AALTO UNIVERSITY SCHOOL OF SCIENCE AND TECHNOLOGY
Faculty of Electronics, Communications and Automation
Department of Telecommunications

Ontologies and Computational Methods for Traditional Chinese Medicine

Master's Thesis

Joonas Jokiniemi

Espoo, 27th of October 2010

Supervisor: Prof. Jyri Hämäläinen
Instructor: M.Sc. (Tech.) Timo Korhonen

Author: Joonas Jokiniemi

Title: Ontologies and Computational Methods for Traditional Chinese Medicine

Date: 27.10.2010 Language: English Number of pages: 11 + 56

Faculty of Electronics, Communications and Automation

Department of Telecommunications

Professorship: User-Oriented Technology

Supervisor: Prof. Jyri Hämäläinen

Instructor: M.Sc. (Tech.) Timo Korhonen

Abstract:

Traditional Chinese Medicine (TCM) has been used for thousands of years in China for the purposes of health maintenance, disease prevention and treatment of health problems. Several published studies support the effectiveness of TCM treatments and the global use of TCM is constantly increasing. In China, Western and Chinese medicine are practiced in parallel.

During the past few decades, the use of information technology in medicine has increased rapidly. The development of information technology has opened up new possibilities for information storage and sharing, as well as communication and interaction between people. Along with the growing use of information technology, a wide variety of patient databases and other electronic sources of information have emerged. However, the information is fragmented and dispersed, and the terminology is ambiguous.

The objective of the thesis is to examine the position of TCM today, and to find out what changes and new opportunities the modern information technology brings for different aspects of TCM. This study describes how ontologies and semantic tools can be utilized when collecting existing knowledge and combining different databases. Also different computational methods and TCM expert systems are introduced. Finally, the most recent projects in the field of TCM are discussed and the future challenges are reflected.

The computational methods for TCM, such as diagnostic tools and expert systems, could be very useful in anticipating and preventing health problems. E-science and knowledge discovery offer new ways for knowledge sharing and cooperation. TCM expert systems can be used to generate diagnosis or automatic clinical alerts. In the future, a comprehensive and easily accessible online health service system could be developed and used to improve the health and well-being of people.

Keywords: Traditional Chinese Medicine, ontology, semantic tools, computational methods, data mining, expert systems, knowledge discovery

Tekijä: Joonas Jokiniemi

Työn nimi: Ontologiat ja laskennalliset menetelmät perinteisessä kiinalaisessa lääketieteessä

Päivämäärä: 27.10.2010 Kieli: Englanti Number of pages: 11 + 56

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Tietoliikennetekniikan osasto

Käyttäjäkeskeinen teknologia

Valvoja: Prof. Jyri Hämäläinen

Ohjaaja: DI Timo Korhonen

Tiivistelmä:

Perinteinen kiinalainen lääketiede (PKL) on tuhansia vuosia vanha hoitomuoto, jonka tarkoituksena on terveyden ylläpito, tautien ennaltaehkäisemisen ja terveydellisten ongelmien hoito. Useat vuosittain julkaistavat tutkimukset tukevat hoitojen tehokkuutta ja PKL onkin jatkuvasti kasvattamassa suosiotaan maailmanlaajuisesti. Kiinassa PKL ollut suosittu hoitomuoto jo pitkään ja nykyään sitä harjoitetaan rinnakkain länsimaisen lääketieteen kanssa.

Viime vuosikymmeninä tapahtuneen tietotekniikan kehityksen ja yleistymisen myötä myös PKL:n menetelmät ovat muuttuneet ja tietotekniikkaa on alettu hyödyntämään PKL:n tutkimuksessa. PKL:n tietoa on tallennettu digitaaliseen muotoon, minkä seurauksena on syntynyt suuri määrä erilaisia tietokantoja. Tieto on jakautunut eri tietokantoihin, joiden terminologia ei ole yhtenevää. Tämä aiheuttaa ongelmia tiedon löytämisessä ja tietoa hyödyntävien sovellusten kehittämisessä.

Tässä työssä selvitetään, mitä PKL on, ja mikä sen asema on nykyään Kiinassa ja muualla maailmalla. Työn tarkoituksena on tutkia PKL:n tietoteknisten sovelluksen kehittämistä ja siihen liittyviä haasteita. Työssä perehdytään PKL:n ontologioiden ja semanttisten työkalujen toimintaan, sekä PKL:n laskennallisiin menetelmiin ja niiden tarjoamiin mahdollisuuksiin. Lisäksi kerrotaan uusimmista kansainvälisesti merkittävistä projekteista ja pohditaan tulevaisuuden näkymiä.

Jo kehitetyt PKL:n tietotekniset sovellukset tarjoavat uusia mahdollisuuksia tiedon etsimiseen ja parantavat tutkijoiden mahdollisuutta jakaa tietoa ja tehdä yhteistyötä. Tietokoneavusteiset diagnoosityökalut ja asiantuntijajärjestelmät tarjoavat mahdollisuuksia lääkärin tekemän diagnoosin varmistamiseen. Tulevaisuudessa laskennallisia menetelmiä hyödyntäen voitaisiin tarjota terveyttä ja hyvinvontia edistäviä palveluja verkossa.

Avainsanat: Perinteinen kiinalainen lääketiede, ontologia, semanttiset työkalut, laskennalliset menetelmät, datan louhinta, asiantuntijajärjestelmät, tiedon löytäminen

Foreword

This Master's thesis was carried out at the Faculty of Electronics, Communications and Automation of the Aalto University of Science and Technology. I would like to thank my instructor Timo Korhonen for giving me inspiring ideas and helping me with the thesis.

Otaniemi, 27.10.2010

Joonas Jokiniemi

Table of Contents

Research questions

General:

1. What is Traditional Chinese Medicine (TCM)?
2. How is TCM used in China and worldwide today?

TCM and society:

3. What is the position of TCM in Chinese health care system today?
4. What kind of problems are TCM industry and research facing today?

Technology aspects:

5. How can information technology be used in the field of TCM?
6. What are the problems and challenges in developing IT applications for TCM?

Service and application development:

7. Why knowledge gathering, knowledge discovery and knowledge sharing can be challenging in the field of TCM? What kinds of solutions are available for these problems?
8. What is the role of ontology development in TCM?
9. What kind of applications and expert systems exist for TCM and what are their functions?
10. What are the most recent and future trends in TCM IT applications?

1 Introduction

Traditional Chinese Medicine (TCM) has been used for thousands of years in China for the purposes of health maintenance, disease prevention and treatment of health problems. In TCM the patient-doctor interaction has always been important. To define the correct approach for treatment the physician usually interviews the patient thoroughly. On the contrary, Western medicine generally focuses on prescribing medication to deal with the patient's symptoms as effectively as possible and the patient interview is done more quickly. TCM treatments intend to restore the balance in patient's body and thereby eliminating the cause of the disease or other health problem.

The user interface of TCM has therefore always been the patient-doctor interaction, and the appropriate treatment has been prescribed based on that interaction. Training of new TCM physicians has traditionally happened through interaction between student and teacher, as well as reading books. During the past few decades, the use of information technology in medicine has increased rapidly. Along with the growing use of information technology, a wide variety of patient databases and other electronic sources of information have emerged. These developments have changed the traditional methods of TCM. The development of information technology has opened up new possibilities for information storage and sharing, as well as communication and interaction between people.

Today, the global use of TCM is increasing rapidly and over 130 countries in the world are using Chinese Herbal Medicine (Jerry, 2007). In China, Western and Chinese medicine are practiced in parallel. There are tens of thousands of different plants and herbs that are used to produce TCM medicines. Especially in China, but also in the West, there has been growing interest to invest in studying TCM herbal medicine. However, different TCM herbs and medicine formulas have not yet been studied systematically and thoroughly enough, and the existing data is highly dispersed.

This thesis includes TCM expert Yulan Nius' opinions, which were obtained through an interview. Yulan Niu graduated as a TCM physician in 1976. She has worked for 16 years in China in different departments of a hospital. Niu moved to Finland in 1991 and since 1995 she has practiced TCM in her own clinic offering acupuncture, massage, herbal medication, diet consultation and Gua Sha. She also holds an academic degree of herbal medicine and acupuncture. Niu has also studied Western medicine.

The objective of the thesis is to examine the position of TCM today, and to find out what changes and new opportunities the modern information technology brings for different aspects of TCM. This study also explains the difficulties encountered when developing computer applications for TCM, and what kind of solutions are available to solve these problems. The biggest difficulty of processing information in the field of TCM is that the existing knowledge is fragmented and dispersed, and the terminology is ambiguous (Zhou, et. al. 2004). Therefore, TCM ontology development has been necessary in order to develop other applications that make use of ontology in their functions. This thesis describes how ontologies and semantic tools can be utilized when collecting existing knowledge and combining different databases.

The study introduces different computational methods that have been developed for many distinct purposes, such as enabling knowledge extraction from text and aiding TCM physicians in conducting diagnosis. Also a few of the most promising TCM expert systems, capable of reliable diagnosis, are discussed in detail. Some computational methods can also be used to help researchers to identify required information more efficiently and discover new relationships which are often not found otherwise (Lukman, 2007). A knowledge discovery system capable of such functions is introduced. Finally, the most recent projects in the field of TCM are discussed and the future challenges are reflected.

2 History and basic principles of Traditional Chinese Medicine

2.1 History

The basic knowledge of Traditional Chinese Medicine was unraveled in China over 3000 years ago (Lao, 1999). The oldest written classic of TCM is Huangdi Neijing, which was written 2000-3000 years ago (Xu, 2009). However, the theory of TCM has been evolving continuously over the centuries. TCM has been carried over for thousands of years for variety of clinical treatments for different diseases and health problems. The most common theory used nowadays is about 2000 years old. The majority of basic principles in TCM originated from philosophy, which also contributed to the development of Taoism and Confucianism.

2.2 Basic theory

The theory developed by ancient Chinese scientists says that all natural phenomenon can be related to Yin and Yang, which are two opposite, complementary, interdependent and exchangeable aspects of nature (Tang, 2008). According to this theory everything in this universe consists of five basic elements: wood, fire, earth, metal and water (Figure 1). The theory of Five Elements is used to interpret the relationship between the physiology and pathology of the human body and the natural environment. The visceral organs, as well as other organs and tissues, have similar properties to the five elements and they interact physiologically and pathologically as the five elements do. The doctrine of five phases describes two cycles, a generating or creation cycle, and a cycle of overcoming or destruction. The universe is constantly changing towards dynamic balance or harmony. In TCM such knowledge is applied to understand, prevent and cure disease.

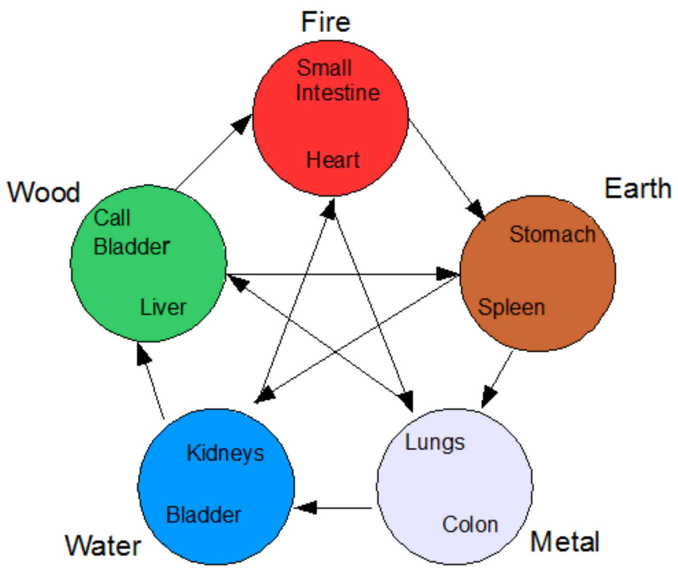

Figure 1: Five elements of TCM

In TCM, Yin refers mainly to the material aspects of the organism while Yang refers to its functions (Tang, 2008). There is a circulation of Qi energy and blood, which is controlled by the organs. The organs regulate and preserve Qi and blood through channels and collaterals. Disease occurs when there is a disturbance in Yin-Yang or flow of Qi or blood. The cause of disease can also be disharmony in the organs caused by pathogenic or climatic factors. Pathogenic factors can be emotional difficulties and unbalance or inappropriate lifestyle. Climatic factors can be heat, cold, dryness or dampness which disrupts the natural harmony of the organs. TCM treatment aims at restoring the balance by expelling or suppressing the cause of disturbance.

Modern Western medical science attempts to isolate merely physical factors as the cause of all diseases **(**Huang et al., 2007). Nearly every illness is believed to originate from germs, bugs, viruses, and other tangible factors. In TCM many of these factors are seen as symptoms instead of cause of the disease. When a certain organ is already weak and unable to resist outside invasion, it makes it more exposed to be attacked by bacteria.

2.3 Diagnosis

In Traditional Chinese Medicine diagnosis is based on overall observation of the patient's symptoms. The imbalance in patient's body is assessed by four examination methods: inspection, auscultation and olfaction, inquiring and palpation (Lukman, 2007). In the inspection approach, TCM practitioner observes abnormal changes in the patient's vitality, color, appearance, secretions and excretions. The pathological changes of internal organs are predicted by the inter-relationship between the external part of the body and the internal organs. Observations of patient's pulse, face, hair, tongue, urine and stool provide information. Auscultation refers to listening of the patient's voice, breathing, and coughing and is used to judge the pathological changes in the interior of the patient's body. Palpation of the patient's radial artery pulse and asking various questions are also important part of the diagnosis. The asking often includes various detailed questions about patient's family, living environment, diet, sleeping habits and amount of physical exercise.

Conducting a reliable TCM diagnosis requires considerable diagnostic skill. Usually a training period of many years is necessary for a TCM practitioner to understand the complexity of symptoms and dynamic balances. According to TCM expert Yulan Niu, diagnosis made by feeling the patients' pulse is among the hardest methods of diagnosis and usually takes three years of practice to master. Diagnosis conducted by looking at patients' tongue is much easier to learn.

2.4 Methods of treatment

The most common methods of treatment in Traditional Chinese Medicine include acupuncture, herbal medicine, food therapy, cupping, massage, meditation, and breathing exercises (Lao 1999). Other TCM treatment methods include Gua Sha, which is a treatment method where upper layers of skin are removed using a hand-held scraper in purpose of removing external pathogens, and moxibustion that means burning of dried Chinese herb on acupoints (Huard et al., 1977). Acupuncture is a technique in which several fine needles are inserted into specific points on the patient’s body. The aim of acupuncture is to increase circulation and balance of Qi energy. Acupuncture can be used to treat many health problems, such as dysmenorrhea, back pain and headache, and it can also be used as anesthesia (Kaptchuk, 2002). Also the aim of herbal and food therapy is to restore the balance of Qi and harmony to patient’s body.

2.5 The interaction between patient and doctor

According to TCM expert Yulan Niu, the interaction between patient and doctor has a very important role in Traditional Chinese Medicine when defining the right treatment. It is common to conduct a thorough interview with the patient to discover the underlying reason for the disease or other health problem. Conducting the interview may take several hours. This is contrary to the methods of western medicine, and the fast pace of modern life, where efficiency is the priority and the time of doctor per each patient is very limited. However, in the long run a thorough interview is very useful and it helps the doctor to receive important information about patient’s lifestyle and history of health problems. Based on this information the doctor can give comprehensive treatment instructions which may include medication and also recommendations for diet and exercise. Yulan Niu says that a TCM diagnosis can anticipate future problems much better than diagnosis in Western medicine. For example, flu can be diagnosed from patients’ tongue a few days before it actually starts and can possibly be prevented by using the right methods.

3 The current state of Traditional Chinese Medicine

3.1 In China

The current Chinese health care system is undergoing major reforms (Xu, 2009). The government is aiming to increase health insurance coverage, improve quality of care, reduce health care costs and minimize inequality of health care access between rural and urban areas. In this chapter the situation and changes of TCM in the Chinese health care system are described. We also take a look at the challenges and opportunities facing TCM in China.

Western medicine and Traditional Chinese Medicine are the two mainstream medical practices in China today (Xu, 2009). Western medicine was brought to China by missionaries especially during the Qing Dynasty in the nineteenth century. Integrative medicine, which attempts to combine the best practices of TCM and western medicine, is also practiced in China, as well as Mongolian and Tibetan medicine. In the 1950s, Chairman Mao Zedong authorized and attempt to formalize TCM and make it an academic and systemized form of medicine. This included eliminating everything that was considered to be superstitious.

In the past, TCM knowledge and skills were passed on only to family members or to a few carefully selected students through apprentice-master relationships (Xu, 2009). Starting from 1950s, TCM education has been formalized into an academic training and most of the TCM medical professionals are educated at medical or pharmacy schools. The education is similar to western medical professional's education and it usually takes 3-8 years of training to get to get a degree. In addition to TCM theories and methodology, the fundamental TCM curriculum includes a lot of western medical science such as physiology and molecular science.

The Chinese health care system is based on the three-tier system, which was developed in the 1950s (Xu, 2009). It includes hospitals, health centers and clinics. Hospitals have the best facilities and resources while health clinics and health service centers provide most of the health services, especially for treating patients with common illnesses and less severe health problems. In 2006 only 2% of Chinas health institutions consisted of hospitals while 67% of the institutions were village clinics and 23% were other clinics.

The number of TCM professionals dramatically decreased from 800 thousand to 500 thousand between 1911 and 1949 due to the growing popularity of western medicine (Jia, 2005). The number of TCM professionals dropped even further after the medical professional licensing requirements were implemented by the Chinese government in 1999 as shown in Figure 2 (Xu, 2009). Currently only 12% of the licensed doctors in China are TCM doctors and only 6% of pharmacists are licensed TCM herbalists.

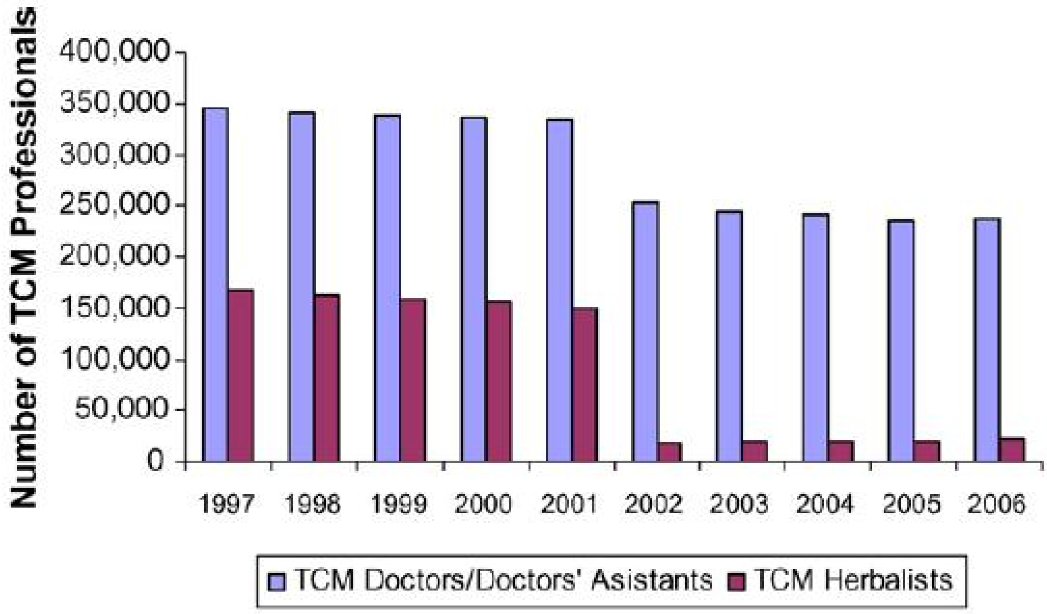

Figure 2: Number of TCM professionals in China from 1997 to 2006 (Xu, 2009)

Before the economic reform conducted in 1978, the central government was responsible for both health care financing and delivery (Xu, 2009). After 1978 several health care reforms were conducted to decentralize and privatize health care organizations. The government has dramatically reduced their financial investment in health care services, but it still maintains tight price-control in the health sector. Most of the hospitals are still government owned but many private clinics have opened in both rural and urban China.

Chinese herbal medicine is categorized into three groups: raw herbal medicine, sliced herbal medicine and patent medicine (Xu, 2009). Herbal medicine production, distribution, pricing and utilization are under the regulations of different government agencies, which have also implemented some quality control on medical production. Due to the complexity of Chinese herbal medicine, the critical component combinations are sometimes not clear and there is still need to improve the quality control in the production of the herbal medicine.

Chinese herbal medicine and herbal extracts have been exported to about 163 countries in the world (Xu, 2009). TCM herbal medicine export was almost US$1.2 billion in 2007, which was 8% higher than in 2006. Chinese herbal medicine represents 20-50% of the worldwide herbal medicine market share.

During the SARS (Severe Acute Respiratory Syndrome) epidemic in 2003 the death rate in China was 6.5% while the average international death rate was 9.5% (Chen, 2004). In Guangzhou where the use of TCM treatments for SARS was started earlier than elsewhere the death rate was only 4% and in Beijing the death rate decreased 80% after integrating TCM treatments. However, evaluating the clinical effects of TCM is complicated because it is more often used for chronic conditions than acute diseases. It might take months or years for the patient to recover from chronic illness, which makes the evaluation of long-term outcomes problematic. Efforts on conducting systematic reviews on clinical studies of TCM efficiency and safety have been put on in China.

TCM hospitals in China represent only 13.8% of all hospitals and integrated TCM and western medicine hospitals represent only 1.1% (Xu, 2009). Yet, about 90% of the general hospitals have TCM departments, although their outpatient visits represent only 8% of the total. Also almost all of the TCM hospitals are practicing western medicine as well as TCM. Most of the TCM hospitals implement western technology in diagnosis and treatment. Also surgeries for diseases, especially cancer, are common practice in TCM hospitals.

While TCM is expanding in USA its use may be contracting in China as a result of dramatic health care reforms and changing cultural values (Burke, 2003). In the late 1970s, elements of the market economy were introduced into China by Deng Xiao. As mentioned before, this led to significant health care reforms which were initiated during the 1990s. These reforms prioritized profitability, economic autonomy for health facilities, and decentralization of public health services. This led to demise of the rural cooperative medical system. These reforms also caused fundamental changes in health care financing. Free universal health care was replaced with free-for-service and private insurance strategies. As a result the medical costs and out-of-pocket expenses were increased and the new policies encouraged overpricing of drugs and high-tech services. The inequity of access was also growing and prevention programs in poor areas were reduced. This kind of reforms must have impacted the ancient healing system which consisted largely of TCM treatments.

In China there has lately been a cultural shift towards the use of western medicine (Burke, 2003). TCM is often considered outdated, nonscientific and based on superstition, especially among the younger and educated people. The fast pace of modern life is making people too busy for time consuming interviews and treatments of TCM. On the other hand, many people still use TCM because it is effective and offers more personalized and pleasant experience for patients. Many patients need the encouragement and support they can get from TCM staff.

Majority of Chinese believe in TCM, but only a relatively small portion of them prefer TCM treatment to western medical treatment when they are sick (Xu, 2009). The results of a survey conducted among 1161 out-patients, says that 25% of the respondents use TCM treatment because they have chronic conditions and another 17% use TCM treatment because western medicine failed to cure them. Another larger survey says that 54% of the population prefers western medical treatment, 25% prefers TCM/western integrated treatment, while just 12% is using TCM treatments only.

Although TCM is well integrated in the Chinese health care system it is still facing many challenges in being a part of the goal of improving public health in China (Xu, 2009). TCM philosophy is an important part of the Chinese culture, but most of its theories and methodologies are not yet confirmed by modern science. Also the economic reform in the past decades has made the concepts of modernization and westernization popular among Chinese, while the negative adverse-effects of western medical treatment are not well known by the majority of Chinese. One problem is that due to the historical evolution and cultural changes, some of the centuries old TCM knowledge is lost. The poor financial situation of many hospitals may also increase the shift for using western medical treatment instead of TCM because it has lower profit margin.

In response to all the challenges facing the growth of TCM use, the Chinese government is putting a lot of effort in protecting and expanding the role of TCM in the health care system (Xu, 2009). Many government agencies in China are increasing their investment on TCM research and efforts on integrating TCM with western medical treatment. While government is increasing its investment in clinical research and regulation, it is also important to encourage and create new means of collaboration to achieve better research results and share the existing knowledge.

Chinese government is also putting effort in standardized growing and production of medicinal plants and compounds and has developed more than 100 research units and encouraged private enterprises to build over 600 standard planting bases for herbs in great demand (Patwardhan, 2005). Selection of the correct germplasm using modern DNA fingerprinting and chemoprofiling techniques is also a priority.

3.2 Worldwide

Increased side effects and lack of curative treatment for several chronic diseases, as well as high cost of new drugs, microbial resistance, and emerging diseases are resulting in increasing public interest in complementary and alternative medicines (Patwardhan, 2005). The global scenario presents both promise and challenges for the growing market of TCM herbal products and treatments. Traditional Chinese medicine is recognized as an effective complementary and alternative medicine modality by the National Institutes of Health in United Sates (Burke, 2003). Currently TCM is widely used by consumers and its popularity is growing constantly.

For example in Australia there are an estimated 3 million TCM consultations annually (Barnes, 2003). Australia has developed a two-tiered system for regulation of complementary medicines based on risk. ‘Listed’ medicines are considered to be lower risk than ‘Registered’ medicines. Listed medicines must comply with the principles of Good Manufacturing Practice (GMP), and undergo a safety assessment which includes references to published literature. The level of evidence required to support efficacy depends on the claims made. The safety concerns for herbal medicinal products present several unique challenges compared to conventional medicines.

In 2001, US $4.2 billion was spent in the United States on botanical remedies (Patwardhan, 2005). In India the value of botanicals related trade was about US $10 billion with annual export of US $1.1 billion. China’s annual herbal drug production was worth US $48 billion with export of US $3.6 billion. Japan, Hong Kong, Korea and Singapore are the major importers of TCM products taking 66% share of China’s botanical drugs export.

On 2003 Hong Kong Chief Executive has made a plan to make Hong Kong “the international center for Chinese medicine”, and his government is funding 18 TCM research projects including clinical trials, developing quality standards, and basic pharmacological studies (Normile, 2003). On 2002 Taiwanese President proposed spending as much as $1.5 billion over the following 5 years to develop Taiwan's Chinese medicinal herb industry.

In 2007, 21.5% of the Taiwan's population used Chinese Herbal Medicine (CHM) and the usage is likely to rise, as Taiwan is rapidly transcending into an aging society (Jerry 2007). The value of the Taiwan's CHM market was estimated to be around $1.5 billion in 2005.

Traditional herbal products still suffer a credibility gap in the West because the claims made for them rest largely on anecdotes and clinical observations instead of randomized, double-blind, placebo-controlled trials (Normile, 2003). Clinical trials reported both in China and Western countries are rarely sufficient for publishing in high-quality medical journals. Many experts agree on the fact that establishing and applying stronger clinical trial methodologies in TCM is essential for determining its usefulness in a modern medical context (Barnes, 2003).

Globally, there have been centralized efforts for monitoring quality and regulating the growing business of herbal drugs and traditional medicine (Patwardhan, 2005). Active interest in providing standardized botanical medications has been taken by health authorities and governments of various nations. United States Congress has encouraged rapid growth in the nutraceutical market with passage of the Dietary Supplement Health and Education Act in 1994. The National Center for Complementary and Alternative Medicine has been established as the United States Federal Government's lead agency for scientific research in this field of medicine. Its mission is to explore complementary and alternative healing practices in the context of rigorous science by supporting sophisticated research and training researchers. One of its main goals includes disseminating information to the public on the modalities that work and explain the scientific fundamentals underlying discoveries. The center is exploring and funding all such therapies for which there is sufficient preliminary data, compelling public health need and ethical justifications. Also World Health Organization (WHO) is enthusiastic regarding traditional medicine and has been active in creating strategies, guidelines and standards of botanical medicines.

The practice of Western medicine has shifted from curative medicine to preventive medicine due to the rising number of the aging population, especially in the developed world (Jerry, 2007). Currently, over 130 countries in the world are using Chinese Herbal Medicine. Today, more than 120 countries have CHM research institutes studying the composition and effects of natural remedies. The reason for increasing popularity of CHM is due to the rise of many non-fatal chronic diseases among the elderly, such as hypertension, chronic bronchitis, and arthritis (Zhou Y, et al., 1995). It is important to provide suitable drugs for the elderly, who require three to four times more medication than the younger generation. CHM appeals to the older people because the focus of TCM treatments is in maintaining the overall balance of the human body and increasing its natural defense against diseases.

The use of Chinese Herbal Medicine has recently gained much popularity in the West (Jerry, 2007). A study carried out from 1990 to 1997 in the United States suggests a substantial increase in the number of people seeking alternative medicine. In fact, the total amount of out-of-pocket expenditure for alternative medicines is comparable to the total US physician services in 1997. In 2000, the United States Food and Drug Administration (FDA) introduced a Draft Guidance for Industry Botanical Drug Products, which recognizes the possibility of botanicals to be marketed as prescription drugs in the United States. According to some estimates at least two-thirds of the United States population will be using one or more of the alternative therapeutic approaches before 2011 (Patwardhan. 2005). Growing popularity of TCM shows as rapid increase in number of licensed Chinese medicine providers in the United States.

According to report of The World Bank trade in medicinal plants, botanical drug products and raw materials has an annual growth rate between 5 and 15% (Patwardhan, 2005). Botanical medicine represents an important share of the pharmaceutical market also within the European community.

In 2002, 1141 different TCM plant drugs were registered for their therapeutic activities. This includes several new single compounds from plants as arteannuin (antimalarial), indirubin (anticancer), etc. (Patwardhan, 2005). One of the main TCM related goals of China’s Ministry of Science and Technology is to modernize research in TCM. Controlled clinical trials have been initiated at several hospitals and research organizations to prove the safety and efficacy of the Chinese medicine. However, recent reports on adverse effects of drugs like ma huang and gingko have sounded a cautionary note that promoting traditional medicine from conception to commercialization will not be easy.

4 Ontologies in Traditional Chinese Medicine

Traditional Chinese Medicine is a complete and complex system of medicine and its effectiveness is supported by thousands of scientific studies (Zhou et al., 2004). Along with the development of information technology and growth of Internet, large number of medical databases have emerged. The large amount of ambiguous and polysemous terminologies causes problems for those who wish to search and use the information in the different databases. Semantic Web technologies, which are based on ontologies, can be utilized to address the challenges of TCM knowledge management and to integrate heterogeneous relational databases from both TCM and Western Medicine (Yu, 2008).

'Ontology' is the term used to refer to the shared understanding of some domain of interest, and it can be used as a unifying framework to solve the problems caused by ambiguous terms (Uschold, 1996). A simplified model of ontology functions and dimensions are visualized in Figure 3. An ontology is a formal representation of concepts within a certain domain, and it also defines the relationships between concepts. This way, an ontology provides a shared vocabulary for modeling a domain and its' hierarchies and other relations.

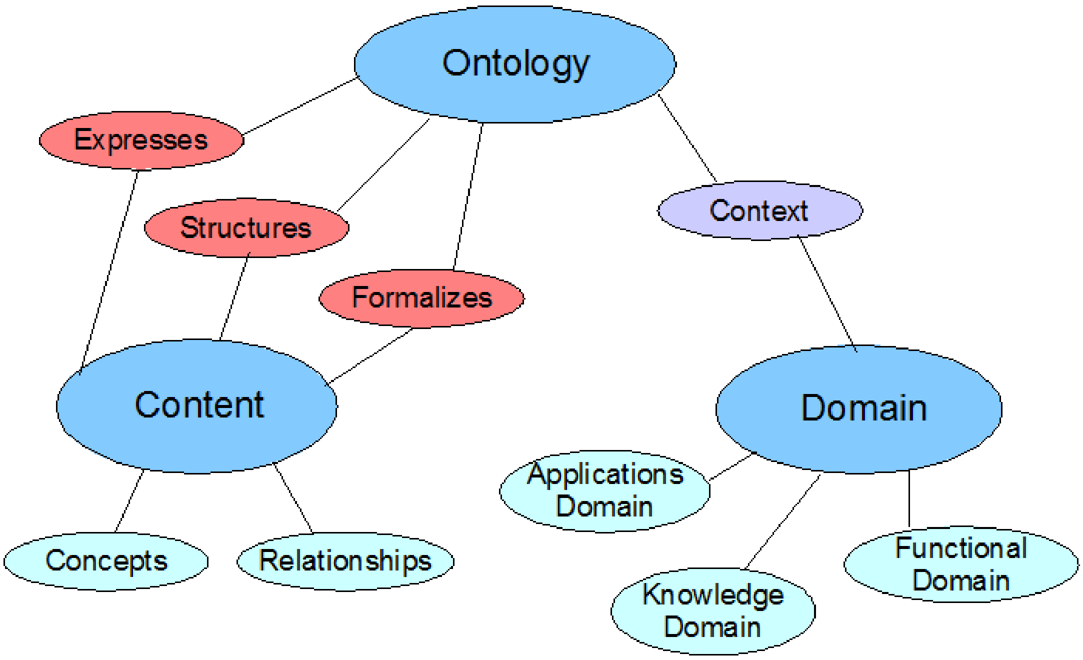

Figure 3: Ontology functions and dimensions

An ontology is a formal and explicit specification of a shared conceptualization (Yu, 2008). Conceptualization refers to an abstract model that identifies the relevant concepts of a given domain. Ontology is explicit because the type of concepts used and the constraints on their use are explicitly defined. Being formal means that the ontology should be machine understandable, which allows intelligent agents to infer new statements from existing ones based on several rules. Ontology is shared because it captures consensual knowledge of a community. An ontology can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms can be represented by an ontology. Mechanism for the dynamic fusion of ontologies from different parties is needed because usually one version of ontology cannot fit the needs of different parties with independent schedule of ontology production and consumption.

It is now commonly recognized that constructing a domain model, or ontology, is an important part in the development of a knowledge-based system **(**Huang et al., 2007). The role of ontology is to capture domain knowledge and offer it in a generally understandable form, which brings many advantages, such as better engineering of knowledge-based system with respect to acquisition, verification, and maintenance. Ontologies can also make sharing and re-use of knowledge much more efficient.

One aim of ontology development is to address different levels of data heterogeneity (Yu, 2008). Domain-level heterogeneity is caused by interdisciplinary communication. It can be resolved by using the ontology alignment technique for different, yet overlapping ontologies. Organization-level heterogeneity comes from system interoperation across organizations within one domain or community. It can be resolved through the semi-automatic mapping of different database schema to a shared domain ontology. Department-level heterogeneity is caused by inconsistency of data engineering protocols and quality-assurance policies among departments of an organization. It involves inconsistency of data formats, flaws in data quality and heterogeneity of engineering systems. Department-level heterogeneity can be resolved through data warehousing method which includes physical transformation or ontology-schema mapping.

4.1 Ontologies in biomedicine

An ontology defines types of biomedical entities as classes like drug, herb and disease (Yu, 2008). Binary relationships between biomedical entities are defined as properties such as treat, composite and interact. An ontology serves as the schema of semantic graphs and can be viewed as a directed labeled graph where nodes are classes and arcs are properties. A new biomedical entity can be introduced to the knowledge base as the instance of one or more classes.

To achieve a higher level of cognition and intelligence, standardized languages and mechanisms for synchronizing dynamic and evolving ontologies are needed (Yu, 2008). Such languages and mechanisms have already been developed for ontology platform. Ontology fusion is the mechanism of integration and alignment of distributed ontologies. Ontology reuse is the mechanism to maintain a cache of evolving sub-ontologies for application-oriented reuse.

4.2 Existing TCM ontologies

To address the problem of information retrieval from distributed and heterogeneous databases, China Ministry of Science has funded the unified traditional Chinese medical language system (UTCMLS) project (Zhou et al., 2004). The goal of this project is to study the terminology standardization, knowledge acquisition, and integration in TCM. The UTCMLS project started in 2001 and one of the main goals of the project is to design a reusable and refinable ontology to integrate and accommodate all the complex knowledge included in TCM. The estimated number of concepts in this TCM ontology is several hundreds of thousands.

In the United States the National Library of Medicine has compiled a large-scale multidisciplinary team to work on the unified medical language system (UMLS), aimed at reducing the barriers to use and develop computer applications for medicine (Lindberg et al., 1993). The development of this language has inspired the Chinese UTCMLS project, and the structure of the TCM ontology is heavily influenced by the semantic network of UMLS.

Constructing a TCM ontology is a complex and labor-intensive task that includes combining and synchronizing different versions of existing medical terminology vocabularies (Zhou et al., 2004). This can cause problems because of the existing heterogeneity and indistinctness in the terminology. The two main tasks in content development for TCM ontology are knowledge acquisition and conceptualization from different sources of information and formulization and implementation of ontology schema.

UTCMLS is a large-scale ontology that supports concept-based information retrieval and integration (Yu, 2008). Developing a TCM ontology with formally specified concepts and relationships is a challenging task due to TCM's vast amount of complex and non-standard knowledge. Ontology editors, such as Java Ontology Editor, Eclipse-based DOME, and Protégé, are applications designed to assist in the creation or manipulation of ontologies (Denny, 2004). Protégé 2000 is a frame-based knowledge base development and management system, which offers an intuitive graphical user interface (Figure 4) and constructs the ontology with RDFS as the underlying representation language. Protégé 2000 software was used in the UTCMLS project for the ontology development.

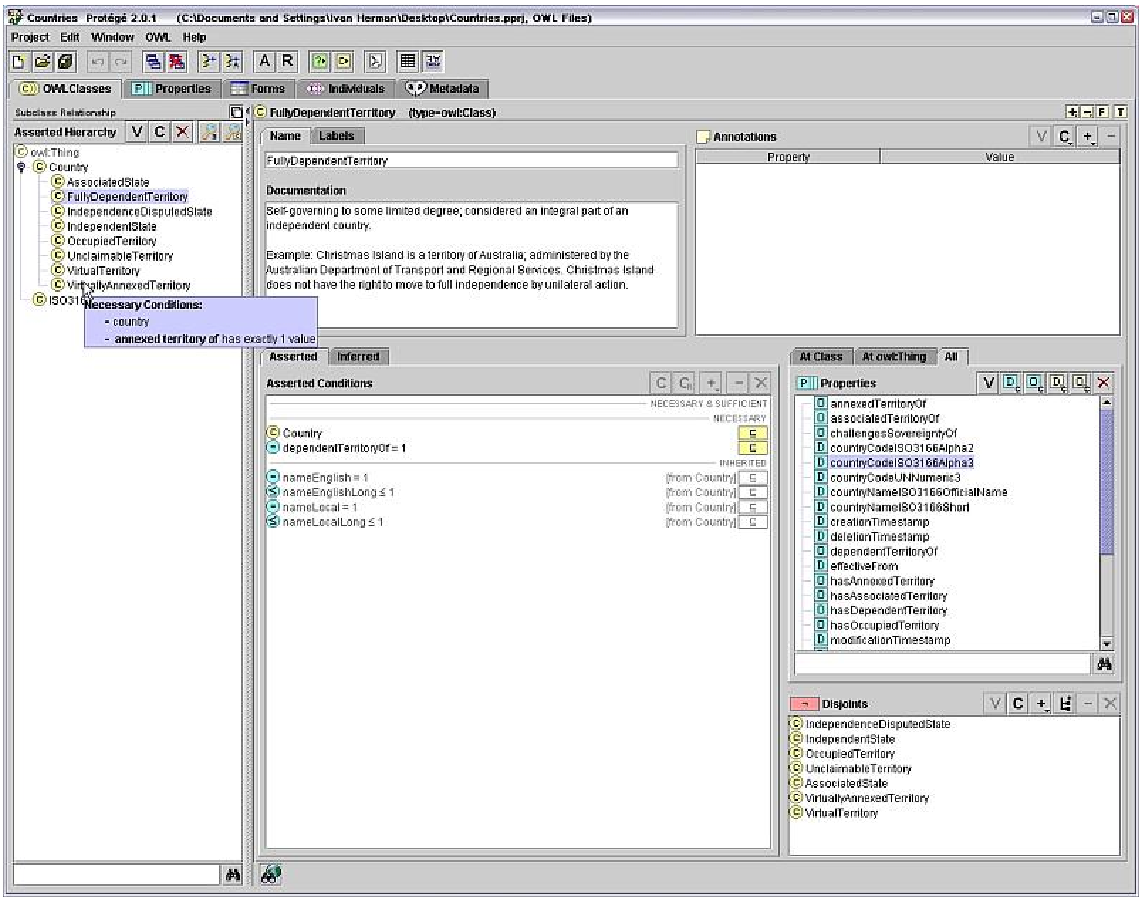

Figure 4: Protégé 2000 user interface

A project in Zheijang University is aiming at constructing a high-level TCM domain ontology including pointers to related biomedical ontologies (Yu, 2008). The focus is on designing a schema-level ontology, which includes high-level classes and properties, and optionally instance enumerations. These connection points will be discovered through domain analysis and implemented by semantic links. One aim of the project is to invent a generic methodology that can be applied to other situations beyond this project. The methodology of the project is mainly in TCM pharmaceutics such as the knowledge about herbs and herbal formulae, but should be scalable upwards to apply to any sub-domain of TCM and the whole TCM knowledge and practice system. The methodology should also be scalable downwards to single topic level.

In Yu's (2008) project the relational databases of UTCMLS are stored and the goal is to publish everything on the Semantic Web using OWL. OWL is a family of knowledge representation languages for authoring ontologies, which are most commonly serialized using RDF/XML syntax. The semantics of classes and properties in web documents can be formally described with OWL. UMLS repository is used as a source of information for core biomedical knowledge consisting of semantic types and relationships.

Another two high-level ontologies were developed when Cao et al. (2004) were designing a new method for extracting knowledge from semi-structured TCM text. Their knowledge acquisition method required them to develop Traditional Chinese Drug ontology and Traditional Chinese Formulae ontology. The development of these ontologies will be explained in detail in this chapter. The application of the ontologies will be discussed later on Chapter 5.1.2.

Next, the structure of TCM related knowledge will be discussed based on the UTCML project, since it is the largest existing TCM ontology and can be considered as the foundation for many other TCM ontologies. The whole TCM knowledge system can be divided into two components: concept system and semantic system (Zhou et al., 2004). The semantic system (Figure 5) includes the semantic type and semantic relationships of the concepts, which form the foundation of semantic level knowledge of a concept. The semantic types (Figure 6) provide categorization of all the concepts in UTCML. The semantic relationships (Figure 7) define the relations between the semantic types. Each concept must have at least one semantic type. The semantic structure allows construction of an abstract semantic network including all the concepts in UTCMLS and is vital to TCM ontology.

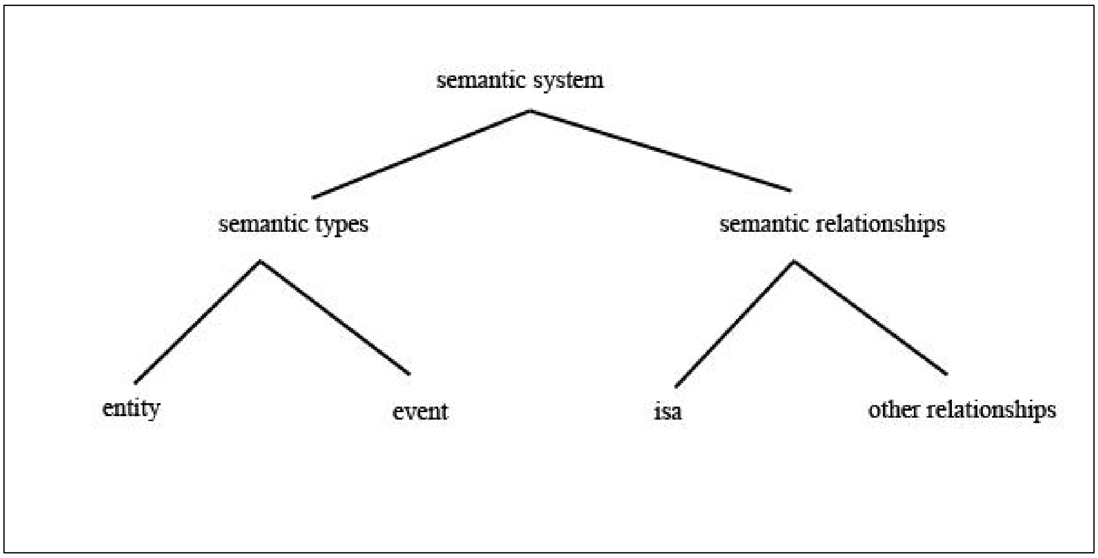

Figure 5: Semantic system (Zhou et al., 2004)

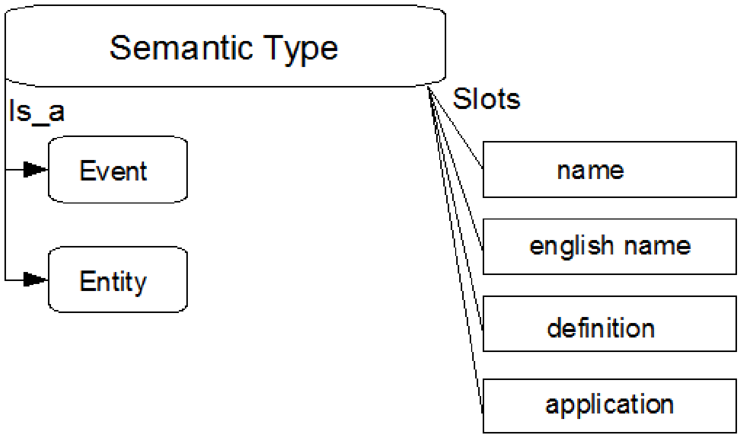

Figure 6: Semantic types (modified from Zhou et al., 2004)

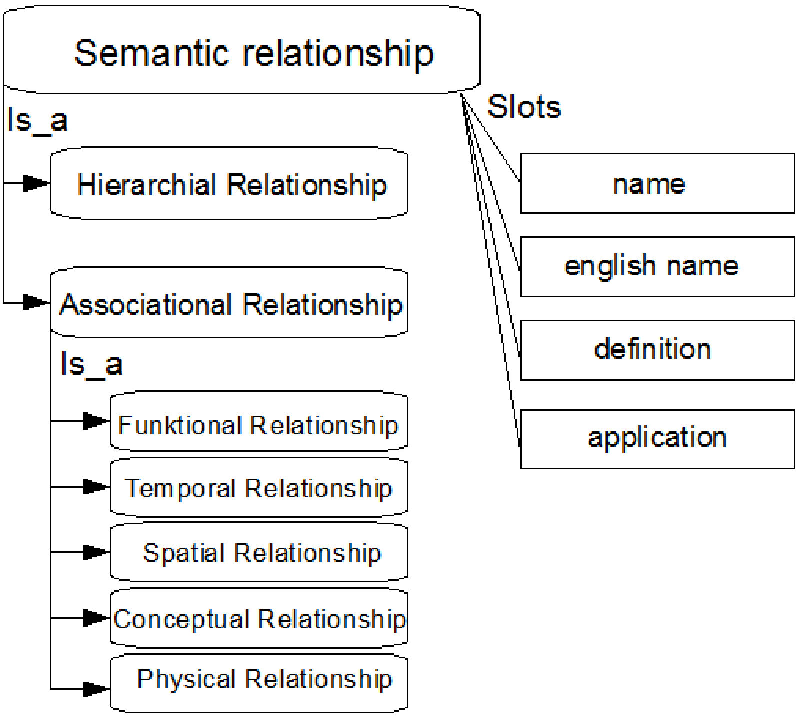

Figure 7: Semantic relationships (modified from Zhou et al., 2004)

The concept system contains 14 sub-ontologies (Figure 8) such as acupuncture, formula of herbal medicine and the basic theory of TCM (Zhou et al., 2004). These sub-ontologies and their hierarchial structure reflect the content and organization of TCM knowledge. Each of the sub-ontologies has six basic slots: concept name, concept definition, concept interpretation, relevant term, interrelated concept and semantic type. Concept name slot contains the standard name of a concept, concept definition slot and interpretation slot contain the descriptive content about the meaning of a concept. The relevant term slot defines the different terms of a concept which are used in other vocabulary sources. This helps to construct the relations between UTCMLS and the other terminology sources. The interrelated concept slot contains the concepts, which have some kind of semantic relationship with a concept. The semantic type slot has the semantic type definition of a concept. If necessary, it is also possible to define more slots for a sub-ontology.

Every TCM concept consists of three basic attributes, concept definition, concept interpretation and concept name (Zhou et al., 2004). The concept definition class involves the definitions of essential meanings of a concept. The concept name class defines the synonyms, abbreviations and lexical variants of a concept. In other words the concept name gives the relevant terminological names of a concept from different controlled vocabularies. The concept interpretation class gives the explanation of a concept. Together with the concept name slot of each sub-ontology, these three classes give the lexicon level knowledge of a concept while the semantic structure aims at the semantic level knowledge of a concept.

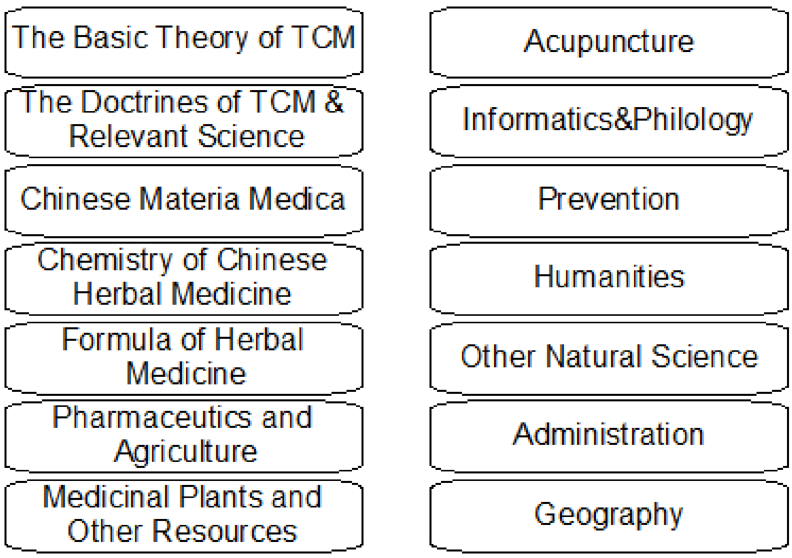

Figure 8: TCM sub-ontologies (modified from Zhou, et. al. 2004)

While Cao et al. (2004) were designing a new method for extracting knowledge from semi-structured TCM text, their knowledge acquisition method required them to develop traditional Chinese drug ontology and traditional Chinese formulae ontology. To represent a Traditional Chinese Drug (TCD) three aspects need to be considered (Figure 9): a TCD hierarchy, a vocabulary for describing the properties of a TCD, and a formal language for representing the constraints of the properties. There is no unambiguous classification available for TCDs, although most TCDs can be classified according to their curative effects and natural properties. The classification used in the TCD ontology was mainly adopted from well-known TCM related textbooks and encyclopedias.

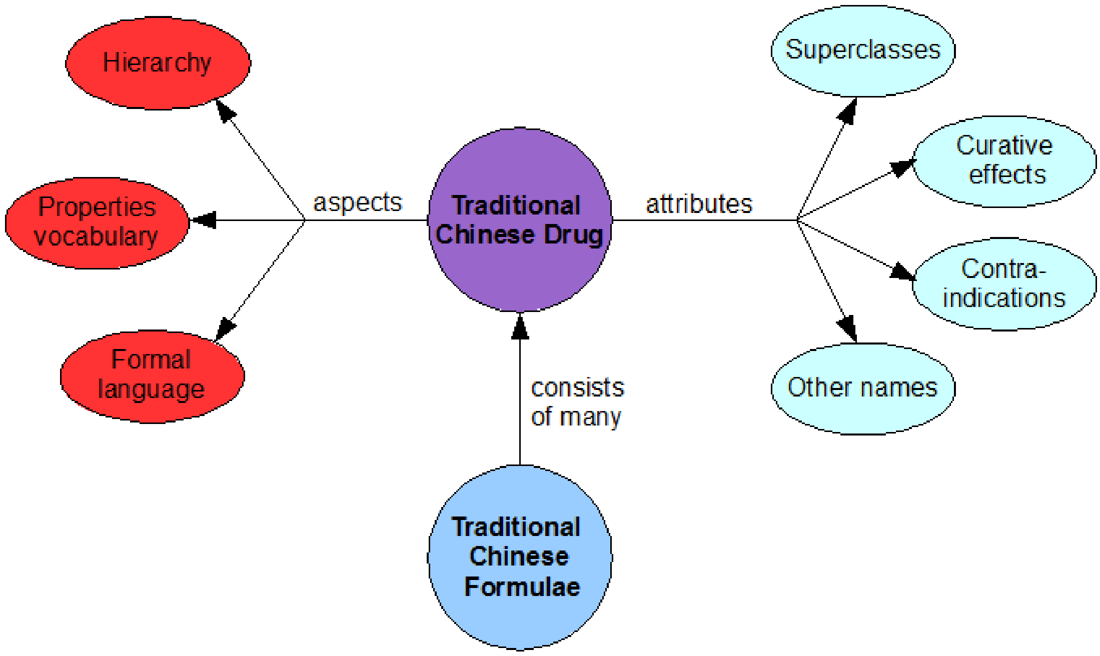

Figure 9: Traditional Chinese drug, aspects and attributes

In TCM theory, the curative effects of a drug are primarily defined by its natural properties (Yaniv et al., 2005). The Four Natures, The Five Tastes and The Channel Tropisms, that will be discussed later, play and an important role in prescription to cure diseases. In order to describe these properties Cao et al. (2004) designed a frame-oriented representation. All the TCDs are subclasses or instances of the superclass Drug, which has list of attributes and relations. Each TCD’s first attribute is superclasses, whose values indicate the superclasses to which that particular drug is an instance or subclass. The second attribute denotes the curative effects of a drug instance or a class. The other attributes include the information about drug contraindication and other names of the drug. There are various contraindicative relations between drugs and patient’s possible conditions. The most important contraindications are drug incompatibilities and contraindication during pregnancy.

Drugs are often combined with each other for better effects, but some drugs cannot be combined into a formula, because it might cause toxicity and side effects (Cao et al., 2004). In the TCD ontology, there are seven relations which can be used to represent the possible combinations of different drugs. Some drugs can be used in *synergy* to strengthen their curative effects, while other drugs can *antagonize* the effects of some drugs. Sometimes one drug is the main drug, and the other drug can be used to *assist* the effects of the main drug. Some drugs are unique in the way that they are the *only remedy* to some disease. Some drugs can *restrain* the poisonous effects of other drug and in contrary some drugs can *increase the toxicity* of some other drugs.

Cao et al. (2004) also developed an ontology for Traditional Chinese Formulae (TCF). Each TCF consists of several different drugs with indicated dose. A class called TCFormulae is used in the ontology to model the TCFs. TCFormulae can use all the slots defined in superclass Drug to describe a TCF. TCFormulae also has its own specific slots, such as ingredients and decocting-method. The important attributes in TCFormulae include principal drug, assistant-drug, adjuvant-drug and dispatcher-drug. Some of these attributes have logical associations with other slots. For example, when a drug is defined as principal-drug in a formulae, the same drug cannot be an assistant-, adjuvant- or dispatcher-drug.

5 Computational methods for Traditional Chinese Medicine

Developing computational methods for TCM is important, because they can help researchers to identify required information more efficiently and discover new relationships which are often not found otherwise (Lukman, 2007). Computational methods are also useful in bridging the gap between TCM and Western medicine. Data mining, which includes text mining and knowledge discovery, offers a solution for dealing with excessive amount of information. Data mining techniques can be used to explore active ingredients in the effective TCM formulations for individual diseases. One very important aspect of computational methods is that by creating formalized knowledge for expert systems, they can be used as a second diagnostic opinion. It is also important to highlight the differences and similarities of TCM and Western medicine, so that the biomedical mining of TCM can be focused on problems addressable by TCM but not Western medicine.

Modern Western medicine tries to identify the individual factors that are causing the illness while TCM is viewing the causes of illness as symptoms of diseases (Lukman, 2007). Diseases are regarded as the result of abnormal interactions or imbalances in the human body. If a particular organ is unhealthy, it will be vulnerable to external invasion. In TCM, the balance and interaction of all the components are regarded as more important than the effect of individual component, which is in contrast to the principle of Western medicine. In TCM other components can be added to suit the patient’s Yin and Yang conditions or to nullify the toxicity or side effects of the main components. However, there are also some similarities between TCM and Western medicine. For example, in both, TCM and Western medicine, the interactions between the clinically active molecules and their biological targets at the molecular level are the foundation of disease treatment.

5.1 TCM herbs and formulations

TCM herbs and formulations can be categorized in several different ways: The Four Natures, The Five Tastes and The Channel Tropisms (Lukman, 2007). The Four Natures are related to the degree on Yin and Yang, varying from cold, which is extreme Yin, to cool, neutral, semi-warm and hot, which is extreme Yang. The Five Tastes of TCM formulations include pungent, sweet, sour, bitter, and salty. The Channel Tropisms refer to the target organs of the formulae.

5.1.1 Databases

Thorough data on TCM formulations is needed to answer the challenge of acceptance and development of TCM due to the lack of comprehensive understanding on TCM progression and disease treatment at the molecular level. Efficient data mining should be possible when considering an ideal database (Lukman, 2007). Next, some of the most important databases, which also allow data mining, are introduced and analyzed in the context of overall usability and, if possible, integration with other databases or different approaches such as Western medicine.

Acupuncture.com.au (http://www.acupuncture.com.au/education/herbs/herbs.html) is a database which provides a list of TCM formulations grouped according to their actions. Both English and Chinese names of the herbs are available which is useful in cross-studying TCM herbs using traditional and modern experimentations. Also the tastes and contraindications of the herbs are included.

The *Dictionary of Chinese Herbs* (http:// alternativehealing.org/Chinese_herbs_dictionary.htm) includes information on the toxicity and side effects of TCM herbs as well as information about treatments and samples of TCM formulas for diseases such as diabetes, cancer and dengue fever. An important addition is a compilation on TCM herbs which are not compatible with certain drugs. The compilation provides explanations for the incompatibility of TCM herbs and drugs to the drug designers.

The *Plants For a Future* (http://wwwpfaf.org) is a database which allows the user to retrieve TCM herbs with specific medicinal usage. A query of a herb returns information on its name, potential side effects, physical characteristics, and medicinal uses accompanied by relevant scientific citations.

A *3D structural database of the biomedical components* is a database which presents active compounds extracted from TCM herbs. This database contains 10,564 herbal components records and 2073 TCM herbs entries from 296 different families. Every entry includes detailed information such as basic molecular properties, optimized 3D structures, the herbal origin with English and Latin name, and clinical effects. This kind of database can be used for designing new drugs through finding the common structural characteristics among bioactive components extracted from TCM herbs with same curative effects on a specific disease.

There are also other databases available, and some are not specifically for TCM herbs or formulations, but it is possible to retrieve information relevant on TCM herbs.

5.1.2 Data mining tools

Extracting meaningful information and knowledge from free text has been actively researched in the fields of machine learning and data mining and text data mining has become one of the most active research sub-fields in data mining (Zhou et al., 2010). The new methods for biomedical text mining can help scientists in developing novel hypotheses and new knowledge from the biomedical literature. TCM knowledge obtained from clinical practice has become a significant complementary source of information for modern biomedical sciences. The main information sources for TCM text mining applications and the most important TCM data mining tools are introduced in this section.

The main information sources that can be useful to text mining applications in the TCM domain are presented in Figure 10 (Zhou et al., 2010). One of the most important TCM information sources is a clinical data warehouse collected in TCM hospitals in the major cities of China, as it has been recognized that the electronic medical records for both inpatients and outpatients are a significant data source for TCM research. Structured databases are the basic TCM information sources, such as herb and formula databases. The data is represented in a structured relational database, and it is mainly collected from books, published literature and publicly available books. The structured databases are important for TCM knowledge discovery, for example, the TCM herb related databases have demonstrated great potential for chemical drug discovery when utilized by a TCM knowledge discovery system as explained in Chapter 5.4.

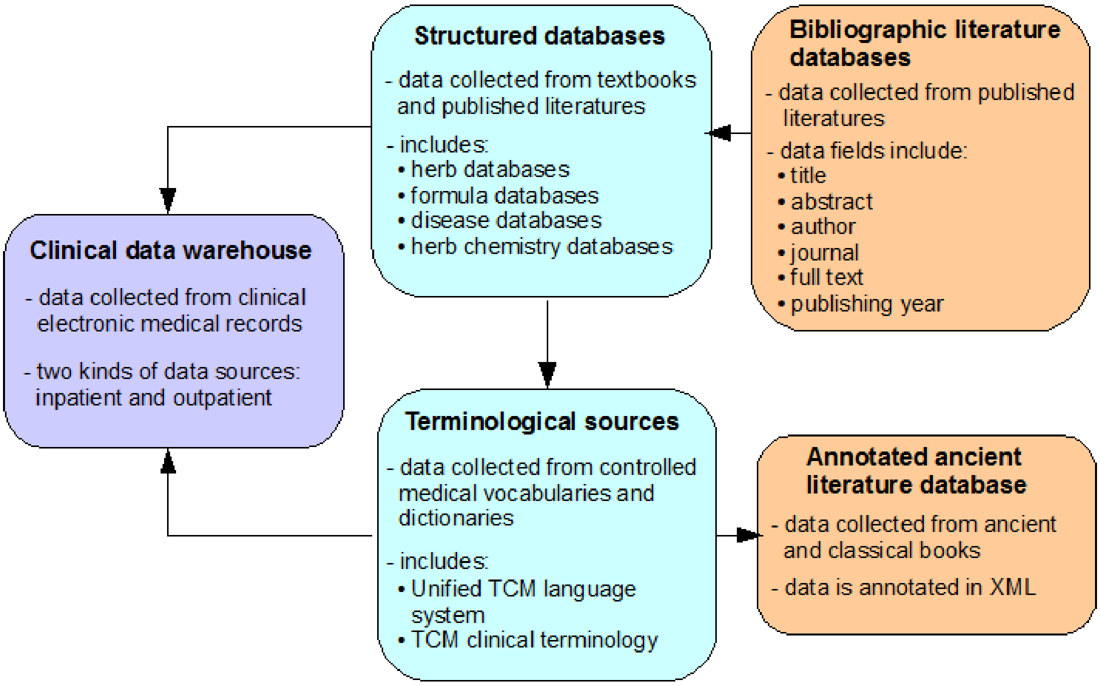

Figure 10: TCM text mining information sources (modified from Zhou et al., 2010)

Because of the various expressions, synonyms and phrases used in the clinical literature, it is challenging to perform automatic extraction of structured data from free-text (Zhou et al., 2010). One of the terminological sources is the earlier introduced (Chapter 4.2) ontology framework UTCMLS which offers a standardized terminology system with an appropriate hierarchial structure for obtaining reliable results for data extraction. There are also other terminological systems available, such as SNOMED-CT, a clinical terminology for TCM diagnosis and treatment. TCM bibliographic databases are one of the main data sources for text mining applications. They are manually collected and maintained databases, which contain citations for journal articles. The annotated ancient TCM literature database is presented in a semi-structure linked to that of ancient TCM books. The database indexing is based on the knowledge elements of TCM and the annotation is supported with tags in structured XML documents. The arrows in Figure 10 represent the relationships between two information sources, for example, the terminology sources are used for annotating the ancient literature database.

Medical Discover for Traditional Treatment inTelligence (MeDisco/3T) is a text mining system designed to extract the clinical Chinese medical formula data from literature, and discover the combination knowledge of herbal medicine by frequent item set analysis (Zhou et al., 2005). MeDisco/3T extracts structured TCM formulation information, for example TCM herbal names and Chinese herbal medicine components from literature based on the bootstrapping approach and achieving 95% name extraction precision. On the next stage the association rule mining algorithm is used to analyze the TCM formulation data, so that the frequent Chinese herbal medicine pairs and family combinations are found. More than 18,000 TCM formulations are stored by the system.

Compound innovation method integrates the possibility construction space theory (PCST) and the TCM formulation theory in order to design a novel TCM compound (Lukman, 2007). The idea of PCST is to use thinking operator to build Possibility Construction Space (PCS) based on the Realistic Space (RS), and to constitute a new space using some of the points and sets by selection and construction activity as presented in Figure 11 (Xie et al., 2009). When using this method it is very important to select useful elements from element sets P and determine the attribute set A and the rule set C correctly. Set A is a set of attributes that can best reflect the effects of a drug. Set C is a rule set that is based on TCM knowledge that can be used to judge the compounds' features. The goal of this method is to make use of artificial intelligence and inference engine to combine single drugs into a new useful drug formula. Using this method requires that the similarities between two individual drugs are measured based on the drugs targets and functions as well as on the TCM formulation characteristics, such as their natures and tastes.

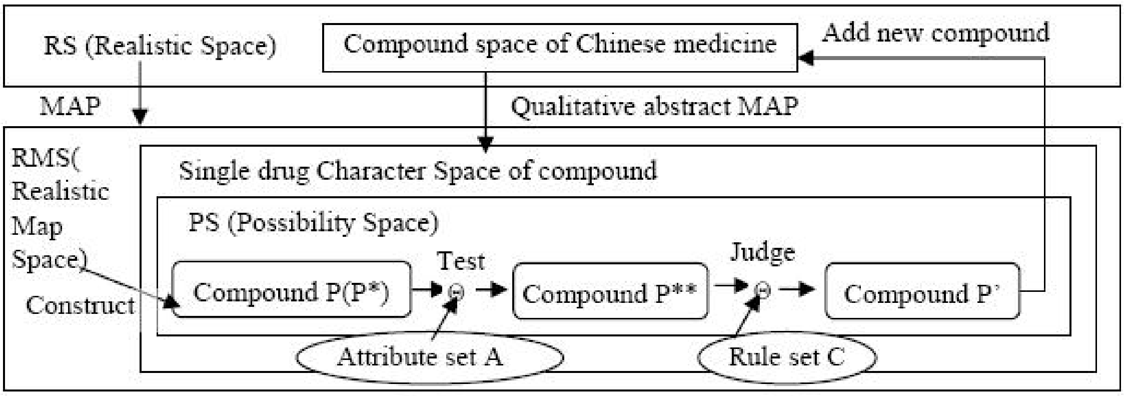

Figure 11: Construction process of innovation formulations in Chinese medicine (Xie S, et al., 2009)

An ontology-based system for extracting knowledge on TCM drugs and formulations was developed by Cao et al. (2004). The related ontologies were introduces earlier in chapter 4.2. Next, this system will be discussed in further detail.

Cao et al. (2004) developed a new method for extracting knowledge from semi-structured TCM text. Semi-structured means that although the text is written in natural language, it follows a syntax that can be easily summarized and formulated, and that the concepts of interest and their instances are explicitly given in the text, or can be identified. The method depends on the earlier introduced ontologies of traditional Chinese drugs (TCDs) and formulae (TCFs).

The language in handbooks and dictionaries containing TCD and TCF texts is quite canonical (Cao et al. 2004). This motivated the design of an executable knowledge extraction language (EKEL) which can be used to write declarative programs for extracting the knowledge from TCD and TCF texts. An EKEL program includes many agents and each has two major components. The first component defines the context in which the agent can be activated. The context is defined by boundaries, and each context has two boundaries. In the second part of each agent are the knowledge-extracting operations, and each may have a condition. These conditions represent specific situations where the operation can be activated to perform knowledge extraction while the context of each agent represents a general situation for the agent and its operations to be activated. An agent always checks if its condition is met before executing and operation in the agent’s context. If the condition is met, the agent performs the pre-defined action. Otherwise the next operation will be executed until the agent has carried out all of its operations. Errors in EKEL knowledge-extracting program are hard to detect, so instead of identifying errors in the EKEL errors are detected in its output and all the detected errors in the EKEL programs output are reported to the user after the program stops.

In the user interface of the EKEL system there are three different areas (Figure 12) (Cao et al. 2004). The left half of the interface is a window for showing the EKEL program, while the upper-right area loads and displays the textual source file from which the knowledge is extracted and the lower-right area shows the extracted knowledge frames. The knowledge extraction process consists of two stages: EKEL compilation and multi-agent knowledge extraction. During the first stage the EKEL compiler checks whether agent specifications are legal according to the EKEL syntax. If the compiler detects errors, it informs the user so the errors can be corrected. The compiler then generates and internal structure, called syntax-tree, for each agent’s contextual syntax. The syntax-tree and operations of each agent are linked, so that when there is a match in the input text, the correct operations are executed. In the second phase the knowledge extraction is done by each agent. The agents are triggered by the input text according to the agent’s boundaries. When there is a successful match, the agent performs its operations one by one in the pre-specified order.

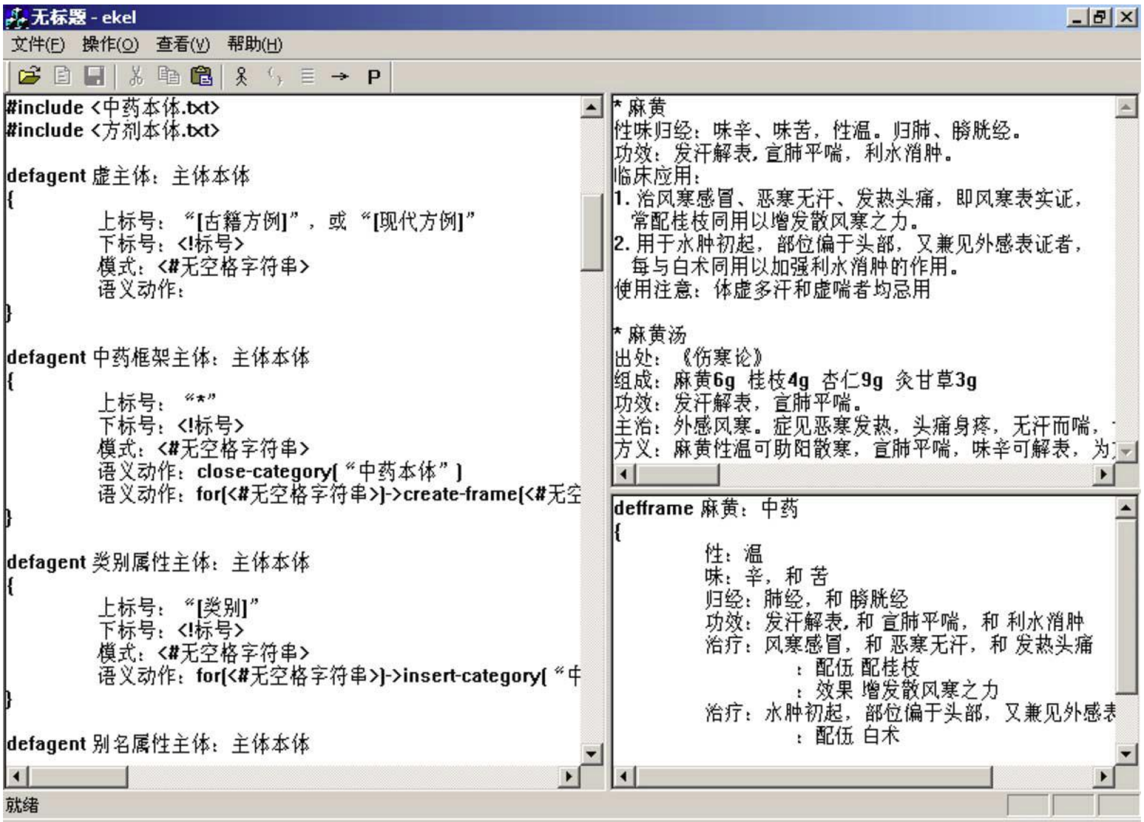

Figure 12: The user interface of the EKEL system (Cao et al. 2004)

The EKEL system has been tested with several TCD and TCF text sources, and it has been used to acquire knowledge of more than 2710 TCDs and 5900 TCFs (Cao et al. 2004). The EKEL is a flexible language which can be used to specify agents for extracting knowledge from semi-structural TCD and TCF texts. EKEL can also be used to extract knowledge from semi-structured western medicine texts and musical texts. The methods main limitation is that the input text must be semi-structured. If a free text is used as an input it must be first manually tagged to become semi-structured.

5.2 TCM diagnostics

As described earlier in Chapter 2, TCM diagnosis generally includes four different approaches: inspection, auscultation and olfaction, inquiring, and palpation. A computational approach for assisting TCM diagnosis is a very useful way of rechecking the result of diagnosis, since the results of TCM diagnosis may vary among individuals due to different experience and environmental factors (Zhang et al., 2005). Next, the most relevant computational methods for inspection, auscultation, and pulse analysis are reviewed.

5.2.1 Inspection

The patient's tongue color is one of the most important features that are traditionally observed in the diagnosis through inspection. According to TCM theory, health information of the internal organs of a human body can be obtained by observing the corresponding parts of the tongue (Lukman, 2007). Tongue image analysis can also be used to evaluate the therapeutic effects of TCM formulations (Zhu et al., 2006). Efforts in developing computational visualization techniques for tongue analysis are still at an early stage of development. However, there has been progress in the development of various tongue visualization techniques and it is motivated by clinical evidence verifying the relationship between some diseases and abnormalities in the patient's tongue and tongue coating.

There are still several issues in computerized tongue image analysis that need further research (Lukman, 2007). The methods to capture the tongue image and the following process of segmentation and color calibration need to be considered. The two quantitative features that are generally extracted from tongue images are chromatic and textural measures. Also selection and representation of tongue images are issues that need to be resolved.

In Traditional Chinese Medicine, tongue diagnosis has been based on the capacity of the eye for detailed discrimination and has therefore been dependant on the subjective analysis of the examiner (Chiu, 1999). The first experiment for the quantitative analysis of color identification of the tongue was carried out in 1985 by the Anhui Traditional Chinese Medical College and the University of Science and Technology of China. The results of this experiment showed that there is potential in the quantitative analysis of tongue diagnosis based on image processing techniques. Recently, a structural texture recognition algorithm was proposed by Chiu et al. (1995). The algorithm adopted the RGB model for mapping the tongue colors and used features such as spatial gray-tone dependency matrices (SGTDM) and Fourier power spectrum to verify or identify certain properties of coating on the tongue. The results obtained in this experiment were encouraging for developing a standardized computerized tongue examination system (CTES).

The major factors relating to traditional tongue diagnosis and the conceptual processes of diagnosis are presented in Figure 13. Tongue diagnosis is usually done by observing the substance and coating of the tongue (Horng, 1992). The CTES focuses mostly on the coating of tongue which can be distinguished by its color and the textural properties shown in Figure 14.

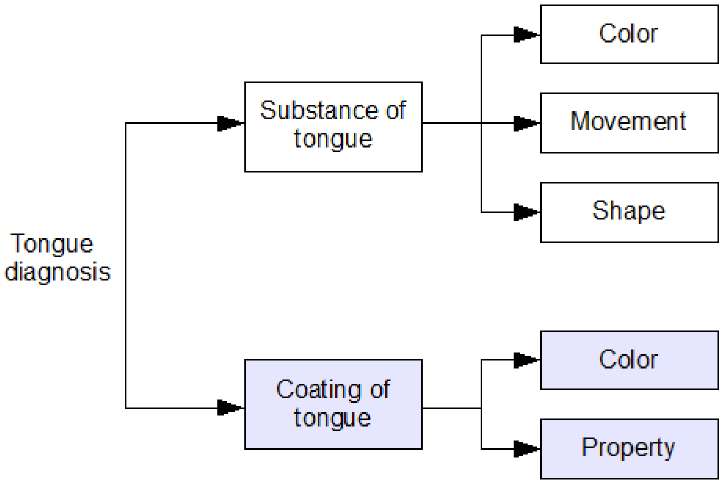

Figure 13: The major factors relating to traditional tongue diagnosis (modified from Chiu CC. 1999)

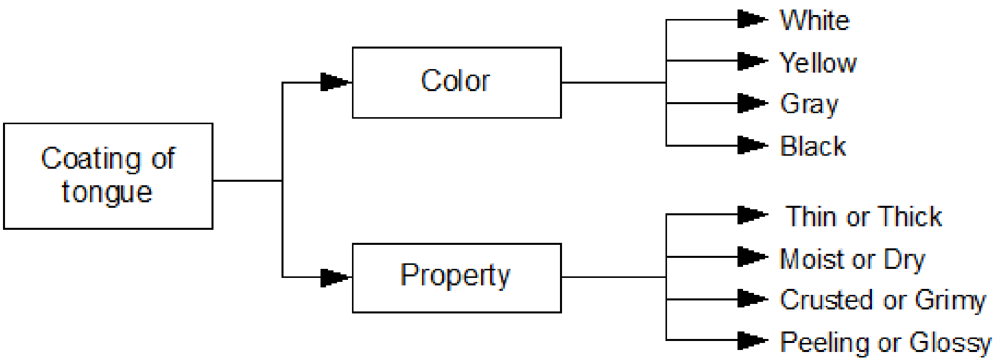

Figure 14: Coating of tongue, color and textural properties (modified from Chiu CC. 1999)

The hardware implementation of a computerized tongue diagnosis system proposed by Chiu (1999) consists of a color CCD (charge-coupled device) camera, a standardized light source with daylight quality, a head supporting mechanism and a computer. The correctness of tongue diagnosis is easily affected by environmental factors, such as light source and brightness. This standardized hardware implementation is aiming to reduce the environmental effects in order to improve the diagnostic results.

The software implementation of the CTES includes several modules: a structural recognition approach module, an image database management module, a tongue diagnosis rulebase and the reasoning module (Chiu, 1999). The structural recognition approach module diagram is presented in Figure 15. The structural recognition approach can be based on color properties or spatial textural properties of tongue. Algorithms based on color properties of tongue use the HSL (hue, saturation, luminance) color model, while algorithms based on spatial textural properties use monochrome images and utilize features of SGTDM. These approaches are developed for collecting particular information about the substance and coating of tongue, such as colors of the coating and substance and all the properties of the coating excluding moisture and dryness.

As mentioned earlier, based on TCM principles different, health information of different organs of a human body can be obtained from the corresponding parts of tongue. Hence, each tongue image is divided into four sub-regions as presented in Figure 16 (Chiu, 1999). Health condition of liver and gall bladder can be obtained from regions (a) and (c), while region (b) represents stomach and spleen, and region (d) corresponds to heart and lungs.

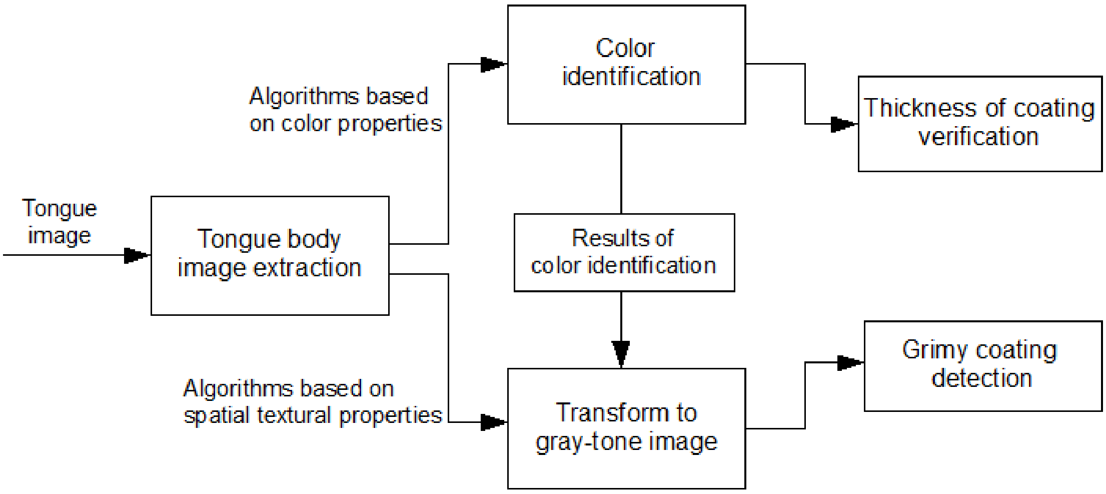

Figure 15: Coating of tongue, color and textural properties (modified from Chiu CC. 1999)

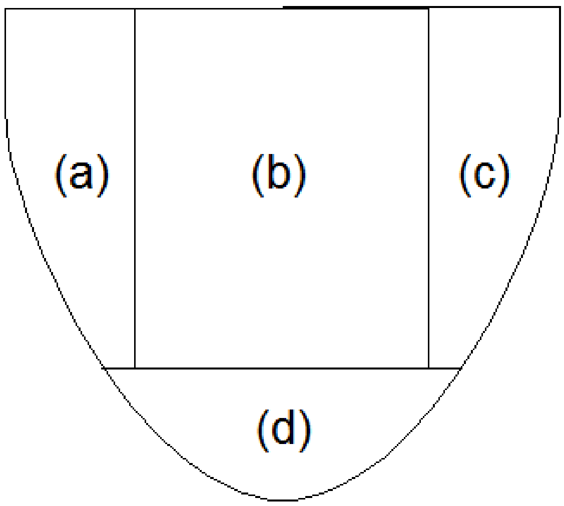

Figure 16: Tongue sub-regions (modified from Chiu CC. 1999)

The CTES performance was evaluated according to the rates of correctness for recognizing the major colors of substance and coating, the levels of thickness of coating and the existence or nonexistence of grimy coating (Chiu, 1999). The diagnostic results of CTES were compared to the inspection results obtained by a group of three physicians using naked eyes. The tongue properties of each sample are classified as being correct if the result of each property identified from CTES is consistent with the results obtained by the physicians. The rate of correctness has a slight variation depending on the indentified property and sub-region of tongue. The overall rate of correctness for identifying the colors of tongue, verifying the thickness of its coating and detecting any grimy coating is more than 86%. The limitation of this study is related to the difficulty of identifying the properties of wetness and dryness on the tongue when using computerized methods. However, these properties can be easily identified when the inspection is done by a TCM practitioner. The overall correctness for the tongue property examination by CTES exceeds 86%, and this result shows a great potential for computerized tongue diagnosis.

Watsuji et al. (1999) developed a systematic diagnostic supporting system of tongue inspection. The system uses several logic scores based on fuzzy theory and is used to diagnose syndromes suck as excess syndrome, deficiency syndrome, coldness syndrome, and heat syndrome. When the system is used to analyze excess syndrome or condition of deficiency, information about the shape of the tongue and the thickness of fur are used. However, the system mainly focuses on coldness and heat syndromes. When analyzing coldness or heat syndrome, the system uses five grades of color, five grades of color of fur, and five grades of wetness-dryness. Input from TCM experts is used to label seven grades of diagnostic outcomes for each condition. Next, fuzzy rules are constructed and used to connect different diagnostic labels and tongue diagnostic outcomes. The relevance of the system's diagnostic was verified on 20 healthy volunteers for coldness and heat syndromes and the results were convergent on 90% of the patients. The system has not been tested with sick patients so it is hard to tell how accurate the system would be in real use.

Jang et al. (2002) used a multi layered feed forward neural network and error back-propagation learning rule with RGB histograms as inputs to analyze patient's tongue colors. This method uses tongue color, tongue thickness, fur thickness, degree of wetness, and fur shape for the analysis. The neural network performance for analyzing the tongue color was fairly good, but this approach was also tested only on healthy patients.

5.2.2 Auscultation

According to the TCM theory, changes in Ying and Yang correspond to changes in the functioning of entrails, which might also cause changes in voice (Lukman, 2007). Currently the available computational methods for auscultation are very limited due to the lack of objective assessment and accurate methods for quantification. However, some studies in this field have been done. The idea behind these studies is that any malfunction of an internal organ would change the pattern of voice, which could be detected using pattern analysis.

Chiu et al. (2000) discovered that the most commonly used voice parameters, the power of the voice, the speed of speech, and pitch frequency or pitch variations cannot be used in analyzing the changes in patient's voice for purpose of auscultation in TCM diagnosis. Appropriate acoustic parameters of voice had to be proposed in order to provide a quantitative analysis of auscultation. Four different parameters were proposed: the average number of zero-crossings, the variations in local peaks and valleys, the variations in first and second formant frequencies, and the spectral energy ratio. The temporal parameters, the average number of zero-crossings and the variations in local peaks and valleys outperformed the other two parameters and were used in this study.

The results of this study are only adequate (Lukman, 2007). However, the results and data obtained from this study can be very useful in the future. Also, this study is the first one done in the difficult field of computational auscultation and much more research is needed to obtain reliable results. More groups of patients could be classified based on this study. Finding additional parameters for distinguishing patients with different conditions may lead into better results in the future.

On another study (Chiu et al., 2002) the fractal dimension (FD) parameter was proposed to compute the speech characteristics related to production irregularities. Each test sentence was pronounced three times by patients and the difference in complexity of the sentences was compared. The difficulty of using this method is that when the same sentence was read by the same patient three times, the distribution of FD values was not exactly the same each time. In this study the classification result between the deficient and normal patients exceeds 85% and the authors claimed that the accuracy of classification using FD values is more effective than the acoustic parameters used in the earlier study.

5.2.3 Pulse analysis

Subjective pulse diagnosis is very difficult to learn and it often takes years of practical experience for a TCM expert to be able to make an accurate subjective pulse diagnosis, as mentioned by TCM expert Yulan Niu. Hence, developing a computational method for making accurate pulse wave diagnosis is very beneficial. The rationale for pulse analysis is based on the fact that blood travels at different pace at different organs (Tang et al., 2005). This will allow characterization of the health condition of a particular organ. Pulse waves can be measured at six points located near the wrists of both hands of the patient. The different points correlate to different organs. For example left hand has three measurement points one corresponding to heart, one to gall and liver, and one to small intestine, bladder and right kidney.

The first step in analyzing the pulse wave data with computational approaches is to use measurement equipment to record pulse waves (Lukman, 2007). Two inputs are used when recording the pulse wave. The first one is used to measure the pulse waves with a condenser microphone with amplifiers and Bessel filters. This allows the transmission of the pulse waves to be recorded with computer via an A/D converter. The second input is used to weigh the pressure when the condenser microphone is pressed against the patient’s wrist. Using the condenser microphone is based on the same principles as the phonocardiogram which is used in the western medicine to analyze the heart condition. After recording the input waves are analyzed using graphical expression and the derivative method. The visual representation of the pulse waves can be used to teach inexperienced TCM practitioners to distinguish patients with disease from the normal ones. Usually a derivative method is used to normalize the pulse waves to allow clearer comparisons.

When comparing pulse analysis with other methods of examining intestines its advantage is that it allows correlation study between different organs (Lu et al., 1999). While blood tests and ultrasound scanning enable detection of liver problems, TCM pulse analysis is the only approach which gives an insight of how liver problems affect other organs such as spleen, stomach, lung and gall bladder. This kind of information is very important in forming a comprehensive diagnosis of the patient’s intestines. For example, in case of severe liver problems such as liver cirrhosis the related organs such as spleen and spleen meridian are also affected.

5.3 TCM expert systems

For a long time scientists and writers have been envisioning of computer systems capable of delivering accurate answers for questions over a broad domain of human knowledge (Ferucci, 2010). Open domain question answering holds enormous promise and can be used in applications in many different fields, such as business intelligence, healthcare and enterprise knowledge management. IBM has developed a computer system called Watson, which can successfully compete against top human players at the Jeopardy! quiz show. To achieve top-level performance in Jeopardy! the system must be able to answer rich open-domain questions rapidly, and to predict its own performance. The key to success is automatic learning and accurate combining of confidences across an array of complex algorithms and over different dimensions of evidence. These features are important also when developing expert systems for Traditional Chinese Medicine. The most important TCM expert systems are discussed in this chapter.

TCM expert systems can be very useful to medical practitioners and common people, and that's why the design of such systems has been actively researched for years (Lukman, 2007). Expert systems can be used to generate diagnosis or clinical alerts and accurate interpretations. According to World Health Organization there is 80% error rate for gastric ulcer treatment when TCM is used compared to 45% when Western medicine is used, which means that reliable expert systems for TCM could be useful for rechecking the diagnosis. Usually TCM expert systems use the four diagnostic methods: inspections, auscultation and olfaction, inquiring, and palpation. One important phase in the development of TCM expert system is the choice of algorithm for knowledge acquisition or data mining. The most important step in the development process is to create abstract rules from a large number of cases. Also a number of factors, such as problem definition, feasibility study, and user's demand investigation, must be taken into account when designing an expert system. The most typical techniques used for TCM expert systems include Bayesian approach, weighted summation, and ontology. Instead of using just one technique many systems use hybrid approaches for better performance.

Rapid evolution of internet technologies has opened new possibilities for enhancing traditional decision support systems and expert systems **(**Huang et al., 2007). Some web-based expert systems for the purpose of medical diagnosis exist already, but there are only few expert systems capable of Chinese medical diagnosis which can be implemented and used as a web service. Most of the existing expert systems are not web-based and are designed for stand-alone purposes, such as differential diagnosis of heart valve diseases, diagnosis of parotid tumors disease, or to combining expert knowledge and data mining.

Most of the TCM expert systems developed in the last 20 years are built incorporating totally or partially rule-based reasoning model and later extended with fuzzy logic (Wang et al., 2004). Fuzzy logic is a problem-solving control system methodology, which can be implemented in hardware, software, or a combination of both. Fuzzy logic variables may have a truth value that ranges between 0 and 1 and it can provide a simple way to arrive at a definite conclusion based upon imprecise or vague information. However, rule-based systems are not capable of implementing all possible inference by chaining rules. Also, completely expertise-dependent nature of diagnosis and vagueness of medical terms and in TCM makes the knowledge acquisition very challenging for TCM expert systems.

A Bayesian network (BN) is a probabilistic graphical model that has a powerful capacity of handling uncertainty and efficiently performing reasoning tasks (Wang et al. 2004). BNs are based on a solid mathematical theory and they constitute causal models which can be used to obtain almost all possible inferences. With these assets, BNs have become an attractive tool for representing knowledge with uncertainty. For example, a BN can be used to represent the probabilistic relationships between diseases and symptoms. Therefore, BNs can be very useful in the field of diagnosis and prediction.

Huang et al. (2007) have listed some useful advices related to the functions of an expert system aiming to help physicians in the diagnosis.

- The possible causes of disease should be listed after the disease has been identified. This will help patients to avoid getting the disease again.

- The treatment of disease, such as prescription, herbal effect, and dosage, should be provided and integrated into the reference resource after the disease has been identified.

- Pictures of typical symptoms should be provided so they can be shown for the patients.

- There should be a discussion board, so the system administrator and physicians can answer users' questions and advices.

- The physicians using the system should be able to update new clinical diagnosis experiences and knowledge into the system knowledge base.

The Electronic Brain Medical Erudite (EBME) is a large medical consultation system that uses the algebraic sum method, a kind of weighted summation method (Xiao, 1995). The purpose of the EBME system is to aid in the prompting of typical diseases and the teaching of diagnosis. EBME was originally implemented on a low-cost popular microcomputer, written in BASIC. Earlier the system used mostly Bayesian methods, which caused some limitations such as low performance in solving the contradiction between the low frequency of a disease manifestation and the high specificity of the manifestation. The use of weighted summation was found to be more suitable for this kind of system and provided improved diagnostic accuracy compared to Bayesian algorithms.

EBME can use the Enormous Knowledge Base of Disease Diagnosis Criteria which represents approximately 4000 items of diagnostic information for 1001 diagnostic entities (Xiao, 1995). The description of each item of diagnostic information utilizes the most commonly used names and terms. This will help in automating the knowledge engineering processes, avoiding repeated labor, and standardizing medical information processing. Each diagnostic item also has a unique one-to-one code which is connected to the diagnostic information base (DIB). DIB is a preliminary prepared database which has about 5000 items of diagnostic information. One useful aspect of the EBME system is that it allows a disease group diagnosis before a more accurate disease diagnosis is done. Each disease group contains a maximum of 30 diseases. When a preliminary evaluation of EBME was performed if was found out that its performance on a series of trials with a total of 815 cases appeared qualitatively similar to that of the hospital clinicians.

Wang et al. (2004) constructed a novel self-learning expert system for diagnosis in Traditional Chinese medicine by using several data mining techniques. This system uses diagnosis concepts abstracted from the key elements of routine TCM diagnosis theories. These elements can be grouped into two categories: the places where the disease occurs, such as liver or lung, and pathological states of the body or causes for diseases, such as cold and dampness. The key elements and symptoms have certain correlations, which can be used when performing a diagnosis. The system also uses a standard syndrome-name database, which has standardized names for different syndromes and also combines the syndromes with the key elements. Each syndrome always includes at least one key element indicating the place of disease occurrence and another key element about disease causes or pathological state. The diagnosis process in this system can be divided into two phases: first the diagnosis is performed according to given symptoms and the present key elements are activated, and in the second phase the standard syndrome-name database is inquired for corresponding syndromes that are consistent with the activated key elements.

A sample database was constructed using over 800 patient records that had been collected from a TCM hospital (Wang et al. 2004). In this database attributes are grouped into three categories: symptoms and signs, key elements, and syndromes. The sample database was used to formulate classifiers for diagnosis and to discover dependency relationships among attributes. The standard syndrome-name database was separated from the sample database so that the syndrome-name database could be dynamically modified whenever new syndromes emerge.

The system design is based on modularizing techniques and it has three main parts (Figure 17) (Wang et al. 2004):

1) Input of the system includes the sample database, clinical observation database, and key-elements database. All three databases have two kinds of attributes: symptoms and key elements.

2) Analysis and diagnosis part of the system includes one module for each of the following functions: variable selection, discovering dependency relationship, learning classifiers, syndrome differentiation, mining frequent sets among key elements, and modification by experts.

3) User interface, which allows the users to interact with the system.

Next, the second part, which is the kernel part of the system, will be described in more detail. The module for variable selection selects attribute subsets and prepares for the following analysis performed in the other modules (Wang et al. 2004). With help of this module the users can individually select variables they want to study in the module for discovering dependency relationship.

The module for discovering dependency relationship uses the GBPS* algorithm to learn BN from data, which represents the dependence and independence relationships of the symptoms and key elements (Wang et al. 2004). The original GBPS algorithm is one of the most effective hybrid learning algorithms and it uses the combination of conditional independence test and scoring metrics based methods (Spirtes et al., 1995). However, this system uses a new algorithm called GPBS*, which is more accurate and efficient that the original GBPS because of the modified search procedure algorithm.

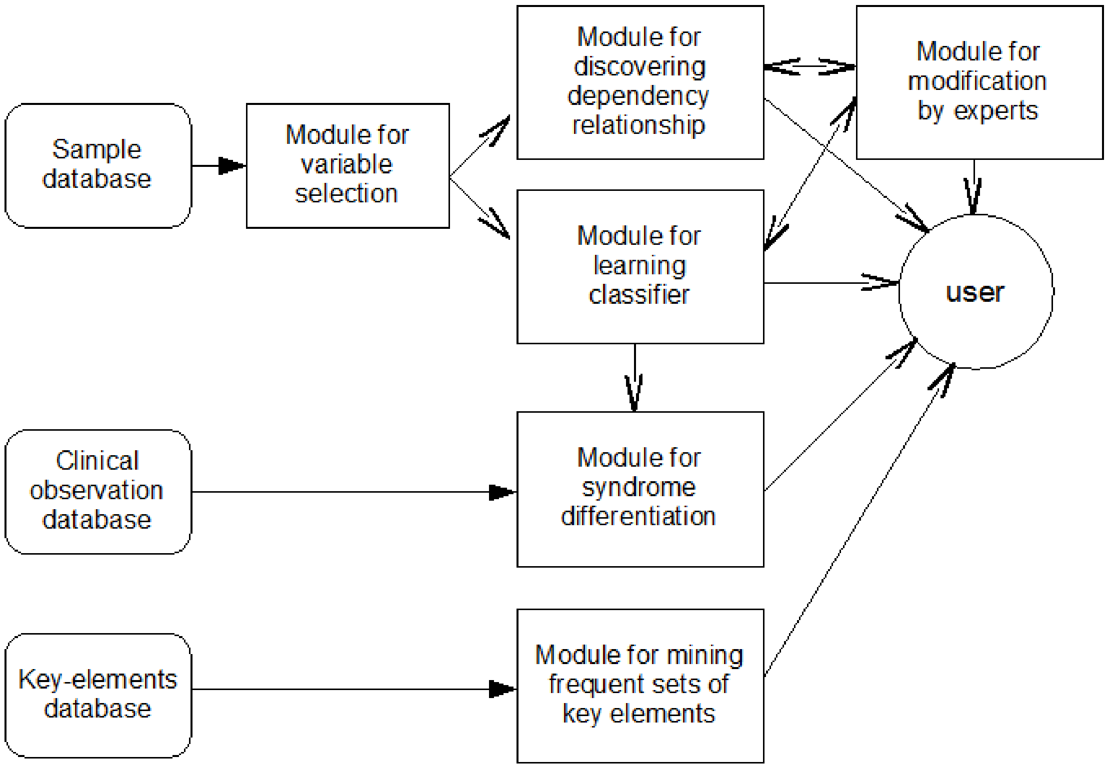

Figure 17: Architecture of the self learning expert system (modified from Wang et al. 2004)

The module for learning classifier transforms the problem of reasoning correct diagnoses on key elements into pattern recognition problem where the diagnoses are obtained by classification (Wang et al. 2004). This way the problem can be formulated strictly and unambiguously, and the solution will be clear and effectively computed.

The module for syndrome differentiation is used to differentiate the patients' syndromes based on his/her symptoms (Wang et al. 2004). The module uses the obtained classifier (given as input from the module for learning classifier) to compute the posterior probability of every key element based on Byes theorem. The module chooses the elements whose probability exceeds the threshold set by experts and outputs the names of those syndromes containing also the related key elements queried from standard syndrome-name database. Finally, the module computes occurrence probability of each syndrome and selects the syndromes with the probability excess diagnosis threshold.

The module for mining frequent sets incorporates FPC algorithm, a novel algorithm for mining constraint association rule, to periodically mine all the necessary association rules of key elements from the key element database (Dong et al., 2002). The developers of the expert system found out that computer aided determination on routine TCM syndromes can be formulated into a constrained association rule mining problem.

The module for modification by experts allows the users to browse, edit, and perform reasoning on the resulting BN (Wang et al. 2004). Experts can also modify the parameters and structure of the learned BN based on their expertise to improve the systems performance. The fact that this expert system includes over 1000 symptoms and 61 key elements makes it extremely difficult to recognize all of the interdependency relationships between symptoms and key elements. Even TCM experts cannot recognize all of the relationships. However, the system was able to learn a family of directed acyclic graphs by using the GBPS* algorithm. This enables encoding dependence and conditional independence relationships between symptoms and key elements.

The system was tested with a group of attributes selected by experts (Wang et al. 2004). The group consisted of four key elements and eleven symptoms. The GBPS* algorithm was able to discover the direct and indirect dependence relationships between these symptoms and key elements. This demonstration showed that the results obtained by using the expert system were consistent with the results of TCM experts. It also proves that Bayesian network learning is effective and reliable in the purpose of discovering dependency relationships between symptoms and key elements.

The system was also tested with 135 symptoms and 26 key elements selected by the module for variable selection (Wang et al. 2004). Five representative patient records were used to evaluate the system's performance. All other patient records were used as training instances to construct a classifier for each selected key element. Then the key elements of the five records were diagnosed and the module for syndrome differentiation was used to combine key elements and diagnosing the corresponding syndromes and their probabilities. The syndromes with probabilities greater than the chosen threshold were output as the final result. The syndromes diagnosed by the system were chiefly consistent with the expert's diagnostic results. These results indicate that the system was able to perform reliable diagnosis and could be useful in TCM diagnosis. This system promotes essential improvements in the field of automatic diagnosis, but it is still a prototype and needs some further development.

Chinese Medical Diagnostic System for digestive system (CMDS) is a unifying framework for intelligent disease diagnosis system **(**Huang et al., 2007). The system uses a web interface and expert system technology to diagnose a number of digestive system diseases, such as stomachache, vomiting, dysentery, constipation, jaundice, and tympanites.

Ontologies provide a source of precisely defined terms that can be communicated across people and applications **(**Huang, 2007). In the development process of CMDS the medical ontology was constructed using METHONTOLOGY methodology, which divides the process of ontology construction into following phases: specification, conceptualization, formalization, integration, implementation, and maintenance (Fernandez et al., 1997). In addition, the supporting activities of knowledge acquisition, documentation, and evaluation are carried out during most of the other phases. Updating the medical ontology is often necessary since the medical science is still evolving. This kind of evolving prototype life cycle allows the developers to go back to a previous phase if some definition is missed or incorrect, which makes inclusion, removal or modification of definitions possible at anytime during the ontology development.

An expert system can be a useful tool in helping physicians, family doctors or medical school students in solving the problems related to various diseases they encounter in diagnosis processes **(**Huang, 2007). In the design of an expert system for digestive system diseases a number of factors must be considered. Problem definition, feasibility study, and users' demands investigation are important parts of the design process. Also, a questionnaire was designed to investigate physicians' requirements and demands for the system. Some very useful advices related to system functions were obtained through the questionnaire and interviews.

Thousands of years of clinical medical experience has allowed the Chinese physicians to build up domain knowledge on digestive system disease and how to make an accurate diagnosis. In knowledge-based systems like CMDS, knowledge acquisition and knowledge representation are the fundamental building blocks. Efficient gathering of knowledge from experts and transforming it into a machine usable format are among the hardest and most time consuming tasks when building up this kind of system (Gierratano et al., 1998).

Huang et al. (2007) propose an efficient knowledge acquisition scheme, which allows fast conceptualization of medical domain knowledge by using a hybrid method consisting of the patients' interviews, expert interviews, knowledge elicitation by object-oriented knowledge acquisition, and CMDS system. Patients and experts can be interviewed and also questionnaire surveys can be conducted for hospital patients to identify the common disease problems in digestive system. Object oriented knowledge elicitation by an automated graphical knowledge acquisition editor provides several solutions to knowledge representation problems including transportability, knowledge reuse, and knowledge growth. Finally, the web-based CMDS system itself can be used for knowledge elicitation from experts. The interface can gather data on digestive system disease symptoms, diseases related information, treatments, and prevention methods.

The architecture of the CMDS expert system from the viewpoint of its functionalities is presented in Figure 18 **(**Huang et al., 2007). The three main components of the CMDS system are Java Expert System Shell (JESS), Database Set (DS), and Knowledge Extractor (KE). CMDS also has a user interface and a physicians' information system.

Java Expert System Shell is a rule engine and scripting environment written entirely in Sun's Java language **(**Tseng et al., 1999). JESS allows developers to use its forward chaining capability to extract knowledge which is expressed in the form of declarative rules. JESS is composed of an inference engine, XML-based knowledge descriptor, and an explanation system. The inference system makes it possible for users to inquire the system that automatically matches facts against patterns and concludes which rules are fired in CMDS **(**Huang et al., 2007). For example, physicians first input the main symptoms of a patient and then CMDS will check the description. It is possible that the main symptoms don't match digestive system diseases characteristics and the system will give the "No Results" message. Otherwise, system will move on to request more detailed information of the symptoms through four TCM diagnostic processes: inspection, auscultation and olfaction, inquiring, and palpation.

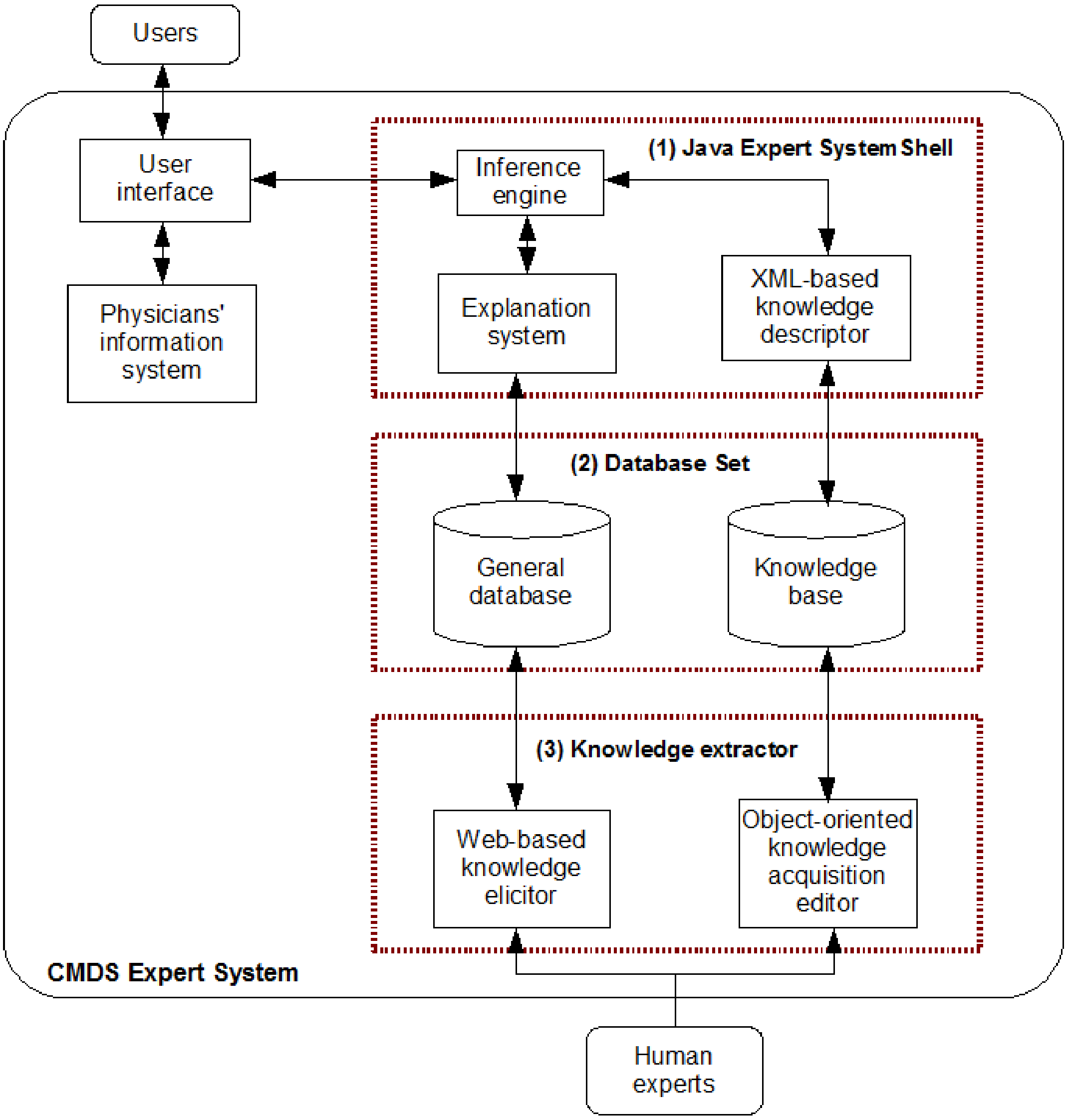

Figure 18: CMDS architechture (modified from Huang et al., 2007)

After successful matching is completed CMDS will identify the disease and generate conclusions of the diagnosis, treatment methods, prescriptions, and explain the inference processes (Huang et al., 2007). Next, the XML-based knowledge descriptor is used to transform the inference requests into uniform rule-based knowledge representation for the knowledge base. CMDS will acquire related information by mining digestive system disease databases through Database Set and Knowledge Extractor.

Database Set includes general database and a knowledge base (Huang et al., 2007). The general database is used to store all the information needed for disease diagnosis. The databases include a symptom database, a Chinese herbal medicine information database, a digestive system disease database, a disease prevention and treatment database, and an image database. The knowledge database represents the CMDS expert systems' knowledge as rules. The rules are in "IF – THEN" format and all the disease diagnosis are based on these rules. For example, if the patients' symptoms are "abdominal pain with accumulation of cold" and "weakness pulse condition", then the diagnosis would be a stomachache.

The Knowledge Extractor is used to extract the knowledge from experts and then store the knowledge into the general database, and the knowledge base of CMDS (Huang et al., 2007). The web-based knowledge elicitor provides the experts a natural language description of the diseases and the elicitated knowledge will be stored into the general database. The object-oriented knowledge acquisition editor has a graphical interface which provides a modeling platform that can be easily understood by experts and knowledge engineers.

CMDS user interface was constructed using compound web application techniques, such as HTML and JSP, to represent the web-based user interface (Huang et al., 2007). Also several graphics designing software including Adobe Photoshop and Dreamweaver were used. The system allows use of text and graphics and has different interfaces for disease interrogation enquiry process, conclusions and recommendations, and discussion board. Figure 19 and 20 are screenshots of the interfaces for disease interrogation enquiry and diagnosing results. The online physicians' information system offers widespread information on digestive system and its related issues such as domain knowledge on digestive system, frequently asked questions, Chinese herbal medicine, disease prevention measures, discussion board, and contact information on experts.

The CMDS expert system was tested thoroughly with 20 different cases where the results were obtained from TCM diagnosticians and the CMDS independently (Huang et al., 2007). The test results were promising, since the answers given by CMDS and diagnosticians were almost identical. In the future the research will focus on integrating related medical diagnostic systems with CMDS by using Java Expert System Shell and refining the ontology.

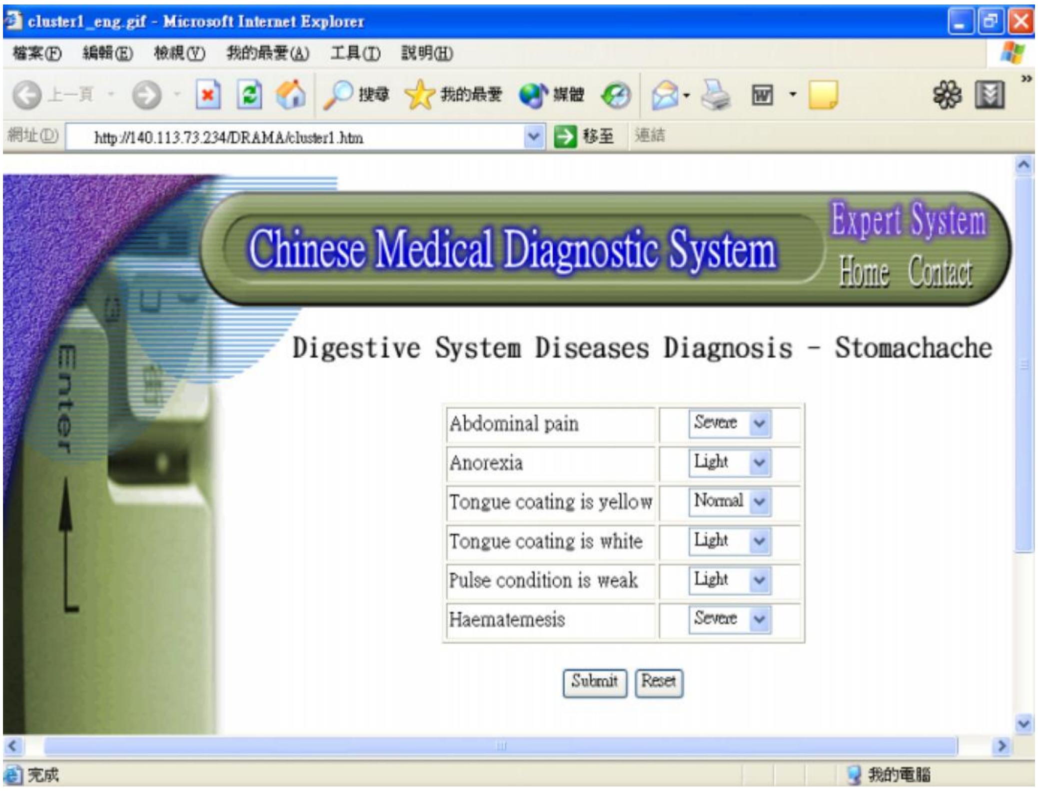

Figure 19: User interface of stomachache diagnosing (Huang et al., 2007)

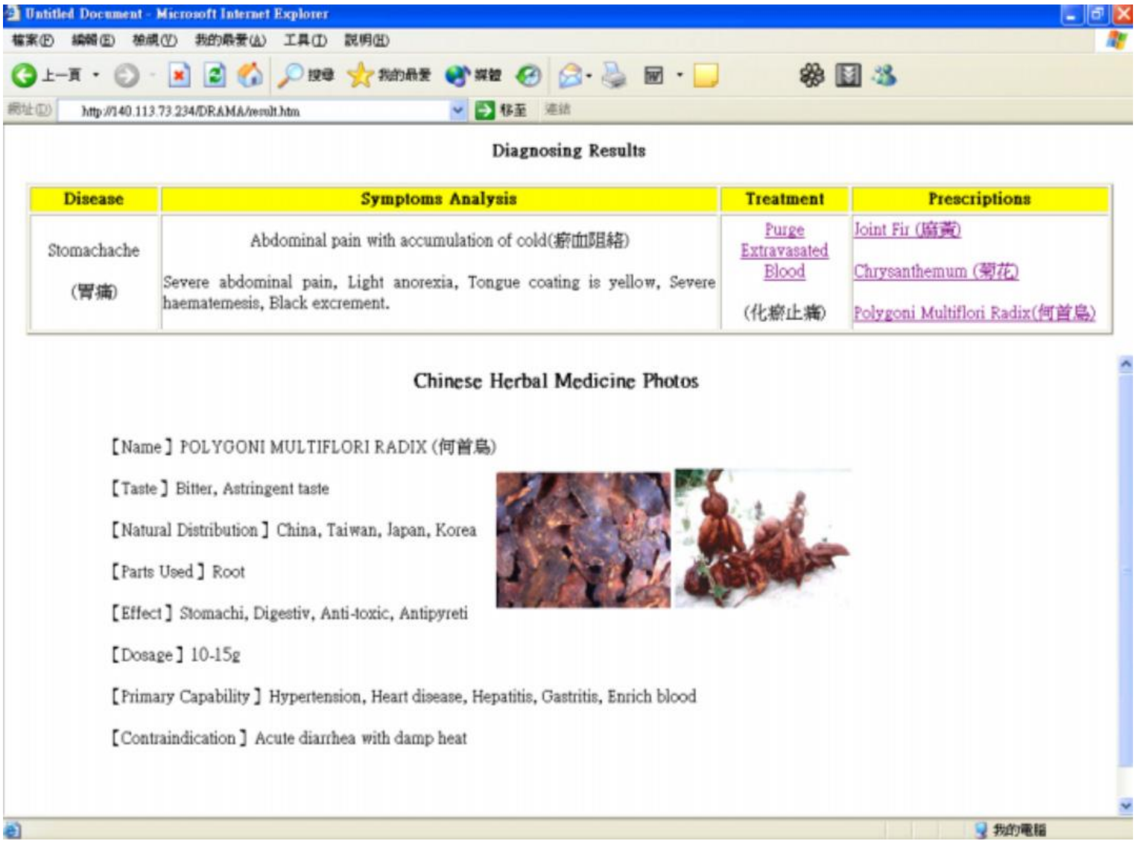

Figure 20: user interface of diagnostic results (Huang et al., 2007)

5.4 Knowledge acquisition and sharing in TCM

Web offers us a large-scale and universal information space, but its' enormous amount of data makes it increasingly difficult to discover and access useful information because the information content is presented primarily in natural languages for human browsing (Wu et al., 2003). The web can achieve its full potential only if the information can be shared and processed by machines as well as humans. This problem is also immanent in the Traditional Chinese Medicine, since the vast amount of TCM information resources is distributed among many specialized databases (Chen et al., 2007). More and more biology databases have been introduced into TCM research since many TCM scientists are now using bioinformatics methods to analyze TCM contents. It is necessary for TCM scientists to be able to reuse the information in these databases. This creates an increased need for integration of heterogeneous information resources.

Service-oriented science has the potential to increase individual and collective scientific productivity by making powerful information tools available to everyone (Chen et al., 2007). Different organizations are already offering e-learning services, information analysis services and data mining services. There are also many bioinformatics services available that allow TCM scientist to improve their research from the point of biology. This also creates an increased demand for coordination of TCM services to support collaborative scientific activities.

E-science is a term meaning the use of advanced computing technologies to support global collaboration for scientists (Chen et al., 2007). The main obstacle in achieving complete and seamless TCM e-science is caused by the heterogeneity and distribution of the information service resources. Encoding domain knowledge using ontologies provides a possible solution for overcoming the problems of semantic heterogeneity of both information and service resources. Formal semantics in ontologies offers a way to integrate scientific information resources. The
e-science system developed by Chen et al. (2007) will be introduced later in this chapter.

Knowledge discovery in databases (KDD) is a growing research field that includes many types of knowledge discovery methods, such as quantitative approaches, visualization techniques, and classification approaches (Bath, 2004). KDD techniques can be used to increase human medical knowledge by discovering underlying rules that are unknown to domain experts. Later in this chapter a TCM knowledge discovery system (Qiao et al., 2008) will be introduced.

However, there are some features in TCM knowledge that can make the development of knowledge discovery system challenging (Qiao et al., 2008):

- Symptom terminology is often ambiguous, which means that a single symptom may have distinct names, and different symptoms may also have the same name.
- Symptom records are sometimes incomplete. Due to the limitation of medical cases, TCM data resources do not always describe each symptom of a disease.
- TCM information can be related to individual characteristics, such as the state of an illness. Hence, the same disease can be treated with different recipes depending on the current state of the illness.

In 2003 Chen et al. planned a TCM-Grid in purpose of aiding the development of distributed systems that help health professionals, researchers, enterprises, as well as personal users to retrieve, integrate and share TCM information and knowledge. The project included developing a Database Grid for discovering and accessing TCM database resources that are geographically distributed to different locations. Also a Knowledge Base Grid was developed to support global TCM knowledge sharing (Wu et al., 2003).

A huge amount of domain specific TCM information is stored in existing databases (Chen et al. 2003). Through years of work on collection, translation and compilation of ancient TCM books, hundreds of TCM databases have been formed. These databases contain precious information such as clinic cases, medicine prescriptions, and diseases records and form a basic information platform that can be useful in many TCM research activities. The purpose of the Database Grid is to make each of these database resources accessible on the Grid. The Database Grid should also enable efficient and coordinated use of these content-related databases.

Looking from a service-oriented perspective, it is important to build large-scale knowledge-based services such as self-health consultation services (Chen et al. 2003). Building this kind of services usually requires the use of a variety of TCM knowledge base resources such as TCM Ontology and TCM compounding rule bases. The Knowledge Base Grid was designed to help integrating the decentralized knowledge base resources to support constructing intelligent services. The semantic web technologies can be used to overcome the heterogeneity of the web by creating a standard method for knowledge representation.

The proposed infrastructure for the TCM-Grid includes three levels (Figure 21) (Chen et al. 2003). The first level consists of various data and knowledge resources, on the second level there are high-level database and knowledge-base services, and on the third level are the research institutes, hospitals, and enterprises that are using the services. The Database Grid could be used to manage database resources allowing efficient information retrieval, while the Knowledge Base Grid would be used to facilitate knowledge sharing among the different users of the system.

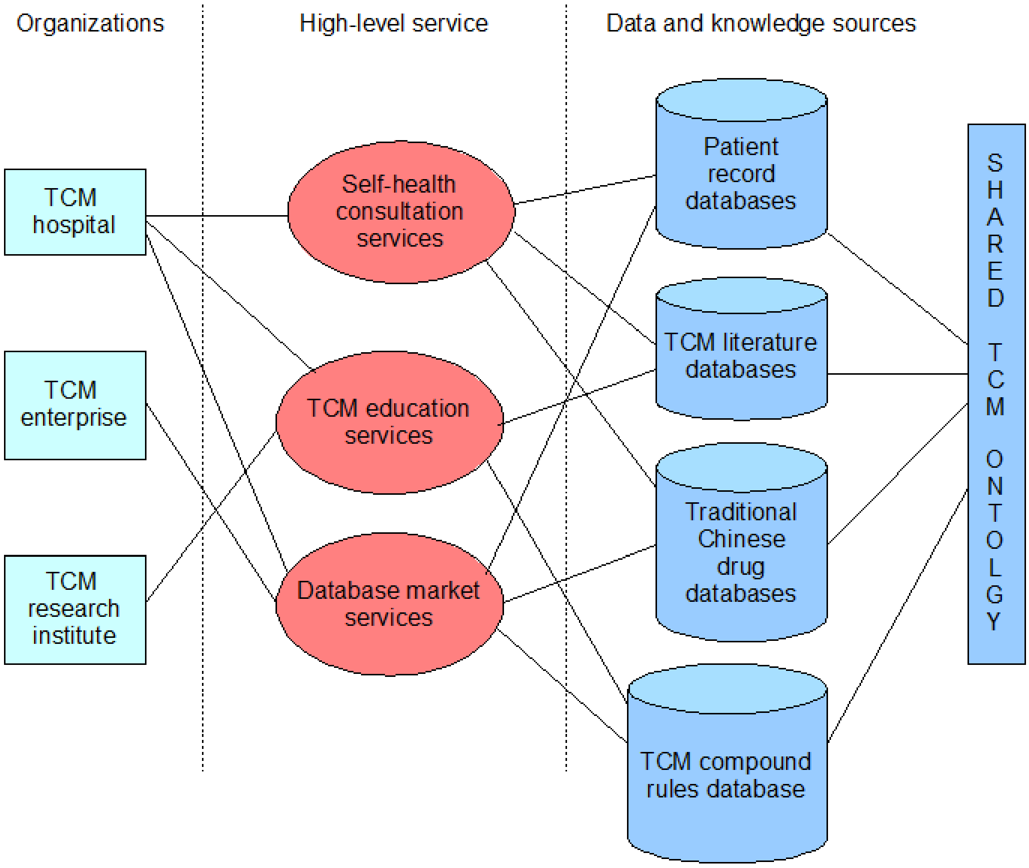

Figure 21: TCM-Grid infrastructure (modified from Chen et al. 2003)

A typical scenario for using TCM-Grid could be the design process of a new herbal medicine (Chen et al. 2003). This process could include a TCM research institute, a few hospitals for clinical testing, and some enterprises for producing and selling the new medicine. In the first phase, the research institute would initiate a new medicine design process. A new TCM-Grid based medicine design support service would help in analyzing the TCM compound rule base and TCM therapeutic principle knowledge base and based on the acquired information a new TCM medicine formula could be designed. The hospitals would then carry out the clinical testing process and then record the results using a clinical decision support service. Finally, the new medicine could be put into production by the involved enterprises.

In 2007 Chen et al. proposed an approach to address the problems of knowledge discovery and sharing. This semantic e-science system applies semantic techniques and standards such as RDF and OWL to enable database integration and service coordination. It includes modeling of domain knowledge and developing large-scale domain ontology, interconnecting distributed databases as one huge virtual database by using richer semantics, and coordinating scientific services by semantic-driven workflow. The used ontology is the UTCMLS ontology introduced in Chapter 4.2. The architecture of the system consists of client side and server side (Figure 22). The server side has a layered structure that includes resource layer, semantic layer and function layer, while the client side consists of semantic toolkits DartQuery, DartSearch, DartMapping, and DartFlow.

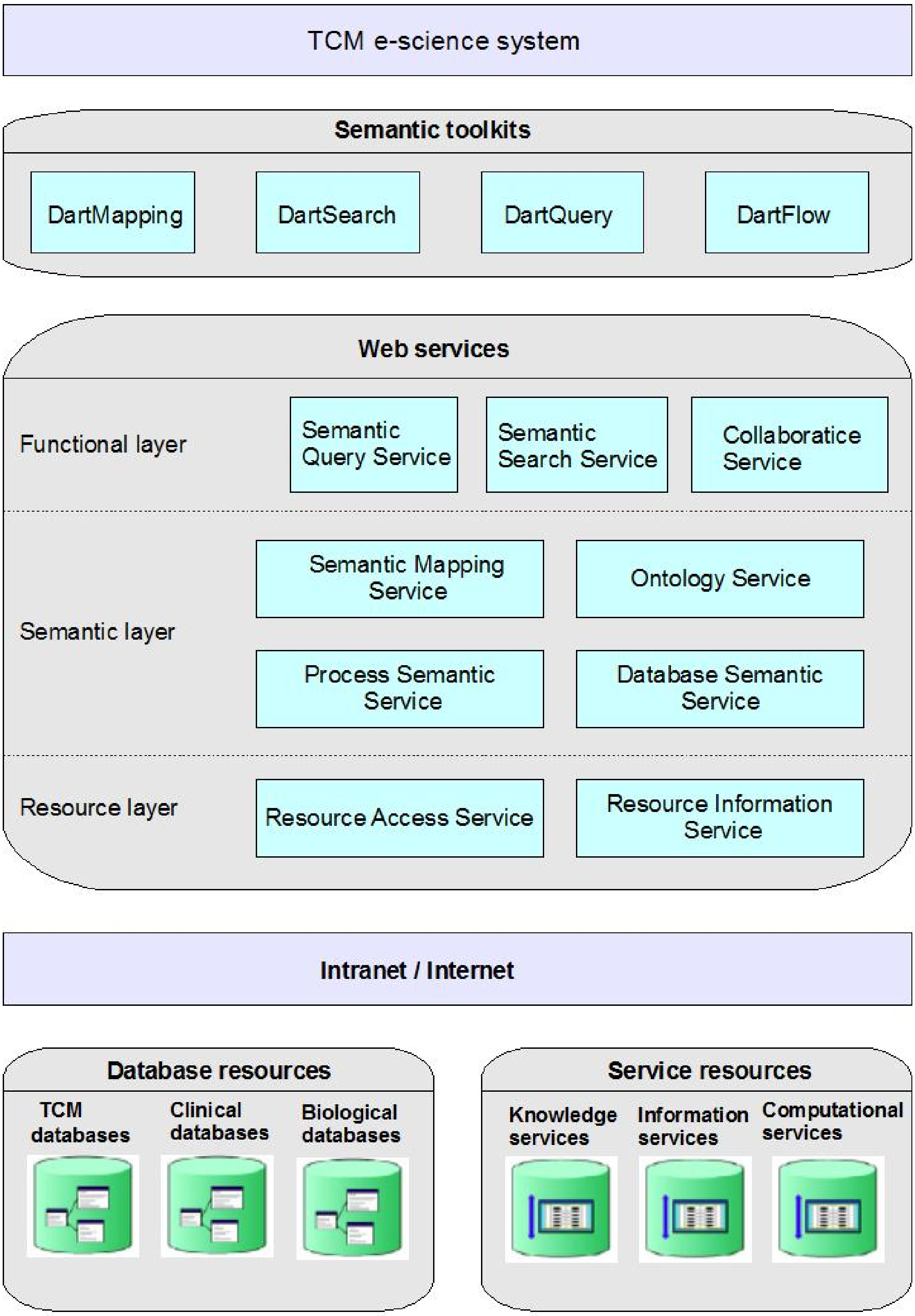

Figure 22: The abstract architecture of TCM e-science system (modified from Chen et al., 2007)

The resource layer is based on Grid services and it mainly supports the typical remote operations on the contents of resources on the Web and querying the meta-information of databases and services (Chen et al., 2007). This layer has two services: Resource Access Service sustains the typical remote operations on the contents of databases and execution of services, while Information Service supports inquiring about the meta-information of database or service resources.

The semantic layer is composed of two sub-layers and is mainly designed for semantic-based information manipulation and integration (Chen et al., 2007). The lower sub-layer has two services: Process Semantic Service is used to export services as OWL-S descriptions, and Database Semantic Service is used to export the relational schema of databases as RDF/OWL semantic description. The upper layer also contains two services: Ontology Service exposes the shared TCM ontology and provides basic operations on the ontology, and Semantic Mapping Service id used to maintain the mapping information.

The function layer supports scientific collaborative research by delivering semantically refined functions (Chen et al., 2007). Semantic Query accepts semantic queries and then inquires Semantic Mapping Service to determine which databases are capable of providing the answer. The semantic queries are converted into a set of SQL queries and finally the results are wrapped using semantics and returned as triples. Semantic Search Service indexes all ontology mediated databases and accepts semantic-based full-text-search. Collaborative Service discovers and coordinates various services in a process workflow to support research activities.

The three main types of computerized services in the field of TCM are computation services, information services, and knowledge services (Chen et al., 2007). Computational services can execute computational jobs and analyze scientific data. Information services, such as Semantic Query Service and Semantic Search Service, provide manipulation and retrieving of specific information. Knowledge services are able to apply information to solve domain-specific problems or discover facts.

DartMapping is a tool that provides two major functions: establishing semantic mapping from heterogeneous relational database to a mediated ontology semi-automatically, and converting relational databases schema to ontology statements based on the semantic mapping information (Chen et al., 2007). DartSearch is a database search engine that provides a Google-like search interface and allows full-text search over distributed databases. The search process is performed based on the semantic relations of the ontology, which enables more accurate constraints definition for getting more relevant information from the search results. For getting even more exact information, it is necessary to perform querying instead of searching in the semantic layer. DartQuery is a web-based query tool that enables dynamic queries over distributed databases. DartFlow supports service coordination for scientists by allowing them to register, query, compose, and execute services in the semantic layer. A fundamental e-science platform providing access to over 50 databases and 800 services was built based on the proposed approach.

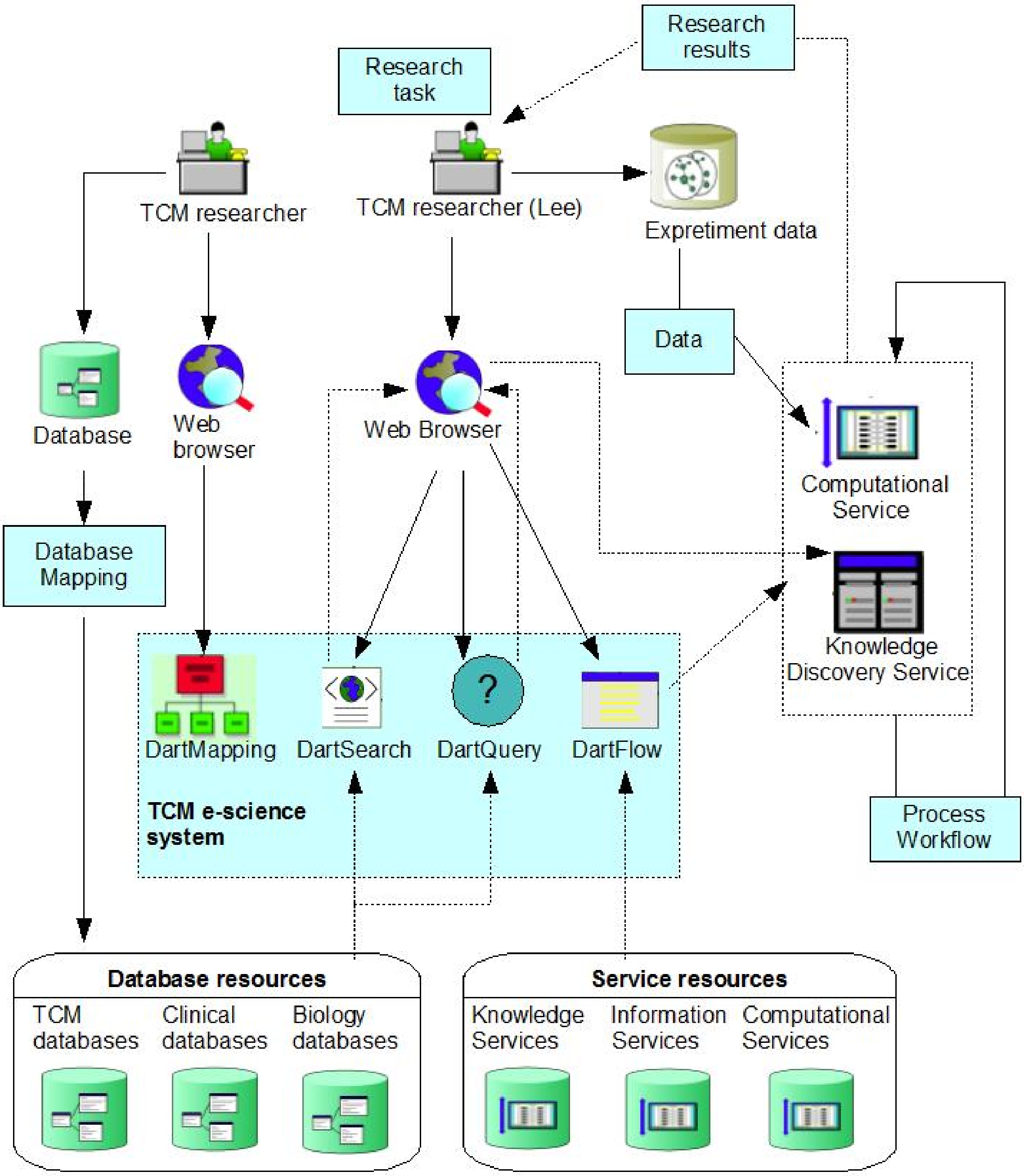

Figure 23: A scenario of using the TCM e-science system for research (modified from Chen et al., 2007)

Figure 23 represents an example scenario of using the TCM e-science system. The purpose of this example is to demonstrate what opportunities the TCM e-science system offers for a researcher, and how it can be utilized in practice. A TCM researcher named Lee is performing a research task about the possible kidney related adverse effects of a certain herbal medicine. However, Lee does not have enough knowledge about the kidneys and their biological functions so he has to search for more information before starting to conduct the experiments. Using DartSearch Lee can perform a semantic search over distributed databases. The search will return useful information about kidney functions and their relations to TCM theory and treatment methods. If Lee wants more exact results about some certain area he can perform a semantic query using DartQuery. The semantic search in DartSearch implies that the required information is located in two categories: intestines and herbal medicine formula. Based on this information Lee can perform a semantic query targeting databases that contain information under the two categories. Lee constructs suitable semantic query statements in DartQuery and the query returns a set of literature about intestines and herbal medicine formulas.

After studying the relevant literature, Lee notices that he still needs more information on the relation between kidney functions and herbal medicine formulas. However, he is unable to find the required information with DartSearch and DartQuery, which means the required information does not exist within the system. Lee communicates with other TCM researchers by utilizing the virtual environment provided by the e-science system. Lee finds out that an institution has a database containing the information he needs. Followed by Lee's request the institution registers the database into the system be utilizing DartMapping. Now Lee is able to obtain the required information by querying the database.

Lee selects the appropriate services according to his research requirements and designs a service flow in DartFlow. First Lee uses knowledge discovery service to discover some underlying rules about the relevant information. The result of knowledge discovery indicates that a herbal medicine that closely resembles the target medicine of Lee's research has adverse effects related to kidney functions. Utilizing all information obtained so far, Lee conducts the experiments on the adverse effects of the herbal medicine with kidney functions. Next, the experiment data is submitted to the computation services. The final results of the service flow show that the herbal medicine Lee is researching indeed has some adverse effects related to kidney functions.

Qiao et al. (2008) developed the first knowledge discovery system for TCM (KISTCM). KISTCM is designed for the purpose of analyzing TCM data and finding hidden rules in TCM formulae. The system consists of four components:

1) System input, that includes a TCM sample database with attributes of medicines.

2) A preprocessing module, which is used to sample, standardize, quantize, smooth, and visualize the TCM data.

3) A knowledge discovery and data analysis component. This is the central part of the system and it includes four application parts: discovering dependency relationships, mining formula properties, efficacy dimension reduction, and mining formula-syndrome relationships.

4) The user interface, which enables easy and convenient use of the system.

The interior workflow of KISTCM is presented in Figure 24 (Qiao et al., 2008). First, the user sends requests via the request transformer and then the validation tester checks the requests for syntax errors and internal contradictions. Next, the requests are added to the task queue and the experimental data extracted from TCM database is input to the formula analysis machine. The formula analysis machine is the central component of the system. It includes the preprocessing module, the knowledge discovery module, and the data analysis modules. During the data preprocessing phase, the formula analysis machine uses domain knowledge from the TCM database and KDD algorithms to perform data analysis. Then, the result evaluator checks the validity of the discovered relationships and patterns and if they are valid and appropriate, they will be converted to the required data format in the result transformer. Finally, the results will be displayed in a user-friendly manner.

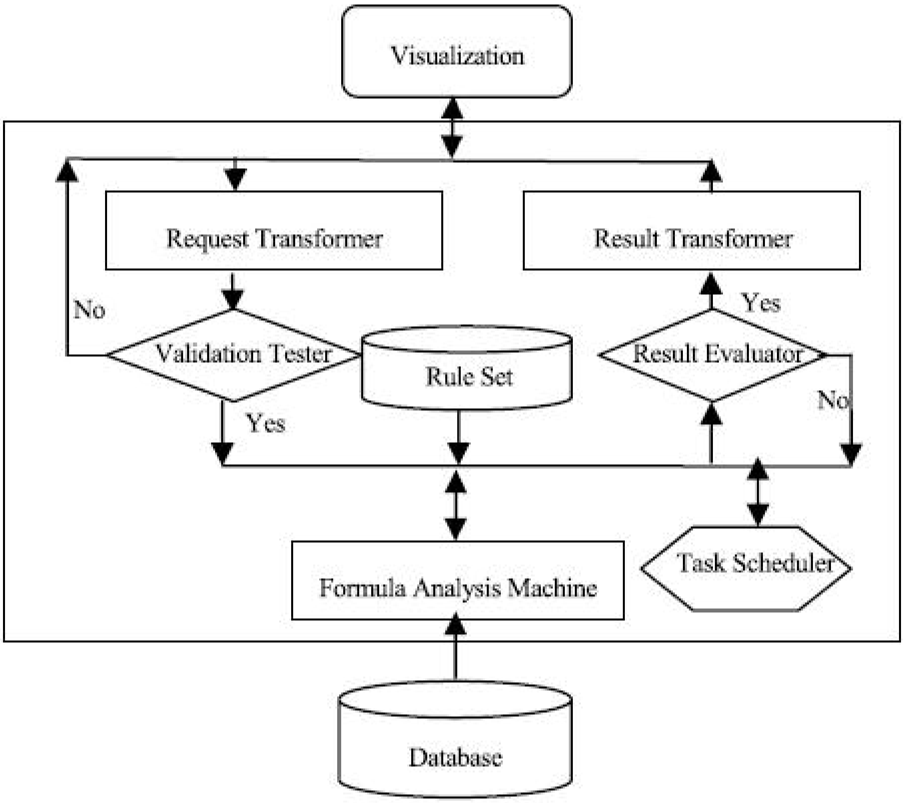

Figure 24: Workflow of KISTCM (Qiao et al., 2008)

KISTCM includes many new knowledge discovery methods that can be applied to real-world data (Qiao et al., 2008). MDRE (Medicine dependency Relationship Evaluation) algorithm is a new algorithm used to determine the dependency relationships between medicines. A TCM formula is composed of distinct medicines and choosing the most efficient medicines is a crucial phase in treating diseases. Exploring synergy principles of medicines is important and by using KISTCM new combinations of effective medicines can be found. A formula consists of various kinds of medicines, but the effects of a formula are not simply the sum of its parts since there are several redundant effects between different medicines. The effects of a formula can be determined by reducing the redundant effects. This system introduces the first efficacy dimension reduction method based on neural networks. It includes calculating the similarity between efficacies by using attribute-distance vector matrix and using the shortest attribute distance to compute the efficacy similarities.

Mining the relationships between syndromes and formulae in TCM is a difficult but very important task (Qiao et al., 2008). KISTCM uses GEP algorithm to discover these relationships. GEP is a new type of evolutionary algorithm, which combines the advantages of genetic algorithms (GA) and genetic programming (GP). First the relationships between syndromes and formulae are quantized and then GEP is used to mine formula-syndrome relationships. Most of the knowledge discovery studies in TCM analyze formulae without considering dosages. However, a slight change in dosage of a single herb might affect the curative effects of a formula. In KISTCM the impact made by each component is determined by calculating the effect degree which includes the dosage of the component.

KISTCM was implemented to test the performance of the knowledge discovery algorithms introduced earlier (Qiao et al., 2008). Spleen and Stomach formula database was used for experiments and the real-life data about TCM herbs was derived from Chinese Medical Formula Dictionary. The performance of MDRE algorithm in dependency relationship discovery was tested against another algorithm called MMDE (Multidimensional Medicine Dependence Evaluation). MDRE clearly outperforms MMDE in both prediction accuracy as well as execution times. The experimental error using MDRE was perceived to be less than 20%, which is considered to be a minor error from TCM experts' clinical experience.

Neural networks method was employed to perform efficacy reduction in order to evaluate the effectiveness and efficiency of proposed dimension reduction algorithm and to compare the methods' performance with other artificial neural network algorithms (Qiao et al., 2008). The proposed neural network method had high precision and recall but had the disadvantage of high time consumption for training the networks. The effectiveness of formula-syndrome relation discovery using the proposed GEP-based method was tested with 255 formulae to estimate the effectiveness of the proposed method. The precision accuracy was determined to be higher than 60%, which is surprisingly good and is a substantial improvement in this research area. The estimation of discovering formula properties was greater than 70%. According to the TCM experts this is an improvement, since the prediction error is generally higher than 50% when the properties are manually extracted from documents by TCM experts.

5.5 Most recent and future trends in TCM

The modernization of TCM represents a huge and complicated systematic engineering task, which involves many units including basic and applied theory, basic research, growth and process of medicinal materials, and improvement of Chinese patent medicine (Li et al., 2007). TCM has become one of the most potential medications after chemical drugs, and its modernization will provide the traditional medicines of the world with useful reference information. The modernization of TCM means the combination of TCM with modern technology, modern academic thoughts, and modern scientific culture.

China has had than a tenfold increase in its gross domestic product from 1979 to 2005 making it the second-largest economy in the world on a purchasing-power parity basis (Allen P, 2007). The Chinese government's plan to spend billions on health-care reform has multinational technology companies like International Business Machines Corp. (IBM) and Dell Inc. competing to secure a slice of a huge potential market (Chao L, 2010). Thus, China is fast emerging as a pioneer of data-centric healthcare.

Currently, IBM is deploying technology that will help hospitals standardize their patient records in a group of hospitals located in China’s southern Guangdong province (Chao L, 2010). This will enable the use of patient records to make statistical analyses of Traditional Chinese Medicine treatments. The statistics drawn from patient records will be able to help doctors in indentifying treatment plans that combine Western and Traditional Chinese Medicine. IBM hopes its solution will be one of the first to provide evidence for the effectiveness of non-Western remedies. The deployment is called Clinical and Health Records Analytics and Sharing, or CHAS, and is expected to be ready in Guangdong Hospital of Traditional Chinese Medicine by the end of the year.

The semantic capability of CHAS allows the system to understand and analyze the scientific meaning of specific terms even when other terms are used in patient records (Hunter, 2009). CHAS is able to process terms in a variety of languages including Chinese and English, ultimately processing all of the information, regardless of format, terminology or language, into one standardized document. These features enable easier sharing of information within the hospital as well as, eventually, outside the hospital to other healthcare facilities. CHAS is capable of eliminating many redundancies and inefficiencies currently found in the healthcare system such as the unnecessary repetition of tests simply because two hospitals use different testing equipment with different reference ranges.

Microsoft is also working with Chinese hospitals as the company has announced a partnership with the Affiliated Hospital of the Medical College of Qingdao University (Chao L, 2010). The project aims to deploy a hospital information system that integrates electronic medical records, patient and bed management and laboratory and other data management. Also Dell has been expanding its services in China by offering consulting services to hospitals.

6 Interview with TCM expert Yulan Niu

The information obtained from interview with TCM expert Yulan Niu has been utilized in different parts of the thesis. In this chapter all the information acquired from the interview is compiled and further analyzed.

“Conducting a reliable TCM diagnosis requires considerable diagnostic skill. Usually a training period of many years is necessary for a TCM practitioner to understand the complexity of symptoms and dynamic balances. According to TCM expert Yulan Niu, diagnosis made by feeling the patients’ pulse is among the hardest methods of diagnosis and usually takes three years of practice to master. Diagnosis conducted by looking at patients’ tongue is much easier to learn.” (Chapter 2.3)

Conducting a thorough and solid TCM diagnosis takes years to learn. Computerized diagnostic solutions offer a way to support and recheck the diagnostic results of a TCM physician. Further advance of these solutions could also enable development of a computer system capable of reliable and comprehensive TCM diagnosis independent of an actual human expert.

“According to TCM expert Yulan Niu, the interaction between patient and doctor has a very important role in Traditional Chinese Medicine when defining the right treatment. It is common to conduct a thorough interview with the patient to discover the underlying reason for the disease or other health problem. Conducting the interview may take several hours. However, in the long run a thorough interview is very useful and it helps the doctor to receive important information in patient’s lifestyle and history of health problems. Based on this information the doctor can give comprehensive treatment instructions which may include medication and also recommendations for diet and exercise.” (Chapter 2.4)

In the Western culture it is common to seek doctor’s consultation only when there is an acute need for a treatment caused by some disease or other health problem. Having more regular check-ups and consultation by a health professional could be an efficient way of improving the general well-being and health of the population in Western countries. Also a comprehensive and easily accessible online health service could be used for self diagnosis and for finding an appropriate method of self-treatment.

“Yulan Niu says that a TCM diagnosis can anticipate future problems much better than diagnosis in Western medicine. For example, flu can be diagnosed from patients’ tongue a few days before it actually starts and can possibly be prevented by using the right methods.” (Chapter 2.4)

Prevention of diseases and serious health problems could significantly reduce the costs produced to society by sick leaves and health care costs. This means that more emphasis should be put on detecting the health problems before they become serious. Computerized TCM diagnostic solutions could be very useful in anticipating and preventing health problems. There is definitely a need for openly available web-application for reliable TCM self-diagnosis.

“TCM expert Yulan Niu says that computerized methods for diagnosis offer interesting possibilities, but she has experience of a program that often gave wrong diagnostic results. She also says that there are some reliable diagnostic tools available, but most of them are only available for professionals working as TCM physicians or researchers.” (Chapter 7)

There are different kinds of computer applications available in the field of TCM. All of the applications introduced in this thesis are only available for professional use, and thus, could not be tested and analyzed in practice. Other solutions which are available for free evaluation period or purchasing for the public are not high-level solutions capable of delivering reliable results.

7 Conclusions

In this chapter each research question is reviewed separately, and the summary of the answer is presented with reference to the chapter where the more detailed answer can be found. Some of the research questions are not directly answered in the thesis, so the answer in this chapter is based on overall information in the thesis.

“What is Traditional Chinese Medicine (TCM)?”

Traditional Chinese Medicine is thousands of year’s old medical system that can be used to cure various diseases and other health problems, and help people to maintain good health. TCM treatments intend to restore the balance in patient's body and thereby eliminating the cause of the disease or other health problem. (Chapter 2)

“How is TCM used in China and worldwide today?”

In China, TCM has been the main form of medical treatment for a long time, but today both Western and Chinese medicines are practiced in parallel. With the development of industrial society, many civil diseases have emerged. The growth of TCM market around the world is fueled by the growing understanding with the fact that the natural medicines are an efficient way of maintaining one’s health. Several studies published each year support the effectiveness of TCM treatments and it is constantly becoming more popular worldwide. Especially the global TCM herbal medicine market is growing at fast pace. (Chapter 3)

“What is the position of TCM in Chinese health care system today?”

The cultural significance and popularity of TCM in China has reduced due to pressures to modernize, Westernize and compete globally (Xu, 2009). This might cause problems as the health costs are rising while China’s population is still growing, as is the economic inequality among Chinese people. TCM’s capability to offer effective, low-cost treatment for people living in the poor rural areas may become a critical resource in the future. (Chapter 3)

The loss of more traditional ways of medicine, such as TCM, through unsupportive health care policies may remove a valuable resource. The alternative and traditional medical services are often very important in completing the modern system and providing care for individuals who may not be able to use the conventional services. Currently, Chinese government is investing heavily in modernizing and revitalizing TCM.

“What kind of problems are TCM industry and research facing today?“

The biggest problems of TCM enterprises are that their size and capital are quite weak, and most of them still focus on their traditional manufacturing capability (Chen, 2005). There is also lack of R&D ability, because TCM researchers in academia focus too much on the chemical and pharmacological studies of purified ingredients only. Also, the only source of medical herbs is from China, where there is little assurance of consistent quality control.

“How can information technology be used in the field of TCM?”

Information technology is a very important aspect in modernizing TCM. Digitalizing patient data and TCM knowledge, developing expert systems and knowledge sharing systems, and developing tools for computerized TCM diagnosis, are the most important functions that TCM IT applications can offer.

“What are the problems and challenges in developing IT applications for TCM?”

Along with the development of information technology and growth of internet, large number of medical databases, containing huge amount of ambiguous and polysemous terminologies, have emerged. This causes problems for information processing and developing IT applications for TCM. (Chapter 4)

“Why knowledge gathering, knowledge discovery and knowledge sharing can be challenging in the field of TCM? What kinds of solutions are available for these problems?”

The vast amount of TCM information resources is distributed among many specialized databases and it is necessary for TCM scientists to be able to reuse the information in these databases. The most significant problems for knowledge discovery and sharing in the field of TCM are caused by ambiguous terminology and large number of dispersed databases. Ontology based e-science provides a possible solution for overcoming the problems of semantic heterogeneity of both information and service resources. Knowledge discovery in databases is a growing research field that can be used to increase human medical knowledge by discovering underlying rules that are unknown to domain experts. (Chapter 5.4)

“What is the role of ontology development in TCM?”

TCM ontology development has been necessary for development of many TCM applications. Ontologies can be used to store information about concepts and their interrelations in a machine understandable way and to address the problems caused by ambiguous terminology. China Ministry of Science has funded the unified Traditional Chinese Medical language system, which is a large-scale ontology that supports concept-based information retrieval and integration. Also, several other ontologies have been developed for distinct purposes, such as automatic knowledge extraction for semi-structured text. (Chapter 4.2)

“What kind of applications and expert systems exist for TCM and what are their functions?”

Many different computational approaches for aiding in TCM diagnosis have been developed. A computational approach for assisting TCM diagnosis is a very useful way of rechecking the result of diagnosis, because the results of TCM diagnosis may vary among individuals due to different experience and environmental factors. (5.2)

Design of TCM expert systems has been researched for years, because they can be useful to medical practitioners and common people. Expert systems can be used to generate diagnosis or automatic clinical alerts and accurate interpretations. Some expert systems show great potential since they are capable of giving reliable diagnostic results and treatment advice based on the symptoms given as input. (Chapter 5.3)

TCM expert Yulan Niu says that computerized methods for diagnosis offer interesting possibilities, but may not be always accurate for she also experienced getting wrong diagnostic results from a program. She also mentioned that there are some reliable diagnostic tools available, but most of them are only available for professionals working as TCM physicians or researchers.

Prevention of diseases and serious health problems could significantly reduce the costs produced to society by sick leaves and health care costs. This means that more emphasis should be put on detecting the health problems before they become serious. Computerized TCM diagnostic solutions could be very useful in anticipating and preventing health problems. A comprehensive and easily accessible online health service system could be used for self-diagnosis and for finding an appropriate method of self-treatment. There is definitely a need for openly available web-service for TCM self-diagnosis. (Chapter 6)

“What are the most recent and future trends in TCM IT applications?”

The modernization of TCM represents a huge and complicated systematic engineering task and it requires combining of TCM with modern technology, modern academic thoughts, and modern scientific culture. Currently, the Chinese government is spending billions in modernizing TCM, and is working with companies like IBM and Dell to develop new data-centric healthcare systems. (Chapter 6)

The future success of TCM requires strong cooperation between industry, government, academia and R&D (Chen, 2005). One important task is to provide the public with more information about TCM, including herbal nutrition and different TCM herbs. This will presumably increase attention and attraction of TCM resulting in more customers to different kinds of TCM treatments. Also, technology sharing between academic and R&D units is essential for speeding up the innovation of new products. More coordinated cooperation between industry and government is required to set up more definite and restrictive directives, laws and regulations, especially for the requirement of clinical evaluation for product quality.

References:

Allen P. 2007. Firms Ready to Dive Into China's Financial Markets. Wall Street & Technology. [online] 23 January. Available at: <http://www.wallstreetandtech.com/showArticle.jhtml?articleID=196903052> [Accessed 11 October 2010].

Alonson AR. 2001. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001:17–21. Available at: http://mmtx.nlm.nih.gov/

Bath PA. 2004. Data mining in health and medical information. Annual Review of Information Science and Technology. Vol 38:1 331-369. doi: 10.1002/aris.1440380108

Barnes J. 2003. Integration of Chinese and Western Medicine. Conference Report. doi:10.1016/S0965-2299(03)00027-X.

Blumenthal D, Hsiao W. 2005. Privatization and Its Discontents — The Evolving Chinese Health Care System. The New England Journal of Medicine. Vol. 353:11 1165-1170. Available at: http://content.nejm.org/cgi/content/full/353/11/1165

Bodenreider O. 2004. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267-70 http://www.ncbi.nlm.nih.gov/pubmed/14681409

Burke A. 2003. Traditional Medicine in China Today: Implications for Indigenous Health Systems in a Modern World. American Journal of Public Health. Vol. 93:7. 1082–1084. PMCID: PMC1447912.

Cao C, Wang H, Sui Y. 2004. Knowledge modeling and acquisition of traditional Chinese herbal drugs and formulae from text. Artificial Intelligence in Medicine, Vol. 32:1 3-13 doi:10.1016/j.artmed.2004.01.015

Chao L. 2010. IBM Works on China's Medical Records. The Wall Street Journal. [online] 31 August. Available at: <http://online.wsj.com/article/SB10001424052748703369704575462653035645256.html> [Accessed 11 October 2010].

Chen, C. 2005. Using the Internet to Integrate Government, Academic and Industry Resources for Improving the Traditional Chinese Medicine Industry. International Journal of Management. Vol. 22:2. 266-272. http://findarticles.com/p/articles/mi_qa5440/is_200506/ai_n21375249/

Chen H, Wu Z, Huang C, Xu J. 2003. TCM Grid: Weaving a Medical Grid for Traditional Chinese. ICCS 2003, LNCS 2659, 1143-1152. doi=10.1.1.111.2897

Chen, H, Mao Y, Zheng X. 2007. Towards Semantic e-Science for Traditional Chinese Medicine. BMC Bioinformatics. Vol. 8:3. doi:10.1186/1471-2105-8-S3-S6.

Chiu CC, Lin HS, Lin SL. 1995. A structural texture recognition approach for medical diagnosis through tongue. Biomed. Eng. Appl. Basis Commun. Vol. 7. 11-16

Chiu CC. 1999. A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue. Comput Methods Programs Biomed. Vol 61:2 77-89. Available at: http://www.cmpbjournal.com/article/S0169-2607%2899%2900031-0/pdf

Chiu CC, Chang HH, Yang Ch. 2000. Objective auscultation for traditional Chinese medical diagnosis using novel acoustic parameters. Comput Methods Programs Biomed. Vol 62:2 99-107

Chiu CC, Yang MT, Lin Cs. 2002. Using fractal dimension analysis on objective auscultation of traditional Chinese medical diagnosis. J. Med. Biol. Eng. Vol. 22 219-225.

Denny M. 2004. Ontology Tools Survey, Revisited. http://www.xml.com/pub/a/2004/07/14/onto.html

Dong Y, Cheng Y, Pan Y. 2002. Mining constrained associations rules based on frequent pattern tree. Journal of Zhejiang University. Vol. 445-450.

Feng Y, Wu Z, Zhou X. 2006. Knowledge discovery in traditional Chinese medicine: State of the art and perspectives. Artificial Intelligence in Medicine. Vol. 38:3. 219-236. doi:10.1016/j.artmed.2006.07.005

Fernandez M, Gomez-Perez A, Juristo N. 1997. METHONTOLOGY: From ontological art towards ontological engineering. Workshop on ontological engineering. Spring symposium series. AAAI97 Stanford, USA.

Ferucci D. 2010. Build Watson: an overview of DeepQA for the Jeopardy! challenge. PACT’10. Vienna, Austria. ISBN:978-1-4503-0178-7.

Gierratano J, Riley G. 1998. Expert systems – Principle and programming. 3rd ed. Boston, MA: PWS Publishing Company.

Horng CH. 1992. The principles and methods of tongue diagnosis. Tongue Diagnosis. Lead Press, Taipei, Taiwan. 2-12.

Huang, JP. 1995. Methodology of traditional Chinese medicine. Beijing: New World Press. ISBN: 9787800052668

Huang MJ, Chen MY. 2007. Integrated design of the intelligent web-based Chinese Medical Diagnostic System (CMDS) – Systematic development for digestive health. Expert Systems with Applications. Vol. 32:2. 658-673. doi:10.1016/j.eswa.2006.01.037.

Huard P, Wong M.1977. Oriental Methods of Mental and Physical Fitness: The Complete Book of Meditation, Kinesitherapy, and Martial Arts in China, India, and Japan, Funk & Wagnalls. ISBN 0-308-10271-1

Hunter J. 2009. IBM Press releases. [online] 26 February. Available at: < http://www-03.ibm.com/press/us/en/pressrelease/26789.wss> [Accessed 11 October 2010].

Jang JH, Kim JE, Park KM. 2002. Development of the digital tongue inspection system with image analysis. Engineering in Medicine and Biology. Vol. 2 1033 – 1034. ISSN: 1094-687X

Jerry I, Hsiao H. 2007. Patent Protection for Chinese Herbal Medicine Product Invention in Taiwan. The Journal of World Intellectual Property Vol. 10, no. 1, pp. 1–21

Kaptchuk T. 2002. Acupuncture: Theory, Efficacy, and Practice. Annals of Internal Medicine. Vol 136:5 374-383. http://www.annals.org/content/136/5/374.full

Lao L, 1999, Traditional Chinese Medicine. In *Essentials of Complementary and Alternative Medicine.* Jonas WB, Levin JS, Eds. Baltimore, Md., Lippincott Williams and Wilkins, 1999, p. 216–232

Li WF, Jiang JG, Chen J. 2008. Chinese Medicine and Its Modernizations Demands. Archives of Medical Research. Vol. 39:2 246-251. doi:10.1016/j.arcmed.2007.09.011

Lindberg DA, Humphreys BL, McCray AT. 1993. Methods Inf Med. Vol 32:4 281-91

Lu WA, Wang YY, Wang WK. 1999. Pulse analysis of patients with severe liver problems. Studying pulse spectrums to determine the effects on other organs. IEEE Eng Med Biol Mag. 18:1 73-75. PMID: 9934604

Lukman S, He Y, Hui SC. 2007. Computational methods for Traditional Chinese Medicine: A Survey. Computer Methods and Programs in Biomedicine. Vol. 88:3 283-294. ISSN:0169-2607

Normile D. 2003. The New Face of Traditional Chinese Medicine. Science. Vol. 299:5604. 188-190. doi: 10.1126/science.299.5604.188.

Patwardhan B, Warude D, Pushpangadan P. 2005. Ayerveda and Traditonal Chinese Medicine: A Comparative Overview. Oxford Journals. doi:10.1093/ecam/neh140.

Qiao S, Tang C, Jin H. 2008. KISTCM: knowledge discovery system for traditional Chinese medicine. Applied Intelligence. doi:10.1007/s10489-008-0149-4. ISSN 1573-7497.

Smith B. 2005. Relations in biomedical ontologies. Genome Biology. Vol 6:R46. doi:10.1186/gb-2005-6-5-r46

Smith B. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology. Vol. 25. 1251-1255. doi:10.1038/nbt1346

Spirtes P, Meek C. 1995. Learning Bayesian networks with discrete variables from data. Proceedings of the first international conference on knowledge discovery and data mining. Menlo Park, CA: AAAI Press.

Tang JL. 2008. Traditional Chinese medicine. The Lancet. Vol. 372:9654. 1938 – 1940. doi:10.1016/S0140-6736(08)61354-9.

Tang WC, Liu CY, Zhao YP. 1999. Research on the Cun-Guan-Chi Pulse Detecting System. Zhongguo Yi Liao Qi Xie Za Zhi. 29:3 164-166. PMID: 16124618

Tong, Y. 2008. Ontology Engineering for Chinese Medicine. Technical Report of CCNT DartGrid Group in Zhejiang University.

Uschold M, Gruninger M. 1996. Ontologies: Principles, methods and applications. Knowledge Engineering Review. Vol 11:2. http://www.upv.es/sma/teoria/sma/onto/96-ker-intro-ontologies.pdf

Yaniv Z, Bachrach U. 2005. Handbook of Medicinal Plants. The Haworth Medical Press. ISBN 1-56022-994-2

Wu ZH, Chen HJ, Xu JF. 2003. Journal of Computer Science and Technology. Vol 18:4 462-473. doi: 10.1007/BF02948920

Wang X, Qu H, Liu P. 2003. A self learning expert system for diagnosis in traditional Chinese medicine. Expert Systems with Applications. Vol 26:4 557-566. doi:10.1016/j.eswa.2003.10.004

Watsuji T, Arita S, Shinohara S. 1999. Medical application of fuzzy theory to the diagnostic system of tongue inspection in traditional Chinese medicine. Fuzzy Systems Conference Proceedings, 1999. FUZZ-IEEE '99. 1999 IEEE International. Vol 1 145-148 doi:10.1109/FUZZY.1999.793222

Xie S, Li D, Nie H. 2009. Proceedings of the 2009 Second International Conference on Intelligent Computation Technology and Automation. Vol. 2 788-791. ISBN:978-0-7695-3804-4-2

Xu J, Yang Y. 2009. Traditional Chinese medicine in the Chinese health care system. Health Policy. Vol. 90:2. 133-139. doi:10.1016/j.healthpol.2008.09.003.

Xueqin G. 1999. Between two worlds: the use of traditional and western health services by Chinese immigrants. Journal of Community Health. Vol 24:6 421-437. doi: 10.1023/A:1018742505785

Zhang G, Lee W, Bausell B. 2005. Variability in the Traditional Chinese Medicine (TCM) Diagnoses and Herbal Prescriptions Provided by Three TCM Practitioners for 40 Patients with Rheumatoid Arthritis. The Journal of Alternative and Complementary Medicine. 11:3 415-421. doi:10.1089/acm.2005.11.415

Zhou X, Wu Z, Yin A. 2004. Ontology development for unified traditional Chinese medical language system. Artificial Intelligence in Medicine. Vol. 32:1. 15-27. doi:10.1016/j.artmed.2004.01.014.

Zhou X, Liu B, Wu Z. 2005. Text Mining for Clinical Chinese Herbal Medical Knowledge Discovery. Lecture Notes in Computer Science. Vol. 3735 396-398. doi: 10.1007/11563983_41

Zhou X, Peng Y, Liu B. 2010. Text mining for traditional Chinese medical knowledge discovery: A survey. Journal of Biomedical Informatics. Vol. 43:4 650-660. doi:10.1016/j.jbi.2010.01.002

Zhou Y, Liu X, Hao A. 1995. Study on the health status of the elderly in some areas of Beijing. Chinese Journal of Ceriatrics. Issue 03. doi: CNKI:SUN:ZHLN.0.1995-03-019

Zhu FL, Gu C, Zhang ZF. 2006. Evaluation of therapeutic effects of Chinese materia medica by tongue image analysis software 1.0 based on tongue colors. Zhong Xi Yi Jie He Xue Bao. 4:2 152-155. http://www.ncbi.nlm.nih.gov/pubmed/16529691