Real-Time Enterprise Ontology Evolution to Aid Effective Clinical Telemedicine with Text Mining and Automatic Semantic Aliasing Support

Jackei H.K. Wong, Wilfred W.K. Lin, and Allan K.Y. Wong

Department of computing, The Hong Kong Polytechnic University,
Hung Hom, Kowloon, Hong Kong S.A.R.
{cshkwong,cswklin,csalwong}@comp.polyu.edu.hk

Abstract. A novel approach is proposed in this paper to aid real-time enterprise ontology evolution in a continuous fashion. Automatic semantic aliasing (ASA) and text mining (TM) are the two collaborating mechanisms (together known as ASA&TM) that support this approach. The text miner finds new knowledge items from open sources (e.g. the web or given repertoires), and the ASA mechanism associates all the canonical knowledge items in the ontology and those found by text mining via their degrees of similarity. Real-time enterprise ontology evolution makes the host system increasingly smarter because it keeps the host system's ontological knowledge abreast of the contemporary advances. The ASA&TM approach was verified in the Nong's mobile clinics based pervasive TCM (Traditional Chinese Medicine) clinical telemedicine environment. All the experimental results unanimously indicate that the proposed approach is definitively effective for the designated purpose.

Keywords: enterprise TCM onto-core, automatic semantic aliasing, text mining, real-time enterprise ontology evolution, clinical telemedicine, D/P system.

1 Introduction

The *enterprise ontology driven information system development* (EOD-ISD) approach has contributed to the successful development of the Nong's TCM (Traditional Chinese Medicine) clinical telemedicine diagnosis/prescription (D/P) system [Lin08]. There are many D/P system variants customized from the same Nong's master proprietary enterprise TCM ontology core (TCM onto-core). Telemedicine refers to administering medicine over the mobile Internet that supports wireless and wireline communications intrinsically [Kaar99, Lacroix99, JWong08]. For example, the Nong's telemedicine environment, which has been deployed successfully in the Hong Kong SAR for clinical practice and treats hundreds of patients daily, is made up of many mobile clinics (MC) that collaborate wirelessly. On board every Nong's MC the essential elements include: a pervasive TCM D/P system, a physician to operate it, a paramedic, a dispenser, and local pharmacy. The physician treats patients locally in a computer-aided manner via the GUI (graphical user interface) of the D/P system and

R. Meersman and Z. Tari (Eds.): OTM 2008, Part II, LNCS 5332, pp. 1200–1214, 2008.

dispenses the prescriptions electronically. The GUI supports only the “select & key-in” procedure, and in this way all the keyed-in words come from the system and are therefore “standard”. These standard words are canonical terms that were extracted from formal available TCM classics, treatises, and case histories and enshrined in the Nong’s master enterprise ontological vocabulary by TCM experts in a consensus certification process [Rifaieh06]. They enable the experience of every medical/patient case to be potentially used as feedback to enrich the D/P system. Since the TCM onto-core of the system is considered “local” and was customized from the proprietary Nong’s enterprise TCM onto-core, new information (including feedback of case experience) cannot be incorporated directly into the MC D/P system. This prevents ad hoc, instant incorporation that would render the Nong's MCs incompatible and lead to uncontrollable collaboration failures. At this moment all clinical cases are only treated as physicians’ personal experience. To allow such experience and other new scientific findings data-mined over the web to be incorporated into the local TCM onto-core, we propose in this paper a novel approach to support automatic real-time enterprise ontology evolution. This approach has two components: automatic semantic aliasing, and text mining (ASA&TM). It was verified successfully in the Nong’s MC based clinical TCM telemedicine environment in a controlled fashion.

2 Related Work

The Nong’s MC based clinical TCM telemedicine D/P system was developed by the enterprise ontology driven information system development (EOD-ISD) approach [Uschold07]. Traditional software engineering usually starts with the user/functional specification and follows through the Waterfall model, which may involve the fast prototyping process that allows user input in different development phases. From the functional specification the design specification is derived, verified and implemented. Effective change control may be achieved by using entity-relationships so that variables and functions would not be inadvertently modified, causing errors in system operation. Yet, the semantics of the program modules are not explicit making their reusability and debugging difficult. This is particular so if the software is for distributed processing because traditional debugging tools designed for sequential processing are not applicable.

EOD-ISD is a new paradigm to take advantage of the intrinsic explicit semantics expressed in the enterprise ontology. The success of the Nong’s D/P telemedicine system is an evidence of the EOD-ISD effectiveness. The D/P system development process differentiates three conceptual levels of ontological constructs: i) the global ontology; ii) the local enterprise ontology core (onto-core); and iii) the local system onto-core. In this light, the details of the three conceptual levels for clinical TCM practice are as follows:

a) Global TCM ontology: This is a school that includes all the relevant TCM classics, medical theories (syllogistic – logically valid but may not have been verified yet), case histories, and treatises. It is an explicit conceptualization of concepts, entities, and their associations [Gruber93], and this conceptualization may be organized into the corresponding subsumption hierarchy of sub-ontology constructs [Guarino95].

b) Local TCM enterprise ontology (or simply enterprise ontology): This is a subset of the global TCM ontology extracted with a specific clinical purpose in mind. For example, the Nong's TCM onto-core is an extraction for clinical telemedicine deployment and practice. Therefore, the Nong's vocabulary reflects only clinical facts and no syllogistic assumptions.
c) Local MC D/P system TCM onto-core: It is customized directly from the Nong's proprietary or "master" TCM onto-core for the meta-interface (MI) specification given; thus it is a subset of the master TCM onto-core. The customization of a D/P system is automatic and the first step is to create the MI specification by selecting desired icons from the Nong's master icon library. Therefore, the local TCM onto-cores for all the MC systems customized for a particular client would be the same, but would differ from those of other clients. Neither the Nong's master TCM onto-core nor the local MC D/P system TCM onto-cores evolve automatically. At their present forms they all risk the danger of being stagnated with ancient TCM knowledge derived from old classics, treatises, and case histories. The proposal of the novel ASA&TM (automatic semantic aliasing and text mining) approach in this paper will remove this stagnation danger.

2.1 Nong's D/P Operation Walkthrough

Figure 1 depicts the EOD-ISD details as following: i) the MI specification of selected icons is first created by the client; ii) the EOD-ISD generator extracts the portion of the master/enterprise TCM onto-core of Nong's to form the D/P TCM onto-core to be customized for the local system; iii) the EOD-ISD generator constructs the semantic net (or DOM tree) and inserts the standard ready-made Nong's parser; and iv) the EOD-ISD mechanism constructs the GUI, which is the syntactical layer of the semantic net.

The semantic net is the machine processbale form of the local TCM onto-core and the syntactical layer is the query system, which abstracts the semantic net, for human understanding and manipulation. For the Nong's MC system variants the syntactical layer is a graphical user interface (GUI), which has the same appearance as the MI specification. All the symptoms keyed-in via the GUI by the physician (e.g. s_1, s_2, s_3) are captured as actual parameters for the query (e.g. $Q(s_1, s_2, s_3)$) to be implicitly (i.e. user-transparently) constructed by the GUI system as input to the parser. The parsing mechanism draws the logical conclusion from the DOM tree (i.e. the corresponding illness for the $Q(s_1, s_2, s_3)$ query). To be precise, the customized local system's ontological layer defines the ambit of the D/P operation. This layer is the vocabulary and the operation standard of the entire customized system. If a MI-based local D/P system has been correctly customized by the EOD-ISD approach, the three layers should be intrinsically and semantically transitive [Ng08]. With semantic transitivity, given an item from any layer the corresponding ones in the other layers always surface consistently.

Figure 2 is a GUI example which was customized from the given MI specification (the GUI and MI have the same appearance). In this example the master enterprise TCM onto-core is the Nong's proprietary version (used with permission). The GUI

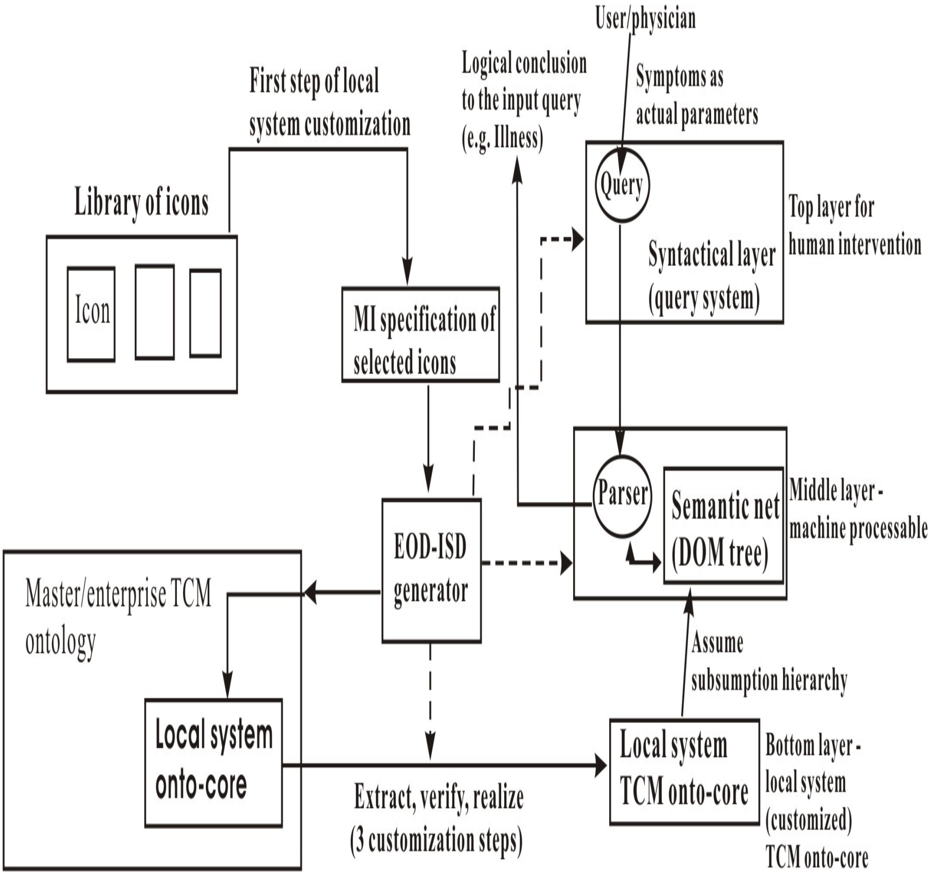

Fig. 1. Local system customization flow with enterprise ontology support

has several sections (every section was generated automatically by the EOD-ISD mechanism from the specific icon included in the MI specification) as follows:

a) **Section (I) –** The bar of control icons.
b) **Section (II) –** Patient registration number (or diagnostic Identifier) (i.e. MX6060303001) waiting for treatment and the important fields to be filled later: i) patient's complaint ("主訴"), ii) diagnosis ("診斷"): illness/type ("病"/"証"), and the treatment principle ("治則治法").
c) **Section (III) –** Symptoms ("現病史") obtained by a standard TCM diagnostic procedure that has been crystallized from eons of clinical experience.
d) **Section (IV) –** Pulse diagnosis ("脈診").
e) **Section (V) –** Prescription(s) ("處方") for the diagnosis filled in section (II); printing the final prescription and dispensing it directly in the MC.
f) **Section (VI) –** Experience window (repository) entrance of the logon TCM physician with unique official medical practice registration number (e.g. 003623 as shown).

g) **Section (IX) –** Specific questions (e.g. Do you loathe cold ambience conditions (“惡寒/怕冷”)?), and general physical inspection (e.g. complexion (“面色”) – pale, red or dark).

h) **Section (X) –** Tongue diagnosis (“舌診”) (e.g. texture and coating color).

In Figure 2 the “Symptoms (現病史)” window echoes the symptoms obtained from the patient by the normal four-step diagnostic procedure: Look (望), Listen&Smell (聞), Question (問), and Pulse-diagnosis (切). Table 1 shows an example of how these four steps would be applied by the physician to reach a diagnostic conclusion in the traditional and manual way; in this case Influenza (感冒) is concluded. With the D/P interface a physician follows the same four steps, but in a “key-in”, computer-aided fashion. The “key-in” D/P operation is standard and potentially allows the D/P results to be used as immediate feedback to enrich the local TCM onto-core. This is possible because all the TCM terms in the “select & key-in procedure” are standard in the

Fig. 2. A D/P GUI generated automatically from the given MI by EOD-ISD

Table 1. A traditional “望, 聞, 問, 切” diagnosis example (manual conclusion)

Look (望)	Listen&Smell (聞)	Question (問)	Pulse-diagnosis (切)	Illness Concluded
pale face	cough, bad breath	headache, fever, loathe cold ambience conditions (惡寒/怕冷)	taut and fast	Influenza (感冒)

enterprise ontology or vocabulary and thus the local system ontology customized by the EOD-ISD process. All the translations of Chinese terms into English are based on the WHO (World Health Organization) TCM standard [WHO07].

The master Nong’s TCM onto-core is annotated in the XML metadata format. Figure 3 is the partial XML display that pertains to the D/P operation in Figure 2.

```
<?xml version="1.0" encoding="Big5" ?>
<咳嗽>
<風寒襲肺>
<主證>
        <咳嗽>
                <咳嗽聲重>id="1"
                </咳嗽聲重>
        </咳嗽>
        <咯痰>
                <稀薄色白或中等易咯>id="1"
                </稀薄色白或中等易咯>
        </咯痰>
</主證>
<兼證>
        <風寒束表證>id="1"
        </風寒束表證>
```

Fig. 3. Partial XML annotation pertaining to the D/P operation in Figure 2

Although all the known MC D/P system variants customized from the master Nong’s enterprise TCM onto-core are successfully deployed, they all risk the danger of being stagnated with the ancient TCM knowledge enshrined in the master enterprise TCM onto-core. The desire to remove this danger prompts the proposal of the novel ASA&TM approach, which aids real-time ontology evolution of the host D/P system.

3 The Novel ASA&TM Approach

In the process of automatic semantic aliasing (ASA) the similarity between two entities are computed. For example, $A(x_1, x_2)$ and $B(x_1, x_3)$ mean that the entities

A and B are defined by two separate sets of attributes, (x_1, x_2) and (x_1, x_3) respectively. These two entities are semantically similar because they have the common parameter x_1. In the ASA context two entities are semantically the same if there are defined by exactly the same set of parameters, and they became aliases to each other if they were redefined later by only some common attributes such as x_1. In TCM this kind of aliases is rife because the same illness, which was canonically enshrined in ancient formal classics, may be redefined by some different parameters due to geographical and epidemiological needs and differences. For example, if A was canonically enshrined (i.e. a canonical term or context) and now regarded as the reference, B is then its alias. If both A and B are canonical, the context not being taken as the reference is the alias. The ASA mechanism computes the degree of similarity or relevance index (RI) of an alias to the reference context (e.g. the similarity or RI of A to B if the latter is regarded as reference context). Since the local TCM onto-core of a customized Nong's D/P system was customized from the master enterprise onto-core, which is canonical in nature (constructed by consensus certification from formal TCM text, treatises, and case histories), all its entities are canonical. But, this local TCM onto-core risks the danger of stagnation with ancient canonical TCM knowledge because it is not equipped to evolve automatically with time. The proposed ASA mechanism removes this danger by absorbing new TCM information that includes new treatment cases by physicians and scientific reports found on the open web. This can only be achieved with text mining (TM) in an incessant, real-time manner. The tasks by the ASA mechanism include:

a) Master Aliases Table (MAT) construction: If the local D/P system is equipped with the ASA&TM capability, then the MAT table will be built as part of the system initialization phase. For every canonical reference context/illness (RC) four tables/vectors are built as shown in Figure 4: i) CAT (context aliases table) to record all the aliases (canonical in nature) – a reference context may be an alias to another; ii) CAV (context attributes vector) to record all the defining symptoms; iii) RIT (relevance index table) to record the RI that defines the "non-transitive" degree of similarity of the alias to the RC; and iv) the PPT (possible prescriptions table) to expand the number of usable prescription beyond the canonical ones enshrined for the RC; achieved with the "SAME" principle (to be explained later). Therefore, the MAT serves as the entry point (or directory) to the complete set of four RC vectors/tables in the local D/P system. These vectors are the catalytic structures, which facilitate real-time onto-core evolution in the ASA&TM context. Initially the set size is equal to the number of canonical RC enshrined in the local TCM onto-core of the customized D/P system. In fact, the ASA&TM is directly applicable to the Nong's master enterprise TCM onto-core if the customized system adopts the whole master enterprise onto-core. Thus, in a generic sense the ASA&TM approach aids any enterprise onto-core for real-time evolution.
b) Text mining invocation: The text miner, namely, the WEKA is invoked by default once the ASA&TM mechanism has finished its MAT initialization. The reason for

choosing the WEKA will be explain later. The WEKA ploughs through the open web as well as the given "repertoire" of treatment cases from different TCM physicians over time to find aliases for the canonical RC in the MAT. For every new aliase the corresponding information in the four catalytic structures will be filled. The RI for the new alias with respect to the RC will be computed by the designated algorithm (explained later). Via text mining and automatic semantic aliasing the local D/P system would become smarter and clinically more effective, in light of the "SAME" principle (to be explained later).

The ASA mechanism and the text miner can be switched off with the following implications:

a) If they were switched off right before the MAT initialization, the local D/P system operates with only the original canonical knowledge extracted from the master enterprise TCM onto-core.
b) If text miner is switched off after running for some time, then the onto-core evolution respective to the MAT contents stopped and the system operates with the original canonical knowledge plus whatever new information acquired before text mining was stopped.

The ASA mechanism computes the similarity of two terms with the given algorithm. For the two terms Ter_1 and Ter_2, $Ter_1 = Ter_2$ logically indicates that they are synonyms. But, for the $P(Ter_1 \cup Ter_2) = P(Ter_1) + P(Ter_2) - P(Ter_1 \cap Ter_2)$ expression, where $\cup$ / $\cap$ for union/intersection, $Ter_1 \neq Ter_2$ means that Ter_1 and Ter_2 as aliases (not synonyms). $P(Ter_1 \cap Ter_2)$ is the probability or degree of the similarity; $P(Ter_1)$ and $P(Ter_2)$ are probabilities of the multi-representations (other meanings). For example, the English word "errand" has two meanings (multi-representations): "*a short journey*", and "*purpose of a journey*". Figure 5 summarizes the ASA rationale. The illnesses **Illness (A, a)**, **Illness (A, b)** and **Illness (A, c)** are basically aliases to another because they have some common attributes (e.g. x_1 and x_3). If the attributes are weighted, the degree of similarity between any two of them in terms of the RI value can be computed. In the ASA&TM convention, if **Illness (A, a)**) is the referential context (RC), **Illness (A, b)**) is an aliais, and similarity by the respective RI (e.g. $RI = 0.7$) indicates that **Illness (A, b)** is 70% similar to **Illness (A, a)**. If the attributes are categorized, for example: primary attributes (PA) of weight 0.5, secondary attributes (SA) of weight 0.3, tertiary attributes of weight 0.2, and nice-to-know attributes/ones (NKA/O) of weight 0.0, the RI scores of all the aliases (i.e. **Illness (A, b)**, **Illness (A, c)** and **Illness (X, x)** (as a new alaises found by the text miner over the open web)) for the chosen RC **Illness (A, a)** can be computed; this is shown by Table 2. The total set of attributes for the four illnesses in Figure 5 is called the *corpus* in this research. The second attribute **b** (e.g. in **Illness (A, b)**) is our covnetion to show geographical and/or epidemiological significance; the illness **A** can be defined by somewhat different attribute(s) in different locations (e.g. **b** and **c**).

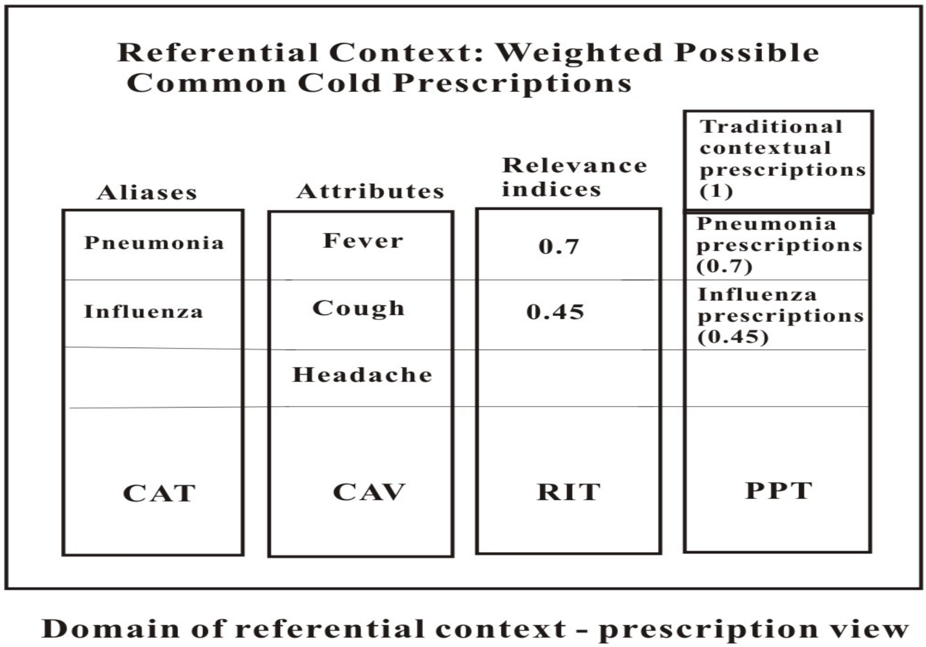

Fig. 4. A set of our catalytic data structures for the Common Cold RC

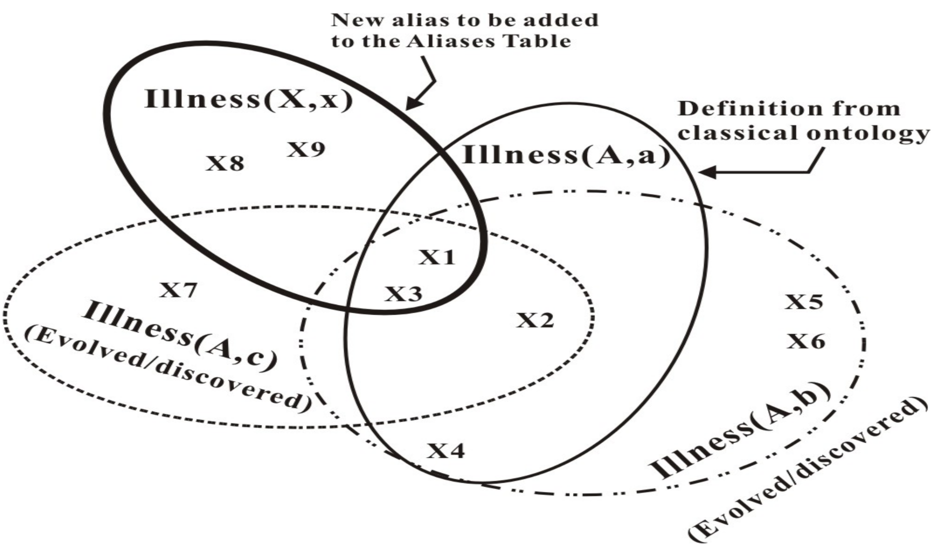

Fig. 5. Illness (X, x) is a new alias for the Illness (A, a) RC

Table 2. Aliases' RI scores for the Illness (A, a) as the RC in Figure 5

Illness/context	Attributes, corpus or alias's set	Attribute classes (for the referential context)	RI scores, { 50% -PA, 30% -SA, 20%-TA, 0%- NKA}	Remarks
Illness (A, a); Common Cold	*corpus* : $\{x_1, x_2, x_3, x_4, x_8, x_9\}$	PA - $\{x_1, x_2\}$, SA- $\{x_3\}$ TA- $\{x_4\}$, NKA - $\{x_8, x_9\}$	$RI = \frac{0.5}{2}(1+1) + 0.3(1) + 0.2(1) + \frac{0}{2}(1+1) = 1$	**Referential context (RC)**
Illness (A, b)	*alias's set:* $\{x_١, x_٢, x_٣, x_٤, x_٥, x_٦\}$	PA - $\{x_١, x_٢\}$, SA- $\{x_٣\}$ TA- $\{x_4\}$, NKA - $\{x_5, x_6\}$	$RI = \frac{0.5}{2}(1+1) + 0.3(1) + 0.2(1) + \frac{0}{2}(1+1) = 1$	**Alias of 100% relevance**
Illness (A, c)	*alias's set:* $\{x_1, x_2, x_3, x_7\}$	PA - $\{x_1, x_2\}$, SA- $\{x_3\}$ NKA - $\{x_7\}$	$RI = \frac{0.5}{2}(1+1) + 0.3(1) + 0.2(0) + \frac{0}{2}(1) = 0.8$	**Alias of 80% relevance**
Illness (X, x)	*alias's set:* $\{x_١, x_٢, x_٣, x_٧\}$	PA - $\{x_١, x_٢\}$, NKA - $\{x_٨, x_٩\}$	$RI = \frac{٠.٥}{٢}(١) + ٠.٣(٠) + ٠.٢(٠) + \frac{٠}{٢}(١+١) = ٠.٢٥$	**Alias of 25% relevance**

Text mining is fundamental to the success of the proposed ASA&TM approach. It addresses the issue of how to find useful information patterns from a preprocessed text effectively [Bloehdorn05, Holzman03]. Preprocessing involves structuring the input text in the predefined way dictated by the technique/tool used. During the structuring process parsing may be applied along with the addition of derived linguistic features and removal of others. The successfully preprocessed input text (e.g. in the Attribute-Relation File Format (ARFF) for the WEKA text mining tool) is saved in the database for subsequent and repeated uses. Text preprocessing may involve the following tasks, in an alone or combined fashion: text categorization, text clustering, concept/entity extraction, granular taxonomy, sentiment analysis, document summarization, and entity relationship modeling. A "high quality" text mining process usually yields useful results in terms of relevance combination, novelty, interest, and discovery. A common text mining usage is to measure and quantify statistically how important a term/word is in a document or a corpus (a bundle of documents). The interpretation of importance, however, remains with the domain experts. For example, the following two quantification parameters are commonly used in various problem domains including our research with modifications:

a) *Term frequency* (or simply tf): This weighs how important the i^{th} term t_i is by its occurrence frequency in the j^{th} document (or d_j) within the corpus of K documents. If the frequency of t_i is tf_i and $tf_{l,j}$ is the frequency of any other term t_l in d_j, then the relative importance of t_i in d_j is $tf_{i,j} = \frac{tf_i}{\sum_{i=1}^{L} tf_{l,j}}$, for $l = 1,2,..i,..L$ and $\sum_{i=1}^{K} tf_{l,j}$ includes tf_i.

b) *Inverse document frequency* (idf): This measures the general importance of the t_i term in the corpus of size K as $idf_{i,k} = \frac{K}{\{d : t_i \in d\}}$, where d is the set of documents that contain t_i.

Our in-house experience shows that the WEKA is more effective than other in mining textual patterns, and therefore it is adopted for the proposed ASA&TM prototypes to be verified in the Nong's mobile clinics based diagnosis/prescription (D/P) system environment. Table 3 compares WEKA with some popular text mining tools in the field.

The RI of an alias with respect to the given reference context (RC) is computed based on the idf concept with some modifications; it is redefined as $RI_i = \frac{SAC}{\{MAS \in [V]\}}$. In practice, RI_i, its inverse, $\frac{\{MAS \in [V]\}}{SAC}$ or the normalized form $0 < RI_i \leq 1$ can be used. RI_i is the weight or degree of similarity of the i^{th} alias in terms of the ratio defined by "*size of the attribute corpus (SAC) of a reference context (e.g. illness) over the mined attribute set (MAS)*". The attributes in MAS are conceptually canonical or standard terms in the background ontology. The RI values help improve the curative chance of a patient because it associates usable prescriptions by the "SAME' principle. This principle, which was enshrine in the core of TCM sine its dawn, is defined as: "*If the symptoms are the same or similar, different conditions could be treated in the same way medically, independent of whether they come from the same illness or different ones* [WHO07]"; in Chinese terminology it is the "同病異治, 異病同治" principle. For example, if the three different sets of prescriptions for the **Illness (A, a)**, **Illness (A, b)** and **Illness (A, c)** illnesses are PAa, PAb and PAc respectively, then by the SAME principle the total set usable prescriptions in the context of automatic semantic aliasing for **Illness (A, a)** is $PAb_{byRI} = PAa \cup PAb \cup PAc$, where $\cup$ means union of sets. This is based on the presence of common attributes among the three illnesses. The RI values computed for the aliases PAb and PAc indicate their curative efficacies for PAa; for example, from Table 2 their efficacies are 100% and 80% respectively.

Table 3. Strengths and weaknesses of the some text mining tool examples

Tools	Strengths	Weaknesses
Clementine	• Visual interface • Algorithm breadth	• Scalability
Darwin	• Efficient client-server • Intuitive interface options	• No unsupervised algorithm • Limited visualization
Data Cruncher	• Ease of use	• Single Algorithm
Enterprise Miner	• Depth of algorithms • Visual interface	• Harder to use • New product issues
Mine Set	• Data visualization	• Few algorithms • No model export
CART	• Depth of tree options	• Difficult file I/O • Limited visualization
Scenario	• Ease of use	• Narrow analysis path
Neuro Shell	• Multiple neural network architectures	• Unorthodox interface • Only neural networks
S-Plus	• Depth of algorithms • Visualization • Programmable or extendable	• Limited inductive methods • Steep learning curve
Wiz Why	• Ease of use • Ease of model understanding	• Limited Visualization
WEKA	• Ease of use • Ease of understanding • Depth of algorithms • Visualization • Programmable or extendable	(not obvious for our purpose)

4 Experimental Results

Many experiments were carried out in Nong's MC based pervasive TCM telemedicine D/P system environment. Different local D/P system prototypes were

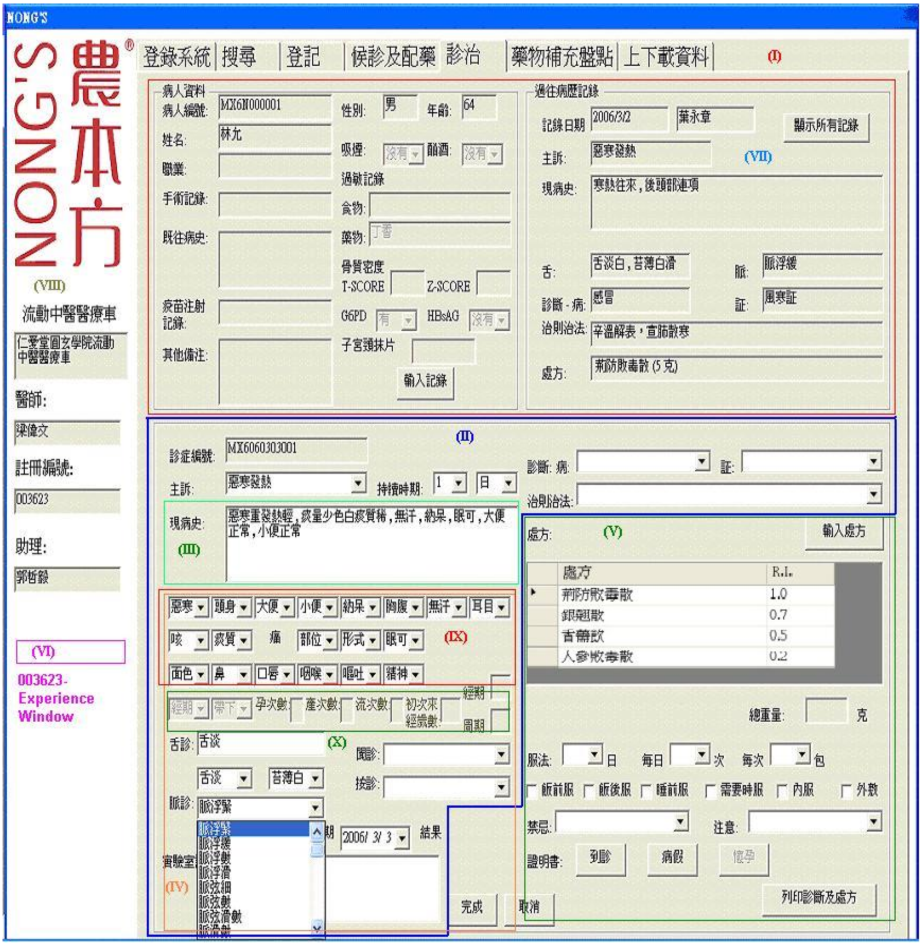

Fig. 6. ASA has established a larger set of prescriptions for treating the RC

customized from the Nong's master enterprise TCM onto-core and equipped with the ASA&TM mechanism with the WEKA text miner. The different experimental results unanimously indicate that the novel ASA&TM approach can indeed drive real-time enterprise ontology evolution for clinical telemedicine practice effectively. In this section one of the experimental results that show how the SAME principle is actually realized for clinical practice is shown in Figure 6. The key to this realization is automatic semantic aliasing that computes the RI values. From RI scores of the aliases the total set of prescriptions for treating the RC is concluded (e.g. $P_{A,a}^{total} = PAa \cup PAb \cup PAc \cup PXx$).

The essence of this experimental result is as follows:

a) *Patient's complaint (主訴):* The patient complained of "loathing cold ambience and had fever – 惡寒發熱".

b) *Symptoms (現病史):* The eight symptoms provided by the patient were keyed-in via the D/P interface (e.g. “no perspiration – 無汗” symptom).

c) *Prescriptions (處方):* Four usable prescriptions were suggested by the D/P system, as a result of ASA; they had different RI scores: 1.0 (directly from the RC), 0.7 (from 1st alias of RC), 0.5 (from 2nd alias of RC), and 0.2 (from 3rd alias of RC). The RI scores indicate the relative efficacy of the alias’s prescription for treating the RC, as shown in the (IV) section of Figure 6. In the original Nong’s MC telemedicine medicine system, which does not has the ASA capability, the set of prescriptions for treating a RC is restricted to the one that was initially established by the consensus certification process in the master enterprise TCM onto-core (e.g. the PA_a set RC **Illness (A, a)**).

The union of the RC’s and aliases’ prescription sets produces a much bigger usable set for curative purposes (e.g. $P_{A,a}^{total} = PAa \cup PAb \cup PAc \cup PXx$). In the ASA&TM context this is a clinical intelligence discovery, for the $PAb \cup PAc \cup PXx$ subset suggested by the D/P system to treat the RC might have never been enshrined in any TCM classics.

5 Conclusion

In this paper the novel ASA&TM approach is proposed for aiding real-time enterprise ontology evolution in a continuous fashion. This approach is supported by two collaborating mechanisms: automatic semantic aliasing (ASA) and text mining (TM); therefore, it is called the ASA&TM approach. The text miner finds new knowledge items from open sources such as the web and given repertoires of medical cases. Then, the ASA mechanism associates all the canonical knowledge items in the ontology and those found by text mining by their degrees of similarity, known as the relevance indices. The capability of real-time enterprise ontology evolution makes the host system increasingly smarter because it keeps ontological knowledge abreast of contemporary advances. The ASA&TM approach was verified in the Nong’s mobile clinics based pervasive TCM (Traditional Chinese Medicine) clinical telemedicine environment. All the experimental results unanimously indicate that the proposed approach is indeed effective for the designated purposes. The next logical step is to let physicians evaluate the ASA&TM approach vigorously in a controlled environment of clinical telemedicine practice.

Acknowledgments. The authors thanks the Hong Kong Polytechnic University and the PuraPharam Group of Hong Kong SAR for the respective research funding, A-PA9H and ZW93.

References

[JWong08] Wong, J.H.K., Wong, A., Lin, W.W.K., Dillon, T.S.: Dynamic Buffer Tuning: An Ambience-Intelligent Way for Digital Ecosystem. In: Proc. Of the 2nd IEEE International Conference on Digital Ecosystems and Technologies (IEEE-DEST 2008), Phitsanulok, Thailand (February 2008)

[Lin08] Lin, W.W.K., Wong, J.H.K., Wong, A.K.Y.: Applying Dynamic Buffer Tuning to Help Pervasive Medical Consultation Succeed. In: Proc. of the 1st International Workshop on Pervasive Digital Healthcare (PerCare), IEEE PerCom 2008, Hong Kong, March 2008, pp. 184–191 (2008)

[Lacroix99] Lacroix, A., Lareng, L., Rossignol, G., Padeken, D., Bracale, M., Ogushi, Y., Wootton, R., Sanders, J., Preost, S., McDonald, I.: G-7 Global Healthcare Applications Subproject 4. Telemedicine Journal (March 1999)

[Kaar99] Kaar, J.F.: International Legal Issues Confronting Telehealth Care. Telemedicine Journal (March 1999)

[Rifaieh06] Rifaieh, R., Benharkat, A.: From Ontology Phobia to Contextual Ontology Use in Enterprise Information System. In: Taniar, D., Rahayu, J. (eds.) Web Semantics & Ontology. Idea Group Inc. (2006)

[Uschold07] Uschold, M., King, M., Moralee, S., Zorgios, Y.: The Enterprise Entology, Artificial Intelligence Applications Institute, University of Edinburg, UK, `http://citesee.ist.psu.edu/cache/papers/cs/11430/ftp:zSzzSzftp.aiai.ed..ac.ukzSzpubzSzdocumentszSz1998zSz98-ker-ent-ontology.pdf/uschold95enterprise.pdf`

[Gruber93] Gruber, T.R.: A Translation Approach to Portable Ontology Specifications. Knowledge Acquisition 5(2), 199–220 (1993)

[Guarino95] Guarino, N., Giaretta, P.: Ontologies and Knowledge Bases: Towards a Terminological Clarification. In: Towards very large knowledge bases: Knowledge building and knowledge sharing, pp. 25–32. ISO Press, Amsterdam (1995)

[Ng08] Ng, S.C.S., Wong, A.K.Y.: RCR – A Novel Model fro Effective Computer-Aided TCM (Traditional Chinese Medicine) Learning over the Web. In: The International Conference on Information Technology in Education (CITE 2008), Wuhan, China (July 2008)

[WHO07] WHO International Standard Terminologies on Traditional Medicine in the Western Pacific Region, World Health Organization (2007) ISBN 978 92 9061 248 7

[Holzman03] Holzman, L.E., Fisher, T.A., Galitsky, L.M., Kontostathis, A., Pottenger, W.M.: A Software Infrastructure for Research in Textual Data Mining. The International Journal on Artificial Intelligence Tools 14(4), 829–849 (2004)

[Bloehdorn05] Bloehdorn, S., Cimiano, P., Hotho, A., Staab, S.: An Ontology-based Framework for Text Mining. LDV Forum – GLDV Journal for Computational Linguistics and Language Technology 20(1), 87–112 (2005)