

Automatic mapping of ICPC-2 PLUS terms to the SNOMED CT terminology

Jon Patrick¹, Yefeng Wang¹, Graeme Miller², Julie O'Halloran²

¹University of Sydney, School of Information Technology, Australia ² University of Sydney, Family Medicine Research Centre, Australia

Abstract

Achieving interoperability in sharing and exchanging data between health information systems requires the support of standard medical terminology. To integrate standardised terminology into information systems, there is a need to map legacy interface terminology to a reference terminology. In this study, we mapped ICPC-2 PLUS, the interface terminology developed in Australia and classified to the International Classification of Primary Care Version 2, to the SNOMED CT terminology. We have developed a series of automated mapping algorithms to assist humans to perform the mapping. The Unified Medical Language System (UMLS) metathesaurus mapping, which utilises the links between ICPC-2 PLUS and SNOMED CT terms in the UMLS library mapped 46.5% of ICPC-2 PLUS terms to SNOMED CT. Lexical mapping explored the lexical similarities between terms in these two terminologies, and mapped 60.3% of ICPC-2 PLUS terms overall. Post-coordination of remaining unmapped terms was performed, allowing one ICPC-2 PLUS term to be mapped into composition with two SNOMED CT terms, which gives an increase of about 20% in mapped terms. Overall we have mapped 80.58% of ICPC-2 PLUS terms. A manual review of the mapping shows that about 90% of string-based mappings are accurate. Unmapped terms and mismatched terms are due to the differences in the structures between these two terminologies. Also, terms contained in ICPC-2 PLUS but not in SNOMED CT caused a large proportion of failures in the mappings.

Keywords: Medical terminology, terminology mapping, ICPC-2 PLUS, SNOMED CT

1. Introduction

A big challenge in many health organisations is the sharing and exchanging of data among its multiple systems and with external organisations. An Electronic Health Record (EHR) system will allow for interoperable health data exchange. Interoperability of healthcare data is an essential component of a national health information infrastructure that enables accurate access to important elements of patient data such as diagnosis, procedure and medication. It will significantly improve the quality of healthcare, improve efficiency and reduce costs. To achieve interoperability, a standard terminology is needed to support the health information system [1-2].

Using standard terminologies often requires that system developers migrate away from legacy interface terminologies by mapping them to a standard reference terminology [3]. An interface terminology is a collection of healthcare-related terms that supports clinicians' entry of patient-related information into computer programs, such as clinical notes entry. These terminologies generally contain a rich set of flexible, user-friendly phrases used in stand-alone applications used for data entry and display text interfaces, while a reference terminology represents clinical concepts and their relationships for computer storage, retrieval, manipulation and analysis. They have broader coverage and deeper hierarchies. As the legacy interface terminologies are specifically designed for a particular application, or for a particular sub domain in healthcare, the terminologies are lack of standardisation. To achieve interoperability and allow standardised terminologies to be used in conjunction with legacy applications or information systems, mapping of interface terminologies to reference terminologies must be established.

The *electronic Journal of Health Informatics* is an international journal committed to scholarly excellence and dedicated to the advancement of Health Informatics and information technology in healthcare. ISSN: 1446-4381 © Copyright of articles is retained by authors; originally published in the *electronic Journal of Health Informatics* (http://www.ejhi.net). This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5 License (http://creativecommons.org/licenses/by-nc-sa/2.5/au).

The process of terminology mapping refers to an identification of identical concepts or relationships between different terminologies. Terminology mapping is an important step to achieving knowledge sharing. There has been a large amount of effort spent on medical terminology mapping [4-10]. The main approaches to automated terminology mapping include lexical matching, concept matching and structural matching. A number of linguists have attempted to make use of linguistic information such as lexical similarity and semantic similarities [4-6]. Other approaches have been developed recently using structural information to map between terminologies. Mork and Bernstein [7] modified a genetic terminology mapping algorithm for mapping human anatomy, using lexical similarities and structure similarities. Fung and Bodenreider [8] derived an algorithm to find candidate mappings between any two terminologies inside the UMLS, making use of synonymy, explicit mapping relations and hierarchical relationships. Post-coordination [9] can be used to map terms to compositions of two or more concepts. Elkin and Brown [10] developed a technique for discovering and formalising the implicit semantic relationships between SNOMED Reference Terminology (SNOMED RT) and International Classification of Disease Version 9 Clinical Modification (ICD9-CM)

Imel and Campbell [11] provide a strong motivation to map medical terminologies, and they firmly believe that the mapping will increasingly become automated, leading to increased efficiency and effectiveness. However, the nature of this task makes it is very difficult to automate, because heterogeneous terminologies may reflect subtly different conceptualisations of domains by the creators of these terminologies.

In this paper, we proposed an algorithm that automatically maps an interface terminology used in Australian general practice (ICPC-2 PLUS, the interface terminology developed in Australia and classified to the International Classification of Primary Care Version 2) into a comprehensive reference terminology (SNOMED CT, The Systematized Nomenclature of Medicine Clinical Terms). In the process of mapping, we utilised three different mapping strategies to match terms lexically and performing term composition. This mapping process is semi-automatic, because it requires humans to verify the results at the end of mapping, but it transforms the time-consuming searching and mapping task into an easier selection and validation task. We also have evaluated our mapping algorithm.

2. Methods

2.1 Overview of the ICPC 2-PLUS and the SNOMED CT Terminology

ICPC-2 PLUS [12] is an interface clinical terminology classified to the International Classification of Primary Care Version 2 (ICPC-2), used in Australian general practice. ICPC-2 is a classification designed for primary care. It classifies data relating to patient reasons for encounter, problems managed, non-pharmacological treatments, referrals and orders for pathology and imaging. It is a biaxial classification system with seventeen chapters on one axis and seven components of the other. The chapters, each with a letter, are based on body systems with the inclusion of two chapters for psychological and social problems. The second axis consists of components that provides terms for symptoms and complaints (component 1), common process codes (component 2-6) and diagnosis or disease terms (component 7). ICPC-2 PLUS is an extended vocabulary of more specific terms classified in ICPC-2, which provides greater specificity for data input than the ICPC-2 classification. It currently contains over 7,000 terms that are commonly used in Australian general practice. ICPC-2 PLUS is primarily used in Australia. It is installed in various software packages and used by approximately 1,500 GPs in electronic health record systems throughout the country. The terminology is maintained and regularly updated by the Family Medicine Research Centre (FMRC), at the University of Sydney. Users of ICPC-2 PLUS are actively involved in the ongoing development of the terminology.

The Systematized Nomenclature of Medicine Clinical Terminology (SNOMED CT) [13] is developed and maintained by the College of American Pathologists. It is a comprehensive, controlled clinical reference terminology. The basic elements of SNOMED CT are concepts, descriptions, relationships and hierarchies. The SNOMED CT contains more than 360,000 concepts, about 1 million descriptions and 1.4 million relationships. Concepts are basic units of SNOMED CT terminology; each concept has a unique concept id in the terminology. Descriptions are terms of names that are assigned to specific SNOMED CT concepts. Each concept has at least three descriptions including one preferred term, one fully specified name and one or more synonyms. The synonyms provide rich information about the spelling variations of a term, and naming variants used in different countries. The concepts are connected by complex relationship networks that provide generalisation, specialisation and attribute relationships, for example, "focal pneumonia" is a specialisation of "pneumonia". Each concept in SNOMED CT is classified into one of the 18 top categories, including clinical finding, aetiologies, procedures, body part, substances, qualifiers etc. It has been used in coding clinical records and the Australian government is proposing to adopt the SNOMED CT terminological system for describing certain aspects of clinical encounters. This decision creates a need to map ICPC-2 PLUS to SNOMED CT. To complete this task in a reasonable amount of time, some computationalbased methods of matching concepts are needed to assist humans to do the job.

2.2 Mapping using the UMLS Metathesaurus

The Unified Medical Language System (UMLS) [14] is a useful knowledge source that can provide the mapping between different terminologies, and this is done by incorporating different medical terminologies into a Metathesaurus organised on the basis of a *concept*. One of its primary purposes is to connect different names for the same concept from many different vocabularies. Similar terms in different vocabularies are implicitly connected by a unique concept identifier. There are over one million concepts and 2.8 million distinct strings from over 100 source terminologies. The 2005AB version of the UMLS contains ICPC-2 PLUS terminology and SNOMED CT terminology, which are the terminologies we are about to map. The ideal of this approach is to find the terms in these two terminologies that share a common unique concept identifier (CUI) in UMLS.

The Concept Names and Sources File in UMLS Metathesaurus contains all link information among different terminologies. Every concept name in the UMLS has an entry in this file. Each row in this file includes several fields, which are used to connect the concept name to its language, source vocabularies and its concept identifier.

All concepts in the Concept Names and Sources File were extracted into a database. Each record contains CUI (Unique identifier for concept), LAT (Language of term), SAB (Source vocabulary) and CODE (the code in source vocabulary) fields. We first select all terms from both ICPC-2 PLUS and SNOMED vocabulary where the source language is English, then we use SQL query to extract all terms that have the same CUI. For example, the term Ache; generalised has source vocabulary field equals to ICPC2P and CUI equals to C0281856 and its CODE is A01001 in UMLS; the term Generalized aches and pains' SAB value is SNOMEDCT and its CODE is 82991003, and also it has a CUI of C0281856, we therefore map the ICPC-2 PLUS term Ache; generalised (A01001) to SNOMED CT concept Generalized aches and pains (82991003) according to the same CUI value.

2.3 String-Based Matching

The obvious way to identify mapping between concepts is to compare the name of the terms. The theory behind string-based matching is that most terminologies have lexical similarity in their vocabularies, describing the same concepts because the natural languages underlying the vocabularies are the same. This linguistic connection exists naturally since all terms are developed by humans and are required to be understood by humans. Four string-based mapping techniques are used.

2.3.1 Normalised String Matching

Before comparing the string, the terms from both terminologies are normalised. The aim of the normalisation process is to reduce the lexical variations when performing the string comparison. First, the words within parentheses are removed. These words are usually the attributes of the terms in SNOMED CT terms, for example Channel catfish virus disease (disorder) and the plural form of the terms in ICPC-2 PLUS, e.g. Lump(s); behind ears. Then the terms are broken into tokens or their constituent words and converted into lowercase. Stop words such as a, the, of, NOS etc., punctuations and duplicated tokens are removed. A morphological process is performed on the remaining terms to remove the inflections. For example, the term lymphocytes may be reduced to lympho or the terms inflection and inflect to inflect. This process allows some verbs to be mapped to nouns, hence increases the coverage of mapping. The disadvantage is that it can also reduce precision. After morphological process, some common spelling variations of the terms are generated, for example, haemocyte to hemocyte. Finally, the remaining words are sorted in alphabet order. Using this method, the SNOMED CT term Disease of liver (disorder), for example, is normalised to disease liver and can be mapped to ICPC-2 PLUS term Disease; liver. After normalisation, the token in remaining string are compared

using exact string match.

2.3.2 Expanded String Matching

The second string-matching algorithm is expanded string matching. This process aims to expand the abbreviation in any terms to its full form then perform the normalised string matching. If the term is not matched in normalised string matching, the abbreviation expansion will be processed. This process expands abbreviations in both ICPC-2PLUS terms and SNOMED CT terms. There are two types of abbreviated terms found in ICPC-2 PLUS terminology. The first type is acronym such as *IUCD* which stands for Intra-Uterine Contraceptive Device and the other type is abbreviations due to space limit in ICPC-2 PLUS terms. When the ICPC-2 PLUS terms are created, each term is defined to have no more than 30 characters, therefore some long tokens are written in short form, for example, musculo and musculosk are both for musculoskeletal. In the first case, we compiled an abbreviation dictionary using a list of abbreviations provided by ICPC-2 PLUS user's guide. In the second case, we created a list of short words to full words mapping using natural language description of the term in ICPC-2 PLUS. To expand the abbreviations in SNOMED CT, we use the synonymous descriptions for the term to get the full name. We didn't use any external resource, because we are about to map between controlled terminologies and we don't want to introduce ambiguity. All possible abbreviations in the term are expanded and then normalised string matching is performed.

2.3.3 Substring Matching

To increase the matching coverage, substring matching is also performed. The pair of terms is matched if the normalised and expanded ICPC-2 PLUS term is a substring of the SNOMED CT term. The premise behind the match is to consider matches where the source term and target term are similar, but not equal. The similarity is calculated using the total proportion of words that are common in both the source term and target term. The more words the source and target term have in common, the higher their similarity. Since the matches are not considered to be exact, we could possibly produce a large number of incorrect matches. We ranked the matches by the similarity score.

2.3.4 Synonym Matching

Synonym matching uses a thesaurus to explore the semantic meaning of the word constituents. We aim to transform the original term to its synonyms or lexical variations using an external dictionary. There are a number of works that use WordNet [15] to provide complementary information for lexical matching. We use the WordNet synset to provide synonym information about the term. A synset in WordNet is a collection of synonyms of a concept. Using synonyms allows, for example, the mapping of "heart disease" into "cardiac disease". We use the wn tool in WordNet to generate synonyms for each atomic term in the term, and we produce a permutation of each synonym. WordNet also provides the derivationally related terms for a given word which can be used for searching. For example, the word "fever" is linked to its related adjectives "feverish" and "feverous". Synonyms and derivationally related terms for concepts in both ICPC-2 PLUS and SNOMED CT are generated, then the terms are mapped using normalised string matching.

2.4 Post-Coordination Mapping

The post-coordination mapping process aims to map a pre-coordinated ICPC-2 PLUS term to compositions of two or more SNOMED CT concepts. To achieve domain coverage, terminology developers usually create new concepts using pre-coordination and post-coordination.

A pre-coordinated term is a composed, clinically meaningful term that has already been determined or declared in the terminologies using a distinct concept to fully describe the attributes of the term. ICPC-2 PLUS terms have a lot of low level pre-coordinations. Most ICPC-2 PLUS terms are pre-coordinated under the *Medication, treatment, procedure component* ICPC-2 hierarchy, for example, "Medication; request; blood", "Referral; dentist" and "Test; cytology; respiratory."

In contrast to pre-coordination, a post-coordinated term is a novel composed term constructed by developers which did not exist in the terminology using composition of atomic concepts in the terminology. A post-coordinated term can be composed by several atomic concepts in the terminology and linked by a relationship. A user can dynamically construct a term "Referral; dentist" using two atomic terms "referral" and "dentist".

The post-coordination mapping process consists of three steps, firstly, break the terms into atomic form; secondly map each atomic term to SNOMED CT concepts; finally, identify the relationships between the SNOMED CT concepts. We first break the ICPC-2 PLUS term into atomic terms. An atomic term is a term that cannot be completely expressed or defined in other terms and links already in the overall terminology. This step includes term normalisation, term expansion and breaking the text into separate words.

Then we map each atomic term to the SNOMED CT atomic concepts. The atomic term mapping is based on the longest string match, for example the term Test; blood; ear will be broken into three atomic terms test, blood and ear. The term Blood test is mapped to Test blood (procedure) in SNOMED CT concepts rather than mapping the terms Test and Blood separately, since *Blood test* is the longest string that has a mapping in SNOMED CT. Finally, we find the relationship between the SNOMED CT concepts by matching the relationship patterns we discovered in SNOMED CT. We aim to map two kinds of post-coordination in SNOMED CT, the Qualification and Combination. The following table shows some examples of post-coordination.

2.5 Evaluation

The mapping results were evaluated by two experts from the Family Medicine Research Centre in the University of Sydney. They are responsible for developing and refining the ICPC-2 PLUS vocabulary. All mapping candidates generated by our algorithms are exported into a spreadsheet. The experts use CliniClue SNOMED CT browser to verify the mappings. All the matches are selected on a one-to-one map of "best-fit". The "best-fit" means the most desirable and suitable match among all matching candidates. All context-dependent concepts in SNOMED CT were excluded as well as legacy concepts. Some matches are of questionable validity, due to inappropriate ICPC-2 PLUS mapping to UMLS concepts, and some mappings are reasonable lexical or concept matches but they are categorical mismatches. Only the "best-fit" matching candidates are considered as the correct matches. The remainder of the matches are incorrect mappings.

Source Term	Post-coordination
Pain;mouth	Pain (Clinical Finding) + attribute = "Finding Site" + Entire mouth region (Body Structure)
Dislocation;knee;simple	Traumatic dislocation of knee joint (Clinical Finding) + attribute = "Onset" + Simple (qualifier value)

Table 1: Example of post-coordination mapping.

3. Results

This section reports the mapping results and evaluation of our mapping algorithm. The UMLS mapping results and string-based mapping results had been evaluated by human experts. A total number of 7,410 active ICPC-2 PLUS terms were selected for the experiments. The following subsections outlined the results of each matching method and the evaluation of the mapping.

3.1 UMLS Mapping Result

A total of 3,448 (46.53%) ICPC-2 PLUS terms have been mapped to SNOMED CT terms using UMLS mapping. The mapping algorithm found 6,557 mapping candidates. 3,326

majority of matched terms are singleworded terms and multi-word expressions. Some terms with different spelling variations are also mapped. The Expanded String Matching further mapped 304 terms. It effectively increased the number of mappings in chapter L (Musculoskeletal) of ICPC-2 PLUS, because most of the terms in this chapter were compressed to short form. However, the average mappings per term increased to 1.33. Synonym Matching is not very effective and only gave a 1% increase in mapping coverage. Most of the Substring Matching results were one-to-many, and the average number of matches per term increased to 24.37. Overall, the string-based term mapped 60.34% of ICPC-2 PLUS terms to SNOMED CT concepts. The string-based mapping

3.3 Post-Coordination Mapping Result

We excluded the terms that had been mapped in the previous mapping and performed post-coordination mapping on the remainder of the terms. The remaining set consisted of 3,840 terms.

Qualifications are the post-coordinations that have at least one qualifier value concept. The post-coordination type of combinations occurred when the relationship between the concepts could be identified, but did not include qualifications, while undetermined post-coordinations are those with atomic concepts that had been matched to SNOMED CT concepts, but the relationship between the concepts could not be determined. Overall, 1,500 terms in the reminder set were terms

Matching Method	Matched Term	Candidates	Percentage	Newly Mapped Term
Normalised String Matching	3,266	3,770	44.08%	-
Expanded String Matching	3,570	4,731	48.18%	304
Synonym Matching	3,662	5,321	49.42%	92
Substring Matching	4,471	108,953	60.34%	809

Table 2: String-based mapping results.

Post-coordination Type	Number of Mappings	Percentage
Qualification	343	4.63%
Combination	902	12.17%
Undetermined	255	3.44%
Total	1,500	20.24%

Table 3: Post-coordination results.

(50.72%) mapping candidates were best-fit mappings, and 96.49% of the mapping candidates have at least one best-fit mapping.

3.2 String-Based Mapping Result

A total of 3,266 ICPC-2 PLUS terms were mapped to SNOMED CT terms using normalised string matching. This matching method generated a total of 3,565 mapping candidates, on average, 1.2 matches per matched term. The results are tabulated in Table 2.

The normalised string matching results were evaluated by two experts. Similarly to the UMLS mapping evaluation, only the "best-fit" matches were considered as correct matches. 3,031(92.8%) terms have at least one correct mapping candidate. Among the 3,770 mapping candidates, 3,565(94.25%) were correct mappings. mapped using post-coordination, which is 20.24% in the whole ICPC-2 PLUS terminology. 343 terms are of qualification relationship, 902 terms are of combination relationship and the relationships of 255 post-coordinations cannot be determined. The undetermined relationship requires humans to identify at the end of the mapping.

4. Discussion

The UMLS mapping requires the latest UMLS Metathesaurus version to get the best performance, since the content of the UMLS and its source vocabularies are refined and updated frequently. Current experiments were conducted on the 2005AB version which contained a version of ICPC-2 PLUS from 2000, and a version of SNOMED CT from 2002. The version of ICPC-2 PLUS in the UMLS accounts for only 87% of terms currently available in the terminology. ICPC-2 PLUS has since been updated in the UMLS to the most current version, and we therefore expect that a larger number of mappings will be discovered when we use the latest version.

The mapping provided by UMLS Metathesaurus can be considered as a golden standard. By observation, a large percent of the mappings provided by UMLS Metathesaurus are lexical mappings. However, the mapping still produces on average 1.9 mappings per term. Some of the mappings are still ambiguous. Using the preferred term in SNOMED CT descriptions as the one-to-one mapping reduces the accuracy of mappings because the SNOMED CT terminology is developed in America and the preferred terms are in American English. As the preferred term and synonyms for the same concepts are used differently in Australia, we still need humans to validate the mappings in the end.

On evaluation, the normalised string matching and expanded string matching were accurate and useful for about 50% of the ICPC-2 PLUS terms. Several mismatched terms were due to coordination of the terms: the term is connected with conjunctions or slashes, such as the term "Splint/immobilise; nerve". Categorical mismatches occur when the source term and target term have strong lexical similarity but belong to different categories. For example the ICPC 2-PLUS term "A59007: Pain management" is mapped to SNOMED CT concept 394882004 pain management (speciality), whereas it should be matched to 278414003 pain management (procedure).

The substring matching had broader

coverage, but resulted in a huge number of mapping candidates. The matches returned are based on string similarity, and thus are not considered to be exact. Upon normal inspection, a lot of substring mappings were imprecise. Nevertheless, roughly 10% of the mappings were still accurate. We have to suggest a number of possible correct matches to increase the chances of one being the correct target concept. One possibility for reducing the superfluous mapping candidates in string-based mapping was to use the semantic information and categorical information in the SNOMED CT hierarchy terminologies to eliminate the irrelevant mappings.

The results of post-coordination mapping have not yet been evaluated. Nevertheless, the system has demonstrated its ability for automated term composition using a combination of string-based mapping techniques. One important phenomenon in post-coordination is the identification of relationships between the mapped terms. This may require description logic generation and more detailed semantic analysis to make sure the composition of two concepts make sense. We believe that the post-coordination mapping is a way to solve the content completeness problem among different terminologies.

5. Conclusion

As the number of medical terminologies increases, so too does the need for terminology integration. As a result, the demand for rapid and effective computer-assisted terminology mapping has arisen. Computerised mapping systems could significantly reduce human effort, especially when mapping large terminologies, such as SNOMED CT.

In conclusion, we have mapped about 80.58% of ICPC-2 PLUS terms to SNOMED CT concepts with differing levels of accuracy via three automated mapping approaches. This research has demonstrated that automated mapping can perform different levels of terminology mapping. The results have shown that some of the mapping methods produce very reliable mapping, while some methods yield broarder coverage but less convincing selections. The mapping results provide an opportunity to analyse the differences in these two different terminologies. Further refinement of the mapping methods could be done to reduce superfluous and incorrect mapping using structural and categorical information, for example, the elimination of synonym ambiguity. Also, more sophisticated post-coordination mapping could be developed in order to provide more reliable mapping.

References

1. Lau L, Shakib S. Towards Data Interoperability: Practical Issues in Terminology Implementation and Mapping. HIC 2005: Thirteenth National Health Informatics Conference, Australia, 2005.

2. Bowman S. Coordination of SNOMED-CT and ICD-10: Getting the Most out of Electronic Health Record Systems. AHIMA - Perspectives in Health Information Management, 2005.

3. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc; 13(3), 2006.

4. Sun JY, Sun Y. A System for Automated Lexical Mapping, J Am Med Inform Assoc 13(3), 2006.

5. Kannry JL, Wright L, Shifman M, Silverstein S, Miller PL. Portability issues for a structured clinical Vocabulary: Mapping from Yale to the Columbia Medical Entities Dictionary. J Am Med Inform Assoc 3, 1996.

6. N. Noy, M. Musen. Prompt: algorithm and tool for automated ontology merging and alignment, in: Proc National Conference on Artificial Intelligence, 2000.

7. P. Mork, P. Bernstein, Adapting a Generic Match Algorithm to Align Ontologies of Human Anatomy. In: 20th International Conference on Data Engineering; 2004.

8. K. Fung, O. Bodenreider, Utilizing the UMLS for Semantic Mapping between Terminologies AMIA 2005.

9. Rector A, Clinical Clinical terminology: why is it so hard? Methods Inf. Md. 38 (1999) pp 239-252. 10. Elkin P., Brown S., Automated enhancement of description logic-defined terminologies to facilitate mapping to ICD9-CM J Biomed Inform. 2002 Oct-Dec;35(5-6):p 281-8.

11. Imel, M. and Campbell JR., Mapping from a Clinical Terminology to a Classification. AHIMA's 75th Anniversary National Convention and Exhibit 2003.

12. ICPC-2 PLUS, An interface terminology classified to the International Classification of Primary Care Version 2. http:// www.fmrc.org.au/icpc2plus/. 13. SNOMED International. http://www.snomed.org/snomedct/

14. National Library of Medicine, UMLS Unified Medical Language System, Website: http://umlsks4.nlm.nih.gov.

15. C. Fellbaum, WordNet: An Electronic Lexical Database, MIT Press, 1998.

Correspondence

Professor Jon Patrick Centre for Health Informatics Research & Development School of Information Technologies University of Sydney NSW 2006 Australia

Phone: +61 (0)2 9351 3524 Fax: +61 (0)2 9351 3838 http://www.it.usyd.edu.au/~jonpat