The British Medial Acupuncture Society was established in 1980, which included practitioners from the physician and nursing professions (Baldry 2005).

In the last two decades, acupuncture practice has grown rapidly in the UK both in the NHS and in the private sector. The practice has enormous diversity and frequently the ‘traditional’ and ‘western’ approaches are the most applied. However, the current acupuncture practice has inherited influences from both paradigms. Dale (1997b) asserts a complex interface between biomedicine and Chinese medicine. A national survey by Dale (1997a) to look at how acupuncture is practised in the UK confirmed differences
between perceived ‘western’ and ‘traditional’ style of acupuncture, however, it also highlighted the ‘permeable boundaries’ in practice and showed that most practitioners practiced both OM and CM styles.

Trigger points had been first described by Travell and Rinzler in 1946 (Birch 2003). They are characterised by reactive points on palpation or stimulation which would trigger local and referred pain. The stimulation such as stretch, needle insertion or injection with saline were used for treatment (Travell and Simons 1992). Conceptually, trigger points are a good match to the specific class of acupuncture point-Ashi points, rather than meridian points or extra points (Birch 2008).

Segmental acupuncture is defined as a method of point selection by western medical practitioners (Filshie and Cummings 1999). Its treatments are based on charts of dematomes, myotomes, sclerotomes and viscerotomes. In this method, sensory stimulation is applied to the relevant spinal segment, restoring function to the muscles and autonomically innervated organs in that segment (Lewis and Halvorsen 2003). The trigger point and segmental point theories and practices demonstrated orthodox medicine extends theories of neurophysiology knowledge into acupuncture practice.

1.4.3 Current practices and regulations

The current practices have different approaches, either integrated or ignoring one or the other. The integration between CM and OM has been debated over a few decades and will remain a controversial subject for some time. Giarelli and Tognetti (2006) identified two main models of strategic integration: the American style market-driven integration and the European style state-driven integration. In the first model, the health services are integrated by demand led process. The other model is driven by the state to integrate the CAM into the mainstream OM systems.

There are different types of acupuncture practice under the integration model. The most used and researched is conventional acupuncture which relies on dry needling and classic acupuncture points. It involves penetrating the skin with thin solid metal needles into specific locations (acupuncture points) on the body, and then needles are manipulated by hands or by electrical stimulation. There are varieties of acupuncture in practice. The researcher categorised them based on observation of practice as follows:
• By theory behind the practice: Five elements, Yin Yang, Meridian, Pattern, nerve stimulation, holograph theory etc.
• By selection in specific body part and holograph theory:
  Auriculo-acupuncture, head acupuncture, hand acupuncture, foot acupuncture, etc.
• By equipment or material used for acupuncture: dry needling acupuncture, electro-acupuncture, laser acupuncture, TENS (Transcutaneous Electrical Nerve Stimulation), injection of sterile water, procaine, or vitamins at an acupuncture point, clubs (plum) acupuncture, fire needling, etc.
• By methods: single needle, implant needle, injection of substances
• By acupuncture practice in different style, often reflecting culture influences, e.g. China, Japan, Korean, France and other counties.

In this research, our focus is the classic acupuncture points and dry needling acupuncture in both Chinese and Orthodox traditions.

In the global context of acupuncture practice, there are different degrees of legislation in different regions and countries.

Acupuncture as part of Chinese medicine was legalised in China in the 50s’ and mainly delivered in hospitals. Most Chinese medicine practitioners have training in both biomedicine and traditional medicine. There are over 2000 traditional Chinese medicine hospitals at provincial level and acupuncture is on offer in most general hospitals with 38 million inpatients are being treated with CM every year (Robinson 2006), where different degrees of integration are seen in the delivery of healthcare.

In the UK, the report from the House of Lords Select Committee on Science and Technology (2000) on CAM was particularly concerned with regulation on public safety of acupuncture and herbal medicine. This led to a response from the Department of Health (2004) who launched a consultation document on the statutory regulation of herbal medicine and acupuncture practitioners. It is intended to define appropriate mechanism for regulation of acupuncturists and herbalists, in discussion with their professional bodies, both medical and non-medical.

In Europe, the CAM therapies are widely used in Germany and France (Ernst 2000). Both countries only allow acupuncture to be practised by medical doctors.
Belgium started the process of legalising acupuncture in 1999 in European Traditional Chinese Medicine Association (ETCMA) report. Netherlands and Portugal also moved towards legislation of acupuncture as reported by European Council for Classical Homeopathy and the European Shiatsu Federation in 2004.

In 2002, Italy's national medical association, the National Federation Doctors and Dentists can follow the 'Guidelines on Non-conventional Medicines and Practices' in their practice (Sarsina 2007).

In the US, acupuncture moved toward medical legitimacy in 1996 after the U.S. Food and Drug Administration reclassified acupuncture needles from the Class III or 'investigational' category to the Class IIb, the category that means 'safe, effective, but with special restrictions' (FDA 1996).

In Canada (Boon 2002), the harmonization of regulations and scopes of practice for CAM practitioners across Canada is seen as one of the biggest future challenges. Currently, acupuncture is regulated in some provinces, but not in others.

1.5 Requirements for knowledge representation

In this chapter, the motivation and requirements of the study have been explained. The challenges are discussed in the next chapter. The requirements for clinical practice, and research are discussed from the perspective of knowledge sharing. The requirements for research into the interface between CM and OM are restated along with discussion of two key objections.

1.5.1 Requirement for clinical acupuncture practice

Acupuncture is seen as a relatively safe and minimally invasive procedure, and one which can specifically benefit patients with certain conditions. A survey of 9408 acupuncture patients in the UK by MacPherson and his colleagues (2006) showed musculoskeletal problems to be the main conditions for the treatment. Acute and chronic musculoskeletal conditions for acupuncture treatments have found greatest acceptance in the USA (Siepina and Frenkel 2005). Some studies from randomized controlled trials have suggested acupuncture is effective in treating conditions such as osteoarthritis, headache and migraine (Kwon et al. 2006; Berman et al. 2004; Vickers et al. 2004). There is increasing CAM use and its popularity is growing in all regions of the world (WHO 2003; Young and Pennick 2007; Trinh et al. 2006; Vincent 1989).
Acupuncture practiced by conventional health care professionals in the UK has largely dispensed with CM theories and concepts, while some practitioners can entirely base treatment on trigger points, some use a mixture of OM and CM approaches (Vickers and Zollman 1999; Colin Lewis 2003) and their treatments ‘tend to be more formulaic’ (White 2006).

The trend to share knowledge and integrate practice between the two camps has been suggested by a few surveys. Patients expected that their doctors to have knowledge of CAM or be able to refer to CAM when appropriate (Ben-Arye et al. 2008). Furthermore, it can be argued that the many patients who receive acupuncture treatment from health care practitioners deserve comprehensive information so as to inform their choice. Increasingly, physicians agree that some CAM therapies hold promise for the treatment of symptoms or diseases but need better designed research for evidence, and some physicians show willing to learn and some already provide CAM therapy in their practice (Penman et al. 2005; Hsiao et al. 2005; Wahner-Roedler et al. 2006).

It is suggested that over 30 per cent of GPs (General Practitioners) in England referred their patients to CAM and acupuncture is one of the most common therapies (Thomas et al. 2001). Dual-trained practitioners, such as physician acupuncturists, advocated an open-minded perspective to practitioners of other paradigms and treated patients with both CAM and OM (Hsiao et al. 2005).

However, varied techniques, theories and styles in practice have always existed within acupuncture with many of the details of acupuncture practice differing remarkably. Hence, the challenge for knowledge and information sharing within the diversity of acupuncture theories and practices is significant.

There is a need that the information regarding the treatment and conditions can be effectively shared by other health care professionals and among the acupuncture practitioners. It is not easy for other health care professionals to make judgment of what acupuncture practices and prescriptions were useful or not and for what conditions. It is even harder to make informed decision for, or even by the patients. The formalisation of acupuncture knowledge representation will provide the potential platform for clear semantics in communication. The following is a scenario to illustrate the situation.

Some practitioners use the Chinese style based on classic theories such as yin and yang, patterns and meridians, while others rely purely on Western neurophysiology. Needles may be inserted deeply or just pierce the skin for a period of 30 seconds or 30 minutes with or without manual or electricity stimulation. When a GP considers referral
of a patient for acupuncture, the information from reviews on MEDLINE may not be sufficient to recognise what type of acupuncture treatment had been given, e.g. prescription, methods of stimulation, etc.. Moreover, it is difficult to know if a local practitioner can provide the same treatment. Further, dissemination of good practice and sharing knowledge with other professions or practitioners is desirable but difficult. A shared representation of acupuncture knowledge will enable OM practitioners to recognise the diversity of acupuncture approaches when making a patient referral.

1.5.2 Requirement for academic research

Scientific research on acupuncture has grown since the second half of the 20th century. Despite that, the value of acupuncture remains inconclusive; many diverse theories and differences in evidence acupuncture effects have been suggested and are still debated in terms of scientific approach and validity.

There are a number of different objectives when evaluating traditional medicine through clinical research, as when using clinical research to evaluate conventional medicine (WHO 2000) such as:

- evaluate traditional medicine in its own theoretical framework
- evaluate traditional medicine in the theoretical framework of conventional medicine
- compare the efficacy of different systems of traditional medicine and/or conventional medicine
- compare the efficacy of different traditional practices within a system of traditional medicine

Indeed, researches into acupuncture have yielded many theories and explanations as to the nature, effects and mechanism of acupuncture. OM knowledge and modern science and technology were widely deployed to explore the biological effects of acupuncture. It is still not clear how it works though a few theories and hypotheses have been suggested, such as Gate theory, electronic-magnet theory, nerve and lymph system theory, etc. (Sims 1997; Filshie and Cummings 1999). It is reported that the biological effects of acupuncture included the variety of neurohumoural factors and growth control factors (White 2006; Shang 2007).

In light of the technology advancement, we can further explore a phenomenon and its theoretic or observational assumptions. For example, acupuncture treatment
effect investigated by functional MRI scan of human brain (Cho et al. 1998; Wu et al. 1999). Their research suggested it was possible that acupuncture stimulated or activated the corresponding brain cortex via the CNS, thereby controlling the chemical or hormone release to achieve the treatment or acupuncture analgesia. Kim (2006) argued in his ‘beyond paradigm’ that the scientific translation of acupuncture ‘is not a direct theoretical correspondence between acupuncture and science’. It should be ‘an active and emergent production in the exchange of two cultures, not a reproduction from one culture to another’.

Most reviews report that trials of acupuncture efficacy are equivocal or contradictory. Kaptchuk (2002) asserted that lack of knowledge by researchers may aggravate the problem with evaluating the efficacy of acupuncture.

The complexity in clinical testing of non-pharmacological intervention modalities like acupuncture has been a heated debate. For example, key debates are related to the use of valid sham acupuncture as control group; embedding appropriate research methodology for allowance for the emergent nature of acupuncture diagnosis and treatment; the heterogeneity of diverse acupuncture styles and methods in practice (Kaptchuk 2002; Birch 2004).

1.5.3 Requirements for knowledge sharing

Arguably, a key obstacle to making progress about ascertaining the value of acupuncture is a lack of understanding at a conceptual level. This is an acute problem because the two paradigms’ representation of knowledge makes it difficult to share and to harmonize. There are ethical, as well as practical reasons to overcome the objections that are raised against the harmonisation of knowledge representation for acupuncture (Gao and Kay 2005):

*Different standards already exist for CM and OM, therefore there is no need to represent both paradigms.*

This objection suggests that both systems can co-exist independently, but in today’s global, multi-cultural environment, the two will inevitably meet. Both systems of medicine have the ultimate goal of delivering effective patient care, so it becomes an ethical imperative to find out which works best for patients’ benefit. To understand the difference of concepts and terms in each tradition will enable us to conduct relevant
studies to obtain the proper evidence for better clinical practice. More importantly, clinicians need to be aware different paradigms in health care that may benefit their patients, in particularly, understand when and equally important when not to apply them.

*CM and OM are too alien, e.g. 'chalk and cheese', it is not possible to represent both views in a sensible means.*

The difficulties of communication between CM and OM is partly because the differences in concept, theories and philosophies. Lack of knowledge is recognised as one of the greatest barriers to appropriate use of alternative medicine, and as Berman (2001) argues it is important to present the therapies in the context of their own philosophies and models of health and illness. It may not, however, be a simple choice between one or the other, but the context may require some fusion of both.

Clinicians have to deal with heterogeneous knowledge derived from many resources and multiple paradigms. It is the challenge for clinical informatics to investigate and provide solutions to help clinicians to deal with overloaded information.

### 1.6 A guide to the thesis

**CHAPTER 1** introduces the thesis. It presents the research aim and objectives, the contributions of this research, and the motivations and requirements for it. The chapter strongly supports the need for this thesis.

**CHAPTER 2** reviews research background on acupuncture in CM and OM, and ontology in health informatics. The terms and concepts used within this thesis are also clarified. The challenges were discussed for constructing a single ontology across paradigms.

**CHAPTER 3** presents the theoretical and practical principles guiding this research. It examines the assumptions underpinning the research paradigm and methodology.

**CHAPTER 4** discusses ontology development regarding tools, languages, methodology.
CHAPTER 5 presents the acupuncture ontology, a major outcome of this thesis. Ontology is presented and discussed in context of CM and OM.

CHAPTER 6 presents critical evaluations validate acupuncture ontology and data collection and analysis.

CHAPTER 7 completes the thesis by presenting a summary and highlights the scope for future research.

1.7 Summary

This chapter has suggested that there is a need for the research described in this thesis – the development of an acupuncture ontology, especially the integrated representation which is under-investigated, and remains to be tested.

This chapter has introduced the thesis, with the aim, objectives, and contributions. A brief introduction on acupuncture and its relations to CM and OM was presented. The question as to 'why' the aim and objectives were of importance was addressed. We outlined the importance for clinical practice and research, in particular, with the focus of knowledge sharing across paradigms.

Finally, this chapter outlined a road map to the entire thesis. Chapter 2 explores the literature relevant to this thesis, presents the detail on background of ontology and challenges for representation.
Chapter 2

Research Background

The aim of this chapter is to review the relevant background to this research. First we consider current developments in ontology and review existing medical ontologies. Next we develop a more specific focus concerned with the challenges of representing acupuncture knowledge. An important purpose of this chapter is to introduce the concepts and terms used throughout the remainder of the thesis.

2.1 Introduction

The goal of the research is to construct an ontology; more specifically a formal knowledge representation of aspects of acupuncture. This chapter is intended to answer the following questions as a starting point for our study:

- What is an ontology?
- How can we develop an ontology?
- What systems, languages, and tools are available for ontology development?
- What ontologies have been developed in medicine?
- What are the challenges for acupuncture knowledge representation for Chinese medicine (CM) and orthodox medicine (OM)?

The following sections address these questions in turn.

2.2 Ontology

2.2.1 Ontology – A philosophical perspective

Ontology was born out of Philosophy. The term ‘ontology’ (or ontologia) was coined in 1613, independently, by two philosophers, Rudolf Göckel (Goclenius), in his Lexicon philosophicum, and Jacob Lorhard (Lorhardus) and it first appeared in Bailey’s
dictionary of 1721 in English, which defines ontology as 'an account of being in the abstract' (Smith 2001). Regardless of its name, what we now refer to as 'philosophical ontology' has sought the definitive and exhaustive classification of entities in all spheres of being; for philosophy, ontology is the 'branch of metaphysics that concerns itself with what exists' (Blackburn 1996, p.269).

It was Aristotle who is credited with asking the question 'what is being?'. He answered it by introducing categories of being. 'Being', in the study of ontology, is Aristotle's 'being que being' that anything can be said to be. However, as it is the main argument in metaphysics, 'being' is complex and still a debatable concept of what exists in reality. Predicate is also from Aristotles' Greek. Category and predicate have slight different meaning in English, as Sowa (2000a, p98) explained 'The word predicate is closer to Aristotle's meaning of what may be said of something. The word category is associated with the process of classifying entities according to some monadic predicate'.

Categories of being in philosophy have been developed but few agree with each other. The categories are incomplete; indeed they may never be complete, as Smith and Brogaard (2003) states:

'...more than 2000 years philosophy research of ontology formed a few different categories of being from philosophic grounds. Although little agreement exists about what the categories are or how one could decide among competing systems, much can be learned and applied to ontology study in informatics.'

This paves the way for further exploration. As different categories have been proposed, the nature of being varies with different philosophical theories of universals and these have profound impact on ontological distinctions and categories. It was argued by (Cocchiarella 1996, 2003) that it is important to recognise conceptual
realism and its relations to conceptualism and realism, which are showed in figure 2.1.

![Figure 2.1 Conceptual Realism](image)

The key terms (with gray back ground) are briefly explained:

**Realism:** There are real universals, i.e., universals in reality, which are predicated of things, and the function of predication in language is to represent predication in reality.

Aristotle's approach to categories is of categorical realism (Carr 1987). As Ackrill (1963) writes that Aristotle's *Categories* 'are not primarily or explicitly about names, but about the things that names signify.' They are categories of entities instead of linguistic names. Language was just a cue to the truth about the world. In the modern realism approach, as Gochet (2002) argues, reality is no longer to be sought in language, but in the theories about the world which scientists hold to be true, and language is only needed to formulate them.

**Conceptualism:** There are conceptual universals. The predicable concepts underlie predication in thought which is base for predication in language.

Kant shifted from realism approach to categorical conceptualism (Carr 1987). Kant treated concepts as the route to categories of objects of possible cognition. As Paton (1936) puts it, for Kant ‘we can have *a priori* knowledge by means of categories, only if the categories are due to the nature of the mind and are imposed by the mind on the objects which it knows’.

**Conceptual realism:** A system contains both conceptual intentional and conceptual natural realism. It 'provides a theory of the nature of predication in reality through an analogical theory of properties and relations' (Cocchiarella 1996). As Cocchiarella argues ‘Conceptualism without any associated form of realism, is at best only a truncated ontology - a socio-biologically based theory of the generic capacity
humans have for language and thought' and natural realism need some form of conceptualism to explain how concepts and languages can be formed in theories and description of the world.

In contrast to Aristotle and Kant who promoted ‘dualistic’ thinking, Peirce proposed ‘triadic’ and ‘anti-inductive’ ways of reasoning, first practiced by experimental methods and evolutive sciences (Deledalle 2000). Peirce’s writings are pervaded by triadic divisions, which he expressed most basically in numerical form as Firstness, Secondness, and Thirdness (Halton 1995, p.169). Peirce’s three categories arise from mathematical ideas about the reducibility of n-adic relations. According to Peirce, triads are necessary because genuinely triadic relations cannot be completely analysed in terms of monadic or dyadic predicates, yet larger polyadic relations can be analysed in terms of triadic and lower relations.

Peirce’s theory went beyond conceptualism and realism. ‘The opposition between nominalism and realism has no longer any sense since pragmatism defined the idea as what it does. The mind is in the world and in continuity with it. The law is a natural as well as a logical process’ (Deledalle 2000, p.13).

2.2.2 Ontology - The semiotic and linguistic perspective
In our study, the purpose of ontology is in knowledge representation. Guarino (1995) emphasises the necessity of a strong interdisciplinary perspective within the knowledge representation community. Disciplines like philosophy and linguistics can offer a concrete contribution to the practice of knowledge engineering. How to represent knowledge faithfully and provide clear semantics for interpretation is directly relevant to our study.

Semiotics is a general study of the use of sign systems. It investigates the types of relationship that hold between a sign and the object it represents (Saeed 2003, p5). There are two major traditions in semiotics: Saussure’s semiology and Peirce’s semiotics.

Saussure's approach was a generalization of formal structuralist linguistics. Saussure used the semiology as a science which studies the life of signs at the heart of social life and this science could teach us ‘what signs consist of, what laws govern them’ (Saussure 1971, p.33). The sign and the object it represents are discussed in term of signifier and signified.
Peirce grounded his logic in a systematic account of meaning and reference. His semiotics provided a novel framework for the understanding of language, thought and all other kinds of representation. It is an extension of reasoning and logic in the natural sciences; a broader view that includes every aspect of language and logic within three branches of semiotics (Sowa 2000b):

1. Syntax is the formal relation of signs to each other.
2. Semantics is the formal science of the conditions of the truth of representations. The study relates signs to things in the world.
3. Pragmatics relates signs to interpreters, agents who use them to refer to things in the world and to communicate their intentions about those things to other agents who may have similar or different intentions concerning the same or different things.

Two particularly important semantic theories are the referential (or denotational) approach and the representational approach (Saeed 2003, p24). In the referential approach, meaning is the relation between words and the world. However, there must be more meaning than simply denotation. A traditional philosophical example is that of the morning star and the evening star. According to Frege's distinction between sense and reference (Frege 1948), the morning star and the evening star have different senses to a speaker but they have the same reference to planet Venus. Sense places a new level between words and the world, a level of mental imaging. The representational approach emphasises the link between language and conceptual structure but the referential (or denotational) approach stresses the relation between language and external reality.

Pragmatic relationships are explored in the meaning triangle of symbol, concept and object (Ogden and Richards 1923) as given in figure 2.2.

![Figure 2.2 the meaning triangle (Ogden and Richards 1923)](image_url)
From the perspective of semiotics, ontologies have to deal with three branches of semiotics: ontology has a syntax, no matter what language; ontology seeks to codify the world; and ontology can only be recognised by an interpreter who has committed to it.

2.2.3 Ontology – The informatics perspective

The study of ontology has attracted many different disciplines; Philosophy, Artificial Intelligence, Knowledge representation and knowledge sharing, and Informatics. In Health Informatics, ontologies are used to represent general or domain knowledge in information systems. The ontological arguments in philosophy discussed above cross over into informatics, resulting in different interpretations and implementations.

2.2.3.1 Vocabulary view of ontology

Although language is the cue for Aristotle to talk about reality, vocabulary as a unit of language is an essential part of ontology in knowledge representation. Ontology is a formal method to support development and alignment of terminology which is a system that organises words and phrases for a community.

It was identified that barriers for knowledge sharing ‘arise from lack of consensus across knowledge bases on vocabulary and semantic interpretations in domain models’ (Neches et al. 1991). They proposed the definition that

‘an ontology defines the basic terms and relations comprising the vocabulary of a topic area as well as the rules for combining terms and relations to define extensions to the vocabulary’

To support knowledge sharing and formal knowledge representation, ontology provides the common terms and relations for some domain or subject matter (Chandrasekaran et al. 1999). The emphasis on vocabulary can also by found by Gruber’s definition that ‘A specification of a representational vocabulary for a shared domain of discourse - definitions of classes, relations, functions, and other objects - is called an ontology’ (Gruber 1993a).

It was argued that the medical vocabulary used in clinical information systems must be more than a simple list of terms and such a vocabulary must have synonymy, domain completeness, and multiple classifications, providing consistent views and explicit relationships, while remaining unambiguous and with no redundancy (Cimino et al. 1989). Most related efforts and activities in health informatics are directed towards
the generation of controlled terminologies, or reference ontologies (Gennari and Oliver 1995; Musen et al. 1995; Oliver 1998; Beissbarth 2006). The terminologies or ontologies developed seek to support consistent usage of terms and to enable information sharing and system cooperation. The importance and challenge to create and maintain a controlled vocabulary have been well recognised (Sitting 1994; Musen et al. 1995; Hammond 1997; Rector 1999).

2.2.3.2 Conceptualisation view of ontology

The most quoted definition of ontology is (Gruber 1993a) ‘an ontology is an explicit specification of a conceptualisation’. As he explains

‘For knowledge-based systems, what “exists” is exactly that which can be represented. When the knowledge of a domain is represented in a declarative formalism, the set of objects that can be represented is called the universe of discourse. This set of objects, and the describable relationships among them, are reflected in the representational vocabulary with which a knowledge-based program represents knowledge. (p.200)

This conceptualisation view has gained broad acceptance within informatics although refinements to the definition have been proposed. It was refined by Studer and his colleagues (1998) that an ontology is ‘a formal, explicit specification of a shared conceptualisation’. Their ‘conceptualisation’ refers to an abstract model of some phenomenon in the world, having identified the relevant concepts of that phenomenon. ‘Explicit’ means that the type of concepts used, and the constraints on their use are explicitly defined.

Guarino (1998) goes further and states that ‘an ontology is a logical theory accounting for the intended meaning of a formal vocabulary’. He argues that an ontology is concerned the intended models which reflects the commitment to a particular conceptualisation of the world. Guarino explains the conceptualisation in ontology is different from extensional relations which reflect a particular state of affairs.

Although Cimino did not mention the term of ontology in his desiderata (Cimino 1998), there was a shift towards conceptualisation through the call for “concept orientation”. In addition, Cimino (2006) argued that the

‘realist approach is not away from concepts, but towards recognition that concepts and universals must both be embraced and can coexist peaceably in controlled terminologies. To that end, additional Desiderata are defined that
deal with the purpose, rather than the structure, of controlled medical terminologies'.

The link between concept and vocabulary is that conceptualisation is an activity that ontology developers structure the domain knowledge in a conceptual model which describes the problem and its solution in terms of the domain vocabulary (Fernández-López et al. 1997). The advantage is that an ontology concerns concepts instead of vocabularies of natural language therefore 'avoid the problems such as homonymy, metonymy, polysemy, etc.' (Bézivin 1998).

2.2.3.3 The reality view of ontology

The views of vocabulary and conceptualisation presented in previous sections differ from the Aristotellean meaning of ontology, which centres on reality. Ontology is described by (Smith 2001; Smith and Brogaard 2003; Welty 2003) in terms of realism, saying that

'ontology is the science of what is, of the kinds and structures of objects, properties, events, processes, and relations in every area of reality'

Ontology in realism is not a software implementation or a controlled vocabulary; rather it is a theory of reality. This differs from the vocabulary and conceptualisation views of ontology, such as 'an explicit specification of a conceptualisation' (Gruber 1993a). As Smith (2004) asserted that ontology is not about concepts as linguistic or computer artefacts but rather of universals. Universals signify their corresponding instances i.e., a symbiotic relationship: the one cannot exist without the other. In addition, Smith (2006) criticised the concept-based approach, arguing that it is wrong to anchor terms in terminologies to referents in reality. He proposed a realist approach by saying that the goal of alignment is not on concepts but on universals categories in reality and their corresponding instances.

In the bioinformatics field, it is argued that 'most bio-informaticians work only with terms and concepts, they cannot altogether disregard the question whether these terms and concepts have any real referents' (Johansson 2006). However, we need be aware of the revolution of the knowledge and theories of reality. As Johansson states that “the history of science tells us that it is no longer possible to believe that science progresses by adding one bit of truth to another in the way brick houses are built by laying bricks on top of each other.' Therefore, it is unavoidably that the new conceptualisation will be introduced after the theories of reality have been revised.
2.2.3.4 The semiotic view of ontology

The Semiotic view emphasises the relationships among vocabulary, conceptualisation and reality. Although we have discussed the three views previously, their distinction and interrelationship can be further analysed and described by semiotics.

Andersen (1991) in his paper *Computer Semiotics* argues that semiotics is a global perspective on computer systems and “semiotic theory can only make scientific statements about phenomena that are used as signs to stand for something else for somebody”.

The definition of ontology by Sowa reflects Pierce’s semiotics of syntax, semantics and pragmatics. Sowa contends that

‘the subject of ontology is the study of the categories of things that exist or may exist in some domain. The product of such a study, called an ontology, is a category of the types of things that are assumed to exist in a domain of interest \( D \) from the perspective of a person who uses a language \( L \) for the purpose of talking about \( D \)’ (Sowa 2000a).

For him, semiotics provides guidelines for organising, and using signs to represent something to someone for some purpose. Although Gruber leads the conceptualisation approach, he recognises ‘pragmatically, a common ontology defines the vocabulary with which queries and assertions are exchanged among agents. Ontological commitments are agreements to use the shared vocabulary in a coherent and consistent manner’ (Gruber 1993a). The reality view of ontology holds by Smith and colleagues (2006) also draw a distinction of three levels in biomedical ontology: the level of reality; the level of cognitive representations of this reality; the level of textual and graphical artefacts.

In addition, Sowa (2000a) gave a statement that the purpose of the ontology is to provide a framework of distinctions that can be used to discriminate and classify things and define words.

To summarise the different views of ontology, the philosophical metaphysic view of ontology is to discover the fundamental categories of the world; the semiotic view offers contribution to how to represent knowledge faithfully and provides clear semantics for interpretation. In health informatics, we follow Sowa’s definition of ontology which supports three different views of vocabulary, conceptualisation and reality from a semiotics approach.
2.3 Principles and methodologies of ontology development

Manual construction of an ontology is a time-consuming task and involves domain experts and knowledge engineer working together. The most popular methodologies, languages and tools for constructing ontologies were reviewed by Corcho and colleagues (2003) and concluded that much more effort is still required to further the practice of ontology development. No matter what approach is taken, some proposed principles and criteria may apply, such as completeness, consistency, extendibility, reliability, reusability and expressiveness, etc. (Gruber 1993b; Uschold 1996; Rector and Rogers 2002).

2.3.1 Principles

Some principles and criteria have been proposed by different authors for ontology development. The following is a list of principles and criteria which is mainly based on Gruber’s design criteria for ontologies (Gruber 1993b).

Clarity: Definitions should be clear and objective to reflect the intended meaning of concepts. Ontology should restrict the possible interpretations and the definition can be stated in logical axioms or natural language. A complete definition (defined by necessary and sufficient conditions) is preferred over a partial definition (defined by only necessary or sufficient conditions).

Completeness: Principle of completeness of an ontology in terms of the partitions (disjoint and exhaustive) defined between its classes. An exhaustive subclass partition adds a completeness constraint to the represented subclasses.

Three types of incompleteness have been identified: incompleteness of concept classification: classes are not declared disjoint while they are disjoint in reality; classes form an exhaustive partition but this is not defined in the ontology (Gómez-Pérez and Benjamins 1999).

Consistency and Coherence: The defining axioms should be logically consistent (Gruber 1993b) and permit inferences that are consistent with the definitions. It is possibility to check for contradictory knowledge from valid definitions.
Extendibility: ontology should offer a conceptual foundation for a range of anticipated tasks and the representation so that it can be extended and specialised (Gruber 1993b; Uschold and Gruninger 1996).

Minimal encoding bias: The conceptualisation should be specified at the knowledge level without depending on a particular symbol-level encoding (Gruber 1993b).

Minimal ontological commitment: An ontology should require the minimal ontological commitment sufficient to support the intended knowledge sharing activities. Ontology should make as few claims as possible about the world being modelled, allowing the parties committed to the ontology freedom to specialise and instantiate the ontology as needed (Gruber 1993b).

Explicitness and orthogonal taxonomies: it is possible to classify any concept on any of its properties independently – e.g. it should be possible to re-arrange the ontology along any axis. Pure hierarchy means making forced choices which are not represented explicitly and the choices will be appropriate for one application but not another (Rector and Rogers 2002).

Reliability: Informal ontologies serve as a basis for manually checking the design against the specification. Using formal ontologies enables the use of semi-automated consistency checking of the software system with respect to the declarative specification (Uschold and Gruninger 1996).

Reusability: Ontology can be important export modules among different software systems. Ontologies provide an "easy to re-use" library of class objects for modelling problems and domains (Uschold and Gruninger 1996).

Logical approximations, labels of concept representations: Any logical model for knowledge representation is at best an approximation of the concepts represented as used in human language and thought (Rector and Rogers 2002).

Separation of categories and instances: Categories are abstract model of instances and categories can be specialised. Instances are individuals which can not be specialised (Rector and Rogers 2002).

Normalisation and canonical forms: reducing an expression to a standard or simplest form by the logical or pragmatic process (Rector and Rogers 2002).

Formality of concept and property definition The definition of concepts and properties can have different forms, from informal natural language to formal first order (Poli 2003; Roche 2003).
Apart from the principles and criteria identified by different authors. Some useful requirements and characteristics of ontology are also proposed by other authors, such as a list of basic, typical and desirable properties of an ontology proposed by McGuinness (2003)

'Basic properties for a simple ontology:

- Finite controlled (extensible) vocabulary
- Unambiguous interpretation of classes and term relationships
- Strict hierarchical subclass relationships between classes

The typical properties but not mandatory:

- Property specification on a per-class basis
- Individual inclusion in the ontology
- Value restriction specification on a per-class basis

The desirable properties but not mandatory nor typical:

- Specification of disjoint classes
- Specification of arbitrary logical relationships between terms
- Distinguished relationships such as inverse and part-whole' (McGuinnes 2003)

The characteristics of ontology can be identified by comparing with thesauri and schema. The differences between ontologies and database schema definitions are discussed by Fensel (2001):

- 'A language for defining ontologies is syntactically and semantically richer than common approaches for databases.
- The information that is defining is described by an ontology consists of semi-structured natural language texts and not tabular information.
- An ontology must be a shared and consensual terminology because it is used for information sharing and exchange.
- An ontology provides a domain theory and not the structure of a data container.' (Fensel 2001)

2.3.2 Methodologies

Although principles and characteristics of ontology have been proposed, few methodologies are broadly accepted to guide ontology development. Some reviews have
been conducted by (Jones et al. 1998; Fernández-López 1999; Pérez 2002; Corcho and his colleagues 2003). They compared methodologies for Cyc (Lenat and Guha 1990), ENTERPRISE (Uschold and King 1995), TOVE (Grüninger and Fox 1995), KACTUS (Schreiber et al. 1995), METHONTOLOGY (Fernández-López et al. 1997; Fernández-López et al. 1999), SENSUS (Swartout et al. 1997) and On-TO-Knowledge (Staab et al. 2001). Their conclusion is that the approaches are not mature and approaches used are not unified as different groups applying its own methods.

Although we do not have a single agreed methodology, the shared characteristics of different approaches may inform ontology development. There are three basic models for ontology development:

1. stage-based models, e.g. TOVE, ENTERPRISE,
2. evolving prototype models e.g. METHONTOLOGY
3. mixed models, e.g. On-To-Knowledge

The stage-based model is more appropriate for the development which purpose and requirements are clearly outlined. There are typically separated stages to produce first an informal description of the ontology, and then its formal representation in an ontology language. It is not necessary for evolving prototype model to have purpose and requirements defined at first stage. This model allows ontologists to go back to any stage if some definition is missed or wrong. The mixed model takes the advantages of above both models and has stages and circles as part of development process.

There are varied ways to start ontology development. It is common to start from scratch (Lenat and Guha 1990; Fernández-López 1999). In contrast to starting from scratch, existing ontologies and ontology libraries can provide a starting point for developing a new ontology (Swartout et al. 1997). Some methodologies take a task as a starting point (Grüninger and Fox 1995; Uschold and King 1995). The disadvantage of task-oriented development is the restriction on reusing and sharing ontologies.

There are three common strategies for identifying concepts from a hierarchy view for ontology construction (Uschold and Gruninger 1996):

- Bottom-up: from the most concrete bottom leaf nodes to the most abstract top root nodes
- Top-down: from the most abstract root nodes to the most concrete leaf nodes (top-down)
Middle-out: from the most relevant to the most abstract and most concrete (middle-out).

The middle-out has advantage of providing better stability and commonality than other two approaches (Uschold and Gruninger 1996).

Most ontologies are manually or semi-automatically constructed. Some techniques were proposed to enrich existing ontologies or even construct a new ontology automatically.

The automated approach can use a large collection of documents as the data source to enrich the ontologies. Such as, a study (Agirre et al. 2000) demonstrated retrieving documents from the internet and linking the topically related words for each concept in WordNet. Then the document cluster technique was applied to disambiguate the finest grained senses to enrich the WordNet. A similar study by Du and colleagues (2009) propose to extract the ontology semi-automatically from the web using their ontology extractor tool which assist ontology engineers to construct ontology.

The text-mining techniques can be applied to extract relevant concepts from existing documents and arrange the hierarchies according to detected relations among those concepts to enrich a domain ontology (Velardi and Missikoff 2001).

A large ontology (Suchanek et al. 2008) is automatically derived from Wikipedia and WordNet. They use DBpedia as a taxonomic backbone links the facts extracted from Wikipedia to the WordNet hierarchy.

A range of ontology development patterns have been proposed.

ENTERPRISE methodology emphasises on logical statement of competence question which as foundation for requirement and evaluation (Uschold and King 1995). It includes four steps for ontology development:

1. Identify the purpose and scope of the ontology;
2. Build the ontology by capturing, coding and integrating the knowledge with existing ontologies;
3. Evaluate the ontology;
4. Documentating and developing guidelines for each phase.

METHONTOLOGY includes a set of activities: scheduling, control, quality assurance, specification, knowledge acquisition, conceptualisation, integration, formalization, implementation, evaluation, maintenance, documentation, and
configuration management. The life cycle of ontology development is based on an evolving prototype (Grüninger and Fox 1995).

On-To-Knowledge methodology is mixture of stage based and evolving model approach. It includes five steps and two circles of development. The details of activities within each step are presented in the figure 2.3.

![Figure 2.3 On-To-Knowledge methodology (Staab et al. 2001, p.29)](image)

Apart from the above three types of methodology approach for ontology development, ontological analysis should be considered as part of the methodology. Different approaches to ontological analysis lead to different conceptualisations of reality. As Wonderweb (Claudio Masolo 2003) project report points out that ontology analysis is ‘the motivations and the constraints and that drive our conceptualisation of reality. It comes to no surprise that the design options for building foundational ontologies reflect the main categorical distinctions discussed in philosophy’. Ontology in informatics is different to its original meaning from philosophy. However, ontology is still constrained by its conceptualisation and underpinned philosophical view of the world.
Smith (1998) follows Husserl and insists three basic theories of formal ontology: the theory of part and whole, the theory of dependence, and the theory of boundary, continuity and contact.

OntoClean methodology (Guarino and Welty 2002; Guarino and Welty 2004) applied a meta-level approach for ontology analysis which draws upon analytic metaphysics. The formal meta-properties, such as essence and rigidity, identity and unity, are used to characterise relevant aspects of the intended meaning of classes, properties and relations in ontology. This analysis imposes constraints on the taxonomic structure of the ontology and helps in evaluating and validating the design decisions.

CYC and SUO general upper model and Sowa's ontologies use what Sowa calls the distinction approach (Noy and Hafner 1997). Sowa (2000a) describes Peirce's principle as a meta-level distinction (three-way distinction) for generating new categories by viewing entities from different perspectives. A category of Firstness (Independent) is determined by qualities inherent in something, Secondness (Relative) by a relation or reaction directed toward something else, and Thirdness (Mediating) by some mediation that brings multiple entities into the relationship. They can be represented differently as follows:

Firstness, Independent, Monadic predicate P(x);
Secondness, Relative, Dyadic relation R(x, y);
Thirdness, Mediating, Triadic relation M(x, y, z);

In addition to Three-way distinctions of Independent, Relative and Mediating, Two-way distinctions (or gradations) which are conceptual interpretations of the perceptual contrasts (Sowa 2000a). Sowa explained that all perception begins with contrasts: light – dark, up – down, hard – soft, sweet – sour. The ordinarily term of relations are abstractions from contrasts. Two-way distinctions describe a contrast along one axis or dimension. It can be expressed as either a discrete distinction or a continuous gradation (table 2.1).

<table>
<thead>
<tr>
<th>Two-way distinction</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrete distinction</td>
<td>Physical and abstract; Pathological and Physiological</td>
</tr>
<tr>
<td>Continuous gradation</td>
<td>Range of pressure; scale of pain; spectrum of colour</td>
</tr>
</tbody>
</table>

Table 2.1 Two-way distinctions
Sowa's Top-level ontology was developed on the basis of distinction approach as in figure 2.4 (Sowa 2000a). It demonstrates how two-way distinctions (Physical - Abstract, Continuant - Occurrent) and three-way distinction (Independent, Relative, Mediating) have been used for constructing top level categories.

Figure 2.4 Hierarchy of Sowa's Top-Level Categories (Sowa 2000a, p.72)

### 2.4 Systems, languages, and tools for ontology

The systems and languages are technical commitment for ontology development. The choices of the system and language impose the constraint on how ontology organise and represent entities and their relations, such as an universal/existential role quantification in DL languages or a simple constraint of slot value in Frame-based systems. In addition, the tool is an instance of systems and languages to facilitate the construction process for ontology.

Before exploring the systems, languages and tools, it is worth considering knowledge representation formalism proposed by Brachman (1979) and Guarino and colleagues (1994). Brachman proposed four levels to classify knowledge representation systems and languages by the primitives used. Guarino (1994) introduced the ontological level intermediated between the epistemological and the conceptual level as in following table.

34
At the logical level, the basic primitives are predicates and functions, which have given a formal semantics in terms of relations among objects of a domain. It allows formal interpretation of the primitives, but the interpretation remains arbitrary.

At the epistemological level, the primitives are general concepts introduced as a knowledge structuring primitive. The structuring choice is an implicit effect of any theory that may have cognitive and computational significance.

The ontological level is the level of meaning. At this level, knowledge primitives restrict the interpretation of a logical theory on the basis of formal ontology. The goal of specification is to restrict the number of possible interpretations and characterise the meaning of the basic ontological categories used to describe the domain (Guarino 1995).

Primitives, at the conceptual level, have a definite cognitive interpretation, corresponding to language-independent concepts.

At the linguistic level, it is about natural language and the primitives are verbs and nouns.

### 2.4.1 Systems for ontology

An understanding of formalism of knowledge representation systems can help to make choices of ontology development environment. Knowledge representation systems can be classified as logical or non-logical. Each has their own strengths and tradeoffs. Non-logical formalisms such as semantic networks, frame systems, were mainly based upon cognitive experiments with human beings and close to the representations in human brains (Sattler et al. 2003). In this section, we will only cover relevant semantic networks, frame systems and their descendent Description Logics (DL) systems. The
other representation formalisms, such as concept graph (Sowa 1984), rule based system, neuro network and other logic systems will not be discussed in this thesis.

The concept of frame-based system was proposed by Minsky (1975). In frame-based system, a frame represents a named data object or concept. A set of slots (attributes) attached to the frame represent properties or attributes of object or concept. Slots can have one or more values, some of which may point to other frames. Since each frame has a set of slots that represent its properties, frame systems are usually considered to be more structured than semantic networks.

Frames generally provide a rich set of language constructs but impose very restrictive constraints on how they can be combined or used to define a class. They only support the definition of primitive concepts, and the taxonomy must be hand-crafted (Stevens et al. 2000). When a domain model is particularly complex, frame-based representation requires techniques such as reified relations, meta-classes and meta-class hierarchy, different propagation patterns for template and own slots (Noy et al. 2002).

Frame-based systems are closely related to semantic networks. The origins of semantic networks lie in Aristotle’s associationism (behaviour is controlled totally by associations learned between concepts) and reductionism that concepts are built of more elementary concepts (Randal 1988). Quillian (1968) is generally acknowledged the first to apply semantic network ideas. His semantic memory is based on human associational model of human memory. The idea was to capture the ‘objective’ meanings of words which reflect the structure and capabilities of human memory in an encoding scheme for computation. Since then, a number of semantic networks have been developed. In semantic networks, concepts are represented by nodes and organized as in hierarchies in a graph. Arcs present links between concepts. The meaning of concept is implied by the links among the concepts.

The links in networks have being used to represent implementation pointers, logical relations, semantic relations, arbitrary conceptual and linguistic relations. Most historical network systems have mixed structure of two or more of these levels, which lead to the great difficulty for interpretation (Woods 1975; Brachman 1979). To overcome the inconsistencies and ambiguities of semantic network systems, Brachman developed KL-ONE systems in later 1970’s based on his PhD study. The system was called structured inheritance networks, which was an implementation about the structure of descriptions and their use in reasoning. The representation of KL-ONE is constructed out of epistemological primitives, e.g. ‘description’, ‘attribute’, ‘concept’, ‘role’,
'inheritance' and 'instantiation' (Brachman and Schmolze 1985). The following characteristics became the basis for development of description logics (Baader and Nutt 2002):

- The basic building blocks are concepts (unary predicates), roles (binary predicates), and individuals (constants).
- The distinction between description (a terminology, T-box) and assertion (assertions, A-box)
- A small set of epistemologically adequate constructors for building complex concepts and roles from atomic ones (undefined primitive concepts).
- Implicit knowledge about concepts and individuals can be inferred automatically by inference procedures.
- The taxonomy of defined concepts is automatically constructed by the logic reasoning system, whereas the taxonomy is introduced manually by the user in semantic networks.

The research in description logics systems is centred on decidability and complexity of the inferences against expressive power (Buchheit et al. 1993). The main benefit from Description Logics is sound and complete algorithms for the subsumption, equivalence, and satisfiability which can be processed by a DL reasoner. A few systems have been developed after KL-ONE system, such as KRYPTON (Brachman et al. 1985), LOOM (MacGregor 1991), CLASSIC (Borgida et al. 1989; Brachman et al. 1999), BACK (Peltason 1991), K-REP (Mays et al. 1991). The tableau-based algorithm that has been applied for highly optimised DL systems was employed in CRACK (Franconi 1998), KRIS (Baader et al. 1994). The most recent implementation of systems are FaCT (Horrocks 1998), RACE (Haarslev and Möller 2000; Haarslev and Möller 2001), and Pellet (Sirin et al. 2007).

Description Logics provides facilities to reason and manipulate knowledge bases. The architecture of a description logics system is show in figure 2.5.
Knowledge base constitutes T-Box and A-Box. T-Box, a terminology, is an ontology which describes classes and relationships. A-Box contains assertions of members of the T-box. Inferences about assertions are based on the T-Box. Ontology is formal explicit description of concepts in a domain of discourse (classes, sometimes called concepts), properties of each concept describing various features and attributes of it (properties, sometimes called roles or slots), and restrictions on slots (facets, sometimes called role restrictions).

2.4.2 Languages of ontology

Ontology languages can be classified as ‘traditional’ and ‘web-based’ languages (Corcho et al. 2003; Pulido et al. 2006). Each will be considered:

a) Traditional ontology languages

Traditional ontology languages are diverse and most are closely related to particular systems. Under knowledge sharing initiative, KIF (Knowledge Interchange Format) and OKBC (Open Knowledge Base Connectivity) were developed to facilitate interchange for different knowledge representation system.

KIF has declarative semantics and is based on the first-order predicate calculus. It also provides representation for meta knowledge and allows non-monotonic reasoning rules. It was used for frame based ontology (Genesereth and Fikes 1992).
OKBC specified a protocol for accessing frame-based representation systems for knowledge sharing (Chaudhri et al. 1997; Chaudhri et al. 1998). It provides a set of representational constructs commonly found in frame systems, e.g. constants, frames, slots, facets, classes, individuals and knowledge bases.

Logic languages for knowledge representation have long been intensively studied. Research and developments of Description Logics in last two decades have shed new light in this area. The first DL Attributive concept Language (AL) was proposed by (Schmidt-Schaub and Smolka 1991). DL languages can be differentiated by the additional constructors which are indicated by appending a corresponding letter. For example, the DL ALC stands for Attributive concept Language with Complements that indicates complement operator (\(\neg\)). A good introduction of DL can be found in Description Logic Handbook (Baader and Nutt 2002).

In the simplest format of DL AL language, concept (denoted by \(C, D\)) are built out of atomic concepts (denoted by \(A\)) and atomic roles (denoted by \(R\)) according to the syntax rule. The basic syntax lists as following:

\[
C, D \rightarrow A \mid \text{ (atomic concept)} \\
\top \mid \text{ (universal concept)} \\
\bot \mid \text{ (empty concept)} \\
C \cap D \mid \text{ (intersection)} \\
C \cup D \mid \text{ (union)} \\
\forall R.C \mid \text{ (universal role quantification)} \\
\exists R.C \mid \text{ (restricted existential role quantification)}
\]

b) Web based languages

Web based languages for ontology are XML-based mark-up languages, e.g. XML, RDF, RDFS, DAML+OIL, SHOE, XOL, OWL. The relationships between languages are presented as layers in figure 2.6 (Arroyo et al. 2004).
XML (eXtended Markup Language) was developed under World Wide Web Consortium (W3C) in 1996. XML provides mechanisms for defining document structure and content. The parsed data is made up of character data for content and mark-up for encoding the description of the storage layout and logical structure. XML supports user-defined tag to describe the structure of documents. Currently, XML is becoming the default language for data exchange for different systems.

RDF (Resource Description Framework) is a language based on XML for representing information about resources in the World Wide Web. It is particularly intended for representing metadata about web resources or any thing can be identified by Uniform Resource Identifiers (URIs). RDF model present statements as nodes and arcs in a RDF graph which is a collection of triples of subject, predicate and object (Klyne and Carroll 2004).

RDF-S (RDF Schema) describes the semantic RDF. It provides mechanisms for describing groups of related resources and relationships between these resources. The classes and properties, range and domain constraints, and subclass and subproperty relations are defined by the schema. RDF-S in particular can be recognised as an ontology language, however, it has limited expressive power.

The Ontology Inference Layer (OIL) was layered on top of RDF to provide web-based representation and inference layer for ontologies (Fensel 2001). The main effort was to take description logics to the web for inference. OIL is placed on top of RDF schema.

DARPA Agent Markup Language (DAML) program aimed at enabling software agents to dynamically identify and understand the semantics of information resources,
and providing semantic interoperability between machines. DAML merged with OIL (McGuinness 2001) which explored existing Web standards (XML and RDF) by adding ontological primitives of frame-based systems and formal rigor of expressive description logic (Horrocks 2001).

OWL (Web Ontology Language) is mainly based on DAML+OIL and is a standard language for Web ontologies. OWL facilitates greater machine interpretability than XML, RDF, and RDF-S by providing additional vocabulary along with formal semantics. OWL has three increasingly expressive sub-languages: OWL Lite, OWL DL, and OWL Full. The decision of choice is based on the need of expressiveness and reasoning power (McGuinness and Harmelen 2004).

OWL Lite provides a minimal useful subset of language features. The language constructs provide the basics for subclass hierarchy construction: subclasses and property restrictions.

OWL DL supports maximum expressiveness while retaining computational completeness (all entailments are guaranteed to be computable) and decidability (all computations will finish in finite time) of reasoning systems. OWL DL includes all OWL language constructs with restrictions such as a class cannot be an instance of another class.

OWL Full provides maximum expressiveness and the syntactic freedom of RDF with no computational guarantees. Some features are not allowed in OWL DL, but they can be expressed in OWL full. For example, a class can be treated simultaneously as a collection of individuals and as an individual in its own right.

2.4.3 Tools for ontology development

There were a number of ontology development tools available when the research started in 2002. However, most tools were immature and did not interoperate with others. Moreover, almost no tool supports all activities of ontology life cycle. Among these tools, Ontolingua, OntoEdit, WebODE and Protégé are recognised ahead of others. More details can be found from references and surveys on ontology tools (Duineveld et al. 2000; Pérez 2002; Corcho et al. 2003; Lambrix et al. 2003; Denny 2002; Denny 2004). A brief summary of these tools is given as following:

Ontolingua (Farquhar et al. 1996) was developed as an ontology server that provides collaborative approach for ontology construction through web user interface. It also provides different translators for different representation languages.
OntoEdit (Sure et al. 2002) was developed under On-To-Knowledge project and the semantics are based on F-Logic which provides efficient reasoning on instances and capability to express arbitrary powerful rules. In addition, OntoEdit provides tool support for ontology development methodology developed under the same project (Staab et al. 2001). However, it does not provide means for subsumption as DL does.

WebODE (Arpirez et al. 2001) ontology editor provides collaborative ontology editing on web and the conceptualisation following phase activities based METHONTOLOGY methodology. WebODE is built on an application server which supports other services. Ontologies are saved in a database in WebODE and it provides translation service into and from other ontology specification languages, e.g. RDF-S, OIL, DAML+OIL, F-Logic.

OilEd (Bechhofer et al. 2001) is a prototype ontology editing tool for OIL (and DAML+OIL) that explores the extension of the frame editor paradigm to deal with an expressive language. The use of a highly optimised DL reasoning engine provides sound and complete yet still empirically tractable reasoning services.

Protégé (Gennari et al. 2003) is an open source frame-based ontology editor. It is built on OKBC model and provides flexibility by supporting meta-model (Noy et al. 2000). The plug-in architecture of Protégé and Java Application Programming Interface (API) make it a flexible base for rapid prototyping and application development. It supports a variety of formats including RDF(S), XML Schema, database and Protégé file format. The OWL plug-in is developed by extending meta-model to support OWL ontology development and provides access to description logic reasoners (Knublauch et al. 2004).

2.5 Ontologies in medicine

Ontology has attracted great attention in health informatics because of the tradition of constructing controlled vocabularies so as to cope with complexities and diversities of domain knowledge. Over the years, various standard medical terminologies and classification systems have been developed such as International Classification of Diseases (ICD 9 and ICD 10), Read codes, Unified Medical Language System (UMLS), Systematised Nomenclature of Medicine (SNOMED, SNOMED RT and SNOMED CT). Some ontologies for representing domain knowledge are also developed in recent years,
such as GALEN, Foundation Model of Anatomy (Rosse and Mejino Jr. 2003), Gene Ontology (Ashburner et al. 2000), Cancer ontology (Hartel et al. 2005).

UMLS was developed at National Library of Medicine to integrate many authoritative biomedical source terminologies into a unified knowledge representation. UMLS have two levels of elements. The Meta-thesaurus contains concepts. The Semantic Network contains general categories called semantic types. In each level, there are hierarchical (e.g., IS-A, parent) relationships and semantic relationships connecting pairs of elements. The IS-A hierarchy of the Semantic Network is a structure consisting of two trees that are interconnected by semantic (i.e., non-IS-A) relationships (Perl and Geller 2003). It is identified by Cimino and colleagues (2003) that 24.3% of subsumptions in UMLS are inconsistent, such as missing semantic type, the semantic type of the parent was too specific and children was too general, even incorrect. The new development of Core Unified Medical Language System was intended to be free from the integration and preservation constraints of terminology sources and to solve the semantic inconsistence problem (Perl and Geller 2003).

SNOMED Clinical Terms is a comprehensive terminology emerged from Clinical Terms Version 3 (Read Codes) and SNOMED RT. It is developed in description logics formalism. The goal of SNOMED CT is to provide a standardized clinical terminology that is essential for effective collection of clinical data, its retrieval, aggregation and re-use, as well as interoperability. The terminology is comprised of concepts, terms and relationships with the objective of precisely representing clinical information across the scope of health care (SNOMED CT User Guide). Content coverage is divided into hierarchies (table 2.3):
GALEN is a large project developing terminology servers and data entry systems for medical terminology. The top-level ontology of GALEN is discussed by Rector and colleagues (1996) and states that a basic structure of substances, processes and structures, coupled with a system of modifiers and modalities can be used to model most medical concepts intuitively. Although heavily influenced by the UMLS Semantic Network, GALEN ontology provides fine granularity and a more general taxonomy. One main feature of GALEN is that the GALEN Representation and Integration Language (GRAIL) which was developed to support part-whole and other transitive relations. It provided a basis for practical experiments and set requirements for future clinical knowledge representation language.

FMA (Rosse and Mejino Jr. 2003) was initially developed as an enhancement of the anatomical content of UMLS. It represents the structural organisation of human body from the macromolecular to macroscopic levels by a frame-based approach. The FMA foundation model of anatomy is articulated by its high level scheme (Rosse et al. 1998):

\[ FMA = (AT, ASA, ATA, Mk) \]

The AT (Anatomy Taxonomy) is the semantic structure, which specifies the taxonomic relationships of anatomical entities and assigns them to classes according to defining attributes. The ASA (Anatomical Structural Abstraction) is the scheme of the
model which describes the partitive and spatial relationships of the concepts represented in the taxonomy. The ATA (Anatomical Transformation Abstraction) describes the time-dependent morphological transformations of the concepts represented during the human life cycle. The Mk (Meta-knowledge) comprises the principles and sets of rules according to which relationships are represented in those three component abstractions. FMA followed a clear-principled and logical representation, which has led to conflicts with traditional anatomy textbook representations.

A survey by Stevens and colleagues (2000) reviewed biology ontologies, such as RiboWeb, EcoCyc, MBO, Gene Ontology(GO) and TaO. They discussed the content of ontologies in terms of scope, concepts and relationships, and their use in biology field.

However, few ontologies for Chinese medicine exist. Although some terms of Chinese Medicine have been covered in Read Codes and SNOMED CT, the coverage is, in our opinion, low and some of the terms are inaccurately defined and ill-positioned.

In our specific area, one particular relevant to our research is a national project with a broad collaboration of 16 distributed groups in China (Zhou et al. 2004). It is funded by China Ministry of Science and Technology to support terminology standardisation, knowledge acquisition, information retrieval and data integration in Chinese medicine. They adopted UMLS approach to develop a Unified Traditional Chinese Medicine Language System (UTCMLS). Another study is conducted in UMLS approach for acupuncture point (Zhu et al. 2004).

To address the scalability of large ontologies, Mao and colleagues (2008) propose sub-ontology evolution approach and evaluated the proposed solution with UTCMLS ontology.

A study on methodology for frame-based ontology for Chinese medicine is conducted by Park and colleagues (2004) in Korea. They propose middle concept approach to address the challenges for representation for context, similarity and reducing programming load.

The significant difference between our study and the above is that they aim specifically for coverage of terminology for the CM domain, whereas we want to formally represent acupuncture knowledge for the purpose of knowledge sharing between CM and orthodox medicine. It is unclear if the middle concept approach can achieve our research aim. The UMLS approach does not follow strict is-a relation whereas the ontology we developed in OWL description logic follows the formal is-a
relationship for subsumption. Therefore, the approaches of comparisons between ontologies are less meaningful.

2.6 The challenges of acupuncture knowledge representation

We have given an introduction to acupuncture and its practice in CM and OM, including the requirements for acupuncture knowledge representation in the first chapter of the thesis. In this section we will discuss the challenges for acupuncture knowledge representation across OM and CM.

Despite the similarity of performing acupuncture as a procedure, there are significant differences between CM and OM. The differences are rooted in language, concept, theory and logic.

From the informatics point of view, it is common that we have to deal with heterogeneous information systems which have different characteristics. For example, Kitakami and colleagues (1996) distinguished four types of heterogeneity in knowledge representation:

Paradigm heterogeneity: Two systems express their knowledge using different modelling paradigms, e.g., an object-oriented database and a relational database.

Language heterogeneity: two systems express their knowledge in different representation languages, e.g., horn-clause logic and production rules.

Ontology heterogeneity: two systems make different ontological assumptions about their domain knowledge, e.g., one system assuming houses to be composed of building material, such as bricks and windows, and another system assuming houses to consist of entities relevant for room rental, such as rooms and shower.

Content heterogeneity: two systems represent different knowledge.

In the following section we will discuss the heterogeneities between CM and OM following above categories.

2.6.1 Paradigm heterogeneity

Concepts, theories and logic behind OM and CM fall into different paradigms. The main challenge for us is to represent both paradigms in a single ontology. The concepts and theories underpinning acupuncture are indeed different to mainstream orthodox
medicine practice. These are different philosophies, concepts and terms outside of orthodox medicine, in which the diseases and symptoms are treated in different ways with results that are difficult to interpret. Medical perception in different historical periods could also apply to the difference between Chinese and orthodox medicine (Kaptchuk 2000). Kaptchuk quotes Foucault (1973, p.54) that 'not only the names of diseases, the grouping of systems were not the same; but the fundamental perceptual codes that were applied to patients' bodies, the field of objects to which observation addressed itself'.

We apply the fourth sense defined of the word paradigm in WordNet version 2.1. Paradigm -- (the generally accepted perspective of a particular discipline at a given time;) => position, view, perspective -- (a way of regarding situations or topics;) The definition of paradigm can be put in plain words as Patton (1978) explained in the context of research.

"A paradigm is a world view, a general perspective, a way of breaking down the complexity of the real world. As such, paradigms are deeply embedded in the socialization of adherents and practitioners: paradigms tell them what is important, legitimate, and reasonable. Paradigms are also normative, tell the practitioner what to do without the necessity of long existential or epistemological considerations. But it is this aspect of paradigms that constitutes both their strength and their weakness – their strength in that it makes action possible, their weakness in that the very reason for action is hidden in the unquestioned assumptions of the paradigm."

Traditional Chinese medicine and orthodox medicine are indeed different paradigms, each reflects their own health model and theories. CM views all clinical phenomena as a whole in the world. The health model of CM is based on theories of Yin Yang, Qi, Five elements, Organ, Pattern. The aim of diagnosis is not disease but to recognise pattern of imbalanced interrelationship and the goal of treatment is to regain the balance. The diagnosis and treatment are so called holistic approach to health conditions. In Chinese medicine, a health condition can be characterised by Yin and Yang. In contrast, the dominant model of disease in biomedicine is causal relation. The aim of diagnosis is to identify causal relations, for example, germ, virus, faulty gene etc. and eliminate them in pathological process. For orthodox medicine there are biology,
anatomy, pathology, physiology, etc. The same health condition can be characterised by physiological and pathological criteria. The goal is to conclude a possible disease diagnosis which will lead to treatment by which the physiological order is restored.

The criteria used for defining concepts are based on the theory from a particular domain, discipline or paradigm. For example, acupuncture points are belonging to meridians in CM. In OM, The mechanism of acupuncture explained differently and functions of acupuncture points are related to neurological system. Meridians are commonly not adopted by OM because they do not fit into its framework (Vickers and Zollman 1999). It is described by Kim (2006) in his paper of ‘Beyond paradigm’ that attempts to explore the scientific investigation of acupuncture by neurologic theory and material process through function MRI. He argued that it is an open-ended process to bridge and negotiate among heterogeneous elements in order to translate acupuncture to sciences.

2.6.2 Language heterogeneity

As Kaptchuk (2000) acknowledged, Chinese medical language is intensely metaphorical and has the potential to describe any aspect of human life. ‘Metaphor is in fact no stranger to medicine’ (Hodgkin 1983). It is recognised that ‘medicine is largely about storytelling and interpretation, and narrative, metaphor, and symbol are fundamental tools of the trade’ (Coulehan 2003). For example, in an engineering metaphor, disease is a malfunction, implying human body as a machine. However, OM practitioners have become desensitised to it and consequently find other uses of metaphor to be fanciful, which can devalue confidence in alternative approaches. In CM, some common daily life terms, such as wind, heat, damp, wood, fire, are metaphorically extended and given specific meaning which have greater significance.

In addition, there are some differences in terminology readily recognised between CM and OM. They are:

- Some terms are used in both CM and OM, such as liver, kidney; however, their meanings are differed at different levels and in different dimensions.
- Some concepts have meaning to CM but they are meaningless for OM such as Qi, Yin, Yang, San Jiao. It is almost impossible to find matching concepts in English for those terms.
2.6.3 Ontology heterogeneity

It is clear that Chinese medicine and orthodox medicine make very different claims about the world. A few theoretic entities, e.g. Yin, Yang, Qi, Five-Elements are not part of the ontology of Western science. Conversely, traditional Chinese medicine does not even consider many of the ontological claims of orthodox medicine, entities such as disease-causing by bacteria or virus (Thagard and Zhu 2003).

Chinese medicine is an experience-based medicine with an established theory system based upon solid clinical practice. The theory is one of natural philosophy, dialectic thinking. Orthodox medicine is relying on modern science and towards evidence-based approach. The views of human body in these two paradigms are different. OM has the view of nerve, muscle, bone whilst CM has the view of meridian and acupuncture points. Although organs exist in both CM and OM, they are different entities. CM acknowledges the physical existence of organs e.g. kidney as a physical organ, but significant theoretic framework of kidney is beyond the physical kidney itself. For example, kidney function is related to bone, correspondence to sense of hearing and emotion of fear and has link to kidney meridian.

2.6.4 Content heterogeneity

Content heterogeneities vary at different dimensions and levels between OM and CM. The development of scientific knowledge has advanced OM in the last century. The experimental knowledge of CM is becoming the subject to study and being evaluated from scientific framework. Health conditions are the subject of both CM and OM. However, as a result of different theories, concepts and logics, health conditions are interpreted and explained differently. The challenge is to deal with knowledge which are related but cannot be translated or mapped directly. It is equally important to specify what is different and on which ground. The following is a list of a few.

In CM, the language is descriptive and metaphorical, and health conditions are described by abstract concepts, such as

- Divisions of the Yin and Yang polar opposites, e.g. interior or exterior, cold or hot, deficient or excessive
- Elemental qualities e.g. wood, fire, earth, metal, and water
• External extremes of wind, heat, damp, dryness, and cold, as well as to internal extremes of anger, excitement, worry, sadness, and fear
• The functional influences associated with organs, Heart, Liver, Kidney etc.

In contrast, OM content are composed of solid and concrete concepts, such as
• Anatomical structure of human body, nerve, bone, organ
• Causal agent, e.g. germs, virus and physical forces
• Physiology and pathology processes
• Biochemical tests, measurements

2.7 Summary

The aim of this chapter is to provide the relevant background to this thesis. Section 2.2 explored different perspectives of ontology from philosophy, linguistics and informatics instead of giving a formal definition of ontology. From philosophical perspective, different ontological categories have been proposed and profound argument is between conceptualism and realism. Pierce’s categories of Firstness, Secondess and Thiredness seem to provide a holistic approach to overcome the argument. From linguistic and semiotics perspectives, ontology is an artefact in linguistic form. It has to deal with syntax, semantics and pragmatics to realise the promise of ontology. From informatics perspective, interpretations of ontology are summarised in four different views: vocabulary, conceptualisation, reality and semiotics. Three views emphasised one of semiotic triangle: symbol, concept or referent. However, these three are equally important from a semiotics point of view.

Principles and methodologies for ontology development were given in section 2.3. Most principles and methodologies were proposed in last two decades. Those principles still provide important guidance for ontology development. Three models of stage-based, evolving prototype and mixture of both methodologies have been proposed including life cycle and activities of ontology development. However, there is no one “correct” way or methodology for developing ontologies (Noy and McGuinness 2001). In addition to activity view of ontology methodology, Formal Ontology, Ontoclean and Distinction approach have important influence on ontology construction.
Technical background review for ontology development included systems, languages and tools. Five levels of knowledge representation formalism were summarised by primitives, interpretations and their features. Semantic network, frame-based systems and their descendant description logics are commonly used systems for ontology. The languages of ontology have the trend to move from traditional languages, e.g. KIF, OKBC to web ontology languages and description logics, e.g. OWL. Few tools available for ontology were mature and they have different features. The decision on technical solution has considerable influence on how and what statements can be made in ontology, such as an universal/existential role quantification or a simple constraint of slot value. These are choices of technical implementation of ontology and decision has to be made for individual ontology development requirement.

Some important ontologies and approaches in medicine, such as UMLS, FMA, GALEN and SNOMED CT, are representatives of semantic network, frame-based systems and description logic implementation. However, few ontologies are available for Chinese medicine and none have consideration to share the knowledge between Chinese medicine and orthodox medicine.

Finally, as our aim is to construct a single ontology of acupuncture knowledge from both CM and OM perspectives, it brings challenges derived from paradigm, language, ontology, and content heterogeneities. Yet, the main challenge is to represent underlying concept, theory and logic that form the two systems. In order to achieve knowledge sharing between them, it is important to make explicit distinctions of heterogeneities, while equally important to represent relations of clinical relevancy across the two paradigms.
Chapter 3

Approach to the Research

This chapter discusses the identification and justification of the principles which guide this research. The research assumptions, design and development are discussed in terms of ontological, epistemological, and methodological principles.

3.1 Introduction

Prior to conducting this research, it was important to be clear on the research paradigm which guides and influences how to carry out the research. As Guba and Lincoln (1994, p.119) asserted, the ontological, epistemological and methodological principles determine the boundaries of research paradigms. We discuss the philosophical ground and research assumptions first. The approaches and consideration for the research are explained in terms of the ontological arguments (section 3.3), epistemology arguments (section 3.4) and methodology arguments (section 3.5). The pragmatic issues related for conducting this research are discussed in section 3.6.

3.2 Philosophical arguments of the research

3.2.1 Philosophical ground

It is important that researchers understand the underlying philosophical ground before making any decision on the choices of methodologies. Research paradigms are rather complex because of the degree of commitment to the philosophical distinctions of ontology, epistemology, and methodology. The basic traditions of positivist, postpositivist and constructivism, and their implications for research were outlined (Guba and Lincoln 1994; Lincoln and Guba 2003; Corbetta 2003).
Table 3.1 Contrasting paradigms

<table>
<thead>
<tr>
<th></th>
<th>Positivism</th>
<th>Post-positivism</th>
<th>Constructivism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ontology</strong></td>
<td>Realist; Singular reality</td>
<td>Critical realist; Social reality is 'real' but only knowable in a probabilistic sense.</td>
<td>Relativist; Multiple realities constructed by individuals; Multiple/Holistic</td>
</tr>
<tr>
<td><strong>Epistemology</strong></td>
<td>Objectivist; Dualist; Findings true</td>
<td>Modified objectivist; findings probably true</td>
<td>Subjectivist; Interactive; Researcher and subject are interdependent</td>
</tr>
<tr>
<td><strong>Methodology</strong></td>
<td>Experimental; Manipulative; Verification/falsification of hypotheses; Chiefly quantitative methods</td>
<td>Modified experimental; Manipulative; Verification/falsification of hypotheses; May include qualitative methods</td>
<td>Hermeneutics; Dialectical; Interaction between researcher and subject; Interpretation and interaction</td>
</tr>
<tr>
<td><strong>Outcomes of the research</strong></td>
<td>Context &amp; time independent; Generalisations leading to 'natural' immutable laws or predictions</td>
<td>Context &amp; time dependent; Generalisations leading to models for predictions; Probabilistically true laws</td>
<td>Context &amp; time dependent; Working hypotheses leading to understanding</td>
</tr>
</tbody>
</table>

(Guba and Lincoln 1994; Lincoln and Guba 2003; Corbetta 2003)

It is a much simplified version that we have adopted. Nevertheless, there was a great deal of development following these basic traditions, in particular, the evolution of theories along the social study. The table demonstrates the relationship between philosophical ground and research paradigms, which decides how and what research questions can be asked and answered in the research design and process. Positivism and postpositivism are common philosophical grounds for natural science which is concerned with the discovery of natural laws. They are also used for social study. However, it was argued that natural science is different to social studies which seek explanations of particular social phenomena or historical processes (Benton and Craib 2001). Constructivism underlines the philosophy paradigm for social study.
Constructivists acknowledge the social construction of knowledge and believe that the reality is constructed from community consensus (Lincoln and Guba 2003).

3.2.2 Research assumptions

The assumptions of this study are summarised as follows:

- This research is centred on knowledge representation of meaning and perspective, i.e. how different perspectives can be represented in a single ontology.

- The researcher is the primary research instrument with input from participants. The ontology was developed by researcher and evaluated by the experts. The data was collected, analysed, and interpreted rather than measured from a statistical viewpoint.

- This research is conducted in a research laboratory setting instead of a natural setting, and the products would be considered as prototypes.

The different knowledge about the reality became a ‘reality’ for knowledge representation which is the subject of the study. The objectivity is often difficult to achieve because it is necessary that the researcher has to make decision on the ways that knowledge is to be represented, and has to interact with participants in order to construct the ontology and collect data of evaluations. In the study, the researcher must accept that the ‘world’ in informatics is socially constructed rather than objective and researcher-independent.

The paradigm of this research is close to constructivism, but will not follow it to the extreme. Constructivism refers to the form of research encompassed within the interpretivist paradigm (Benton and Craib 2001). It advocates that knowledge and truth are the result of perspectives (Guba and Lincoln 1994). However, we did not follow the extreme of radical constructivist position which excludes the existence of an objective world completely (Corbetta 2003). This paradigm enabled the researcher to admit ontological commitment, acknowledge the epistemological position, and select the research methodology.
3.3 Ontological arguments of research

To conduct the research, we have to answer the basic ontological question: ‘what kind of things are there in the world?’ Three distinctive entities need to be addressed for this research. They are the knowledge to be represented, the characteristics of the knowledge, and the representation of the knowledge.

3.3.1 The object of study – domain knowledge

We have explained the history of acupuncture, requirements and challenges of its knowledge representation in the previous two chapters. There are different hypotheses related to acupuncture and its mechanism is still largely unknown. Some part of knowledge is recognised in the practice of one paradigm but might not be appreciated by practitioners in the other paradigm. For example, meridian is used in CM practice but it is not accepted by OM practitioners. Hence, the knowledge representation should not be limited to one single view of the world. Furthermore, we have to accept that represented knowledge is constrained by our limitation of what we know.

The different knowledge of acupuncture has been generated from different paradigms and they all involve human construction. Among others, Flick (2006, p79) posts a similar view as Schütz that knowledge is constructed by selection and structuring from a constructionist viewpoint:

‘All our knowledge of the world in commonsense as well as scientific thinking, involves constructs, i.e., a set of abstractions, generalizations, formalizations and idealizations, specific to the relevant level of thought organization’

In the informatics world, it is a challenge to represent different knowledge as Sowa uses the notion of ‘knowledge soup’ to point out the difficulty of matching abstract theories to the physical world (Sowa 2000a).

3.3.2 The characteristics of object – paradigms and disciplines

In our approach, we made a commitment to represent both CM and OM paradigms within which the characteristics of domain knowledge were studied. There were distinctive differences between these two paradigms in the health care domain. The knowledge of the same object of acupuncture practice has been originated from these different paradigms. As we explained in chapter 1 and chapter 2, the knowledge of
acupuncture has been developed in Chinese medicine, and constantly and increasingly enriched by ongoing scientific research and variety of clinical practices in the present day.

The CM and OM knowledge of acupuncture can be viewed as different types of knowledge as Karin Breu (2003) discusses the characteristics of practical and scientific knowledge.

Table 3.2 characteristics of practical and scientific knowledge

<table>
<thead>
<tr>
<th>Audience</th>
<th>Scientific knowledge</th>
<th>Practical knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representation of knowledge</td>
<td>Propositional: formal and explicit</td>
<td>Narrative and visual: procedural and tacit</td>
</tr>
<tr>
<td>Starting point for knowledge creation</td>
<td>Research questions derived from the existing literature</td>
<td>Need to search for solutions to the problems faced in practice</td>
</tr>
<tr>
<td>Method of knowledge creation</td>
<td>Systematic, well-defined procedures</td>
<td>Direct experience and trial-and-error learning</td>
</tr>
<tr>
<td>Language of knowledge representation</td>
<td>Abstract, technical and linguistic precision</td>
<td>Concrete, everyday, ambiguous and informal</td>
</tr>
<tr>
<td>Output</td>
<td>Prescriptive theory; deductive, reductionist and generalised</td>
<td>Tools and frameworks: imaginative and contextualised</td>
</tr>
<tr>
<td>Epistemology and validation</td>
<td>Logic of measurement, proof and results</td>
<td>Ability of actions and tools to produce intended goals and outcomes; Level of adoption; Evaluation of tools and knowledge by users</td>
</tr>
</tbody>
</table>

Karin Breu (2003)

A significant amount of practical knowledge has been generated over the long history of CM practice. It is a simplistic view of knowledge to regard OM as purely scientific and CM as purely practical as both paradigms contain and utilise both forms of knowledge.

The representation of knowledge is studied under two disciplines of health care and informatics. There were distinctively different perspectives between the two disciplines. The concern of health care is the creation and use of knowledge. It can be independent to any interventions of health informatics.

It is vital for health informatics that knowledge is properly represented and stored in suitable forms. To some degree, the theories, logics and concepts of particular knowledge has to be analysed for representation. We feel that it is appropriate to reply on the analysis and input from experts of health care professions to decide what
knowledge to present whereas the matter of how to represent the knowledge should remain remit of health informatics.

**3.3.3 The result of the study – ontology as knowledge representation**

The product of this research is an ontology for acupuncture and the research was conducted in health informatics. It is an artefact constructed by the researcher and it is intended to represent different views of knowledge from both CM and OM. The intention is to utilise the technology to support comprehensive representation of current knowledge in one subject area.

A significant portion of Chapter 2 is dedicated to different views of ontology which were explored from philosophical, informatics, semiotic and linguistic perspectives. In particular, the informatics views have shifted from vocabulary to conceptualisation which is challenged by representation of reality.

We hold the semiotic view of ontology which embraces different views of vocabulary, conceptualisation and reality. The vocabularies are symbolic units for computation and transformation to underlying coding sets and machine languages. The concepts are semantic units for communication among systems and human. The realities are objects that exist independent to the references, representations, or indications in information systems and information systems have to be connected to the reality in someway.

**3.4 Epistemological arguments of research**

It is clear that the researcher is part of the research. The process of research is not independent of input from the researcher and it involves justification of decisions and interpretations.

**3.4.1 Research decisions on system, language, knowledge representation formalism**

The decisions have to be made by the researcher as to what system is to be used, which languages are suitable for representation, and what resources are reliable to elicit knowledge. Exploring the knowledge representation formalism enables us to understand current issues in knowledge representation and suitable systems and languages.
Discussions of systems, languages, and knowledge representation formalism in detail were provided in chapter 2.

Although these decisions are dependent on the researcher, the decisions have to be justified. The justification is discussed in the design and implementation in chapter 4. The choice of system, language, and knowledge representation formalism should support the purpose of the study. The advanced development in informatics provided more choices with which to suit the purpose.

3.4.2 Researcher engagement of development – the interpretation of domain knowledge

If we are to deliver a faithful representation, then we must first encounter the issues of how and what are recognised as knowledge by experts of the health care professions. The intertwined recognitions of knowledge make it harder for health informaticians to make the judgement of what is the knowledge and how faithfully it can be represented.

How do we know the knowledge is the reliable truth of the universe? No matter which paradigm of inquiry, it all requires some degree of interpretation. Positivists assume there is a reality out there apart from the flawed human apprehension of it, whereas constructivists believe that agreements about truth may be the subject of community negotiations regarding of what can be accepted as truth (Guba and Lincoln 1989). This has the implication that decision and judgement of true knowledge can not be error free. Many bio-informaticians seem to shy away from believing that we can have knowledge about a mind-independent biological reality and this tendency is neither well-founded nor harmless (Johansson 2006).

The reality and the knowledge of reality are two distinct entities. From our point of view, informatics engineers should be realists and represent the knowledge truthfully, even if so called knowledge might not reveal the ultimate truth of the reality at a point of time. The unified top entities for ontology may not be easily determined. It may be the same that we do not have a unified theory for all medical entities. The conflict or incompatibility may have to stay with medicine and medicinal informatics for a long time.

However, we have to be fully aware that the judgement of knowledge as part of process for the construction of ontology has added subjectivity into the final product. The knowledge sources were confirmed from health care experts. The different sources
and decisions on the scope may result some degree of differences in ontology. Even the same knowledge resources are used, there might be some differences caused by researcher’s decisions of ontology design. Hence, in this study, justification and validation of validity from both informatics and health care professions are required. They were achieved through experts’ input in ontology design, construction and evaluations. Two groups of experts were employed to minimise the bias, whilst one group provided advices on knowledge sources for ontology development and another group undergone ontology evaluations.

3.4.3 The interpretations of data from participants

The participants of the research are health care and informatics experts and they were involved at different stages of the study. At the early stage, their expertises contributed to define knowledge sources and the scope of the ontology as well as informal discussions during the ontology construction. The evaluations of the ontology in the final stage in forms of interview were formal and conducted in research lab.

The interpretation was inevitable while the researcher received the advices and feedback for ontology from experts. It is upon the researcher’s capacity to form and implement action plans. The feedback from the evaluation interviews during which the researcher interacts with participants were observed, codified and analysed. Equally, it is important to understand the participants’ point of view and examine inputs from experts. Research participants may be more or less reflexive regarding their own perspectives, but the knowledge that emerged from interviews was a result of collective process by researcher and participants.

This conforms to the notion that knowledge and interpretation in a constructivist research paradigm is the result of a collective, not an individual process as Schwandt (2003 p.305) asserts that interpretations are not constructed in isolation but against a backdrop of shared understandings, practices, language. The collective knowledge and interpretation were reflected in the development and refinement of the ontology as part of methodology for ontology construction. The detail of considerations and experts contributions were discussed in ontology evaluation chapter 6.
3.5 Methodological arguments of research

3.5.1 Quantitative vs. Qualitative

Quantitative and Qualitative are two main paradigms in research and they are interchangeable to positivism and phenomenology (constructivism) (Hussey and Hussey 1997). The research is often placed somewhere between these two extremes of positivist and constructivism as underpinning philosophical ground. The difference between quantitative and qualitative is also the logics that underline the sampling approaches.

Quantitative methods typically depend on larger samples selected randomly in order to permit confident generalization from the sample to a larger population (Patton 1990, p.169). The validity would be low in quantitative approach, since it ignores context and the phenomenon’s viewpoint. The reliability is high in positivistic research, as it is associated with measurement and control variables that are not feasible in phenomenological research (Hussey and Hussey 1997).

Qualitative inquiries typically focus on relatively small samples with purposeful selection. The distinction between qualitative and quantitative is that the former produce data which are freely defined by the subject rather than structured in advance by the researcher (Patton 1990). Therefore, the qualitative information-rich samples will illuminate the study question in depth.

The nature of this research is qualitative in approach and the purposive sampling is discussed in chapter 6. The differences between qualitative and quantitative research were summarised in table 3.3 (Blaxter et al. 2001, p.65).
Table 3.3 The differences between qualitative and quantitative research

<table>
<thead>
<tr>
<th>Qualitative paradigm</th>
<th>Quantitative paradigm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concerned with understanding behaviour from actors' own frames of reference</td>
<td>Seek the facts/causes of social phenomena</td>
</tr>
<tr>
<td>Naturalistic and uncontrolled observation</td>
<td>Obtrusive and controlled measurement</td>
</tr>
<tr>
<td>Subjective</td>
<td>Objective</td>
</tr>
<tr>
<td>Close to the data: the ‘insider’ perspective</td>
<td>Removed from the data: the ‘outsider’ perspective</td>
</tr>
<tr>
<td>Grounded, discovery-oriented, exploratory, expansionist, descriptive, inductive</td>
<td>Ungrounded, verification oriented, reductionist, hypothetico-deductive</td>
</tr>
<tr>
<td>Process-oriented</td>
<td>Outcome-oriented</td>
</tr>
<tr>
<td>Valid: real, rich, deep data</td>
<td>Reliable: hard and replicable data</td>
</tr>
<tr>
<td>Ungeneraishopal: single case studies</td>
<td>Generalisable: multiple case studies</td>
</tr>
<tr>
<td>Holistic</td>
<td>Particularistic</td>
</tr>
<tr>
<td>Assume a dynamic reality</td>
<td>Assume a stable reality</td>
</tr>
</tbody>
</table>

Adopted from Oakley (1999, p.156)

3.5.2 Research methodologies for qualitative research

There are a number of different types of qualitative research methodologies (Hussey and Hussey 1997; Blaxter et al. 2001; Oates 2006).

- Survey – a standardised and systematic method to obtain the same kind of data from samples of population
- Action research – investigate and improve working practice by member of communities or organization
- Ethnography – researcher uses socially acquired and shared knowledge to understand the observed patterns of human activity
- Case studies – the extensive investigate of one instance of phenomenon
- Design and Creation – A research strategy focus on developing new information artefacts (Oates 2006)

The choice of methodologies should support the purpose of the research and different methodologies can be combined as appropriate to study a particular set of research questions (Blaxter et al. 2001).
Surveys are widely used for quantitative research but they are also applied in qualitative research. They are applied for investigating patterns and generalisation from the samples for large population which does not suit the purpose of this study.

Action research is suitable for the people conducting research in their workplaces to improve the practice. It is not compatible with our research purpose which concerns modelling and methods, and is conducted in informatics laboratory.

The aim of ethnography is to experience and interpret the social world. The approach requires researcher to gain access and trust from the people and organisation being studied. Observation is the primary method of ethnography. The reflexivity of relationship between the researcher, the participants and the research process addressed by ethnography approach is relevant to all types of research (Oates 2006). This methodology is not appropriate for this study.

Case study in research is concentrated on finding out about the complexities of the real world as Yin (1993, p.3) described ‘it is the method of choice when the phenomenon under study is not readily distinguishable from its context’. Case study is open to mixed methods of data collection (Robson 2002, p.179), and can be used in both the positivist and constructivism paradigms. However, in informatics, case study is often an artificial scenario to which proposed modelling technique, method or program have been applied (Oates 2006, p.143). They are typically much simpler than normally occurs in real life. The formal case study research is more suitable for research into the real-life use of the IT artefacts. In this research, the implied case study in informatics other than its formal sense of research methodology was employed, setting the case study in constructing acupuncture ontology.

The products of researches in informatics are often called artefacts, such as, constructs, models, methods, and instantiations (March and Smith 1995):

- Constructs – the concepts, vocabulary, entities, objects etc. as building block for information system
- Models – combinations of constructs for representation and applications for specific purpose
- Methodology – guidance on development of model or system and process for apply the system for problems solving
- Instantiations – a working system of constructs, models, methods which can be demonstrated and implemented in computer system
Design and Creation is suitable for research in IT development. It commonly involves analysing, designing, developing and evaluating products for computing and information system (Oates 2006, p.108).

The advantages of design and creation research strategy include (Oates 2006, p.121):

- A prototype or working IT artefact as a result of the research
- Rapid evolving technology and requirement from different domain provide wide scope to make contribution to knowledge
- Permit creative and technical advancement for researcher
- Appropriate for research in computer science and software engineering

While the disadvantages include (Oates 2006, p.122):

- Require to justify difference between a research and normal development
- Require necessary technical or artistic skills from researcher
- It can be difficult to generalise to different settings
- Dependence of researcher for the success
- Research is at risk of ‘out of date’ because of rapid advanced technology

The design and creation is the methodology adopted for this research as its aim is to develop an informatics artifact, prototype ontology for knowledge representation. This methodology allows us to apply latest technology and address the requirements and challenges for knowledge sharing across paradigms.

The main activities for design and creation methodology including analysing, designing, developing and evaluating were discussed in different chapters of this thesis. The study involves the researcher identifying the research requirements and challenges (chapter 2). The designing and developing ontology is discussed in chapter 4 which describes the specific methodology for ontology construction. It includes activities under three groups of identify, categorise, and test. The flexible sets of distinctions approach and technical solutions for overcoming the challenges are discussed in chapter 5. The evaluation of the product of the research is a qualitative approach. The methods of observation and interview applied for evaluation are briefly discussed in the following section and detail can be found in chapter 6. We are aware of the disadvantage of this methodology in the context of this research.
3.5.3 The methods for data collection in evaluation

The methods for data collection for this research are combined observation and interview which are commonly applied at the evaluation stage. To apply the triangulation can strengthen the study of the same phenomena or program as Patton (1990, p.187) suggested ‘to achieve the triangulation within a qualitative inquiry strategy by combining of different kinds of qualitative methods, mixing purposeful samples, and including multiple perspectives’. The detail of choices of types of observation and interview are discussed in the chapter 6 of evaluation.

Observation is a method commonly used to obtain descriptive quantitative data of behaviour or events. The advantage of observation is that researcher can directly record human behaviour to gain the first-hand information. However, such accounts may not be accurate due to researcher’s interpretation and personal opinion or even inaccurate record, and it can also be limited because participants may consciously or unconsciously change the way they behave under the observation (Sapsford and Jupp 1996).

The second-hand observation using photographs, films or videos attracted increasing attention (Flick 2006). In this research, the digital video recording was used as an instrument for data collection. It allows capturing the facts and processes that are fast and complex with rich detail of video and audio data available for reanalysis.

Interview enables researcher to ask participants questions in order to find out what they think and feel. In general, the interview is one-to-one but it may involve other people or group interview. It is not possible to avoid the personal, social interaction between the interviewer and interviewee and the whole process can be time consuming as observation does (Foster 1996, p.118). The participants may react to the set up of interview and also how the questions have been asked. Therefore, it is necessary to justify the validity of the data collected.

Interviews permit the researcher to ask more complex questions and through the interaction and follow-up questions to elicit in-depth data which is increasingly richer (Hussey and Hussey 1997, p.158). The in-depth interview with less structure may allow the freedom for researcher to explore respondent’s meanings and beliefs to enhance the validity. However, by doing so it is at the possible expense of reliability that another researcher might not able to use the same methods to gain the same results (Foster 1996).
3.6 Research Pragmatics

The theoretical underpinnings are important guides to the way in which research is conducted, analysed, and presented. Some issues are concerned with pragmatics – factors that cannot be addressed by theory alone, but need to be addressed practically.

3.6.1 Research participants

The study involved experts from two disciplines of health care and clinical informatics, and two paradigms of Chinese medicine and orthodox medicine. They provided expert advices which enable the ontology development and evaluations. The decisions on selection of experts and sampling techniques are discussed in following chapters.

Experts were involved in the discussion of research issues such as scope, methods and knowledge resources for ontology construction. The communications were through informal meetings with frequent follow-up meetings and email discussions. The experts who supported ontology development were not involved in the evaluation in order to provide different perspectives and avoid possible biases.

Experts who involved in the ontology evaluation were informed about the research background, goal and methods for evaluation. Consent was obtained from the participants before recording of evaluation and interview.

3.6.2 Research environment

The research was conducted in the SHIRE (Salford Health Informatics Research Environment) informatics laboratory. The ontology was designed and constructed in a laptop and standard PC running Microsoft Windows XP professional. The open source software Protégé is the platform for ontology constructions and configured with access to the Description Logics (DL) reasoner.

The evaluations were carried out in the SHIRE informatics laboratory. The process was recorded by digital video recording system. Three feeds from three digital cameras at the different angles of observation room and a real time screen capture of action in laptop were recorded simultaneously and burned into DVD. There was a two-way minor between the observation room and control room. Researcher can observe participants from the control room. Participants in the observation room can not see
people in the control room. The layout of the observation and control room is in appendix K.

3.6.3 Research ethics

The ethical considerations of research need to be addressed throughout the research process. The issues related to this research are as following:

The research should be legal and obeys relevant national legislation, such as, intellectual property right of materials and software for research.

The researcher should treat all direct or indirect participants fairly and with honesty in the research following appropriate professional codes of conduct.

Participants have right to be treated with dignity and they should not be used for researcher’s own financial or other advantages. The right of participants are summarised as following (Oates 2006, p56):

- Individuals or organisation have right not to participate the research
- Participants have right to decline answering some questions, or to refuse certain activities, or even withdraw from the research at any time
- Relevant information should be provided to the participants to form an informed consent, including purpose of research, research funding and governance, the activities involved, data collection and analysis, and the rights for participants.
- Participants have right to be anonymous that their personal data are not reported.
- Participants have right to ask confidentiality of data obtained from them. The data should be kept safe and not to be published or reported without their consent.

3.7 Summary

The three traditional philosophical grounds are positivism, post-positivism and constructivism, under which research questions and approaches were determined. This research follows the constructivism approach which supports our research assumptions.
The ontology arguments of research are centred on knowledge, characteristics of knowledge and the artefact. They are subjects of this study and the semiotics view of ontology enables us to embrace different views of reality.

The epistemological arguments reveal the approach in the process of the research is subjective. Researcher is part of research whose tasks involve decision making on systems, language and interpretations of knowledge and data from participants.

The methodology for this research is under qualitative paradigm and follows design and creation methodology which focus on development of new information artefacts. The construction of acupuncture ontology was used as a case study in informatics.

The observation and interview were applied as key data collection methods. Research participants and environment were explained and research ethics were fully considered at the end of this chapter.
Chapter 4

Ontology Design and Development

This chapter presents how we designed and constructed an acupuncture ontology. We describe the development environment first, then the methodology, and finally we introduce the flexible sets of distinctions approach and discuss how it can overcome the difficulties of managing different perspectives across disciplines and paradigms.

4.1 Introduction

We explain the development environment, methodology and flexible sets of distinctions approach in this chapter, leaving discussion of the ontology content in next chapter. We consider the described approach to be a general one, and one which can be adopted by those who face similar challenges. Here the study is in the context of two medical paradigms, but we believe the approach can be applied to other situations which also need to represent different perspectives in a single ontology. Paradigms are, in this same sense as perspectives, interchangeable.

We have discussed the importance of system and language which influence the way we make expressions and statements. Here the discussion of the system is in two parts; first an introduction of the development environment itself, which includes the Protégé system, OWL plug-in for Protégé, and RACER, a Description Logics classifier; and second, a discussion of the system from the ‘distinction’ point of view, i.e., we discuss the flexible sets of distinctions approach and how it facilitates the analysis of system and domain knowledge from different paradigms. The relations among requirements, challenge, the development environment and methodology and the final product of ontology are demonstrated in the mind map of the study in appendix L.
4.2 The ontology development environment

We have noted previously (chapter 2) that few tools were mature enough to support ontology development at the time of our study. Although our choice was limited, there were several good reasons for choosing Protégé for our ontology development. The main reasons include:

- Interoperability and reusability of the development environment
- Open source and extensibility
- Support for both frame-based and logic ontology editing
- Support for what were regarded as the standard ontology languages
- Long term tool development had taken place and continuing academic support
- Previous applications include medical ontology development

We briefly explain the terminology of Protégé and the OWL plug-in system features in section 4.2.1 and 4.2.2. RACER, the reasoner for the study, is discussed in section 4.2.3.

4.2.1 Protégé

Protégé was developed for knowledge-based system in the Stanford University and has been widely used for ontology editing (Cardoso 2007). Protégé is a frame based ontology and knowledge-base editing tool which provides support for the web ontology language (OWL). The underpinning knowledge model of Protégé satisfies the OKBC (Open Knowledge Base Connectivity) protocol. Because Protégé complies with OKBC model, the basic concepts are similar to other frame based systems. Protégé has classes, instances, slots and facets. The terminology of Protégé has been described in detail elsewhere (Noy et al. 2000; Gennari et al. 2003). The following is merely an introduction of basic concepts of Protégé development environment to assist the later discussion.
Frame | Class, individual, slot and facet are all frames  
--- | ---  
Class | a set of entities constitute a taxonomic hierarchy  
Meta-class | a class whose instances are classes  
Individual | an entity which is not a set  
Instance | an entity of a class, both individuals and classes themselves can be instances of classes  
Slot | a set of properties of a frame. Protégé makes distinctions between own slot and template slot  
Own slot | represents property of the class but not its instances. An own slot is defined by template slot of its meta-class. The own slot of a frame is not inherited by its subclasses.  
Template slot | represents properties of instances of a class. A template slot of meta-class becomes an own slot of its instance (a class). The template slot is inherited by its subclasses and propagated to their instances.  
Facet | represents additional information about slot, e.g. constraints of value range of a slot.  

Table 4.1  Key terms in Protégé  

The meta-class architecture in Protégé allows users to customise or define new meta-classes. It has been used for representing the RDF knowledge model (Noy et al. 2000) and the OWL model (Knublauch et al. 2004). Protégé plug-in architecture in Protégé provides an open programming API for extending Protégé for custom-tailored application (Gennari et al. 2002).

The advantages of frame-based formalisms are cognitively simple, intuitive and understandable to domain experts. It has been used for constructing large size medical ontologies, such as Foundational Model of Anatomy (FMA). However, the frame-based formalism has semantic problems and the development of description logics has overcome these problems (See chapter 2). Furthermore, the hierarchies of ontology in frame-based system have to be maintained manually. Although automatic classification may not be necessary for a small ontology, it becomes important when building a large...
ontology to ensure the consistency. The advantage of using Description Logics for an ontology is that the associated reasoning services provide subsumption and consistency checking.

In the next section, we will discuss the OWL plug-in and its features for developing OWL language based ontology.

4.2.2 OWL plug-in

Ontologies play a key role for explicit semantic representation of the Semantic Web which enables understanding for both human and machine. The Web Ontology Language (OWL) from W3C provides a standard language for an ontology.

The OWL plug-in has been developed since late 2003. The OWL plug-in extends Protégé to provide a means for editing an ontology in the OWL language. The implementation of the OWL language is accomplished via a meta-model and plug-in architecture. The OWL-DL model is mapped to frame-based formalism and supports access to external DL reasoners. The classification result of an inferred hierarchy is displayed separately alongside the existing hierarchy until the next classification is called.

The Figure 4.1 is a screen shot of user interface of OWL plug-in in Protégé.
The left side of windows showed hierarchy view of concepts. The right side of windows displayed the specifications of the concepts. It included description logic definition and inherited values from superclass.

There are some changes of terminology from frame-based to that used by OWL, such as Class is changed to OWLClass and Slot is changed to Property. Some OWL-DL specific terminology will therefore be explained to clarify the later discussion.

The OWL expression is represented in standard DL syntax. These symbols are shown in Table 4.2.

<table>
<thead>
<tr>
<th>OWL element</th>
<th>Symbol</th>
<th>Key</th>
<th>Example expression in Protégé</th>
</tr>
</thead>
<tbody>
<tr>
<td>Owl: allValuesFrom</td>
<td>∀</td>
<td>*</td>
<td>∀ hasChildren Female</td>
</tr>
<tr>
<td>Owl: someValuesFrom</td>
<td>∃</td>
<td>?</td>
<td>∃ hasHabitat University</td>
</tr>
<tr>
<td>Owl: hasValue</td>
<td>∋</td>
<td>$</td>
<td>hasGender ⊃ male</td>
</tr>
<tr>
<td>Owl: minCardinality</td>
<td>≥</td>
<td>&gt;</td>
<td>hasChildren ≥ 1 (at least one value)</td>
</tr>
<tr>
<td>Owl: maxCardinality</td>
<td>≤</td>
<td>&lt;</td>
<td>hasDegree ≤ 5 (at most five values)</td>
</tr>
<tr>
<td>Owl: cardinality</td>
<td>=</td>
<td>=</td>
<td>hasGender = 1 (exactly one value)</td>
</tr>
<tr>
<td>Owl: intersectionOf</td>
<td>∩</td>
<td>&amp;</td>
<td>Student ∩ Parent</td>
</tr>
<tr>
<td>Owl: unionOf</td>
<td>∪</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Owl: complementOf</td>
<td>¬</td>
<td>!</td>
<td>¬ parent</td>
</tr>
<tr>
<td>Owl: one of</td>
<td>{…}</td>
<td>{}</td>
<td>{yellow green red}</td>
</tr>
</tbody>
</table>

(Knublauch et al. 2004; Knublauch et al. 2004)

The Acupuncture ontology is constructed using the OWL-DL language. We use standard DL syntax in subsequent discussions in this thesis because of it is easy to read and short in length comparing to OWL syntax. For example, the same expression of ‘Student ∩ Parent’ intersection example in traditional DL syntax is shown in OWL DL as following:

```xml
<owl:Class>
  <owl:intersectionOf rdf:parseType="Collection">
    <owl:Class rdf:about="#Parent"/>
    <owl:Class rdf:about="#Student"/>
  </owl:intersectionOf>
</owl:Class>
```
4.2.3 Description logics reasoner

The Description Logics (DL) reasoners provide reasoning services such as automatic classification, consistency and equivalence checking. A small number of DL reasoners were available at the time of our development and RACER was the first DL reasoner accessible from Protégé OWL plug-in directly. The later release of Protégé OWL plug-in provided access to other reasoners such as FaCT and Pallete which provide the similar reasoning services.

RACER implements the TBox and ABox reasoner and extends basic logic ALC (Schmidt-Schauß and Smolka 1991) by number restrictions, role hierarchies, transitively closed roles and inverse roles. The RACER DL consistency and subsumption services are accessible through a TCP/IP interface based on sockets (Haarslev and Möller 1999). The services are accessible through Protégé OWL plug-in. The implementation of DL services in OWL plug-in is based on ALCOIN(D) as described in following table 4.3 from documentation within Protégé software.

<table>
<thead>
<tr>
<th>AL</th>
<th>Allows concept intersection, full universal quantification, atomic negation and limited existential quantification (i.e. existential restrictions with fillers limited to owl:Thing)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Complex concept negation (e.g. not(A or B)). Note that ALC allows disjunction and full existential quantification, which can be represented with conjunction and full negation, and universal quantification and full negation respectively.</td>
</tr>
<tr>
<td>O</td>
<td>Nominals (Singleton sets, one of, eg {a} - these are also used in hasValue restrictions)</td>
</tr>
<tr>
<td>I</td>
<td>Inverse roles (properties which have inverses specified, or properties that are symmetric)</td>
</tr>
<tr>
<td>N</td>
<td>Number restrictions (cardinality restrictions, also includes functional properties)</td>
</tr>
<tr>
<td>D</td>
<td>Data types</td>
</tr>
</tbody>
</table>

Table 4.3 Description logics language expressivity supported by Protégé OWL plug-in
4.3 Approach to the ontology development

Designing a substantial ontology is complex. It is even more challenging to demonstrate the relevance and value to a broad audience. In part, the discipline of ontological engineering has developed to assist with these tasks. Ontological engineering is concerned with the design principle, modification, application, and evaluation of ontologies. There are three main not exclusive types of approach for ontology development: these are stage-based, evolving prototype and mixture of the two. We have taken the latter approach which has been developed and used by the On-To-Knowledge project (Staab et al. 2001). This approach includes five stages:

- feasibility study
- ontology kick off
- refinement
- evaluation
- maintenance & evolution

There are two cycles of the development:

- refinement – evaluation
- refinement – maintenance & evolution

There are different activities among the different methodologies as we reviewed in Chapter 2. Because of the time constraint of this study, the stage 5 of maintenance & evolution and the second cycle of refinement – maintenance & evolution are not included.

The scope and knowledge content are defined through meetings and informal discussions with clinical experts. The manual construction of the ontology takes a middle-out approach by the researcher following the known principles and criteria in the field which are discussed in chapter 2. The system and knowledge content are analysed by the flexible sets of distinctions approach described with the meta-model representation of different perspectives in chapter 4. The validation of the ontology by the DL services and the review by clinical experts are applied during the development to ensure the ontology’s quality. The final ontology is evaluated through six sessions of one-to-one evaluation by experts from CM, OM and clinical informatics.

The main activities of our ontology development are summarised by three categories: Identify, Categorise and Test.
Identify: scope, people, sources of knowledge, development environment, which are the stages of feasibility study and ontology kick off.

Categorise: concepts and roles into hierarchies by logically defined meanings, which are the stage of ontology kick off and refinement.

Test: evaluate the ontology through evaluations and applications.

4.3.1 Identify - stage: feasibility study and ontology kick off

This stage, i.e., to prepare for the ontology development, is equivalent to the feasibility study and ontology kick off stages in On-To-Knowledge methodology.

Identify the scope of acupuncture ontology

We followed Zachman’s information system architecture framework (Zachman 1987). The following questions are used to identify the scope of the ontology, given here with the associated answers.

What is the content for ontology?

The ontology is intended to represent Acupuncture knowledge from both the OM and CM perspectives. It covers foundation theoretic entities, e.g. acupuncture points, acupuncture meridians, Ying Yang, Five Elements, needle technique etc..

How can the ontology be used?

We plan to use this ontology for applications which will support teaching, learning, clinical practice, information retrieval.

Where is the ontology used?

The ontology is a backend engine for different applications and is intended to be used in both educational and clinical practices.

Who will use and maintain the ontology?

We assume that the end users would be acupuncture practitioners and educators. Ontology developer will maintain the ontology.

When will the ontology is used?

It depends on applications which implement the ontology. It can be used for comparing equivalence of concepts, reference the relationships among concepts such as is-a hierarchy, relations based on the attributes and values etc..

Why develop a single ontology to represent both views of CM and OM?

The requirement was discussed in chapter 1.
The technique of competency questions were applied in a formal logic language (Grüninger and Fox 1995) although we followed the informal approach for defining the competency questions (Uschold and Gruninger 1996; Fernández-López 1999). The following are examples of high level competency questions which the ontology should be able to answer. These questions are extended and applied in a query application in ontology evaluations. The purpose is to demonstrate the competency of the ontology. The following are examples of questions. Please see details of questions in a query language in appendix II.

How do we locate an acupuncture point?
What anatomic structures are related to an acupuncture point?
What is the function of an acupuncture point?
What patterns or diagnosis should practitioners consider when they are treating a condition?

Identify people
Since the author is the sole designer/developer to the ontology, it is impossible to avoid personal bias during ontology development. Two groups are introduced, acting as third parties who are independent of the designer/developer, which are important to contribute critiquing factor and minimise the key researcher's personal preferences while trying to represent a theoretical framework in ontology. Experts in acupuncture and medical informatics are identified to facilitate the development and evaluation of the ontology.

One group comprises of ontology experts from the Salford Health Informatics Research Environment (SHIRE) in Salford University. Their task is a technical one of quality assurance to the system representation. Professor Stephen Kay and Dr. Nicholas Hardiker provided expert guidance and support from medical informatics perspective for ontology development.

The other group includes of domain knowledge experts, lecturers who are teaching and practicing acupuncture in Salford University. University lecturer Linqing Cai and Lei Zhao are experts in acupuncture and have been working in both OM and CM clinical practice and teaching. They provided advice on scope, domain knowledge, reference textbooks and reviews on ontology structures during the development through
informal discussions and meetings. The domain experts commented on the ontology developer’s interpretation and elicitation of concepts from key texts.

The experts involved in design and development did not participate in evaluation in order to obtain independent assessment from evaluation. Identifying experts for the ontology evaluations is discussed in Chapter 6.

**Identify source of knowledge**

The acupuncture theories and concepts in ontology come from international textbooks as the main input knowledge source. The principles of selection for books are:

a) The theories and concepts in textbook are international or national standard;

b) The theories and concepts present the original meanings of acupuncture in Chinese Medicine or Orthodox medicine;

c) The theories and concepts present the latest knowledge development of Acupuncture.

These form the base of the resource, but more can be added during the development. (See book list in appendix I)

Existing ontologies are another important resource for ontology construction. We have considered existing top-level ontologies and domain ontologies. Sowa’s ontology provides top-level ontological distinctions. Instead of constructing an anatomy ontology from scratch, we reused the FMA (Foundational Model of Anatomy) ontology. Two top-level distinctions in FMA are physical vs. abstract and material vs. nonmaterial structure. These distinctions offer clear semantic meaning of human body structures. The distinctions exist in FMA and require a careful analysis of some acupuncture concepts related to human body, such as acupuncture point and meridian.

**Identify development environment**

Ontology development can be greatly facilitated by using ontology editing tools. We have discussed Protégé system and its OWL plug-in at the beginning of this chapter. At the time when we started the ontology development, the OWL languages and OWL plug-in were not available in Protégé. The project started in frame-based approach. Later, we converted our initial stage ontology into OWL language based ontology. The analysis of system is discussed in section 4.4.3.1 from distinction approach.
4.3.2 Categorise - stage: ontology kick off and refinement

The main activities for constructing ontology happened at this stage. The decisions and direction are discussed as responses to the methodology review in chapter 2.

Principles and criteria

The most widely adopted principles and criteria were discussed in chapter 2. These same principles and criteria were applied for our ontology development, and they also form part of ontology evaluation criteria (chapter 6).

Construct ontology manually, semi-automatically, or automatically

It is tempting to construct the ontology automatically or semi-automatically. It has been suggested by literature that the main benefit of automatically or semi-automatically approach is that it advances collection of terms for ontology construction. However, it is important to distinguish ontology which is featured with formal is-a relationships and distinctions among ontology, term glossary, thesauri and schema. We believe the full automatic approach maybe suitable for a glossary which does not require formal is-a relation in hierarchy. Therefore, in this study, it is still a necessary stage to ensure correct semantics of intended meaning being defined or expressed properly. The manual constructing is still necessary although it is time consuming. The DL reasoner provides automatic classifications based on the concept definition from the researcher.

Top-down, bottom-up and middle-out approach

Top-down, middle-out or bottom-up are three different approaches to build up the hierarchies in ontology. Stability and spotting commonality are extremely important for top-down approach. However, there is no well-accepted top-level ontology and it is difficult to achieve particularly when the scope of ontology is broad. The bottom-up approach will be difficult to scope until finally build to top of hierarchy.

We proceed in a middle-out fashion where the most fundamental concepts were defined first before moving on to more abstract and more specific concepts. The approach includes identifying the key concepts and relationships in the domain, terms for representation of concepts and relationships.
Categorise

From the philosophical analysis of ontology to different views of ontology in informatics, different methods have been applied for ontology construction and content analysis, which were discussed in chapter 2. The process of construction of hierarchies in ontology is categorization as Sowa (2000a) argued that

'The categories are first and foremost a classification of the ways people think and talk about the world. The nature of the world itself directly affects human perception and indirectly the categories they use. But the number of possible ways of viewing the world is far greater than the total number of concepts that anyone has ever conceived.' (p.88)

Because the ontology is constructed in description logics, the categorisation of concepts and roles has to respect the semantics of description logics. In turn, description logics services can facilitate the computation of hierarchies. However, the results of classification are depended on how concepts are defined.

We have extended Sowa's distinction approach to provide meta-level analysis for domain knowledge and ontology construction in particular system and language. We name this approach, 'flexible sets of distinctions', which enables the representation of different paradigms in one ontology. The detail of proposed approach is discussed in the following section 4.4.

4.3.3 Test - stage: evaluation

The evaluation consists of software program validation, observations and interviews with domain experts from both disciplines of medical informatics and health care. Because the nature of this evaluation is qualitative with respect to the research, the interview evaluation deployed techniques such as 'talking aloud protocols' to taping audio and video in a laboratory setting. The research findings came from both the researcher's personal experience of developing the ontology and assessments from domain experts who had validated the findings.

One perspective of evaluation is concerned with the correctness of the system and the ontology, i.e. getting it right. Therefore, the role of health informatics experts is
to check whether the constructions of ontology had followed principles of ontology development and the system and language constructs are applied properly.

The other perspective is concerned with validity of the ontology, i.e. whether or not it had represented the knowledge faithfully; whether or not the system and ontology are able to improve understanding of the domain knowledge; and whether or not such a development can assist those who need it for clinical benefit. Experts from both CM and OM evaluated the ontology by inspecting it, observing knowledge browser and query applications.

4.4 Flexible sets of distinctions approach

4.4.1 Flexible sets of distinctions

Peirce and Sowa inspired the idea of a ‘flexible sets of distinctions’ approach. It is Peirce’s semiotics and Sowa’s two-way and Three-way distinctions provided the foundation of this approach. We go further by arguing that the flexibility of applying different sets of distinctions as a means to enhance knowledge sharing overcomes the heterogeneities for better knowledge representation.

The distinctions employed to describe the world is based upon different understanding of the universe. The continuing development of general and particular domain knowledge enriches our understanding of the world and ourselves. In ‘Ontology’, the top representation of the universe, is termed entity or ‘thing’. An ontology is about ‘thing’; distinctions give meaning to ‘thing’. From the whole universe to a concept, there are different sets of distinctions to describe ‘thing’. We make a judgement of what ‘thing’ is upon our understanding, however, by our very current knowledge of the universe. Along with the rapid changing world, the meaning of terms that projects to the real world is changing as well. The human body can be described from different perspective and granularity, from body part and organ to cell, molecular and gene, etc.. Because of the development of knowledge, we are bound to have an increasing number of distinctions about ‘thing’.

Information systems mainly deal with symbolic ‘thing’ which reflects ‘thing’ in the real world. Ontologies are designed to work at the knowledge level (Newell 1982) and the terms or symbols used in an ontology should not be a major issue. Symbols or terms carry the intended meaning of representation, which is Peirce’s sign from a
semiotic point of view. Those symbols and terms are bearers, which deliver the meaning of concept and its attribute in ontology. Whenever we try to define a category or concept, a symbol or term is assigned to that concept. If it starts to mean a different concept, we have to give it another symbol or term from vocabulary. In this way, consistency is maintained. However, reducing inconsistency and redundancy by increased vocabulary could result in considerable effort for future maintenance and integration considering the experience from maintaining and mapping different controlled vocabulary.

In ontology, 'thing' generally appears as a taxonomic tree of categories from very general and domain-independent at the top level to increasingly domain-specific. Different sets of subcategories result from different criteria for categorisation. Categories can be organised by one or more distinctions.

An ontology should reflect the complex world and the incomplete knowledge of it. Several top-level ontologies, such as IEEE SUO\(^1\), Cyc\(^2\), provide a robust starting point for constructing ontology. They reflect different intentions and understandings. It is difficult to have a standard and unified view of 'thing'. However, it is possible that an ontology can present different perspectives and understanding. By choosing different sets of distinctions, we can present the theory and knowledge we are interested in at the time. Thus in our research, the meaning of a concept can be narrower or broader according to the different sets of distinctions in ontology. All subsumptions confirm the 'is-a' relationship. The applications based on an ontology will have their choices of distinctions accordingly.

As it was recognised by Sowa (2000a), no fixed collection of distinctions or categories is likely to be adequate for describing all things for all time. We go further by arguing that flexible sets of distinctions are needed for modelling domain knowledge from different paradigms. In principle, the number of classification criteria and distinct subtypes are unlimited because the number of possible dimensions along which to develop subcategories can not be exhaustively specified (Chandrasekaran et al. 1999).

Although a complete definition can use the full power of DL reasoning service than partial definition, 'it is hard to model complete definitions correctly, especially because it is easy to miss certain aspects of a class when creating the definition, in

\(^1\)IEEE SUO. Available from http://suo.ieee.org
\(^2\)Cyc. Available from http://www.cyc.com
which case the definition is incorrectly labelled as complete. ... Correct and complete modelling cannot be guaranteed in general in such an open and distributed environment as the Web' (Bruijn et al. 2005).

Furthermore, Cocchiarella (1996) contended that concepts do not exist independently of the capacity humans have for language and thought, whereas natural properties and relations do. The traditional method to define concepts is by necessary and sufficient conditions. This theory views concepts as lists of bits of knowledge: the necessary and sufficient conditions for something to be an instance of that concept. One major problem with this approach is the assumption that we can establish a mutual definition of a concept (Saeed 2003 p.35).

Ideally, ontology should be goal free and as general as possible. However, we cannot avoid intention, purpose, and task in hand when designing and constructing ontology. These conditions can make ontology less reusable. To rectify the situation, the flexible sets of distinctions has the advantage to accommodating purposeful or intentional design, allowing room for alteration and development over time.

For example, an ontology of a house can represent different perspectives, such as a home owner and a construction engineer. A home owner is concerned with the numbers of bedrooms, sizes of the lounge, interior of the house, etc.. Whereas, a construction engineer pays attention to the building blocks such as bricks, walls, layout of rooms. Yet, they see the same house. If an ontology can be designed to represent the same entity - the house from both perspectives, then, it is shared between different purposes, making it more reusable.

We use two-way and three-way distinctions as methods to analyse domain knowledge so as to identify the key distinctions. There are general and domain specific distinctions. Both are the content of flexible sets upon which we can choose for ontology construction.

Two-way distinction looks at “thing” from different dimensions or axes. As it is shown in figure 4.2, the concept can be analysed and constructed into an ontology along the axes or dimensions of $A_1-A_1'$, $A_2-A_2'$, $A_3-A_3'$, $A_n-A_n'$. The choices of axis can vary, like physical-abstract, pathological-physiological, Yin-Yang.
Three-way distinction (figure 4.3) analyses concept from perspectives of dependency, relation and context. Three-way distinction is a meta-level distinction for generating new categories by viewing entities from different perspectives (Sowa 2000a p.61). These distinctions are Peirce's principle for Firstness, Secondness and Thirdness. As Peirce (1891) explained, 'First is the conception of being or existing independent of anything else. Second is the conception of being relative to, the conception of reaction with, something else. Third is the conception of mediation, whereby a first and second are brought into relation'.
distinction analysis, there are some common sets of distinction across two paradigms, such as location, lasting time, etc. For theory, the two-way distinctions are physiological and pathological in orthodox medicine theory and Ying Yang in Chinese medicine. However, Yin-Yang and physio-pathological are not related. They exist independently in their own paradigms.

From three-way distinction analysis, pain as a finding is Firstness. There are two sets of distinctions to describe the finding. Headache or back pain are Secondness as they refer to an anatomic entity, head or back. What have caused pain, such as vascular disorder, inflammation, or intervertebral disc prolapse are diagnoses in OM medical theories. In CM, concepts such as Qi stagnation, Blood stasis, are diagnosis. The diagnosis is Thirdness, which brings the Firstness - finding of pain, and the Secondness headache or back pain together. Diagnosis is bound to different theories and is subject to the agents who analyse the condition.

4.4.2 Experience of flexible sets of distinctions approach

The challenges we have discussed in chapter 2 mainly reflect the different distinctions that exist in different paradigms. Before analysing the domain and constructing ontology, we need to analyse the system embedded for developing ontology. As Davis and colleagues (1993) argues that ‘commitment to a particular view of the world starts with the choice of a representation technology and accumulates as subsequent choices are made about how to see the world in these terms’. The distinctions in the system constrain how we represent the domain knowledge.

4.4.2.1 Analysis of system and language

The following analysis can also apply to the OWL plug-in in which the terms are the main difference. The Class and Slot in frame-based Protegé are called OWLClass and Property in OWL plug-in. The Facet in frame-based Protegé is not applied to OWL plug-in.

The analysis of Protégé system in terms of two-way distinctions is shown in Table 4.1. Classes and Subclasses represent a two-way distinction of generalization and specialisation. Class is a generalization of ‘thing’ and its subclass is its specialisation. The distinction of Meta-class and Class is a set and member of a set. Class is also an instance of Metaclass. Same relationship applies to Class and Individual. Slot is divided
into own slot and template slot. There are three axes to distinguish them. One distinction is where they come from. Own slots come from a meta-class' template slot whereas Template slots come from the class to which they attached. The second distinction is what they are describing. Own slots describe the properties of an object represented by the frame (an individual or a class). Template slots describe properties of the class' instances. The third distinction is inheritance. Own slots attached to a class are not inherited by the class' subclasses or propagated to its instances. A template slot attached to a class is inherited by its subclasses.

<table>
<thead>
<tr>
<th>Protégé Concepts</th>
<th>Two-Way Distinctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class(OWLClass) vs. Subclass</td>
<td>generalization vs. specialization</td>
</tr>
<tr>
<td>Class(OWLClass) vs. Individual</td>
<td>set of individual vs. individual</td>
</tr>
<tr>
<td>Metaclass(owl:Class) vs. Class(OWLClass)</td>
<td>set of class vs. set of individual</td>
</tr>
<tr>
<td>Class (OWLClass) vs. Slot (Property)</td>
<td>Entity vs. its attribute or property</td>
</tr>
<tr>
<td>Individual vs. Facet</td>
<td>Individual vs. its constraint, axiom</td>
</tr>
<tr>
<td>Slot vs. Facet</td>
<td>attribute or property of class vs. attribute or property of Slot</td>
</tr>
<tr>
<td>Own slot (Annotation property) vs. Template slot (property)</td>
<td>where they come from; what they are describing; inheritance</td>
</tr>
</tbody>
</table>

Table 4.4 Protégé system and two-way distinctions

Class and Individual in Protégé system are the TBox and ABox of a knowledge-based system which we have discussed in chapter 2. The analysis from three-way distinction may help the understanding of how these Class and Individual are specified in Protégé (Table 4.5).

From a three-way distinction approach, Class, Slot (attribute, property) and Meta-class are analysed at Class level in Protégé. From a modelling point of view, Class, as abstract entity, is Independent. Slot (attribute, property) as Secondness, specifies the characteristics of the entity or relations to another entity. It is the Meta-class which brings Class and Slot (attribute, property) together, therefore, it is Thirdness. In this way, the characteristics of Class are specified by its Slot (attribute, property) and restrained by its Meta-class.

Similar analysis in context of Individual level can be applied to terms for Individual, Individual and Class (concept) in Protégé system from a terminology point of view. Individual, which represents a single thing in reality, is Firstness. Term, as
symbolic representation of individual in Protégé, is Secondness. Class (Concept), specifies relation between the symbolic representation (name or descriptions of individual) and the individual, is Thirdness.

<table>
<thead>
<tr>
<th>Protégé Concepts</th>
<th>Three-Way Distinctions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class (OWLClass)</td>
<td>P (x)</td>
</tr>
<tr>
<td>Slot (Attribute)</td>
<td>R (x, y)</td>
</tr>
<tr>
<td>Metaclass (owl:Class)</td>
<td>M (x, y, z)</td>
</tr>
<tr>
<td>Individual</td>
<td></td>
</tr>
<tr>
<td>Name (Term, Symbol)</td>
<td>R (x, y)</td>
</tr>
<tr>
<td>Class (OWLClass)</td>
<td>M (x, y, z)</td>
</tr>
</tbody>
</table>

*Table 4.5*  Protégé concepts and three-way distinctions

Based on distinction analysis, we apply the solution for implementing flexible sets of distinctions by two methods.

One method is that the choices of distinctions can be presented by slots which specify the common characteristics of Individual. In OWL plug-in, the characteristics of Individual are represented by properties of OWLClass.

Another method is to use slots of meta-class to present the choices of distinctions which specify the common characteristics of Class. In OWL plug-in, the annotation property is used to represent choices of distinctions i.e. paradigm of CM or OM. The result of this method is that the value of own slot (frame-based system) or annotation property (OWL langue system) will not be inherited by its subclass. The meta-class implementation has been used by ONTOCLEAN (Guarino and Welty 2002) for their philosophy and logic analysis of ontology. It facilitates the validation of taxonomies by exposing inappropriate and inconsistent modelling choices.

The above two methods are discussed and examples given regarding how they are implemented in design of ontology in Chapter 5.

4.4.2.2 Analysis of domain knowledge

The development of medical knowledge in one paradigm may have impact in another. So does the understanding of one phenomenon that raises new recognition. The
knowledge of human body has developed from simply naming parts to constructing structured representation of anatomy. Human anatomy is the foundation for all fields of medical study. These ontologies are valuable for anatomy knowledge sharing and reuse. CM covers broad knowledge from theory to practice. Several concepts such as acupuncture point, meridian, organ, etc. indeed need to refer to human anatomy. Here we demonstrate how we apply distinctions from both top-level ontology and domain ontology based on our flexible sets of distinctions analysis.

High level distinction in FMA ontology, such as physical-abstract, gives clear semantic meaning (Rosse and Mejino Jr. 2003). The formal analysis of “is-a” and “part-of” relationships provide support for the alignment of ontologies (Smith and Rosse 2004). To reuse anatomy ontology in our study, we need to confirm related distinctions such as physical-abstract, material-nonmaterial. It requires us to analyse Chinese medicine concepts by using those distinctions, then refer to anatomy ontology. However, these distinctions are not commonly used in CM. Yet, without analysing CM concept by those distinctions, the semantic link between CM and anatomic concept is vague. The analysis of physical-abstract, material-nonmaterial clears the question of which body parts, physical material or abstract nonmaterial, should be referred to.

We treat acupuncture point, meridian as subclasses of ‘thing’ and define them as neutral as possible at this level. All subclasses confirm the ‘is-a’ relationship. Acupuncture point and meridian so far have not been proven in any way that they are of physical existence. Here we consider them as abstract entities. They imply the place where we can locate point or meridian. Through this physical location on human body we can insert needle or apply other stimulation to achieve therapeutic effect. We establish reference to human anatomy for acupuncture point and meridian which can be viewed from different perspectives. For CM practitioner, there are many specific distinctions, such as element property, on or off meridian, etc. to distinguish different kinds of acupuncture points. Anatomical structure of nerves and muscles around the point carry practical meaning to the orthodox medicine practitioner. Although CM and OM have different perspectives of acupuncture point and meridian, the location and function of acupuncture point can be shared and communicated.

4.4.2.3 Meta-model representation of different perspectives

We have discussed the requirements and challenges to represent both paradigms of CM and OM in one single ontology in chapter 1 and chapter 2. Different or similar
knowledge of the same reality of health care needs to be represented in a consistent manner.

We have analysed and observed the Protégé systems from the view point of distinctions. Although the logical constraint on property value by universal and existential restriction provide clear semantics, it is still of little clarity on what ground class and property are connected. The mutual definition of concept is difficult to achieve regarding to different perspectives. The same thing can be defined differently from different perspectives. One may argue that they are different concepts represent the same thing from different perspectives. However, this conventional approach offers little help to our study for knowledge sharing between two paradigms. More importantly, it is not economic to generate significant number of concepts which indeed have to be represented in different term strings, leading to future difficulties to map, merge or align ontologies developed from merely one perspective.

Flexible sets approach is implemented through meta-ontology which represents different perspectives. In this research, the focus is knowledge sharing across CM and OM. Perspectives of CM and OM can be represented as values of paradigm for meta-attribute. Of course, if the perspectives are broader or more complex, a small ontology of paradigms can be applicable. Because the paradigm is own slot of class which is defined by metaclass, the subclass will not inherit the value of paradigm of its superclass. After we moved to OWL language representation formalism, the paradigms of CM and OM are represented by annotation properties. These values mark up paradigm specific concepts or properties. The contradiction can be determined manually or automatically identified by DL reasoner. In this study, the classification by reasoner is against both perspectives.

There are two ways to apply different perspectives in applications. One is to have one single ontology then select appropriate perspectives based on applications and contexts. In evaluations of the research, we demonstrate different views of ontology and answer queries by selecting different paradigms for different user groups.

The other way is to split the single ontology to different ontologies by perspectives. It is technically possible to generate different ontologies based on meta-properties, which can also reduce the size of ontology. The flexible sets of distinctions approach is different to segmentation discussed by Seidenberg and Rector (2006) for large size ontologies in which segmentation is achieved by traversal, property filtering and boundary extraction.
As the analysis of distinctions of two paradigms is discussed in previous section, acupuncture point is defined from different perspectives. The concepts and properties are marked up by CM or OM paradigm. There may be other general and specific distinctions that can be applied to acupuncture point. Those are open to other ontologies and application developers. In an ontology application, concept and attributes can be added or removed according to paradigms. The ontology can be showed in three different views, CM, OM or both as we have demonstrated in evaluation for different experts from CM and OM.

4.4.3 Flexibility and extensibility of model

It is not easy to establish a mutual definition of a concept. Some concepts might never have an agreed definition from different perspectives. It is difficult to define what distinctions are essential for ontologies from top-level ontology to domain ontology. To make ontologies reusable more general distinctions or flexible sets of distinctions need to be applied. One concept can have different meanings according to the different perspectives and the perspective determine the meaning of a concept. The different perspectives can be represented by different sets of distinctions. Categories are generated from different sets of distinctions. As we have discussed, the consistency and formal ‘is-a’ of ontology have to be maintained. The different meanings must not have logical contradictions. Therefore, understanding of the same entity from different perspectives can be shared.

Ontologies can only represent the knowledge of our understanding at one time. It is common to introduce new distinctions and perspectives along the evolution of ontology. To ensure the consistency of ontology, new distinctions and perspectives need to be checked with existing distinctions and perspectives in ontology. Because the flexibility is introduced by flexible sets of distinctions, it is much easier to accommodate new additions. The version control and track of changes of ontology is out of scope of the study.

4.5 Summary

In this chapter we explained the methodology for constructing one single ontology across CM and OM. We followed the approach of On-To-knowledge to which
modifications were made to suit the research. The activities of ontology development were grouped under Identify, Categorise and Test. We focused on the ontology development environment and the flexible sets of distinctions approach. The terminology of ontology development environment was introduced, then the system was analysed, and the distinction approach was illustrated.

Furthermore, we have discussed how to apply flexible sets of distinctions to analyse domain knowledge and system. Our experimental research of constructing acupuncture ontology is to prove the concepts of flexible sets of distinctions approach as a means to analysing and constructing ontology. Knowledge of acupuncture was represented from perspectives of Chinese medicine and orthodox medicine. The dimensions and axes, i.e. sets of distinctions are analysed by two-way and/or three-way distinctions.

The distinctions act as building blocks to ontology, however, they should not be fixed, especially when considering the needs of extending and merging ontologies. It has been recognised that no fixed distinctions can be sufficient to describe all things for all time. Furthermore, it is difficult to establish mutual definition of concept. By keeping our choices of distinctions open, new sets of distinctions can be introduced after checking with existing distinctions. Different distinctions will bring the distinct meaning to the concept from different perspectives at different levels.

It is true that more application-specific decisions of choosing distinction sets make the final ontology less shareable and reusable. Clearly defined generic and flexible distinction sets can assist ontology development and reuse. Distinctions from top-level ontology and different paradigms play different roles for knowledge sharing across paradigms. Top-level distinction provides common understanding whereas paradigm specific distinctions present particular knowledge about the same domain from different perspectives.
Chapter 5

Acupuncture Ontology

The aim of this chapter is to discuss the acupuncture ontology, a major outcome of the research. We introduce the main content of terms, concepts, theories and logics in the ontology. The technical solutions are discussed by examples and are followed by a discussion of how we have overcome the challenges posed by the single representation of the knowledge from the two paradigms.

5.1 Introduction

A few ontologies have been developed for orthodox medicine with some content of acupuncture. However, the original meanings of acupuncture concepts were not well represented in them. Some ontologies have been developed following the UMLS approach for Chinese medicine. However, none to our knowledge represents acupuncture knowledge from both paradigms in a single formal ontology.

We address the requirements, current developments and challenges for acupuncture ontology in chapter 1 and chapter 2. The design and implementation of approach have been discussed in chapter 4. The acupuncture ontology is a major outcome of this research and is intended to provide different views of acupuncture knowledge. We concentrate on the content of the ontology and the technical solutions by giving examples from the ontology. The discussion is related to how the challenges were met.
5.2 Acupuncture ontology

5.2.1 Scope

The purpose of the acupuncture ontology is to represent acupuncture knowledge and to facilitate knowledge sharing between CM and OM. The ontology covers core concepts of acupuncture knowledge which include traditional Chinese acupuncture theoretic entities of acupuncture point, meridian, Yin Yang, Qi, Five elements, and Visceral manifestation. Anatomy and medical acupuncture are the representation from the OM perspective, including segmental point and trigger point in medical acupuncture, and the imported Foundation Model of Anatomy. For the purpose of demonstration and evaluation, some clinical findings relating to pain and depression are included. There are over 1500 concepts and 30 properties in the ontology.

Some classifications of ontologies have been discussed (Falasconi and Stefanelli 1994; Stevens et al. 2000; Jurisica et al. 2004) and they are based on:

- The approach: static ontology; dynamic ontology; epistemic ontology
- Being context dependent or not: domain ontologies, common sense ontologies, meta-ontologies, and task ontologies
- The structure of conceptualisation: Terminological ontologies; information ontologies; knowledge representation ontologies
- The subject of conceptualisation: application ontologies, domain ontologies, generic ontologies (top level ontologies), internal ontologies, social ontologies, and representation ontologies

In this study, the ontology can be classified as a domain ontology.

5.2.2 Content

The contents of the ontology reflect the semiotics view of ontology. Terms are symbolic representations which denote the meaning of concepts. Concept is the representation of intended meaning which denotes or refers to entity in reality. As we have discussed in chapter 2, concepts are underpinned by theories and logics from CM and OM paradigms which determine what is important, legitimate, and is a reasonable view of an entity in reality. Therefore, it is essential to represent knowledge from these three dimensions: terms, concepts, theories and logics.
5.2.2.1 Terms

The vocabulary view of the ontology suggests the importance of symbolic representation in informatics and health care which was discussed in the research background. Standard terminologies or controlled vocabularies should be used in ontologies wherever possible. We therefore followed the standard international nomenclature and terminology from WHO for construction of acupuncture terms.

In 1991, WHO scientific group proposed the standard international acupuncture nomenclature which is based on four elements:

‘(1) The English translation of the Han character name of each meridian.
(2) An alphanumeric code derived from the English translation of the meridian names.
(3) The Chinese phonetic alphabet (Pinyin) names of meridians and acupuncture points.
(4) The Han character names of the meridians and acupuncture points.’ WHO (1991)

The original names of acupuncture meridians and points are in Chinese Han characters. Pinyin is the standard Chinese phonetic alphabet for pronunciation of Han characters and English is the most commonly used language in international communications and publications. For example,

English name of meridian: Large Intestine
Alphanumeric code: LI4,
Chinese phonetic alphabet (Pinyin): hégu
Chinese Han character: 合谷

Extra points which do not belong to any meridians are represented by alphanumeric code consisting of a general prefix, Ex (denoting “extra points”), followed by an alphabetic code indicating the region: HN (for head and neck); CA (chest and abdomen); B (back); UE (upper extremity); LE (lower extremity), etc.. The points are numbered from higher to lower levels for head, neck and trunk regions; from proximal to distal for the upper and lower extremities; and, if at the same level, from the medial to the lateral (WHO 1991).

A recent publication of ‘WHO International Standard Terminologies on Traditional Medicine in The Western Pacific Region’ (WHO 2007) organized terms by
code, English term, Chinese Han character and definition/description. The nomenclature of acupuncture remains the same.

In this ontology, the main language is in English. Terms of classic acupuncture point are presented in alphanumeric code, Chinese Han character and Chinese phonetic alphabet (Pinyin). Some terms, which do not have well accepted mapping terms in English, are only presented in Chinese phonetic alphabet (Pinyin) and Chinese Han character. For simplicity, some letters with pronunciation marks in Pinyin are replaced with English letters, such as e for é, u for ū. Definition/description are represented in the formal language of Description logics and natural language respectively.

5.2.2.2 Concepts

**Acupuncture points (针穴位)**

Acupuncture point is a core concept for both Chinese medicine and Orthodox medicine. Acupuncture point includes classical points on fourteen meridians (经穴), extra points (奇穴), Ah Shi point (阿是穴), segmental points (or Spinal acupuncture points), and trigger points.

There are 361 classical acupuncture points which are belong to 14 meridians. Among these points, some are classified into special function points which are Back Shu point (背俞穴), Cleft point (郗穴), Confluent point of Extraordinary Channel (八脉交会穴), Connecting point (络穴), Five Transporting point (五输穴), Front Mu point (募穴), Influential point (八会穴), Primary point (原穴), and Xia He point (下合穴).

There is no fixed and accurate number of extra points because they have been proposed and evolved over time. WHO (1991) named 48 extra points by the following criteria:

‘(1) The point should be in common use.
(2) They should be considered clinically effective.
(3) They should have a clear anatomical location.
(4) They should be at least 0.5 cm from a classical acupuncture point.
(5) If an extra point has the same name as an existing point, a prefix must be added to it.’
Ah Shi point is a point where the patient feels pain or tenderness when it is pressed. There are no specific name and fixed locations for this kind of points. Ah Shi point does not have the same location with respect to classical acupuncture points and extra points.

The use of myofascial trigger points and segmental acupuncture points underpins the Orthodox medicine approach to acupuncture practices (Macdonald 1998; Filshie and Cummings 1999, p.36).

Segmental acupuncture points are divided by their segmental innervation supplied by peripheral nerves.

Trigger point is understood as 'spots of local tenderness in a palpable tense pad of muscle fibres within a muscle that is shortened and weak' (Simons et al. 1999, p.5). A trigger point produces an area of referred pain or tenderness which is characteristic of that trigger point. The mechanism of the action is not known but it seems that it creates an area of sensitised nerve with increased metabolism and reduced circulation. Eventually the muscle tissues degenerate in the area. The trigger point is further divided into active myofascial trigger point, latent myofascial trigger point, and potential myofascial trigger point (Alvarez and Rockwell 2002). Some clinical observation and research, e.g. saline injections into trigger points to identify where pain refers to, have produced detailed anatomical maps which can be extremely helpful in identifying the best places to look for trigger points.

**Acupuncture Meridians (经络)**

In the ontology, we represented fourteen meridians (经) and different classifications of meridians such as classifications by Yin-Yang, Hand and Foot, etc.. Fifteen collateral vessels (络) which do not have acupuncture points are included in the ontology. Some concepts, which are deemed as being less relevant to the study, are not included because of the time constraint, e.g. the twelve meridian divergences (the divergent passage of the twelve meridians going deep in the body).

**Anatomical knowledge representation**

We have discussed Foundation Model of Anatomy in the last few chapters. Because the FMA ontology is developed in the frame-based approach in Protégé, we can not use it directly in our OWL-DL language based ontology. The issues of how FMA should be
represented in OWL language is discussed and suggested that FMA should be represented in OWL-full (Dameron et al. 2005).

We developed a ‘mock’ FMA ontology in the OWL language. The ontology only recreates a very small portion from several hierarchies which are relevant to the study. It includes part of hierarchies of segmental innervation of nerve, organ, organ part and its subclasses of nerve, artery and vein, dimensional entity and body surface subdivision, whereas the definitions and properties are not represented. The mock FMA ontology is standalone and is imported into the acupuncture ontology. We use ‘fma’ as the prefix for all concepts imported from the mock FMA ontology.

5.2.2.3 Theories and logics

Thagard and Zhu (2003) describe the conceptual and explanatory differences between OM and CM. They argue that the linguistic, conceptual, ontological, and explanatory impediments can be largely overcome, and that the dramatic differences between OM and CM do not pose insurmountable barriers to the rational evaluation of acupuncture. However, it is still paramount to represent the precise theory and logics in the context while holding the importance of knowledge sharing in mind. Some relevant theories are illustrated as following:

Yin (阴) Yang (阳) is, as one of the fundamental theories for CM, highly abstract. Kaptchuk (2000) explains that:

‘Yin Yang theory is based on the philosophical construct of two polar complements, namely Yin and Yang. They are convenient labels used to describe how things function in relation to each other and to the universe. They are of continuous process of natural change. Also in this system of thought, all things are seen as part of a whole.’

Yin and Yang are used to describe polarities of function of all phenomena, including those of psycho-physiological system of the organism (Kerry 1991). However, it has not been given significance within the orthodox medical framework.

The term Five Phases (五行) used to be translated as Five Elements which refers to Wood, Fire, Earth, Metal and Water. They ‘are not basic constituents of nature, but five basic processes, qualities phases of a cycle or inherent capabilities of change of phenomena’ (Maciocia, 1989, p.16). They are metaphoric and describe the qualities and relations of phenomena.
Acupuncture point is a widely used concept in both paradigms. Acupuncture point 'is-a' surface topography which 'is-a' anatomical spatial entity in SNOMED CT. Acupuncture points are seen by orthodox medicine as corresponding to physiological and anatomical features such as segmental nerve innervation (Vickers and Zollman 1999). From Five Phases theory, some acupuncture points correspond to individual phase and they are named Well point, Stream point, Spring point, River point and Sea point.

Acupuncture meridians are divided into Yin meridians and Yang meridians according Yin Yang theory. The meridian, as a distinctive theoretical term, is largely ignored by OM and is not included in SNOMED CT. The relationship between acupuncture points and meridians is not represented in SNOMED.

5.3 Technical solutions

5.3.1 Class, Individual, Property

The notions of 'class' and 'individual' correspond to TBox and ABox in knowledge-based system architecture. In general, class represents concepts of domain knowledge in ontology, whereas individual are assertions of facts in a knowledge base. The characteristics and relationships are modelled as properties of a class. We have discussed the Protégé system at class and individual levels from the distinction approach in Chapter 4. The concepts and characteristics are represented by classes and properties. In the following sections some examples of representations through class, individual and property are discussed.

Acupuncture points and meridians are represented as classes in the Protégé system. The relation of point to meridian in CM is represented as the property 'belong to'. The relation of point to non-physical anatomical entity such as nerve innervation is represented as the property 'local anatomy' for OM. To achieve knowledge sharing between the two paradigms, these two properties are applied at the same time to a single concept. For example, LR3 point, by 'local anatomy' it is a L5 (the 5th lumbar segmental) point supplied by the 5th lumbar nerve innovation; by 'belong to' it is a Liver Meridian Point; and it is also a Primary Point and Stream Point. More discussions of formal description logics definition are offered in the next section.
The relations of points to physical and non-physical entities were represented by properties of 'local anatomy' and 'segment supplied by'. The 'local anatomy' property describes the physical material anatomical structures, such as muscle, nerve, around or under the acupuncture needle when it is inserted into an acupuncture point. The property represents the OM perspective from a pure anatomical view of human body. Hence, the relationship is between non-physical acupuncture point and physical anatomical entity.

The 'segment supplied by' property points to the segmental innervation of non-physical anatomic entity from the FMA ontology. It is about the knowledge of the nerve system concerning which anatomical structure is supplied by which segmental innervation. This property links non-physical acupuncture point to non-physical segmental innervation, which is different to the property of 'local anatomy'.

The feedback from health informatics experts suggested that it might not be necessary to have two properties to relate anatomical entity. Although these properties represent clear distinctions between non-physical and physical anatomical entities, the distinction can be reflected by constraints on values of one single property. We followed the suggestion and replaced 'segment supplied by' with 'local anatomy'. In addition, the range of 'local anatomy' property was changed from fma:Material physical anatomical entity to fma:Anatomical entity which included both physical and non-physical anatomical entity as subclasses. The non-physical anatomical entity includes segmental innervation.

The relationship of diagnosis between CM and OM is one area of our ontology. In our approach, a diagnosis from CM such as a Pattern is not directly related to a diagnosis from OM such as a Disorder. There is no directly mapping, such as Pattern A = Disorder B, between them. Furthermore, they are not directly linked through a property. This is because a diagnosis reflects theories and logics from a particular paradigm which has its own diagnostic classifications. Different paradigms have to be connected through ‘individual’ phenomena of clinical manifestations.

Given the numerous phenomena, there are many ways to classify the phenomena for our clinical purpose. Clinical manifestations are phenomena gathered and grouped based on CM or OM knowledge. e.g. pulses can be grouped differently in CM and OM. Distinction between phenomena and clinical manifestations is that phenomena has no engineered reasoning and justice. However, clinical manifestation is a selective collection of phenomena.
For example, headache can be caused by vascular disorder. However, this diagnosis of vascular disorder does not instantly form any diagnosis in CM. For a CM practitioner, headache can be identified through differentiation of Liver Yang Rising or other patterns.

In general, individual should not be part of the ontology. However, in terms of knowledge-based system, it is necessary to have individuals for applications. To demonstrate the relationship, we included some individuals and used those individuals for queries and demonstrations in the ontology’s evaluations, although only a small number of diagnoses of Pattern from CM and disorders from OM are included in the ontology.

The modelling clearly represents the differences between the two diagnostic approaches and they cannot be directly translated into each other. Clinical manifestations as individual can have different diagnosis from CM and OM. The modelling reflected the importance of respect to the theories underpin the practice as it was emphasised by (White et al. 2001; Kaptchuk 2002; Birch 2004).

5.3.2 OWL

OWL is the standard language to share ontology for the semantic web. It is built on top of RDF which specifies resources on web. All concepts and properties are unique on web as they are present as URI with name space.

The OWL DL provides the power of description logics. The OWL DL makes the automatical classification possible by DL reasoners.

As we have discussed in previous chapters, the Protégé OWL plug-in provides transformation from description logics to OWL DL. The view of OWL DL syntax is less important for our ontology development. However, it is still useful to understand basic OWL syntax. The following is an OWL DL expression for Well point exported from the OWL plug-in.
The advantage of using a standard web ontology language is that it is sharable on the web. However, because OWL is in simple XML text format, the performances for accessing and computing data are not satisfactory for larger ontologies. Some database systems, such as Oracle (Oracle 2007), start to support RDF triple model storage which may improve the performance of OWL.

The lack of standard query languages for OWL is still an issue, although some queries languages have been proposed and developed to support data aggregation. This issue will be discussed in detail in chapter 6 as a query language was required for the ontology’s evaluation.
5.3.3 Description Logics

A brief history of description logics has been discussed in Chapter 2. We have also provided the basic terminology in Chapter 4 where we discussed the development environment. In this study, description logics, as a language for ontology, provides formal expressions for representing the specification of a concept. The followings are some examples of description logics definitions of concepts in the ontology.

5.3.3.1 Formal specifications in description logics

The definition of concept is specified in OWL DL language in Protégé. The OWL plug-in provides a user friendly interface to specify the description logic expression. It transforms the expressions into OWL-DL format. Therefore, users can focus on choosing appropriate DL constructs to represent the semantics of concepts. The following DL expression represents a formal definition of Well point which is a subtype of Five Transporting Point.

Well point ≡ Five transporting point
\n\n\n∩ (((∃ correspond to. Wood) ∩ (∃ belong to meridian. Yin Meridian))
∪ (((∃ correspond to. Metal) ∩ (∃ belong to meridian. Yang Meridian)))

The definition means that a Well point is a Five Transporting point which corresponds to the Element of Wood and belongs to Yin Meridian or corresponds to the Element of Metal and belongs to Yang Meridian.

The DL reasoner automatically classifies all 12 classic acupuncture under Well points rather than manually adding each point under Well point.

For example, LU11 (少商) is a Well point, which is defined as following:

\nLU11 ≡ Acupuncture Point
\n∩ (∃ belong to meridian. Lung Meridian)
\n∩ (∀ belong to meridian. Lung Meridian)
\n∩ (∃ correspond to. Wood)
\n∩ (∀ order number on meridian. Order11)
\n∩ (=1 order number on meridian)

We did not specifically declare that Lung Meridian is a Ying Meridian in definition of LU11. Lung meridian is a subclass of Ying Meridian in Meridian hierarchy.
The reasoner is able to classify LU11 as a Well point, resulting from comparing definitions of Well point and LU11. This is beyond simple term matching because Lung Meridian does not appear in the Well point definition.

Another example is GB41 which corresponds to the same Element of Wood, but it belongs to Gall Bladder Meridian which is a Yang Meridian. Then, GB41 point is automatically classified as a Stream Point defined as following:

\[
\text{Stream point} = \text{Five transporting point} \ni ((\exists \text{ correspond to Wood}) \ni (\exists \text{ belong to meridian Yang Meridian})) \cup ((\exists \text{ correspond to Earth}) \ni (\exists \text{ belong to meridian Yin Meridian}))
\]

5.3.3.2 Necessary and sufficient conditions

The property of ‘local anatomy’ is linked to OM ontology of anatomy as previously discussed. It is used for necessary and sufficient conditions for defining segmental acupuncture point whereas the same property is used for necessary condition for classical acupuncture points. Through this property, we introduce modern OM anatomical knowledge into traditional CM acupuncture. For example:

The necessary and sufficient conditions for LR3 point are expressed as following

\[
\text{LR3 point} = \text{Acupuncture point} \ni (\exists \text{ belong to meridian Liver Meridian}) \ni (\forall \text{ belong to meridian Liver Meridian}) \ni (\forall \text{ order number on meridian Order3}) \ni (=1 \text{ order number on meridian})
\]

The necessary condition for LR3 is expressed as following:

\[
(\exists \text{ local anatomy fma:L5}) \ni (\exists \text{ correspond to Liver Qi}) \ni (\exists \text{ correspond to Earth})
\]

In comparison, ‘local anatomy’ is the necessary and sufficient condition for segmental points such as L5. For example:

\[
\text{L5 point is defined as following:}
\text{L5 point} \equiv \text{Acupuncture point} \ni (\exists \text{ local anatomy fma:L5})
\]

The result of the classification by RACER is that LR3 point is a subclass of L5 point. Liver Meridian Point, Primary Point and Stream Point as show in following figure 5.1.
5.3.3.3 Constraints on properties

When and which constraints should be used is dependent upon the requirements of concept specifications. We discuss some constraints used in the ontology.

Universal vs. Existential

The classical acupuncture point should belong to one single meridian. They can not belong to more than one meridian. For example, Liver meridian point is represented as

\[
\text{Liver Meridian Point} \equiv \text{Acupuncture point} \\
\text{\quad } \bigwedge (\exists \text{belong to meridian.Liver_Meridian}) \\
\text{\quad } \bigwedge (\forall \text{belong to meridian.Liver_Meridian})
\]

By this definition, all liver meridian points can only belong to the Liver meridian, not any other meridians.

It is easy to test this universal constraint. LR3 is deliberately defined as belonging to Liver Meridian and Lung Meridian. We manually put LR3 under Liver Meridian point. It was seemingly correct before running the RACER. But the RACER identified that the definition of LR3 is inconsistent, resulting in an error message of inconsistency. It proved the point that proper definition and DL reasoning services can enhance the consistency of ontology. The manual maintenance of such consistency could be very difficult without DL reasoning services.
**Cardinality**

The order number on the meridian is a unique identifier for classical meridian points. Each classical acupuncture point must have one and can only have one order number. For example, LI4 a classic point is represented as

\[
LI4 = \text{Acupuncture point} \\
\cap (\exists \text{belong to meridian. Large Intestine Meridian}) \\
\cap (\forall \text{belong to meridian. Large Intestine Meridian}) \\
\cap (\forall \text{order number on meridian. Order 4}) \\
\cap (\neq 1 \text{order number on meridian})
\]

Property of 'order number on meridian' is a functional property which means that the value of this property is unique. However, as a functional property, it does not enforce that all elements of the domain have values which means it is valid for this property not to have a value at all. Therefore, to define the cardinality constraint as 1 makes the definition more accurate. It requires the property to have an order number which is the instance of Order 4 and it can only take one instance.

**5.3.3.4 Property characteristics**

There are three types of properties, object property, datatype property, and annotation property in OWL. The object property is the link of the triple of class-property-class. The datatype property also follows the triple relation. However, the datatype replaces the second class and it becomes a triple of class-property-datatype. The datatype property is not supported by the RACER reasoner. Therefore, the reasoner ignores the datatype specification when it classifies the ontology. For example, 'location description' property is a string datatype property which is not used for definition. The value of property is free text to describe the location of acupuncture point.

A property can be specified by five predefined characteristics (transitivity, symmetric, functional, inverse of, inverse functional) in OWL. They provide enhanced reasoning about a property (Smith et al. 2004). In the previous example, 'order number on meridian' is a functional property. It represents the fact that order number is a unique number for a classic meridian point.

The property 'belong to meridian' is specified as the inverse of the 'has point' property. An acupuncture point belongs to a meridian. Conversely, that meridian has the acupuncture point.
The property ‘contrast to’ is specified as a symmetric property. We use this property to describe that Yin and Yang contrast with each other.

### 5.3.3.5 Stated view and inferred view

In the Protégé OWL plug-in, the specification of a concept can be seen in two different views: stated (asserted) and the inferred view. The stated view includes all new properties and constraints of a concept. It also displays the inherited properties and values from its superclass. The inferred view is generated after running the DL reasoner RACER. It provides all inferred superclasses, properties and values. The differences can be easily noticed by following two screen shot of LR3 point before and after running RACER.

![Figure 5.2 Stated (asserted) view of LR3 specification](image1)

![Figure 5.3 Inferred view of LR3 specification](image2)
5.4 Overcoming the challenges

5.4.1 Top level ontologies

The importance of top level ontology has been well recognised and some top-level ontologies have been developed in philosophy and informatics. However, there is no agreement on a standard top-level ontology.

In this study, we took the middle-out approach for ontology development. Therefore, we identify what are the core concepts in the ontology first, then move towards both ends, the root and leaf nodes. The core concept Acupuncture point is not only specified in terms of meridians, but also anatomical peripheral nerve supplies. The foundational model of anatomy is used as reference ontology to bridge acupuncture knowledge in the two paradigms. The top level ontology of FMA is described by Rosse and Mejino Jr. (2003) that anatomical entities are distinguished in terms of whether or not they are physical in nature. All physical entities have spatial dimension, such as volumes, surfaces, lines or points, whereas conceptual entities have no spatial dimension. As discussed in the analysis of domain knowledge in Chapter 4, our approach requires analysis of CM concepts from the distinction of physical and abstract and use it then as our top-level ontology. Distinction is recognised as the central focus of an ontology and this distinction is independent of the observer’s viewpoint (Sowa 2000a, p.68).

Acupuncture points and meridians are core entities in this ontology. There is no scientific evidence as to whether they physically exist. Hence, they are modelled as conceptual entity from the distinction of physical and abstract entity. Acupuncture points and meridians are similar to virtual point and virtual line in FMA. Theoretic entities are abstract conceptual descriptions of phenomena in the real world. Metaphors have been intensively used in CM terminology. Wood, Metal, etc., of Five phases are used as qualitative descriptions for describing the characteristics and relationships of phenomena. These traditional CM concepts are modelled under the theoretic entity.

Clinical manifestations are modelled as physical entities to represent clinical phenomena. The distinctions between disorder, disease, syndrome, and even some clinical symptoms are not clear. Diagnostic findings are conclusions of analytic process based on certain theoretic framework. For example, a physical pathology finding of inflammation of lung confirms the diagnosis of pneumonia which can be categorised by
the agents, such as bacterial pneumonia. The diagnostic finding of ‘pneumonia’ is associated to physical clinical manifestations of cough, chest pain, fever, difficulty in breathing etc. and describes the inner relationships of these common clinical manifestations, the cause of infectious agents as well as body site of lung. The same condition can have a CM diagnostic finding of chest pain with pattern of full-heat of lung.

5.4.2 Name, label and term

From Aristotle’s philosophical discussions to modern informatics applications, the language issue is still important. The issues have been discussed around vocabulary, conceptualisation, reality and semiotics views of ontology, and syntax, semantics, and pragmatics of language in research background in chapter 2. The following two issues require solutions:

- Same term represents different meanings
- Same meaning can be expressed by different terms

Here we discuss the solutions from a syntax and vocabulary point of view. In our solutions, terms and concepts are separated. Therefore, one term can associate with different concepts and one concept can have several different terms. A constraint is set up by the Protégé system that all names for class and slot should be unique because class and slot are first class objects in Protégé. This is also true for OWLClass and the property in the OWL plug-in. The name for class is interpreted as unique identifier for concepts. For better readability, we use the underscore symbol to link words in terms. For example, Acupuncture_Point is the name for concept Acupuncture point.

Because name space is applied for the OWL based ontology, the concepts and properties are represented as URI. For example, the same term ‘heart’ is used in both the FMA and the acupuncture ontology. Within the Protégé system, they are distinguished by prefix of ‘fma’ for concepts from FMA ontology. There are two ‘heart’ s in this ontology, one name is ‘fma:Heart’ from FMA and another ‘Heart’ is created in this ontology for Chinese medicine. When this ontology is published on the web, these two ‘heart’ s will be presented as following:

http://www.shire.salford.ac.uk/ontology/Acupuncture#Heart
http://www.washington.edu/FMA#Heart.
Please note that FMA is not published in the OWL format at the time of our development. This URI is not the real URI for FMA ontology.

We have discussed ‘term’ with respect to the ontology content in earlier section of this chapter. Terms need to be represented in Chinese Han character and in English. The single no duplicated name can not fulfil the requirement. The standard construct rdfs:label in OWL can service the purpose to present different terms in different languages. As it is described in OWL language guide that ‘A label is like a comment and contributes nothing to the logical interpretation of an ontology’ (Smith et al. 2004). A screen shot of labels presented in English, Chinese Han character, and Chinese Pinyin are shown in figure 5.4.

![Figure 5.4 Screen shot of annotation widget](image)

By this approach, Heart from FMA and Chinese Medicine can have exactly same label ‘Heart’. The labels are also served the purpose for display and search. Therefore, a search of ‘Heart’ by label will give you two concepts.

### 5.4.3 Meta model and annotation

The annotation property of OWL DL can be applied to classes, properties, individuals and ontology headers (Bechhofer et al. 2004). The annotation property is designed for providing meta-information for classes, properties, individuals and ontology.

The meta model empowers reconfiguration of the system model. FMA ontology is an example of intensive use of meta-class in modelling. The meta-model approach enables the control of attributes determining which should be and which should not be propagated. It is argued that the differences between description logics and frames are not only syntactic, but also semantic (Rosse and Mejino Jr. 2003; Dameron et al. 2005). They analysed some theoretical and computational limitations of converting the FMA
into OWL-DL and proposed using the OWL-Full approach to make it possible to generate simplified OWL-DL representations.

This study has concluded that the meta-model is more suitable for the system designer and ontologist to specify the characteristics of classes as in the FMA approach. The meta-model should not be exposed to end users as it is different to the annotation property despite the whole OWL language being implemented through the meta-model in Protégé. We took the meta-model in a frame-based approach to represent the different perspectives from CM and OM. However, to avoid using OWL Full and to permit the use of DL reasoning services, the annotation property is used to represent different perspectives. Although the annotation property is not a defining property, the effect can be achieved by classes and properties which have been annotated. Furthermore, the annotation property will not be used for classification by DL reasoners. The ontology can be displayed or reclassified based on choices of perspectives from system designers and end users. The benefit of this approach is that computer takes the burden from human to deal with specifications according the specific paradigms. We have discussed this approach in ontology design and implementation of chapter 4. Here we give an example to demonstrate how this approach serves the purpose in ontology development and evaluation.

The acupuncture point LR3 is a traditional Chinese medicine theoretic entity. The value of the annotation property 'paradigm' for this concept is 'CM'. The property 'belong to meridian' has the same annotation property and value. The property 'has anatomy' is a link for modern anatomy and its 'paradigm' annotation value is 'OM'. L5 point, the corresponding segmental point in OM, is annotated as 'OM' paradigm. For system designer and end user, they can take advantage of the annotation property for system configuration and display. If a single paradigm is preferable for an application, then the system can display only the concepts and properties of that particular paradigm but hide the other paradigm. For OM paradigm users, they would only see segmental points and trigger points under acupuncture point. For CM paradigm users, they would only see classic meridian points, extra points, Ah Shi point under acupuncture point. For both CM and OM views, they will see all the above under acupuncture point. In addition, they can see LR3 is under L5 point. In fact, all classic meridian points and extra points can be classified under segmental point according to their nerve innervations.
In addition, it is technically possible to extract a small subset of ontology based on the annotation property, e.g. an ontology that only includes Chinese medicine concepts and properties and a similar one for Orthodox medicine. The reclassification and review would be required to ensure that they are fit for purpose.

5.4.4 Specifications of concepts

5.4.4.1 Informal vs. formal
Formal specifications of concepts are represented in description logics which provides semantics and access to the DL reasoning services. They are useful for informatics and system designers. However, for end user and domain knowledge experts, natural language definitions and specifications are still important.

In the Protege OWL plug-in, asserted conditions and the inferred view are used to specify and display formal DL expressions. The information description in English are entered in Rdfs:comment in the annotations widget.

5.4.4.2 Partially vs. fully defined
The primitive concepts are difficult or not necessary to be fully defined in ontology, such as acupuncture point, meridian, Yin Yang, etc.. These concepts are partially defined by necessary conditions or even have no definition at all. In this case the informal free text annotation will be helpful for ontology users. For example, Yang is defined as necessarily to be contrasted to Yin. For Acupuncture point, it is necessary to have location description and local anatomy. But it is not sufficient to be said one is an acupuncture point just by these two necessary conditions.

Fully defined concept by necessary and sufficient conditions can make use of the full power of DL reasoning service for comparing equivalence and subsumption. The examples have been given in description logics section in this chapter. However, we should be aware of the issue that a mutual agreement may not be achieved. Furthermore, the concept which has been fully defined in ontology may not be truly ‘fully defined’.

5.4.4.3 Proper use of DL constructs
To harness the full power of description logics, the classes and properties need to be defined properly according to domain knowledge and the semantics of DL constructs.
The Protege OWL plug-in provided the ontology test which helps to identify some common errors, alerts, and suggestions, such as missing disjoint on primitive subclasses, class duplicates restrictions from parents, class have multiple asserted parents, etc.. The errors can be identified by using a few functions based on DL reasoning services, such as classified hierarchy, inferred view, inferred super classes and inferred subclass, class consistency check.

The results of classification can always be compared with the intended meaning to be presented. Sometimes, one class is classified into a hierarchy which it should not be. It may indicate that description logics constructs for that class are not used properly. As we have discussed in a few places in this thesis, the description logics is the formal language for specification. The key is to use the language properly in order to represent the intended meaning of concepts based on the theories and logics.

5.4.5 Multiple classification

5.4.5.1 Type and role classification

The hierarchies in an ontology should be pure ‘is-a’ relationships. The part-whole, kinds of and role relations should not be part of the hierarchy. In medical informatics, Rector and colleagues (2001) criticized the ‘tangled’ taxonomies because, although they are familiar to users and easy access to terms, which the hierarchies mix the notions of kinds, parts, function, etc., they make formal inference almost impossible. The same issue has been discussed as an example of modelling ‘red rose’ by Guarino (1995) and Gangemi and colleagues (2002). Sowa’s analysis of three types of ‘phenomena type’, ‘role type’ and ‘sign type’ entities, distinguishes the differences by which things are classified (Sowa 2000a).

The GALEN approach (Rector et al. 2001) separates out each ‘axis’ into different taxonomies and recombines expressions in the description logic. For example, roles are defined as artifactual concepts, e.g. ‘doctor’ is defined as a ‘person who plays a ‘doctor role’’.

Although ‘Person’, ‘doctor role’ and ‘family role’ can be specified as disjoint classes, the class of ‘doctor’ and ‘father’ represent subclasses of ‘person’ and therefore can not be specified as disjoint classes because an individual person can play different roles.
This principle has been followed in our practice to avoid tangled taxonomies in ontology. Descriptions are modelled in separate hierarchies, such as Qualitative_Value, Quantitative_Value and theoretic entity. For example, the subclasses of theoretic entity are disjoint and Acupuncture_point and Five_Phases are direct subclasses of theoretic entity. Five Transporting point are defined through the property of ‘correspond to’ one of the Five phases (Wood, Fire, Earth, Metal, Water) which describes the quality of the point. The same approach applied to model Pattern of Yin Yang. They are defined by Pattern of relationship between Yin and Yang.

5.4.5.2 **Multiple classifications by flexible sets of distinctions**

We have discussed flexible sets of distinctions in the last chapter. It is possible that ‘thing’ (entity) can be classified by every single distinction as Rector and Rogers (1999) argued that ‘the principle of orthogonal taxonomies is key to re-use. For any specific use, ontologies may be developed using other principles. Anytime choices are made and the basis of those choices left implicit, the choices will be appropriate for one application but not for another’.

The choices of distinctions are dependent upon requirements of ontology. Indeed, to specify all possible dimensions for classifying things is almost impossible and unnecessary. Furthermore, if we treat all dimensions as anonymous classes within the hierarchy we are modelling, the entity that has been modelled is the subclass of all dimensions, ie, the crossing point of all dimensions. It is easy to realise that all distinctions have been defined as an equivalent class of all intersection of anonymous classes. Therefore, it inherits all properties and values from its superclasses. The choices of distinctions will decide its locations in hierarchies. It might be difficult to image the relationship between the entity to be defined and the dimensions. The following is a simple case for illustration.

To be modelled entity: T
Disjoint Distinctions: D1, D2, D3 ...
Anonymous classes: A1, A2, A3 ...
Role: R1, R2, R3 ...

The definition of anonymous as following:

\[ A1 = \exists R1.D1 \]
\[ A2 = \exists R2.D2 \]

...
T has been defined by selective distinctions, for example D1 and D2 in ontology O1:

\[ T \equiv \exists R1. D1 \cap \exists R2. D2 \]

Therefore, \((T \subseteq A1) \cap (T \subseteq A2)\)

‘T’ is the subclass of Anonymous class A1 and A2 in ontology O1. ‘T’ can be fully defined by some of distinctions, such as D1, D2, and D3 in ontology O2. And ‘T’ can also be fully defined by D2 and D3 in ontology O3. All ‘T’ in these ontologies represent the same entity in reality. However, they are defined differently because of interested characteristics from different perspectives or application requirements. ‘T’ in O1 and O3 became siblings of each other and ‘T’ in O2 became the subclass of T in O1 and O3 if we merge or compare these ontologies.

In this study, the requirement is to present different paradigms, CM and OM. Ontologies of O1 and O3 can be any ontologies from CM or OM paradigm. We have no obligation to have different ontologies. However, we believe that it is more important to emphasis what are shared and what are not. Therefore, classes (D1, D2, D3) and properties (R1, R2, R3) are marked up by annotation property ‘paradigm’ in ontology O2. The same entity T is represented by different perspectives in ontology O2.

The above discussed multiple classification inferred from descriptions and definitions. Sometimes, it may be preferable or have to manually configure to more than one parent in some situations, such as a subclass of two primitive concepts. This may result in inconsistency and the specification of disjoint classes which can help to prevent the error. However, if the disjoint is not specified, the inconsistency error will not be detected.

For example, a class may be assigned to have both physical and abstract entities as superclass, e.g. CM visceral manifestation. Because we specified that physical and abstract entities are disjoint classes, no class should be subclass of both. The inconsistent concept will be detected by the DL reasoner RACER. Because visceral manifestation is used to describe clinical manifestation by their relationships to organs, it is easy to interpret as physical entity. However, it might be more suitable to be modelled as relationship under abstract entity. Organ and manifestations are represented under physical entity. Throughout the development, we employ DL reasoner to classify the concepts and check whether the results reflect the intended meaning and the inconsistence have been identified and amended. It is not uncommon to find out that a class have been classified under more than one superclass because their implicit relationship has been discovered.
We have stressed that the disjoint can be applied to subclasses which are derived from superclass by a single distinction. For example, the subclasses of acupuncture point can not be specified as disjoint classes because subclasses are classified by at least two distinctions, anatomy and meridian. Classic meridian point and extra point are distinguished by meridian, therefore, they are disjoint. Segmental point is defined by anatomy and it is possible that one point can belong to one meridian which can also be classified by anatomy. The example such as LR3 is given in earlier discussion. All subclasses of classic meridian point can be specified as disjoint because they are distinguished by meridians they belong to.

5.5 Summary

This chapter introduces the scope and content of ontology and how we overcome the difficulties to formally represent acupuncture knowledge for both CM and OM.

The scope of ontology is limited to core concepts of acupuncture. The content of ontology is discussed from a semiotics view.

The main concern of vocabulary view of ontology is to have standard or controlled vocabulary. We discussed how we followed the standard proposed by WHO to construct terms in the ontology and to satisfy the requirements of representing terms in different languages and code.

We discussed some core concepts in the ontology to illustrate the main contents of ontology. Instead of discussing reality which is actually expressed by vocabulary, concept and the underpinning theories and logics of reality, we discussed those concepts from theories and logics of CM and OM paradigm.

The development in description logics, the OWL language, and the Protégé ontology editing environment enables us to use latest technologies and standards to represent acupuncture knowledge. We demonstrated the ontology through examples which are organised by applying the technologies, such as description logics. However, the whole representation is driven by the requirements of the knowledge to be represented. The system, language only serves as a means to achieve the representation purpose. We highlight solutions used to overcome the challenges for representation.
Chapter 6

Evaluation

This chapter describes how evaluations are used as part of the development process of the ontology. The evaluation by programme and the findings are discussed briefly, then the qualitative evaluation in which the methods of how to evaluate the acupuncture ontology from two disciplines (informatics and medicine), and two paradigms (orthodox medicine and Chinese medicine) are given in detail. We address the importance of recognising distinctions between the above disciplines and paradigms throughout this qualitative evaluation. Finally, the evaluation findings are presented.

6.1 Introduction

Evaluation as part of ontology development is included in most methodologies for ontology construction. In this chapter, we briefly explore the background of evaluation in information systems in general, after which the current approaches for ontology evaluation are discussed. The proceeding explanations provide the background for the settings and approaches of the acupuncture evaluations, the data analysis and the findings of the evaluations. The content of this chapter is organised as following:

6.2 Background of ontology evaluation
6.3 Ontology evaluation
6.3.1 Applications for ontology evaluation
6.3.2 Evaluate ontology programmatically
6.3.3 Qualitative evaluation of the ontology by experts
6.4 Data collection and analysis of qualitative evaluation
6.5 Key findings and discussion of qualitative evaluation
6.6 Conclusions of evaluation
6.2 Brief background of ontology evaluation

6.2.1 Evaluation of information systems

Evaluation itself has long been an independent subject for research. Guba and Lincoln (1989) take a historical view of evaluation development and classify the evaluation approach into four generations, namely, measurement, description, judgement, and responsive. Their observation and analysis reveals the paradigm shift from a positivist to a constructivist stance.

In the early 1990s, the evaluations applied within the life cycle of system development had a primary focus on quantitative methods. More recently, however, evaluation frameworks accentuate the need for qualitative research to be a critical portion of the evaluation process (Currie 2005). The drawbacks of objectivist approaches can be compensated by subjectivist methods for evaluation of information systems in health care (Moehr 2005). Moehr proposed a considerable expansion of evaluation methodology by including an assessment of cognitive and social effect in operational phases after system development. In discussion of evaluation framework of informatics, Currie (2005) emphasises ‘the shift from quantitative to qualitative instead of objective to subjective’.

Scriven (1996) explained the relationship between formative and summative evaluation by saying:

‘Formative and summative evaluation are the two elements in one particular classification of the roles of evaluation. There are many other classification systems for the roles that evaluation plays, and they are not necessarily exclusive of each other.’

Formative evaluation is carried out during the development phase of the research, whereas summative evaluation is made upon completion of a project. Formative evaluations are evaluated in an iterative manner with feedback provided for improvements before the final product is put forth, whilst in summative research, the impact or the outcomes associated with the use of the system are examined (Scriven 1997).

Four basic types of evaluation are proposed by Chen (1996) and he contends that formative and summative evaluations do not have to always oppose each other, but rather that an evaluation can serve both formative and summative purposes. A more
detailed literature regarding formative and summative evaluation history, development and methods can be found in Brown and Kiernan (2001).

There are different evaluation approaches in terms of trade-offs among desired evidence, cost, and other constraints. They are summarised by Almstrum and colleagues (1996) as following:

- Formative/summative evaluation
- Qualitative/quantitative evaluation
- Field studies/laboratory studies
- Experimental/observational evaluation
- Incremental/longitudinal evaluation
- Data-driven analysis/theory-driven analysis evaluation

Although this summary may not be exhaustive, it presents different approaches and their distinctive methods sufficient for evaluation purpose.

6.2.2 Approaches for ontology evaluation

On top of the general issues regarding quantitative vs. qualitative and formative vs. summative approaches, the subject of ontology evaluation still requires more investigation. The knowledge web project (Hartmann et al. 2004) analysed and evaluated existing methods and tools for ontology evaluation and concluded that they are still in the academic realm and not yet suitable for industrial practice. Lack of well-understood notions of ontology evaluation, and validation set back the transition of ontologies from research into industrial applications (Gangemi et al. 2005).

In summary of approaches to ontology evaluation, Rogers (2006) states that evaluation over the quality of ontologies, or their fitness for specific purposes, is improving but remains poor. He suggests that a combination of methodologies is required, but further points out that tools to support a comprehensive quality assurance remain lacking. The followings are common approaches for ontology evaluation (Brank et al. 2005).

- **Evaluate ontology by a set of criteria, standards, and requirements**

These evaluations are part of ontology development and ontology quality control. The evaluation focuses on the formal features of an ontology, such as consistency, and structure of taxonomy (Gómez-Pérez 1995; Grüninger and Fox 1995; Gómez-Pérez
A complex framework ONTOMETRIC proposed by Lozano-Tello and Gómez-Pérez (2004) allowed the users to assess the suitability of existing ontologies, regarding the requirements of their systems. The ONTOMETRIC extends the scope of evaluation criteria over five areas: content of the ontology, language, development methodology, building tools, and usage costs.

Evaluate ontology by a meta-ontological analysis
OntoClean as a methodology has been used for quality control to identify the inconsistency of ontology (Guarino 1998a; Guarino and Welty 2002). The method requires analysing every single concept by OntoClean meta-ontology. It involves considerable effort to understand and assign meta-ontology to concepts in ontology.

Evaluate ontology by application
The evaluation is focused on the utility and usability of an ontology, such as how adequate the ontology is for tasks, how well it represents the domain of interest (Noy and Hafner 2000). This evaluation method is suitable for task based approach and it can only be carried out in respect to a given task at hand, which the specific requirement has to be solved by the ontology (Porzel and Malaka 2004).

Evaluate ontology by quantitative method
This approach applies quantitative methods to measure ontology which is different to qualitative approach. A framework of three distinctions to measure ontology for evaluation, i.e. structural, functional and usability-related measures, is proposed (Gangemi et al. 2005). OntoQA approach (Tartir et al. 2005) develops metrics for both ontology and its instances. The metrics measure relationship, attributes and inheritance of ontology and its potential richness of knowledge representation. The metrics for instances evaluate the placement of instance data within the ontology and the usage of the ontology.

Evaluate ontology by comparison between ontology and source data
Data driven approach (Brewster et al. 2004) evaluates ontology against given textual corpus of the domain through cluster analysis. It provides a means to select the best fit ontology among similar existing ontologies by measurement.
Evaluate ontology by comparing ontologies

This approach has been studied for comparing, merging or selecting ontologies. It is possible to compare overall characteristics of ontologies and focus on particular differences of content and structure of ontologies (Hovy 2002). Some methods for comparing similarities of concepts between ontologies have been proposed (Weinstein and Birmingham 1999; Maedache and Staab 2002). Instead of comparing similar ontologies, CleanONTO approach (Sleeman and Reul 2006) proposes comparing concept definition in ontology with existing definition, such as WordNet.

6.3 Ontology evaluation

Evaluation is seen as an important step within the ontology development process and is essential for ensuring the quality of the ontology. The evaluation conducted in our research is part of ontology development.

The purpose of evaluation in this study has two purposes: the main one is to answer research question that whether a single ontology can enable knowledge sharing across paradigms; and the second one is complementary by enriching the development for current and future applications. Since the ontology is work in process, the results of evaluation would assist us to explore how well the ontology was constructed and was received from experts.

In terms of formative or summative evaluation, we have chosen formative evaluation as it is most appropriate for the nature of this research. The formative evaluation engages a feedback process from participants with data and comments collected to be used for improving a program during its course of development (Scriven 1967).

Amongst the approaches discussed for ontology evaluation, quantitative approaches have been developed recently, which provide measurements of ontology to facilitate selecting and comparing ontologies for particular applications. However, it can not prove whether the ontology design had modelled the knowledge correctly. Given the research question this was inappropriate for this study.

The evaluation for quality assurance of ontology is conducted programmatically by existing function with Protégé OWL plug-in and services of DL reasoner. As Gorovoy and Gavrilova (2007) noted that ‘these instruments for evaluation haven’t reached serious level of maturity and are not widely used’. The reason could be that
most ontology editing environments do not have this functionality (Parsia et al. 2005). However, the ontology editing environment, Protégé OWL plug-in, provides a few functions to perform evaluation programmatically. We will discuss the evaluation by software in section 6.3.2 after the introduction of user interfaces for evaluations.

Assessment by domain experts is often applied as part of the ontology development process (Gómez-Pérez 2001; Blaschke 2005; Burton-Jones et al. 2005). Qualitative approach was adopted in this research for evaluation with methods of observations and interviews. The ontology was evaluated by experts from two disciplines (medical informatics and acupuncture) and two paradigms (CM and OM). Each one-to-one evaluation includes three parts: review of structure and relationship in ontology development, demonstration of different perspectives in knowledge browser, and evaluating ontology by query application. The detail will be discussed in section 6.3.3.

To summarise, the evaluation in this study is formative approach with qualitative evaluations by experts from different disciplines and health care paradigms. The ontology is also evaluated programmatically by the DL reasoning services and the function within Protégé OWL plug-in.

6.3.1 Applications for ontology evaluation

Ontologies are difficult to evaluate and purpose of ontology often cannot be recognised by domain experts without concrete applications. In our study, we had two future applications in mind; a DSS (Decision Support System) for practitioners and a tutoring system for education. However, in the available timescale, they neither existed, even in rudimentary form. We did not have any application to hand for the acupuncture ontology. To develop a full application only for demonstrating the usage of the ontology would be too time-consuming and resource-intensive. In the context of a small scale research project constructing full application is not always a practical proposition, particularly, as in this case, the primary thrust of the research was to do with knowledge representation rather than user interaction. Nevertheless, the intended application is a practical means of decision support and therefore excluding the end-users from the evaluation was not an option. Hence, the open source tools were used to demonstrate and evaluate our ontology.
6.3.1.1 Protégé OWL Plug-in

The Protégé OWL plug-in user interface was used for the evaluation of the ontology. It is much easier to inspect hierarchies through graphical representation in Protégé rather than viewing XML files in text format. The relationships among concepts and properties and definitions (informal and formal) are displayed in a clear layout.

6.3.1.2 Knowledge browser

A knowledge browser was developed through an open API of Protégé as a simple application to present different views of CM and OM in a single ontology. The purpose of developing this browser was to simplify the user interface further and visualise the difference between CM and OM. It is not ideal to show the whole ontology within the Protégé interface; users would be overwhelmed by the scale, the detail and the complexity of the ontology. They would also have to spend time navigating to the main sections of the ontology which would be undesirable given the time available to each evaluation session. The knowledge browser provides a list of the main distinctive branches of the ontology, e.g. Acupuncture point, Meridian, Clinical finding, and Prescription. The rest of the ontology branches, all super classes and sibling classes of selected nodes, were hidden. Furthermore, the browser can display and hide classes based the selection of perspectives of CM, OM or both as show in following three figures which are screen prints of the interface.
Selected section of ontology:
Acupuncture point

Displayed concepts:
- Ah Shi Point
- Extra point
- Fourteen Meridian Point

Hidden concepts (Anonymous-O):
- The subclasses of Ah Shi Point
- Segmental Point

Figure 6.1: CM perspective
Selected section of ontology:
- Acupuncture point

Displayed concepts:
- Trigger Point
- Segmental Point

Hidden concepts (Anonymous-C):
- Ah Shi Point
- Extra Point
- Fourteen Meridian Point

Figure 6.2: OM perspective
6.3.1.3 Sesame query interface

Sesame was another application used to query the ontology. Sesame (Broekstra et al. 2002) and Jena (McBride 2001; Carroll et al. 2004) are two popular open source tools for RDF and OWL. Sesame is a java web application deployed in Apache Tomcat.
application server. It has flexibility and can be deployed on top of a variety of storage systems (relational databases, in-memory, file systems, keyword indexers, etc.). Furthermore, it offers flexible access to API which supports both local and remote access through HTTP or RMI (Remote Method Invocation), and several query languages (SeRQL, RQL and RDQL). The latest version of Sesame also supports the SPARQL language. Jena is a java framework and provides a programmatic environment for RDF, RDFS, and OWL. Jena provides the similar functionality as Sesame does. Sesame was selected for it has an easy to use web interface to import and query the ontology.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Paradigm</th>
<th>Point Selection</th>
<th>SelectedPoints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pattern_of_Liver_Qi_Stagnation</td>
<td>CM</td>
<td>Four_Gate_Points_Prescription</td>
<td>Ir3</td>
</tr>
<tr>
<td>Pattern_of_Liver_Qi_Stagnation</td>
<td>CM</td>
<td>Four_Gate_Points_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Pattern_of_Liver_Yang_Excessive</td>
<td>CM</td>
<td>Pattern_Prescription</td>
<td>Ir3</td>
</tr>
<tr>
<td>Migraine</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir3</td>
</tr>
<tr>
<td>Pattern_of_Liver_Yang_Excessive</td>
<td>CM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Migraine</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Pattern_of_Liver_Yang_Excessive</td>
<td>CM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Migraine</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Migraine</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Major_Depressive_Disorder</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir3</td>
</tr>
<tr>
<td>Major_Depressive_Disorder</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir7</td>
</tr>
<tr>
<td>Major_Depressive_Disorder</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>Major_Depressive_Disorder</td>
<td>OM</td>
<td>Pattern_Prescription</td>
<td>Ir4</td>
</tr>
<tr>
<td>14 results found in 0 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.4 Screen shot of Sesame query interface

There are several query languages available for querying ontology and W3C did not have recommendation on query languages at the time of our ontology development and evaluations. SPARQL (Prud’hommeaux and Seaborne 2008) language has finally become the W3C recommendation for query language for RDF in January, 2008.
The languages can be categorized into SQL like RDF triple languages: SeRQL (Broekstra and Kampman 2004), SquishQL (Miller et al. 2002), RQL (Karvounarakis et al. 2002), SPARQL (Sirin and Parsia 2007), RDQL (Seaborne 2004). A comparison of some query languages can be found from Haase and colleagues (2004). Some query languages that support the description logics have been developed in recent years such as: OWL-QL (Fikes et al. 2003), nRQL (Haarslev et al. 2004), SPARQL-DL (Sirin and Parsia 2007), SAIQL (Kubias et al. 2007). We chose the SeRQL for querying the ontology because the query language was built in the Sesame framework and included the functions required with simple SQL like syntax.

We carried out two modifications to the code of Sesame to simplify the presentation. Conventional Sesame returns results in a fully qualified name format which comprises name space and class or property name, for example, http://www.shire.salford.ac.uk/ontology/acupuncture#LI4. However, this is simply not user friendly. In order to achieve readily-readable presentation of concepts and properties, the first modification to the code was to enable Sesame to display the query results without name space, such as LI4. Please note the later version of Sesame support this function as a user display selection.

The second change to the Sesame code is to exclude the DL expression in the returned results because the SeRQL language in Sesame does not fully support querying OWL-DL language. The result with DL expression is not returned in a meaningful format, e.g. _node:117pod4m8x40, which gives no help to the understanding of either the concept or property. The code of Sesame is recompiled and deployed to Tomcat. Excluding the OWL DL expression will not lead to the loss of inferred relationship because the classification result had been produced by RACER before the ontology is imported to Sesame repository. The OWL file of the ontology was imported into the Sesame repository before the evaluations.

Sesame query interface is used to demonstrate the possible usages of ontology although it is not a full application for end user for specific task.

6.3.2 Evaluate ontology programmatically

The ontology has been continuously and repeatedly evaluated programmatically as part of our ontology development. The test function within the Protégé OWL plug-in is applied for ontology evaluation. The OWL plug-in provide access to DL reasoner - RACER.
6.3.2.1 OWL plug-in tests

The OWL plug-in provides tests which automatically checks if the basic rules of OWL description logic have been followed. Meanwhile, it also provides sanity tests and information for the maintenance. The tests are listed as following:

*Maintenance Tests:*

- Find TODO list items
- List deprecated classes and properties

*OWL-DL Tests*

- No Cardinality Restriction on Transitive Properties
- No Classes or Properties in Enumerations
- No Import of System Ontologies
- No Metaclass
- No Properties with Class as Range
- No Subclasses of RDF Classes
- No Super or Sub Properties of Annotation Properties
- Transitive Properties cannot be Functional

*Sanity Tests*

- Domain of a property should not be empty
- Domain of a property should not contain redundant classes
- Domain of a sub property can only narrow super property
- Inverse of Functional must be Inverse Functional
- Inverse of Inverse Functional must be Functional
- Inverse of Sub property must be Sub property of Inverse of Super property
- Inverse of Symmetric Property must be Symmetric Property
- Inverse of Top Level Property must be Top Level Property
- Inverse of Transitive Property must be Transitive Property
- Range of a property should not contain redundant classes
- Range of a sub property can only narrow super property

As the names of above tests suggested, these tests help to debug common human errors and rules applied for description logics.

6.3.2.2 The DL reasoning services

Description Logics as ontology language were used for ontology design, modification and integration. The reasoning services provided by description logics can ensure
consistency of concept definition, detect unintended redundancies and automatically generate taxonomy based on definition.

The RACER DL reasoner provides the most advanced description logic services and Protége OWL plug-in makes it much easier to access the DL reasoning services. During ontology construction, these functions have been applied to validate the definition with respect to the intended meaning. For example, the function of checking consistency was used to detect errors in definition whereas the function of classifying taxonomy was used to get the concept into its position according its formal logic definition. The classification was also applied through functions of getting inferred subclasses or getting inferred super-classes by right clicking a concept. Some definition errors were avoided after the inspection of classification results. The example has been given in acupuncture ontology chapter.

6.3.3 Qualitative evaluation of ontology by experts

The programmatic approaches can not provide validity of the content of ontology as description logics on their own are not enough and it can not prevent inappropriate definitions of concepts (Ceusters et al. 2003).

The qualitative evaluations by experts were conducted to provide an in-depth view of ontology. Two different methods of data collection, participative observation and unstructured interview, were used for evaluations. Experts represented different disciplines (medical and informatics) and different paradigms (CM and OM). This triangulation approach had benefit of ‘the combination of appropriate research perspectives and methods that are suitable for taking into account as many different aspects of a problem as possible’ (Flick 2002).

Sampling methods in qualitative research differ significantly from those used in quantitative study design. Instead of random, probability-based sampling methods that are features of quantitative research, qualitative studies usually employ non-random methods of sampling. There are four main methods of sampling i.e., convenience sampling, purposive sampling, snowballing and theoretical sampling (Schofield 1996, p.25).

Convenience sampling, as the name suggests, samples those respondents that are easy to recruit, likely to respond or already available.
Purposive sampling, which is also termed selective sampling, is a method which deliberately includes participants who possess the particular characteristic the researcher is interested in.

Snowballing describes the method of identifying several participants of interest and then asking them to suggest other potential participants who share the characteristics the researcher wishes to explore.

Theoretical sampling has been defined as ‘the process of data collection for generating theory whereby the analyst jointly collects, codes and analyses his data and decides which data to collect next and where to find them, in order to develop his theory as it emerges’ (Glaser 1978).

Above sampling methods can be applied simultaneously or at different stages of study. We adopt convenience sampling and purposive sampling to ensure different perspectives are presented by participants. The experts from different disciplines (informatics and acupuncture) and different paradigms (CM and OM) were approached for evaluations. The snowballing sampling can be used to recruit new participants for future development.

6.3.3.1 Setting of qualitative evaluation

All evaluations were conducted in the ‘state of art’ informatics lab at SHIRE (Kay and Hardiker 2005). During evaluation, participant and evaluator sit next to each other in the observation room. There are three high resolution cameras from different angle continuously recording the whole evaluation process. The laptop used for the evaluation is linked to the lab AV system through an AV cable. The video and audio footage recorded from the three cameras and screen contents of the laptop formed four sections of on one single screen which was burnt on to DVD in real time.

The introduction of presentation is displayed on a wall mounted 40 inch flat plasma screen. All the controls, and DVD recording are taken place in the control room next to the observation room. There is a two-way wall mirror between these two room. The informatics lab provides all facilities and minimizes distractions for user evaluation. The layout of both rooms is shown in appendix K.
6.3.3.2 Evaluator and participants

This research was conducted across the disciplines of informatics and health care with a focus on crossing paradigms of CM and OM. Six experts were recruited to participate in six evaluations. Two are doctors trained in Orthodox medicine with expertise in medical informatics. Two acupuncture experts came from traditional Chinese Medicine background. The last two participations are trained in the UK with knowledge of acupuncture practice in health care. The varieties of participants’ backgrounds provided different views for the purpose of evaluation.

Evaluator in this research is the ontology developer, i.e., the author himself due to the constraint of a small project. However, we are aware of the issues that arise from such a method. Firstly, developer as evaluator may produce conflicts in the two different roles, giving rise for introducing a risk of personal bias. Employment of an independent evaluator would be ideal but not forthcoming in the project. The situation is a limitation, and awareness of the problem and its disclosure is the minimum response that can be given here.

Secondly, participants may feel restrained when giving constructive criticism face to face with the developer/evaluator. Tacit approach may be taken to avoid personal conflicts, where positive points are relatively easier to be picked upon but critics fail to deliver or sustain. In this context, the participants did not share the same research autonomy as the researcher does and there is no mandate for them to be more critical than the researcher himself. For this issue, briefing sessions were given particular attention. Briefing session was carried out before interview with verbal consent for video recording. Verbal assurance was given to help maximise conversation for meaningful points and opinions to occur.

The requirement of evaluation is sought firstly in context of informatics where we need to provide evidence that the ontology follows the development principles and criteria. The principle and requirement for ontology are discussed in chapter 2. We have collected and organised evaluation criteria which have been proposed by a few authors into a table in appendix G.

We hold the view that an ontology evaluated by clinical domain experts alone is inadequate because of the different discipline perspectives on the knowledge itself and how knowledge can be represented in an information system.

The ontology is the model of clinical domain entities. It is often the case that ontologies built into information systems are not equivalent to those ontologies in
domain experts' minds; a direct result arising from the different knowledge ontologies in their respective mind sets. They have neither the same perspective nor the shared understanding of the distinctions in each other's domain. Knowledge engineers analyse and define knowledge from the perspective of the information system. They pay attention to issues about completeness, consistency, granularity, etc. These criteria are not appreciated to the same extent nor do they always make sense to the clinical experts. Some can be seen as being common sense decisions to one, yet not to another. Therefore, the concerns and judgements are different between disciplines of informatics and clinical care.

Furthermore, the centre of attention was also given to evaluate ontology from two different perspectives on acupuncture knowledge. The ontology was evaluated by participants came from different paradigms for their specific perspectives. A key contribution of the thesis is that different perspectives are represented in a single ontology which improved knowledge sharing. Clinical experts can feedback on the approaches we adopt as to whether this approach had overcome the challenges which we have discussed the requirements, challenges and solutions in previous chapters.

6.3.3.3 Procedure and methods of evaluation
Six evaluations were conducted on a ‘one to one’ basis and last about 2 to 3 hours each. The orders of evaluation are as following, medical informatics expert evaluations, CM acupuncture practitioner evaluations and OM acupuncture expert evaluations. Consent was obtained from the experts for agreement to participate and being audio-video recorded of the interviews. A introduction of research background, the purpose and procedure of evaluation were given to each participant at the beginning of interview. The evaluator explained the scope and demonstrated the methods and process of the evaluation.

We applied advanced video and audio recording equipments in the lab to record the whole process of evaluation. The "talk aloud protocol" was suggested to experts although it was not always followed by experts who were not used to this method. At times the evaluator had to encourage experts to follow the protocol, at the same time, ask more questions to catch what expert was thinking during the evaluation.

During the evaluation, evaluator and participant sat side by side. The participant is free to ask any questions and the evaluator gives factual answers and discusses significant issues at the end of the evaluation.
All evaluations contain three parts (A-C), however, with different content. The introductions and evaluation contents are different for informatics expert and acupuncture practitioner to reflect their expertise and perspectives. The part C for informatics is not applied for acupuncture experts because it is intended to evaluate approach of ontology construction. The part C of the query application was intended for acupuncture experts to assist their understanding of how ontology could be used for various applications. The contents of ontology were evaluated by asking different questions through the query interface which was designed for the non-informatics experts.

**Three parts of informatics expert evaluations:**

A: evaluating ontology structure in term of classes and relations

B: evaluating different views of Orthodox medicine and Chinese medicine through knowledge browser application

C: evaluating flexible sets of distinctions for constructing ontology

**Three parts of acupuncture expert evaluations:**

A: evaluating ontology structure and relation

B: evaluating different views of Orthodox medicine and Chinese medicine through knowledge browser application

C: evaluating ontology by query application

**Part A: Evaluate ontology structure and relations**

Experts were asked to comment on the following while they were exploring concepts of ontology in the OWL plug-in interface of Protégé:

- Concept hierarchies in term of the concept’s superclass and its subclass
- The siblings of the concept
- The properties of concepts and their values

Informatics experts were also asked to compare informal definition with formal description logic definitions.

A list of concepts was provided to the participants for evaluation. However, they are not constrained by the list. They were free to explore any part of ontology. There were three reasons for providing a list of concepts:
Firstly, the selected concepts presented the core concepts and structures of the ontology. This is to help participants to focus on the content which are significant for the study. To make experts navigate through the volumes of details of all the concepts would neither enhance nor hinder evaluation but merely reduce the available time.

Secondly, the qualities of concepts were not defined to every detail for all. Some areas, such as clinical finding, patterns, were not the key concerns of this ontology. Hence only a small number of concepts were included and defined.

Thirdly, for some concepts which have similar characteristics, we only defined one or two to demonstrate the definitions. The approach and definitions can be applied to many other similar ones. For example, LI4 was fully defined but its sibling LI5 was not. The only difference of definition is the order on the meridian and location.

**Part B: Evaluate different views by knowledge base browser**

The knowledge-base browser was a simple application to present different views of CM and OM in a single ontology.

Participant can select different perspectives of CM, OM, or both CM and OM from the drop down menu. The perspectives were determined by annotation property of paradigm. The concepts in the ontology were displayed or hidden according to the selected perspective in browser. The expert can assess three different views of acupuncture ontology. The branches (sections) of the ontology for evaluation were Acupuncture Point, Meridian, Disorder, Pattern, and Prescription.

**Part C: Informatics expert evaluate flexible sets approach**

Evaluator provides an introduction of flexible sets of distinction approach. The presentation explained two-way and three-way distinctions and how they are applied in analysing knowledge and system, after which experts were asked to assess some concepts in ontology and give comments.

**Part C: Acupuncture experts evaluate ontology by query application**

Evaluation by querying an ontology has been used in FMA project whilst a query interface was developed for evaluation (Mork et al. 2003). We have chosen Sesame as the platform for querying the acupuncture ontology because of its functionality. As well as being open source, Sesame can be easily modified to improve accessibility without losing any information as explained in section 3 in this chapter.
Querying the acupuncture ontology was designed to answer common clinical and teaching questions. Questions were prepared as scripts in SeRQL language (appendix H). The domain experts can ask any questions freely without being constrained by pre-determined SeRQL scripts. Their questions in natural language were transformed by evaluator into SeRQL language at evaluations. Participants were asked to comments on query results.

The ontology was queried by classes, instances, and their properties. The query at class level is performed through a ‘is-a’ relationship which is defined by the relationship between sub-class, class and super-class.

Querying the instance and its value of property is, in effect, to query the populated knowledge base which originated from the ontology. There is no limit in querying across class and instance.

The returned results can base on the paradigms, CM, OM or both. For example, the returned results of disorder could be migraine in OM and patterns, such as Pattern of Liver Qi stagnation, Pattern of Liver Yang excessive etc. for CM. If the user chooses both OM and CM, all of above results will be shown.

It is worth noting that the evaluation did not consider the editing environment and application as part of subject of evaluation. All tools and applications used in evaluation were only for the purpose of supporting ontology evaluation. However, they do have an impact on user experiences particularly for those who are not familiar with informatics. Therefore, we discussed feedback regarding user interface and usability although they are not key contribution of this thesis.

6.4 Data collection and analysis of qualitative evaluation

6.4.1 Methods of data collection

6.4.1.1 Observation

Observation is a method that researcher uses all of her/his senses to gather information about the phenomena under study (Adler and Adler 1994). The process can be audio or video recorded, enhancing sensual observations as an ‘instrumental extension of our senses’ (Collier and Collier 1986, P7) as well as increasing accuracy.
There are many different approaches to observation. They can be identified by the researcher’s role in the study. It is suggested four types of role along this dimension (Gold 1958; Junker 1960, p36):

1. The complete observer: The researcher has no interaction with the subjects during data collection and does not take on the role of participant. It is also called non-participant observation.

2. The observer as participant: The observer interacts with subjects, but does not take on an established role in the group. The observer’s role is that of researcher conducting research.

3. The participant as observer: The researcher taking a participant role facilitates access to the group or institution.

4. The complete participant: The researcher plays an established role in the group and is fully immersed in that participant role, then uses his or her position to conduct research.

The extent of participation is a continuum (Patton 1990, p.206; Ely et al. 1991, p.42; Sapsford and Jupp 1996) that varies from the complete participation with full engagement to complete separation as an outsider.

The benefit of non-participant observation is that it eliminates the reactivity which generates from the immediate physical presence of the observer and allows the observer to remain detached from the subjects and therefore uninfluenced by their views and ideas (Foster 1996, p.73). However, this requires the key condition that the subjects are unaware that they are being observed, hence, limiting on the settings and behaviours that can be observed.

The participant observation is widely used in social research and is characterised by researcher’s engagement to the subjects and role of group. It ‘simultaneously combines document analysis, interviewing of respondents and informants, direct participation and observation, and introspection’ (Denzin 1978, p.183). This gives researcher a close understanding from participants’ point of view. This is an important consideration for this study as representation of experts’ perspectives from different view point is central to the intended ontology. Therefore, we have adopted participant observation for the clear benefits that it enables the researcher to obtain expert participants’ view with better understanding but without dominating the interview where undesirable influence developed upon experts.
All observations in this research were conducted in informatics lab. Participants were invited to give their feedback and judgement from their expertise. The whole evaluation was recorded. Although the researcher stayed with participant in the lab, the researcher as evaluator did not give any judgement and only provided information as participants required.

6.4.1.2 Interview

The most commonly used three types of interview techniques are unstructured interviews, semi-structured interviews, and structured interviews (Bowling 1997; Bryman 2001; Hussey and Hussey 1997; Lofland and Lofland 1995; Robson 2002).

- Structured interviews (or standardised interview) involve the researcher asking each interviewee the exact same questions. The range of answers for questions is fixed and followed by a statistical analysis of the results which is similar to survey research.

- Semi-structured interviews consist of open-ended questions and they are structured around a few topics to guide the interviewer. Semi-structured interviews are commonly used in qualitative research.

- Unstructured interviews usually consist of one or two questions that open a discussion on a specified topic and allowing considerable freedom within the interview. The interview takes the form of a conversation rather than a more structured question-answer session.

In this research, the semi-structured interview was employed during the observations. During the interviews, the researcher followed themes of questions to elicit evaluation comments while interacting with experts to clarify questions and explanations so that meanings were shared and better understanding achieved.

6.4.2 Qualitative data analysis

Amongst the various methods of data analysis are available to qualitative research, Tesch (1990, p.58) presents and discusses major types of qualitative analysis. She concluded that it is impossible to sort qualitative research approaches neatly into ‘types’. Regardless of the methods used, the key principles of the analysis process are the same. The researchers need to fully elicit data and be familiar with all aspects of the
data for analysis. For text data analysis as used in this study, coding is a common method. Codes, consist of a few letters symbolising a category, are used as ‘retrieval and organising devices that allow the analysis to spot quickly, pull out, and then cluster all segments relating to the particular questions, hypothesis, concept, or theme’ (Miles and Huberman 1984, p.56). Thereby the raw data can be processed and interpretation made and presented in appropriate manner.

In qualitative research, using computer software to handle text data has transformed analysis. Qualitative software are developed to enable the steps of the coding, data saturation and theorising process to recovered, documented and displayed as well as help theory-building (Fielding 2001, p.455).

In this study, the common term data collection and coding of text data were employed with the use of open source software Transana to analyse the transcripts of the observation and interviews in order to identify topics and themes from coded categories. Transana provides seamless support for qualitative analysis for video and audio.

The whole process is qualitative with underlying issues that couple tightly with the nature of qualitative research. Firstly, the data elicited together with field notes are in ‘the language in the form of extended text’ and structured on the basis of observation and participation (Miles and Huberman 1984, p.9). As Dey (1993, p15) argued, we can not discount the role played by the researcher as participant observer or unstructured interviewer. The process of eliciting and shaping the data is inevitably influenced by the researcher’s values and beliefs which underpin the context of practice with both the participating experts and the researcher. Meanwhile, transcription process may not be without certain influences.

Furthermore, the processing of the data into interpretations is where Miles and Huberman (1984, p.9) described as from the first order fact (the data) to researcher’s second order conception of what is going on, i.e., ‘the interpretations of the interpretations’ (Noblit 2004, p.191). Essentially, what is being communicated in conversation are terms and meanings. It is bound up with the contrast between what is the asserted and what is implied not to be the case (Talja 1999; Potter and Margaret 1994). Analysis of transcripts and data acts in a manner that dismembers the text in the service of interpretation and generalization out of context, eliminating the sequential and structural features of the narrative accounts (Riessman 1993, p.3).
Hence, we acknowledge that the complexity of the seemingly simple qualitative data analysis. Care and mindfulness is consciously taken as well as self-awareness of the researcher in meticulously capturing of meaning. The approach of data analysis is discussed in next section.

6.4.3 Analysis environment and process

There were six DVDs of evaluation records covering over 12 hour’s video and audio. The data covers all observations and interviews. The richness of video and audio records provides chances to replay and catch details when analysing the records repeatedly. Our analysis has two main points. On one hand, participants’ judgement and comments were closely examined as to how well the criteria and different paradigms has been reflected in the ontology. These are vital contributions to the next stage of revision and further development of ontology. On the other hand, attempts were vigorously sought to crosscheck the process of how their opinions formed with indications and evidence arose from conversations during evaluation to identify and study what evidence exists for their judgement for validation, hence our interpretation to be fairer.

6.4.3.1 Qualitative video analysis tool - Transana

Transana is an open source software for qualitative analysis of video and audio data, which was developed at the University of Wisconsin-Madison Centre for Education Research. It provided various ways to organise and analyse video clips. The software also supports to link transcriptions to video clips seamlessly. A screen shot of Transana is shown in figure 6.5.
Because Transana does not work well with DVD file format such as .vob, we converted .vob files into .avi file format.
In Transana, each original single video files is called an **episode**. User can group episodes into **series** which describe the logical connection among those episodes. In our records, each DVD contains five VOB files. The whole record of a single interview was organised as a series of 5 episodes (There was only one interview which was recorded into 2 DVD, therefore, in total of 10 episodes).

The basic unit for analysis is the **clip** which is a portion of an episode selected by user for analytical purpose. The clips are organised into thematic **collections**. Clips are annotated by **keywords** during the analysis and one clip can have multiple keywords. A keyword represents an analytic concept of the study. Keywords are organized into **keyword groups** (Woods).

The keyword structure reflects the theoretical structure of the qualitative analysis. The methods of coding will be discussed in following section.

### 6.4.3.2 The process of qualitative data analysis

The transcriptions of video/audio records were completed by an external personnel by watching the video and transcribe all into text files. The transcripts were coded by the researcher using Transana to link the codes to the words, paragraphs and actions in the video.
Three coding methods, open coding, axial coding and selective coding, are commonly used (Murphy et al. 1998; Strauss and Corbin 1998; Ezzy 2002). All three coding methods were employed to form a comprehensive picture of evaluations.

The first process is **open coding** which is the process to break down the data by concepts or categories where the text is opened up and broken apart for intensive scrutiny (Corbin and Holt 2005, p.30). The process allows the identification of concepts and the formation of patterns by grouping together similar incidences, claims, and discursive practices. The basic themes were identified through this process. For example, the keywords are grouped into feedback, evaluating concepts, evaluating properties and query concept in ontology.

The second process is **axial coding** which is the process beyond watching video and reading transcript. Within this process, we followed identified theme and employed 'constant comparison' (Ezzy 2002, p.90) to search through the video and transcript for other related comments in order to further develop themes. The evaluation of principles of ontology was one key axis of this coding process.

The last process is **selective coding** which involves further refining the data, bringing together the patterns, consistencies, categories and constructions, and creating meta-themes. The clips were organised into collections by interviews which did not reflect the themes of evaluation. The selective coding was achieved by querying existing collections based on selected keywords and the results of the queries were converted into collections.

The amount of information being analysed was significant. We provided an example of transcript and coding in appendix J. It demonstrates how clips have been analysed and coded, and how clips and transcripts organised and linked in Transana.

### 6.5 Key findings and discussion of qualitative evaluation

In summary, the formative evaluation is the adopted evaluation strategy and the activities of the evaluation are listed as follows:

**During the development stage:**

- The ontology is validated by the DL reasoning services
Informal reviews of the ontology's structure and content were performed by the clinical experts. During the final stage:

Formal evaluations of the final ontology were made by experts from CM, OM and clinical informatics:

- The structure, concepts and relations of the ontology were evaluated through the Protégé editing environment
- The presentation of different perspectives of CM and OM were evaluated through the Knowledge browser
- The ontology content was evaluated in the Sesame query interface

The data collection and analysis of expert evaluations were carried out by:

- Interview and observation recordings on DVDs were the data collection methods for evaluations
- Qualitative coding methods were employed to analyse video and transcriptions from the evaluations

Following the three coding methods, the key issues were identified and coded along the analysis of video and transcripts. We organised summary of feedback by key issues into 6 tables (Appendix A-F). The content of table are hierarchy of content, feedback, brief description of finding, reference of clips. In the table, only the super concepts were presented in the column of content hierarchy. For the readability, the subconcept names, though evaluated, were not included. Expert comments on ontology content were presented in the column of feedback. The brief description of finding described the issues from video clips. Reference clips included the list of clips related to the key issues.

The following were six key issues in the summary tables:

- Knowledge representation and sharing (Appendix A)
- Term representation (Appendix B)
- Principles for ontology construction (Appendix C)
- Top level ontology and FMA (Appendix D)
- Coverage of ontology (Appendix E)
- User interfaces of application (Appendix F)

The following sections are analysis and discussion of the key issues from the summary tables.
6.5.1 Knowledge representation and sharing

The ontology and possible applications demonstrated how to share the knowledge and how it can improve communication of knowledge from different medical knowledge perspectives. The knowledge sharing was achieved through meta-level (annotation), concept level and individual level of knowledge representation.

The meta-level represented the knowledge resources of CM, OM or both. Knowledge browser demonstrated how different views could be available for user as it was demonstrated by screen shots in the knowledge browser section, e.g. a CM acupuncture point LI4 is a ‘Primary Point’ and ‘Large Intestine Point’. It can also be a segmental point under C6, C7 and T1 from OM perspective. Experts were happy to see different views and the clarity it offered to the end user. The annotation property of paradigm was technical solution to indicate the perspectives. One expert suggested to have a small ontology for each paradigm which is possible and has implications for the future development of the ontology. In this study, the target paradigms were CM and OM. However, it does not limit the potential of having an ontology for multiple paradigms to extend the capacity of views.

The concept level represented conceptual understanding and theoretical framework of different paradigms. It was not enough to merely declare where knowledge came from. The conceptual relationships of what was in common or not were represented across the paradigms. For example, the Trigger Point from OM was located under Ashi Point from CM by its conceptual meaning. When being classified by nerve innervations which denote neurological theory of trigger point, the traditional acupuncture points in CM can be located under different segmental points from OM, such as LI4 point in CM is in the region of 6th cervical spinal nerve innervation in OM, making it a C6 segmental point. Clearly, LI4 and C6 point do not have direct relationship by CM theory. Yet, conceptually, LI4 point is a subclass of C6 point by the nerve innervation. These relationships were recognised by experts who valued the meaningful practice in differentiating and clarifying the conceptual relations. They commented as ‘it is good to be related in this way. If someone learned CM, but can also look at OM’ and ‘... (it is) convenient for scholars or practitioners from different background. It is creative, in fact, and it is very clear in this.’
The physical organ in OM and functional organ framework in CM were largely different concepts but were normally associated with the same terms. In this study, the physical organs were represented by FMA ontology whereas the functional organ framework was represented in a different hierarchy.

One medical informatics expert suggested the method for point location could also use method from OM, such as reference to axillary line and midline of clavicle. One expert agreed relationship between segmental point and traditional acupuncture point from theoretic perspective. However, one expert was not convinced and believed experimental evidences were still required. From the comments, we recognised that it is achievable to reflect different theories and understanding in a single ontology. However, to make the ontology more acceptable and practical requires the evolvement of culture, attitude and knowledge integration in practice.

The third level of individual provided link between Patterns from CM and disorders or diseases from OM. Acupuncture experts were interested in seeing the applications of ontology. The example application, such as point combination for different health conditions was demonstrated in Sesame. The method of choosing an acupuncture point could be from anatomy knowledge, e.g., practised by physiotherapist; or from classic Chinese medical theory, Pattern, Five Elements, or Meridian, e.g., practised by traditional practitioner. They could only be represented as individual cases. There was neither direct relationship between Patterns and Disorders, nor prescriptions based upon them. Their relationships were through individual clinical findings of patients.

Experts recognised the challenge for knowledge sharing between CM and OM, hence the sophistication of the ontology construction was pointed out and appreciated by one expert that ‘it has got structures and mapping against the structures, so that is incredible sophisticated. And you can query it. ... And it is sort of... it tells you the details, the diagnosis, the symptom. And it is just perfect.’

Through the application demonstration, it showed the great potential of the ontology in supporting decision-making not only from one single perspective, but also across paradigms. Indeed, ‘this is really (where) the value comes (from)’ for knowledge sharing as expert commented.
6.5.2 Term representation

As we discussed in research background, the vocabulary was the important part of ontology for informatics. It was easily understood that the symbol is important for computation as it is also true for human recognitions.

Most term representations were well accepted, in particular those from WHO standards and FMA. However, there are still controversies around different preferred terms. For example, ‘Five Elements’ was the first version of translation of Chinese literature and has been widely used in practice. ‘Five Phases’ was, however, introduced much later for it reflecting better of the original meaning and it was intended to replace ‘Five elements’ in the standard international acupuncture nomenclature from WHO. Yet, the experts were not convinced to use one not the other partly due to their school of thoughts, even though the original concept ‘五行’ in Chinese does bear both meanings of phases and elements. Hence, the support of different languages in ontology term representation was surely a useful feature as experts commented.

The same terms for organs were used in FMA and Chinese medicine, for example, heart. It is not a surprise that acupuncture experts did not question those because the same terms were commonly used for different CM and OM concepts in published literatures and in their mindset. For example, the same term ‘heart’ was used for different concepts in CM and OM. In chapters of research background and acupuncture ontology, we have discussed the difference between term representation and conceptualisation approach. It is important to represent the conceptual differences from the informatics point of view. These different concepts with the same terms (or label) were belonged to different hierarchies. Therefore, the term was not proprietary to one single paradigm of CM or OM. The same term can have different conceptual meanings, such as ‘heart’ as an organ framework in a collection of body’s physiological functions for CM; and a solid anatomic organ drives blood circulation for OM. By purposefully designating the same term to different concepts, it made the conceptual difference clear and gave the freedom of terms to be used in practice as well. This approach was agreed by informatics experts and also recognised by acupuncture expert by commenting ‘it is good to see they were separated (in different hierarchies)’.

Informatics experts suggested the term for properties such as ‘located on’ and ‘local anatomy’ were not helpful. The intended idea of having different properties was to have different values from non-physical entities (e.g. meridian) and the physical
anatomic entities (e.g. nerve, muscle, organ etc.). ‘Located on’ is not clear to be associated to non-physical entities. We took the advice kept the property of ‘local anatomy’. To avoid ambiguity, the property ‘located on’ was changed to ‘belong to meridian’ which is more specific.

It is clear that many of the comments from informatics experts were valuable for their expertise. For example, a valid rejection from informatics expert was that ‘diagnosis’ is a misleading term for clinical findings, hence, we took the advice and the latter was used. However, some suggestions carried the mark of limitation of acupuncture knowledge that they were unfamiliar with acupuncture knowledge where the feedback needs to be evaluated in careful attention. For example, one expert suggested the name for acupuncture points should not use number, such as LI4, LI5 etc. which were standard terms from WHO nomenclature for acupuncture. Another suggestion was that ‘defined depth’ as name of attribute was better than ‘suitable depth’. However, this suggestion was not recognised by other experts.

It was not only the term construction, but also the order of term representations was considered important or useful by some experts as it could help the intuitive understanding of the context and relations between concepts, such as lung meridian next to large intestine meridian helps to indicate their relationship. As we explained to experts that it is not possible to present the ordered list at the time of development and evaluation.

6.5.3 Ontology principles

The evaluation of ontology principles are addressed by expert evaluation. Some principles are evaluated programmatically during the construction of ontology by using function from Protégé OWL plug-in and DL reasoning services. The focus for evaluation of ontology principles are listed in appendix G.

In the section 6.3.2 ‘evaluate ontology programmatically’, we discussed the classification services provided by the DL reasoner and tests built in with Protégé OWL plug-in. These services and functions helped debugging errors in logical definitions of concepts in ontology. The key distinctions were made within the OWL DL language for ontology. Some cycle definition and inconsistence were identified and amended.

The contributions from acupuncture and informatics experts were significantly different and indispensable for the evaluations. Acupuncture experts were not familiar
with the ontology and the requirement of informatics. The principles and criteria of ontology development were not fully appreciated by the acupuncture experts, nor indeed should they be expected to become familiar with these matters.

The evaluations by medical informatics experts were focused on principles and criteria for ontology construction. Informatics expert examined class and subclass in terms of is-a relationship as well as completeness or disjointness for sibling classes. While inspecting the values of attributes, they went further to explore the classes and definition of those classes. It reflected their awareness that term itself can have different meanings.

Informatics experts evaluated the formal definitions of OWL description logic to identify if they were defined correctly. Informal definition and formal definition of concepts were compared. The problem with informal definition was identified at the first informatics evaluation as the expert commented:

'if we are doing that kind of condition by condition (comparison) ... ,
that is very good because that is completely true ... This, on the other hand, I
guess one would say ... very poor because although this is true, there is nothing
at all up here (informal definition), not even a partial range.'

'... the formal definition appears, based on my understanding of the
domain as it develops, should be more complete. It appears to be correct but
completely concordant with the intentional definition.'

Informatics expert criticised that informal definition included comments and supporting notes which should not be part of definition. Therefore, it was difficult for condition by condition comparison with formal logical definition.

Admittedly, informal definitions were not well documented during the development due to lack of time. Some of those included explanations, research note, etc.. Followed the advice, we restricted the informal definition and improvement was noticeable in the following evaluations. We feel that informal definition should be part of ontology documentation although it is easily overlooked.

Understandably, Informatics experts could compare informal and formal definition of concepts. However, they could not make judgment of whether concept was a proper representation because of lacking of acupuncture domain knowledge. For example, they could not comment on whether domains and ranges were precisely delimited or not, as one expert said,
I have seen so far in my knowledge that should be in that list. ... I can't comment without any domain knowledge....'

The gap was filled by domain experts who evaluated informal definitions. Acupuncture domain experts provided valuable feedback on represented knowledge. Domain experts assessed properties’ domain and range and agreed that they were properly and precisely delimited.

The acupuncture experts saw the value and confirmed whether it was the right answer by their domain knowledge. They more or less relied on term representations instead of the definition of the concept. Sometimes, the judgement was made without further investigating as to the real meaning of terms in ontology. We had to remind them to investigate both the domain and the range of properties.

One informatics expert suggested that special function point should be under acupuncture point instead of Fourteen Meridian Point. We could understand the reason was to have ‘pure’ hierarchy based on types of entities from the perspective of informatics. The special function points are grouped by their functions and correspondences. They should be in a separated hierarchy under acupuncture point to distinguish its classification by functions and correspondences. However, because every special function point is one of fourteen meridian points, they are all under the Fourteen Meridian Point. For this case, we were confident to say they were correctly located and this was approved by acupuncture experts. But this raised an interesting issue about how we deal with hierarchies which were not constructed by phenomena type. In particularly, attention is needed for hierarchy which is based on role type. This issue has been discussed in type and role classification under chapter 5.

It was difficult within the limited time of an evaluation session to fully explain the flexible sets of distinctions approach to expert participants, even though the approach, as a meta-level analysis method, had such an important role in guiding us through the whole design and analysis for the ontology construction. Nevertheless, the informatics experts had suggested that the acceptance of ontology should enable us to prove the approach.

One informatics expert suggested the segmental point should not be sibling of fixed and non-fixed location points. Following the comments, the concepts were analysed again. The fix or non-fixed location was thought to be a useful distinction to group acupuncture points. However, we realised that non-fixed location was not well defined because any acupuncture point should have a location. Therefore, fixed location
point and non-fixed location point were removed. This denotes high level concepts need careful analysis and design to consider its logical implications on both perspectives.

A few two-way distinctions were represented as disjoint concepts by OWL-DL. They were evaluated and accepted by informatics experts. However, some concepts that should be disjoint and which were not stated in logical definition were identified. Although one expert suggested to make disjoint as default, we did not agree with this approach because it could result in some concepts that should not be disjoint being defined as disjoint by default. Therefore, careful analysis to specify the disjointness of concepts has proved to be another important task.

Moreover, expert evaluations picked up some minor data entry errors which should be avoided, such as asserted condition ‘all points from large intestine’ was wrong for lung meridian point. It indicated that a manual quality assurance process is always required even though the classifier can sufficiently raise inconsistency errors.

6.5.4 Top level ontology and FMA

The top level ontology facilitates the alignment of ontologies and their contents. The evaluation showed the top level ontology was too abstract and remote to acupuncture experts. Explanations were required to help these experts to understand the purpose.

For example, acupuncture points were under abstract entity as virtual points in ontology. In some expert mindset, acupuncture point should be physical entities since needles are physically inserted and stimulated for purposeful therapeutic treatment. The explanation had been given to experts. Acupuncture points do not occupy any space and time as mountains or trees do. Furthermore, there is no scientific evidence that acupuncture points have any specific anatomical structure to prove their physical existence. Needles and human body are physical entities involved in acupuncture interventions.

It was easier for experts to recognise that meridian is virtual line instead of physical entity. One expert commented that it was easier to understand that the acupuncture point should be abstract entity after realising meridian is virtual line.

The content from FMA was recognised by informatics and acupuncture experts. As we already discussed in term representation, the physical organ in FMA for OM is distinguished from functional organ framework in CM. The reference to anatomy knowledge from FMA was well received by all experts. This indeed would help both
CM and OM practitioners to understand the relationship between virtual acupuncture points and physical anatomical structures.

6.5.5 Coverage

We clearly stated which concepts (acupuncture points and meridians) were core concepts for acupuncture although they might be important in the Chinese medicine, e.g. Yin Yang, Qi, Blood, etc..

Experts were impressed by the details of ontology. There were still suggestions of missing content. For example, pulse locations and correspondence were pointed out as important but missing. It was difficult to draw clear line of which should be included and which should not be represented, particularly for those considered to be supporting concepts. They might not be core concepts of this ontology. However, without them, we may lack of the concepts for the definitions of core concepts.

In our approach, we declared that the supporting concepts are not necessarily complete. The fact is when core concepts increase, so do their relations to other concepts.

6.5.6 User interface

The Protégé OWL plug-in provided a graphical view of the ontology which is better than a simple text based XML file format. During the introduction, demonstration and time were given for the experts to familiarise themselves with the evaluation process and applications interfaces for evaluation. From our observation, it was evident that acupuncture experts needed more time to get familiar with the Protégé user interface unlike the informatics experts who could get into evaluation tasks immediately. The acupuncture experts did improve after some time and were getting used to the user interface during the evaluation with help from evaluator.

The knowledge browser enabled the choices of different perspectives. The acupuncture experts commented that it was very good to show the different views in knowledge browser.

The use of a query interface transformed the task of evaluating the ontology to application-focussed query and answer. To domain experts, it was a process to assess that appropriate questions were being asked and correct answers elicited. Typically, some training of the OWL language and of description logic is usually required to be
able to understand an ontology. However, querying the ontology is an alternative and, getting related answers directly to the application, has helped the domain experts to understand the ontology better, permitting them to foresee the potential use of the ontology in different applications.

Querying using Sesame has opened up the evaluation of a complex ontology without the need to develop a complete application. It has demonstrated to the non-technical domain experts the potential applications of the ontology for decision-support in practice as well as in education. Furthermore, as a side product, it has also helped them to appreciate the ontology’s purpose without the need to understand the formal knowledge representation structures.

Querying helps to evaluate an ontology by assessing correctness of the outcomes of queries at any point within the ontology and it also assisted the development of ontology. We recognise that the query languages are still ‘techie’ and, although more common in feel for those who are familiar with database applications and SQL language, may still be off-putting to lay users.

Feedback from the domain experts were positive, indicating that they understood the questions of query, and found it relatively easy to judge the correctness of the result. The Sesame query improved understanding of possible applications that ontology can be used. Moreover, the results of queries were used to evaluate correctness of ontology. Therefore, the practice of demonstrating to the domain experts of what the ontology can do enabled them to understand the ontology better.

Some suggestions were common to all user interfaces, such as suggestion to have diagram for user interface and more explanation information within the user interface. Because the three applications used for evaluation were not full applications for end users, they could not fulfil the requirement for user friendly. These suggestions will be useful for the development for real end user applications.

6.6 Conclusions of evaluation

The evaluation as part of development process is formative in nature where experts' feedback were provided on current ontology and reflected in subsequent modification and development of ontology, with further suggestion for future development.
The ontology represented knowledge in a formal and structured way which improved the understanding of acupuncture knowledge. The ontology enhanced knowledge sharing and represented both CM and OM perspectives in a single conceptual framework.

Although the term representation was not the key focus of ontology, it was the important basis for informatics and health care professions to communicate. Deploying standard terminologies in ontology was proved to have gained easy acceptance from users with different perspectives.

The detachment between term and concept gave the freedom to the usage of terms, and clear semantics was achieved by the conceptual structure in ontology.

The clearly defined scope of ontology was the key to achieve comprehensive representation of core knowledge.

The principles for ontology construction had been closely followed. The evaluation by programme, acupuncture experts, and informatics experts provided different contributions. The experts from both disciplines of informatics and acupuncture were indispensable.

The top level ontology played important role for ontology alignment although they were not directly related to the domain knowledge.

A sufficient introduction of what is an ontology and possible usage of ontology can help participants get into the role of evaluator quickly.

Knowledge-base browser and query ontology through Sesame facilitated understanding of possible applications. They provided means whereby the end-user can interact with the ontology directly. For our purpose, the deployment of Sesame was found to be both pragmatic in terms of resource and invaluable in engaging expert domain users in a meaningful evaluation of a complex ontology.

The developer also gained more relevant and faster feedback from DL reasoner service and ontology query to assist the formative design process used to construct the ontology.

6.7 Summary

This thesis has chosen acupuncture knowledge to explore how heterogeneous types of knowledge and information can be represented in a single ontology to facilitate
knowledge sharing and have the potential for implementation for academic research and clinical practice.

The evaluation provided evidence that the research aim and objectives have been addressed and achieved. The qualitative evaluation by experts offered feedback from different perspectives of informatics and domain knowledge. The single ontology formally has represented knowledge and maintained the validity for both paradigms. The ontology has allowed users to access the different views of Chinese medicine and orthodox medicine, which achieved the objective of knowledge sharing.

Evaluation as part of development process was included in most ontology development methodologies, hence, such evaluation is formative in nature. The second role of evaluation is to provide feedback for future development.

Most ontology evaluations employ qualitative approach although some quantitative methods have been proposed. Furthermore, it is equally important to consider the approach of evaluation based on properties of ontology and purpose of evaluation. Lack of standard methodology for ontology evaluations is well recognised.

The services provided by description logic reasoner had been applied at every stage of development. This evaluation process helped ‘debug’ common errors programmatically with supporting tool. The consistency check and automatic classification enable the developer to validate formal description logic definition. The results of classification can also help to confirm whether the intended meanings have been defined properly.

Exploring ontology in Protégé provides a mean for experts to look into every detail. Although acupuncture experts were not familiar with requirements for informatics, they provided valuable feedback on content of domain knowledge. Informatics experts assessed ontology from the perspective of information system. The ontology language, constructs, and formal expression in OWL DL were examined and agreed by informatics experts.

Knowledge-base browser and query ontology facilitated domain experts understanding of potential applications. Querying ontology allows the domain experts to ask questions which enable them to actively engage with evaluation. They preferred query interface so that they can view results and assess the correctness of query results. In this way, querying ontology has helped the domain experts to evaluate the ontology. Meanwhile, the query application demonstrated possible ontology applications for decision support, academic teaching and learning.
However, querying the ontology can not replace the thorough and time-consuming process of following the hierarchy of the ontology to assess the details of concepts and relationship.

This ontology development is a learning process with the ontology in its initial stage. Further development of ontology and knowledge based systems will proceed incrementally.
Chapter 7

Summary and Conclusions

The aim of this chapter is to present the summary of the thesis and to critically evaluate the work so that assessment of the achievement of the aim and objective can be made. The contributions and future research are outlined at the end of thesis.

7.1 Thesis overview

This thesis begins with the outline of the research aim, objects and contributions. It is followed by the introduction to the relations of acupuncture in Chinese medicine and orthodox medicine from earliest history to current practice. The motivation and requirements for a formal knowledge representation of acupuncture were addressed in terms of clinical practice, academic research and knowledge sharing across CM and OM paradigms.

Chapter 2 is concerned with research background in literature. It covers the background of ontology and challenges of ontology as a formal knowledge representation for acupuncture knowledge sharing. There are different views of ontology from philosophy, informatics, semiotics and linguistic perspectives. In particular, ontology was recognised for the importance of knowledge representation and sharing in informatics. The study of ontology has attracted attention in informatics within which different views have been established. The vocabulary view focuses on the terminology aspect of computing. Therefore, coding of shared terms can be drawn to support communications among different systems. This is also true for early medical terminology which emphasised on agreed and controlled medical vocabulary. The conceptualisation view of 'an explicit specification of a conceptualisation' (Gruber
Ontology has been studied since ancient philosophy. Increasingly, there have been growing interest and research over 30 years in informatics. Although principles for ontology construction were proposed over that time, it still lacks of well accepted and established methodologies. The systems, languages and tools for ontology were developed from different areas of informatics. Most of them are at the stage of academic research and unfit for real life applications. Nevertheless, the efforts of international standardisation of language and initiatives on standard and upper level ontologies provide a footing for ontology development. Presently, interest from medical and health informatics are growing with the number of ontologies increasing in different areas and for a variety of applications. However, few ontologies for Chinese medicine have been developed and none address the relations between CM and OM at the time of research. The challenges for developing such an ontology are described at the end of chapter 2.

Chapter 3 presents the methodological underpinnings of the thesis. It focuses on presenting the theoretical aspects of research methodology and data collection methods used. The research adhered to the constructivism paradigm, in particular, it applied design and creation methodology which is specific for informatics research. The data collection methods were participant observation and semi-structured interview along with digital video recording and screen capturing. The detailed data collection by video recording were analysed by qualitative approach and discussed in chapter 6 of ontology evaluation.

In chapter 4, the key activities of ontology design, methods for ontology development, system and knowledge analysis are discussed. The development environment Protegé is an open source frame based system. The plug-in architecture of Protegé enables support of development ontology in three versions of OWL languages: OWL full, OWL DL and OWL lite. The mixture of stage-based and evolving prototype methodology for ontology development were followed and research activities were grouped into Identify, Categorise and Test.

We proposed the flexible sets of distinctions approach for ontology by extending Sowa’s two-way and three-way distinction. The approach was applied for analysing
protégé system and domain knowledge for representation. The technical approach of meta-model in the system and how flexibility and extensibility can be achieved were also investigated.

Chapter 5 presents the ontology of acupuncture, which represents concepts, theories and logics from perspectives of CM and OM. The technical and standard solutions, e.g. description logics, OWL, adopted in the ontology were presented with individual examples. The solutions demonstrate how we have overcome the challenges and highlight the technical detail of the ontology.

Chapter 6 discusses evaluation as part of the ontology development. As well as the regularly run evaluations by the programme of DL reasoning services, the evaluations by experts were further conducted to produce qualitative feedback for ontology development. Three applications were used for evaluation. First, Protégé with OWL plug-in was used for navigation of hierarchies of the ontology. Second, a knowledge browser was developed by the researcher to demonstrate choices of different perspectives and visibility. Third, the Sesame query interface was set up to illustrate the ability and possible applications of the ontology by asking and answering questions.

The evaluations were focused on expert evaluations with data collected by observation and interview. The whole process of evaluation and participant’s actions were captured and recorded onto DVDs. The data was analysed and coded with a qualitative video analysis tool, Transana. It was acknowledged that the ontology enabled a mean for the acupuncture knowledge representation and sharing for different paradigms and the ontology followed well recognised principles. The usage of standard terminology and FMA was well accepted by experts.

7.2 Critical evaluation

Different to the ontology evaluation in the last chapter, the critical evaluation is focused on the aim and objectives of the research itself.

7.2.1 Assessment of the research aim and objectives

The aim and objectives are presented in section 2 in chapter 1. For the convenience of discussion, they are repeated here.
The aim of this research is:

To formally represent the concept, theory and logic of acupuncture from distinct paradigms of both orthodox and Chinese medicine in a single ontology and thereby make implicit relationships explicit and accessible across the paradigms.

The objectives of this research are:

1. To explore the significance of ontology as a means of knowledge representation for acupuncture in health care
   a. To explore the development of ontologies in health care
   b. To examine the challenges of knowledge representation across paradigms
2. To investigate the ontology development methodology so as to respect of representing different perspectives in a single ontology
   a. To explore the characteristics of ontology in informatics
   b. To explore current approaches for ontology construction
   c. To investigate the flexible sets of distinctions and apply it for analysing system and domain knowledge
3. To develop an ontology in a formal description logic language
   a. To develop a single ontology which represents both paradigms
   b. To investigate how to maintain validity for cross-paradigms representation
4. To evaluate the ontology for knowledge representation and knowledge sharing
   a. To validate the ontology programmatically in development environment
   b. To evaluate ontology for knowledge representation from the two disciplines (informatics and health care)
   c. To evaluate ontology for knowledge sharing between the two paradigms (CM and OM)
Evaluation of achievement of the research aim and objectives

In this section, the research objectives are discussed first and are followed by the achievement of the research aim. The references for the achievement of the research objectives in the thesis are indicated by the section and page number.

1. To examine the significance requirements and challenges of acupuncture knowledge representation for different paradigms of CM and OM

Rooted from traditional Chinese medicine, acupuncture has been increasingly practised by OM health care professions in the West. The difference of health concepts and theories of CM and OM in acupuncture knowledge is exemplary. The significance of different views on human body and health conditions was imposed upon the acupuncture practice. It may be easy just to put them as two separate paradigms and study them in isolation. However, we have argued the need of representing different views for knowledge sharing. In the best interest of patient benefits, it can not be denied that there is a requirement for knowledge sharing or at least to be clear as to what benefit they can or cannot offer. (Chapter 1: section 1.5; page: 12)

The unique feature of acupuncture raises great challenges for knowledge representation. After the brief review of the history and current practices of acupuncture, we examined the challenges for knowledge representation across paradigms and discussed heterogeneity of paradigm, language, ontology and content. We concluded acupuncture knowledge representation as a case study in informatics has extended representing different perspectives of same subject to the extreme. (Chapter 2: section: 2.6; page: 46)

2. To investigate the characteristics of ontology and development methodology with respect of representing different perspectives

The characteristics of ontology vary in different disciplines, philosophy, informatics, linguistic and semiotics. It is concluded that semiotics view of vocabulary, conceptualisation and reality was more appropriate for the purpose of this study. (Chapter 2: section 2.2; page: 18)

The commonly used methodologies can be simplified by grouping the key tasks into three categories: identify, categorise and test. (Chapter 4: section 4.3; page 74)

Furthermore, we extended the distinction approach from Sowa and proposed flexible sets of distinctions. It is not easy to have a pure single view of domain
knowledge even within one single paradigm. A degree of flexibility is required to support representation of different views. (Chapter 4: Sect 4.4; page 80)

The extreme case of acupuncture has brought great challenge for supporting different views of health care. The flexible sets of distinctions approach enabled us to represent different theories and views without losing the logic consistence. The approach was explained and demonstrated by analysis of protégé system and domain knowledge (Chapter 4: section 4.4; page 84)

3. To develop an formal knowledge representation in description logic language
The research is the first kind to represent acupuncture knowledge for both CM and OM in the formal OWL-DL language. Different formalisms for knowledge representation exist and ontology is recognised as the key for knowledge representation and sharing. Development of ontology is the key activity and the acupuncture ontology is one of the main outcomes of the research. (Chapter 2: section 2.4; page:34 )

The description logics language offers great advantage for automatic classification and clear semantics. The DL reasoning services have been applied during the development of the ontology to ensure consistency. The relationship between CM and OM was clearly represented by the DL without violating domain theory and logic. The automatic classification enabled reviewing relations across paradigms and comparing concepts for subsumption and equivalence. (Chapter 5: section 5.3.3; page:101)

4. To evaluate the ontology for knowledge representation and knowledge sharing
The methods for evaluation of ontology are still subject of study. The qualitative evaluation is the main approach although some quantitative methods have been proposed. (Chapter 6: section 6.3; page:119)

Experts from two disciplines (health care and medical informatics) and two paradigms (CM and OM) were invited for evaluations of the ontology. It is clear that the attitude of ‘users are always right’ only reveals the fulfilment of users’ requirement. We argued that it is equally important to evaluate ontology from the informatics that the ontological principles have been followed. The evaluations were conducted in three applications, Protégé, knowledge browser, and sesame web application in order to facilitate evaluation and demonstrate the ontology. (Chapter 6: section 6.3.3; page:128)
The evaluation applied observation and interview as methods for data collection. The videos and transcripts were analysed using qualitative video software. The feedbacks from evaluations were positive and encouraging. The questions and different opinions were discussed at interviews and changes made accordingly. (Chapter 6: section 6.4; page:134; section 6.5; page:141)

### 7.2.2 Validity, Reliability, and Generalisability

The key criteria such as validity, reliability and generalisability have been commonly applied to evaluate and justify positivist quantitative research. Qualitative research has raised issues whether such criteria are still applicable in qualitative research and how significant they are. From social study, alternative criteria arise such as credibility, trustworthiness and authenticity proposed by Lincoln and Guba (2003). However, the discussion of validity and reliability offers a helpful starting point for the researcher and it is useful to ensure the quality of the research.

#### Reliability

'Positivist notions of reliability assume an underlying universe where inquiry could, quite logically, be replicated.' (Marshall and Rossman 1989, p.147). It can be argued that reliability may not be a key criterion for qualitative research. However, a researcher can increase the reliability of data and interpretations by ensuring the quality of recording and documenting data (Flick 2006, p.369). To improve the reliability the following methods suggested by Silverman (2001, p.228) were applied in the research:

- Video recording all face-to-face interviews
- Careful transcription and reliable analysis
- Present data in research report
- Computer assisted recording and analysis of data

The methods were described and discussed in chapter 6. The introduction was given to participants before evaluation where the participants were made aware of the process and the freedom of comments. The transcriptions were completed by an expert who provided advices during the study and checked by the researcher for accuracy. The video data was coded to generate findings. However, the suggestion of the coding and data analysis should be done 'blindly' was not followed owing to resource limitation.
Hussey and Hussey (1997) asserted that similar observations and interpretations from different occasions and or by different observers can provide evidence for reliability. In the research, six evaluations were conducted and the similar feedbacks were found.

**Validity**

Validity is that ‘the research findings accurately represent what is really happening in a situation.’ (Hussey and Hussey 1997). However, the claim of truth and knowledge are influenced by theoretical, philosophical, and pragmatic rationales of the inquiry and ‘the issue around validity is the conflation between method and interpretation’ (Lincoln and Guba 2003, p.274).

There is no single method that can deliver the ultimate truth. However, triangulation of different methods (observation and interview applied for evaluation in this study) can enhance the validity of the research and the coherence of data from different methods or even variant of data with reasonable explanations are acceptable evidence (Patton 1990, p.467). In the study, all evaluations were observed and comments were recorded. Then the interview was conducted to discuss observation and comments in order to understand the rationale and verify the comments.

Both qualitative and quantitative researches require interpretation of data. The 'interpretively rigorous' issue has received the most attention (Lincoln and Guba 2003, p.275). There should be clear distinction between the collected data and interpretation by the researcher and further conclusion drawn from supporting findings with adequate evidence for key claims (Seale 1999, p.50). In the thesis, the interpretation and conclusions were clearly presented and linked with video clips and transcripts in the analysis software.

This distinction between data and interpretation provided sufficient information for ‘respondent validation’ (Silverman 2001, p.235) to validate the research findings back to research participants. It would be ideal to have this kind validation at first stage of development or at further development stage. However, the validation itself may not aid accuracy to the data or interpretation although they generate further data for further analysis (Silverman 2001, p.236).
Generalisability

‘Generalisability is a standard aim in quantitative research and is normally achieved by statistical sampling procedures’ (Silverman 2001, p.248). However, these sampling procedures are not suitable in qualitative research. It is common that the sample numbers is small and the selection of samples is not randomised. In this thesis, statistical generalisation is not applicable. The purposive sampling was employed and representative experts from different perspectives of disciplines and paradigms were elected for evaluations.

Another problem of generalisation in qualitative researches is that ‘its statements are often made for a certain context or specific cases’ (Flick 2006, p.391). Therefore, it is important that the research clarifies what degree of generalisation can be made. In this study, how ontology was developed and evaluated was explained in detail. The conclusion is that a formal knowledge representation across paradigms can facilitate knowledge sharing. The ontology satisfied both informatics principles and clinical requirements. The methodology can be applied to other similar studies especially for representing different views in a single ontology.

Having discussed the criteria that justified this thesis, it is also important to make clear its limitations.

7.2.3 Thesis limitations

The scale of the research is small with limited resources. The research was conducted by a sole researcher with participant experts from health care and medical informatics. The ontology only covered a small portion of acupuncture knowledge due to constraint of time and resources. The key challenges and requirement were addressed and the evaluations and demonstrations received positive feedback. However, it should be treated as prototype instead of a final product for real life application.

The experts involved in research were from both academic and clinical practice. The number of experts who involved in the research was small. However, it is not practical to gain access to a large number of experts. It can be ideal to invite representatives from professional bodies through which participant experts were members. To consult and gain support from professional bodies was considered at research design but not implemented because the research aim is to prove the concept in academic research. We would, however, recommend that future research or applications
that build upon this research should consult professional bodies to authenticate ontology content and seek possible broader feedback to minimise bias from individual expert.

The development of the ontology is largely based on academic text books and advice from the clinical experts. It would provide evidence for practical daily usage if clinical records can be included as part of the resources available for building ontology. However, it was not applied in practice due to limited resources and time constraints. It can, however, be included in the future development and refinement of the ontology.

At time of completion, the research only finished the first circle of ontology development with some modifications and changes applied after each evaluation. We consider this to be sufficient for academic research. However, further development is required; for example, some sections of ontology were not well developed and only limited numbers of key concepts were fully defined. Some similar concepts which can adopt same approach were left undefined to save the time on development.

The following section presented a discussion of the thesis in terms of the literature.

7.2.4 The thesis in terms of the literature

This section discusses the research in terms of the literature. The general points identified in research background and other chapters are given below. Each point is followed by a discussion of how this thesis addresses the issue.

1. It is important to present the therapies in the context of their own philosophies and models of health and illness and bridges the gap between OM and CAM can benefit healthcare professionals, patients, and healthcare system (Berman 2001)

Lack of knowledge is recognised as one of the greatest barriers to appropriate use of alternative medicine. This thesis has investigated the requirements and challenges in pursuing better knowledge sharing which serves healthcare.

The formal representation of the acupuncture ontology can facilitate better understanding and communication across both CM and OM. In particular, the representation preserved the concepts, theories, logic of the individual paradigm and provided a mean to share the knowledge.
2. The meaning of ‘ontology’ has changed from its philosophy root in informatics.
The thesis investigated the different views of ontology existed in informatics. We summarised that there were three influential views: vocabulary, conceptualisation, reality. However, these views only stressed the importance of one or another. We believe the view from semiotics can bring the three views into a single framework. It addresses the relationships among vocabulary, concepts, and reality, and should be applied at every step of analysis and design for information system.

3. Lack of methodology for ontology development and it is an art instead of engineering.
We proposed flexible sets of distinctions approach by extending Sowa’s distinction approach. This approach is of generic natural. We applied the approach in the analysis of systems and domain knowledge. The flexibility in applications can only be achieved if the different sets of distinctions were clearly defined and specified. Flexibility is also a vital factor in the design of ontology that the different sets of distinctions can be introduced as knowledge evolves. However, it is necessary that new sets of distinctions need to be checked against existing distinctions. Their relations have to be determined, hence, the consistency ensured. The method to specify different sets of distinctions may vary. This approach can enable representation of different views and preserve the consistency of logic behind.

4. The acupuncture ontology and other representations
This thesis is the first kind that represents different perspectives of acupuncture knowledge from both CM and OM in formal OWL-DL language. There are two projects representing Chinese medicine followed representation format of UMLS. Their focus was terminology for Chinese medicine only and none addressed the interface between CM and OM.

5. Ontology as Knowledge representation
The relationship among vocabulary, concept, and reality has been discussed in a number of papers. The stress of importance of one or another is closely linked to the domain of interests and commitment of philosophy. In a simplistic term, we mean that vocabulary is any symbolic representation in an information system (generally can be divided into representation for human or machine) and concept is the meaning of the symbolic
representation. Without concepts, vocabulary is meaningless. From controlled vocabulary to specification of conceptualisation, the semantic has been emphasised. One key issue with the concept approach is by what degree we have an agreed specification. We argue that it is not easy to achieve even within one single knowledge domain. The reality view does not help in any degree as the fundamental debate on what is reality is subject to commitment to the different philosophical grounds, i.e., basically constructionist vs. positivist.

The researcher is faced with the same issue which inflated the challenges. The thesis investigated the methods to specify concepts from different perspectives. The conventional approach treats them as *different concepts* with different machine symbolic representations at concept level (e.g. concept ID) and associate them with different or same human readable representation at vocabulary level (terms, labels). In this thesis, we argued an entity of reality can be conceptualised and have some degree of agreement or disagreement. If the disagreement is fundamental, the approach will follow the conventional approach to treat as different concepts. If there is some degree of shared understanding, then we treat the agreement of the conceptualisation as a basis and add on different perspectives so they are explicitly stated by meta data. In this way, a single entity can have shared conceptualisation. Differences of conceptualisation for specific perspective can be represented. The logical consistence is ensured by the description logics reasoning services.

An example in the context of this study is acupuncture point. It can have some agreed attributes and values, such as anatomy structure for both CM and OM; and some attributes which are specific to one paradigm e.g. belong to meridian for CM.

6. Evaluations from different disciplines and paradigms

The evaluations were performed by experts from two disciplines of health care and health informatics. The design of evaluation reflected such differences. Informatics expert evaluated the ontology principles, logic expressions, and semantics between informal and formal definitions. Health care experts assessed the content of ontology to obtain whether domain knowledge has been properly expressed as they expected.

7. Technical evaluation and applications for evaluations

The reasoning services provided by description logics were applied during the ontology development. The consistency check and automatic classification were routine
assessment of ontology content and were followed by refinements by the researcher. It proved to be time-saving on development and facilitated identifying and correcting errors.

The applications of Protégé and knowledge browser (based on Protégé API) facilitated the navigation of ontology and display different views. Unlike the approach taken by FMA whereas a query interface was specially developed for ontology evaluation, we applied open source platform, the Sesame, as web application for queries. The ontology was uploaded into the Sesame and the queries were performed in a web interface in SeSQL. Compare to developing a specific query interface to fulfil this task, the above method was efficient and much less time-consuming. Since the ontology is intended for knowledge sharing, to adopt a third party query interface has enabled the researcher to demonstrate that the ontology can be used in different software system apart from its development environment Protégé.

Moreover, by following the standard of ontology, ad hoc approach is intentionally avoided since the latter may run the risk of working perfectly in one system but not others.

7.2.5 Contributions

This thesis makes contributions to health informatics and its practitioners. These are listed below, in order of significance:

- Developed the first validated acupuncture ontology which composed of both orthodox and Chinese medicine knowledge - The background of research - The design and implementation of ontology and ontology evaluations (Chapter 2, Chapter 4, Chapter 6)

- Different paradigms of knowledge were represented, in particular, their relations were explicitly defined (concept level, instance level and meta level). The background of research - The acupuncture ontology (Chapter 2, Chapter 5)

- Contributed original work based on empirical evidence to the academic discipline of Health Informatics by investigating the flexible sets of distinctions approach for ontology construction - The background of research - the design and implementation of ontology - the acupuncture ontology (Chapter 2, Chapter 4, Chapter 5).
• Argued dynamic and flexible definitions of a concept can support different perspectives which enable the knowledge sharing - The background of research - the design and implementation of ontology - the acupuncture ontology (Chapter 2, Chapter 4, Chapter 5)

The findings may be helpful to those who are faced with the challenges of representation for different perspectives because of different theories, logics and philosophical commitment. The flexible sets of distinction approach can enable a dynamic and flexible constructs and systems.

7.3 Future work

The research is the first kind of formal representation in this specific domain knowledge area. As the research is limited by the scope and resources, only the first circle of development was finished and some contents were not completed and further actions are needed according to experts' feedback.

Moreover, involvement and support should be sought from professional bodies to further ensure validity and quality of the ontology.

It will be ideal to examine the applicability of the ontology as core knowledge content for applications. Although the queries of ontology have demonstrated the possible questions can be asked and answered. The suggested areas for applications are:

• teaching and learning applications in education
• data retrieval applications for different databases or internet
• clinical record systems and decision support systems
References


ETCMA. "Legal situation in Belgium." Retrieved 1 Dec 2008, from www.etcma.org/content/view/16/33/.


Young, G. and Pennick, V. E. (2007) "Interventions for preventing and treating pelvic and back pain in pregnancy." Cochrane Database Syst Rev Volume, DOI:


## Appendix

### Appendix A Knowledge representation and sharing

<table>
<thead>
<tr>
<th>Hierarchies of content</th>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acupuncture point</td>
<td>Very good</td>
<td>Knowledge sharing and representation of different perspectives</td>
<td>Knowledge sharing comments, Paradigms, Perspective1, KB browser summary, Different perspectives</td>
</tr>
<tr>
<td></td>
<td>Very good</td>
<td>The relationship between segmental points from OM and traditional acupuncture points from CM.</td>
<td>Comments, C7-LI4, KB C7-LI4, CM OM, Perspective1, Superconcepts of LI4, LI4 and paradigm, CM &amp; OM1, CM &amp; OM2</td>
</tr>
<tr>
<td></td>
<td>Good</td>
<td>The relationship between Ashi point and trigger points.</td>
<td>Perspective3, CM &amp; OM</td>
</tr>
<tr>
<td></td>
<td>Agree</td>
<td>Extra point hierarchy in ontology</td>
<td>Extra point</td>
</tr>
<tr>
<td></td>
<td>Agreed with doubt</td>
<td>CM acupuncture points were linked to OM segmental point by never innervation. Needs experimental proof.</td>
<td>LI4 and SP6, LR3, Demonstration</td>
</tr>
<tr>
<td></td>
<td>Agree</td>
<td>Expert recognised the challenge for knowledge sharing between CM and OM.</td>
<td>Ashi point, Knowledge sharing</td>
</tr>
<tr>
<td>Organ</td>
<td>Organ 1, CM and OM, Perspective</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finding</td>
<td>The clinical finding from different CM and OM perspectives</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prescriptions</td>
<td>Prescription, Point prescription</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method of locating</td>
<td>Use virtual line e.g. axillary line for locating acupuncture point</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organ</th>
<th>Good, Very good</th>
<th>The distinction between physical organ in OM and functional framework in CM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finding</td>
<td>Very good</td>
<td>The clinical finding from different CM and OM perspectives</td>
</tr>
<tr>
<td>Prescriptions</td>
<td>Good, Very good</td>
<td>The usage of point combination for different health conditions</td>
</tr>
<tr>
<td>Method of locating</td>
<td>Suggestion</td>
<td>Method of locating</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

195
<table>
<thead>
<tr>
<th>Content Hierarchy</th>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acupuncture point</td>
<td>Good</td>
<td>Term representation with multiple languages support which based on WHO standard</td>
<td>Query meridian points, Language support, Query LI4</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>Preferred term v.s. synonyms Five element is preferred</td>
<td>Five transporting point, Five phases</td>
</tr>
<tr>
<td></td>
<td>Poor</td>
<td>Acupuncture point named by numbers</td>
<td>Meridian Point</td>
</tr>
<tr>
<td></td>
<td>Suggest review</td>
<td>‘defined depth’ is better than ‘suitable depth’</td>
<td>Suitable depth</td>
</tr>
<tr>
<td></td>
<td>Suggest review</td>
<td>Property of ‘located on’ is not helpful</td>
<td>Located on</td>
</tr>
<tr>
<td>Prescription</td>
<td>Agree</td>
<td>Prescription of acupuncture point</td>
<td>Four gate prescription</td>
</tr>
<tr>
<td>Organ</td>
<td>Agree</td>
<td>Same term used for different concepts</td>
<td>Organ1</td>
</tr>
<tr>
<td>Finding</td>
<td>Disagree</td>
<td>diagnosis is not a suitable name for clinical finding</td>
<td>Diagnosis</td>
</tr>
</tbody>
</table>
## Appendix C Principles for ontology construction

### Hierarchies of content

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good and Poor</td>
<td>The formal definitions were fine. Informal definition included comments and supporting notes which should not be part of definition. It is difficult for condition by condition comparison between formal logical definition and informal definition.</td>
<td>Acupuncture point, Acupuncture point1, Fix location point2, Segmental point, Segmental point1, Ashi point2, Ashi point3,</td>
</tr>
<tr>
<td>Very good</td>
<td>Matching between informal definition and formal logical definition</td>
<td>LU11 definition</td>
</tr>
<tr>
<td>Very good</td>
<td>Formal logical definitions were consistent and domain and range of properties were correct.</td>
<td>Fixed location point, Fix location point2</td>
</tr>
<tr>
<td>Good</td>
<td>The logical definition is correct and consistent which represent domain knowledge.</td>
<td>Fourteen meridian point, Lung meridian point, C7, Trigger point, Transporting point, Located on</td>
</tr>
<tr>
<td>Agree</td>
<td>Agreed domain and range of properties are correct. The constructs of OWL language are used properly in formal logical definition.</td>
<td>Method of locating3, Method of locating4, Fixed location point, Suitable depth, Segmental point, Ashi point, Lung meridian point, C7, Order on meridian,</td>
</tr>
<tr>
<td>Acupuncture meridian</td>
<td>Good</td>
<td>Represent domain knowledge</td>
</tr>
<tr>
<td>----------------------</td>
<td>------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Agree</td>
<td>Domain and range of properties</td>
<td>Local anatomy, Acupuncture meridian</td>
</tr>
<tr>
<td>Agree</td>
<td>The constructs of OWL language</td>
<td>Flow direction</td>
</tr>
<tr>
<td>Error</td>
<td>Asserted condition ‘all points from large intestine’</td>
<td>Lung meridian</td>
</tr>
<tr>
<td>Good</td>
<td>Domain and range of the property</td>
<td>Prescription of four gate points</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prescription</th>
<th>Agree</th>
<th>Relation among disease, disorder, and pattern</th>
<th>Disease disorder and pattern, Disease, Disorder and its subclasses, Disorder and disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disease,</td>
<td>Agree</td>
<td>Distinctions are presented by</td>
<td>Distinction</td>
</tr>
</tbody>
</table>

**Suggested to review**
- Special function point should be under acupuncture point instead of fourteen meridian point
- Properties which related to anatomy were overlapped.
- Informal definition is poor and is not clearly defined.
- Segmental point as sibling of Fixed location point and non-fixed location point
- The disjoint of concepts was not stated in logical definition
- Matching between informal definition and formal logical definition

**Poor**
- Properties
- Trigger point, Fix location point2

**Disagree**
- Segmental point, Segmental point1

**Error**
- Method of puncturing

**Very good**
- Lung meridian
| Qualitative value | disjoint of concepts in OWL |
Appendix D  Top level ontology and FMA

<table>
<thead>
<tr>
<th>Hierarchies of content</th>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acupuncture point</td>
<td></td>
<td>Acupuncture expert feel top level ontology was too abstract and remote to clinical practice. Explanation was needed to understand the purpose distinguish physical or non-physical entity.</td>
<td>Acupuncture point, Acupuncture point2, Space and time, Top level ontology, Anatomy of LI4</td>
</tr>
<tr>
<td>Acupuncture meridian</td>
<td>Very good</td>
<td>It is easier to recognise that meridian is virtual line instead of physical entity.</td>
<td>Meridian1</td>
</tr>
<tr>
<td>Organ</td>
<td>Good</td>
<td>Organ in CM and OM</td>
<td>Organ1, CM and OM, Liver, Heart</td>
</tr>
<tr>
<td>Qi, Blood</td>
<td>Agree</td>
<td>hierarchy of Qi and Blood.</td>
<td>Qi and Blood</td>
</tr>
</tbody>
</table>
### Appendix E  Coverage of ontology

<table>
<thead>
<tr>
<th>Hierarchies of content</th>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Good</td>
<td>Coverage was comprehensive and detailed</td>
<td>Coverage, Migraine</td>
</tr>
<tr>
<td></td>
<td>Suggestion</td>
<td>Missing pulse locations and correspond</td>
<td>Pulse location, Pulse correspond</td>
</tr>
</tbody>
</table>
## Appendix F  User interface of applications

User interface

Three user interfaces of protégé, KB browser and Sesame application were commented by participants in evaluation.

<table>
<thead>
<tr>
<th>Application</th>
<th>Feedback</th>
<th>Brief description of finding</th>
<th>Reference of Clips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protégé, KB browser and Sesame</td>
<td>Suggestion</td>
<td>Suggested to have diagram for user interface</td>
<td>User view</td>
</tr>
<tr>
<td>Protégé</td>
<td></td>
<td>Domain experts were not familiar with protégé user interface</td>
<td>User interface</td>
</tr>
<tr>
<td>KB browser, Sesame</td>
<td>Very good Suggestion</td>
<td>It is very good to show different perspective in KB browser. Need explanation information in user interface</td>
<td>Comments on KB browser, Paradigm, KB browser summary, User interface</td>
</tr>
</tbody>
</table>
### Appendix G  The focus of evaluation on ontology principles

<table>
<thead>
<tr>
<th>Clarity</th>
<th>Programme</th>
<th>Informatics</th>
<th>Acupuncture</th>
</tr>
</thead>
<tbody>
<tr>
<td>The definitions meet the structural criteria for complete definition (a predicate defined by necessary and sufficient conditions is preferred over a partial definition defined by only necessary or sufficient conditions)</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Domain and range of the relations and functions are exactly and precisely delimited</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>The generalization/specialization of a given class exactly and precisely represents the domain knowledge</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Informal definition written in natural language is equivalent to the formal definition</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Key distinctions are made within the language (OWL), so that all conclusions can be drawn from the ontology alone (No Extra-Ontological Distinctions)</td>
<td></td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>All assumptions are made explicitly? (As long as these assumptions remain implicit, the potential for disagreement is present. No Hidden Assumptions)</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Coherence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>It should sanction inferences that are consistent with the definitions. At the least, the defining axioms should be logically consistent.</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Extensibility</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>One can define new terms for special uses based on the existing vocabulary, in a way that does not require the revision of the existing definitions</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Minimal encoding bias</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The conceptualisation should be specified at the knowledge level without depending on a particular symbol-level encoding</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Minimal ontological commitment</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Defined terms are essential to the communication of knowledge consistent with that theory</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Syntax of the definition</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The OWL language has been correctly used to present knowledge</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Constructs of the definitions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The logic expressions have been correctly used to define concepts</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Ontological Distinction Principle</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Classes corresponding to different identity criteria must be disjoint.</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
Appendix H Scripts for ontology queries

Concept level queries:

/* subclass query */
select *
from {point} rdfs:subClassOf {ns7:Lung_Meridian_Point}

select *
from {ns7:LI4} rdfs:subClassOf {Acupuncture_Point} ns7:paradigm {para}
where para like "CM"

select *
from {ns7:LI4} rdfs:subClassOf {Acupuncture_Point} ns7:paradigm {para}
where para like "OM"

select *
from {ns7:LI4} rdfs:subClassOf {class}

/* language query acupuncture point */
select *
from {Acupuncture_Point} rdfs:label {L}
WHERE lang(L) LIKE "en"

Instance Level Queries:

/* list all properties and it value of a instance */
select *
from {ns7:LI4} property {x}

/* meridian */
select *
from {ns7:li4} ns7:belong_to {meridian}

/* local anatomy */
select *
from {ns7:li4} ns7:local_anatomy {anatomy}

select *
from {ns7:li4} ns7:belong_to {meridian} rdf:type {} ns7:paradigm {x}
where x like "CM"

select *
from {ns7:li4} ns7:local_anatomy {anatomy} rdf:type {x} rdfs:subClassOf {y} rdfs:subClassOf {z}

/* location */
select *
from {ns7:li4} ns7:method_of_locating {} ns7:location_description {description}

select method_of_locating, description
from {ns7:li4} ns7:method_of_locating {} ns7:location_description {description};
rdf:type {method_of_locating}

/* puncturing */
select *
from {ns7:lr3} ns7:method_of_puncturing {} ns7:needle_angle_to_skin {angle};
ns7:direction_along {direction}

/* depth of puncture */
select *
from {ns7:li4} ns7:suitable_depth_of_needling {} ns7:value {depth_between}: rdf:type {Measure_type}

/* point to point prescription */
select point, x, y
from {ns7:lr3} ns7:combination {x} ns7:prescripted_point {point}; rdf:type {y}

Select distinct *
from {ns7:lr3} ns7:combination {} rdf:type {Prescription_type}

Select distinct *
from {ns7:lr3} ns7:combination {prescription} rdf:type {prescription_type}:
ns7:prescripted_point {point}

/* search point by label of instance for a particular Concept */
select point, l, meridian
from {point} rdfs:label {l}; ns7:belong_to {meridian} rdf:type {ns7:Liver_Meridian};
where lang(l) like "en" and l like "lr3"

/* search point by lable of different language */
select point, l, meridian
from {point} rdfs:label {l}; ns7:belong_to {meridian} rdf:type {ns7:Liver_Meridian}
where lang(l) like "ch" and l like "taichong"

/* point to symptom */
Select distinct *
from {ns7:li4} ns7:combination {} ns7:for_diagnosis {} ns7:has_manifestation {}
rdf:type {manifestation}
/* Point has treatment effects for the symptoms and its diagnosis. It is a knowledge level question instead of instance level (it is a question of general not particular) and it is presented by its concept type. */

/* point to diagnosis */
Select distinct *
from {ns7:li4} ns7:combination {} ns7:for_diagnosis {} rdf:type {Diagnosis}

/* point for CM diagnosis */
Select distinct *
from {ns7:li4} ns7:combination {} ns7:for_diagnosis {} rdf:type {Diagnosis}
ns7:paradigm {x}
where x like "CM"

/* point for OM diagnosis */
Select distinct *
from {ns7:li4} ns7:combination {} ns7:for_diagnosis {} rdf:type {Diagnosis}
ns7:paradigm {x}
where x like "OM"

Clinical findings

/* Type of Headache */
Select *
from {headache_type} rdfs:subClassOf {ns7:Primary_Headache}

Select *
from {headache_type} rdfs:subClassOf {ns7:Tension-type_headache}

Select *
from {headache} rdf:type {ns7:Chronic_tension_type_headache}

/* headache feature */
Select distinct *
from {} rdf:type {ns7:Chronic_tension_type_headache}: ns7:has_manifestation {}
ns7:has_feature {y}

/* headache and it patterns */
Select *
from {} rdf:type {ns7:Chronic_tension_type_headache}; ns7:pattern_diagnosis {}
rdf:type {pattern}

Select distinct *
from {} rdf:type {ns7:Chronic_tension_type_headache};
    ns7:has_prescription {} ns7:prescripted_point {point}

/* headache, its pattern and its symptoms */
Select pattern, symptoms
from {OM_Diagnosis} rdf:type {ns7:Chronic_tension_type_headache};
ns7:pattern_diagnosis {} rdf:type {pattern}; ns7:has_manif estation {symptoms}

/* headache prescription */
Select *
from {OM_Diagnosis} rdf:type {ns7:Chronic_tension_type_headache};
ns7:has_prescription {} ns7:prescripted_point {point}

/* mental disorder */
Select *
from {X} rdfs:subClassOf {ns7:Mood_Disorder}

Select *
from {X} rdfs:subClassOf {ns7:Major_Depressive_Disorder}

Select *
from {} rdf:type {ns7:Major_Depressive_Disorder}; ns7:pattern_diagnosis {pattern}

Select *
from {} rdf:type {ns7:Pattern_of_Shen_Disturbance}; ns7:has_manif estation {symptoms}

Select *
from {pattern} rdf:type {ns7:Pattern_of_Shen_Disturbance}; ns7:OM_diagnosis {OM_diagnosis}

Select *
from {pattern} rdf:type {ns7:Pattern_of_Liver_Yang_Excessive}; ns7:has_prescription {}
ns7:prescripted_point {point}

Select pattern, symptoms
from {OM_Diagnosis} rdf:type {ns7:Major_Depressive_Disorder};
ns7:pattern_diagnosis {}

rdf:type {pattern}; ns7:has_manif estation {symptoms}

Select *
from {OM_Diagnosis} rdf:type {ns7:Major_Depressive_Disorder};
ns7:has_prescription {}

ns7:prescripted_point {point}

/* pattern and it om diagnosis in label */
Select *

207
from {ns7:pattern_of_shen_disturbance} ns7:OM_diagnosis {} rdfs:label {l}

/* query om diagnosis for pattern_of_shen_disturbance */
Select *
from {X} rdf:type {ns7:Pattern_of_Shen_Disturbance}; ns7:OM_diagnosis {z}

/* get symptoms of pattern */
Select *
from {} rdf:type {ns7:Pattern_of_Shen_Disturbance}; ns7:has_manifestation {symptoms}

/* pattern to prescribed points */
Select pattern, point
from {pattern} rdf:type {ns7:Pattern_of_Shen_Disturbance}; ns7:has_prescription {}
ns7:prescripted_point {prescription} rdfs:label {point}
where lang(point) like "zh*"
Appendix I  Book list for acupuncture


Appendix J  Example of data and analysis of evaluation clip

1. Why do these two have to be subclasses?
2. "Y" because these are in the traditional literature. The five transporting point and well point are most on the far side of the hands and feet. But they belong to different meridians. Meridians are divided by yin meridians and yang meridians. That will make a difference to the property as well. That also corresponds to five elements, wood, metal, etc. These points can correspond to wood but also yin meridian, then that is well point. That is the literature definition of it. Here, I used description logic to translate the sentence into the logical way to describe it. But it works very well.
3. "N" Does it have yin meridian as a class?
4. "Y" Yes. As a class. And it automatically classified it into there. I didn't expect it to work out but actually terrifically (well). That is amazing part. You can double-click the 'logic definition', right click, the edit multiple-line, one line down. You can see this is the logical expression of that. The natural language one yin meridian, yin and yang meridians are disjoint. And all the rest are definitions, for well point. They are all related to yin and yang meridian and what elements they correspond to. And after I defined (in this way), they are automatically classified. Quite amazing.
Appendix K  Layout of lab for evaluation