ECAI 2006
G. Brewka et al. (Eds.)
IOS Press, 2006

Web-Based Tools for Codification with Medical Ontologies in Switzerland[1]

Thorsten Kurz[2] and **Kilian Stoffel** [3]

Abstract. This paper presents three aspects through which code retrieval from an medical ontology can be improved. First, the recall of correct codes can be increased by the application of appropriate morphological operations on query terms and the enhancement of concept names with synonymous terms. Secondly, the efficiency of a manual selection process from a result set can be improved by structuring the result set correspondingly to the original ontology. Finally direct enhancements to the structure and content of the ontologies can be made in order to personalise it. These methods were implemented and evaluated in tools for two ontologies of major importance in the Swiss health care system: the ICD 10 and the TARMED ontology.

1 INTRODUCTION

The use of ontologies has a long history in the biological and medical domain. Currently there are two medical ontologies that are playing a major role for all hospitals in Switzerland and that are used for exchanging information with external entities.

1.1 ICD 10 and TARMED

One of the oldest medical ontologies still in use today is the International Classification of Diseases (ICD) [5]. Its structure and its content have evolved and through the years and its domain of application has widened from a pure classification of death causes to a statistical tool for the surveillance of diseases. The roots of the first version of this ontology go back to the late 19th century, the current version is the 10th revision, the ICD 10. This is the version that is used by the Federal Office of Statistics in Switzerland to record the frequencies of diagnosed diseases, thus hospitals in Switzerland are obliged since 1998 to codify their patient records with codes from the ICD 10 ontology.

The ICD 10 contains 26400 concepts that are organised in hierarchical chapters. For each concept there exists one title and many synonymous names. Each concept has a unique code, by which his place in the hierarchical structure can be found clearly without ambiguity. These codes are used to refer to the concepts. In contrast to ICD 10 the TARMED ontology has been created recently starting from zero [8]. The purpose of TARMED is the exchange of billing information between hospitals and health insurances. Its use is mandatory in the Swiss health care system since January 2004. The TARMED ontology contains about 4600 concepts to codify therapeutic and diagnostic medical acts. These concepts are structured into 38 chapters with sub chapters, some of these concepts can be also found in the 20 service blocks or in the 36 service groups where the concepts are grouped according to different criteria like medical processes or medical qualifications. TARMED concepts have titles, but they have no synonymous names in the versions that have been released so far.

TARMED concepts have unique codes too, although here the codes do not indicate the place of a concept in the hierarchy. Selecting precise codes from ICD 10 or TARMED is crucial for obtaining precise statistics or correct hospital bills. However in discussions with medical doctors, we found a rather low motivation to spend much effort on searching correct codes in the ICD 10 ontology. This is not too astounding, considering that the coding is often done by young assistant doctors without proper training in coding, and the coding is usually done at the end of a busy day with no real perceived need for correct coding. Very much in line with our impressions, a study of the coding quality in the two cantons Vaud and Valais found that the accuracy of the applied principal diagnostic codes is as low as 56.5% in the canton Valais and 65.3% in the canton Vaud [6]. It is to be expected that better results can be achieved by offering tools that make coding easier and more efficient.

The fact that TARMED is used for billing is leading now to an increased interest in correct coding. In contrast to the Federal Office of Statistics, insurance companies are applying a strict validation with the help of various software that ensures that codes are accumulable among each other as well as compatible with the age and the gender of a patient and last but not least with the qualification of the medical staff. For the hospitals coding therapeutic or diagnostic acts with TARMED means now to keep the balance between billing too little due to overseen applicable codes and getting bills refused due to not allowable codes.

1.2 Practical issues with ontologies

While the application domains of the ICD 10 and TARMED ontologies differ, the practical issues with those ontologies are quite similar:

1. Using an ontology, either for coding or for document classification and retrieval necessitates the ability to locate concepts and codes in the ontology. This is not as obvious as it sounds, especially not if a user is not familiar with the structure of an ontology.
2. There are concepts that are mutually exclusive or can only be combined with a limited number of other concepts. Other concepts can only be used in combination with some prerequisite concepts.
3. In different hospitals, there is often a local terminology in use that

[1] This work was supported by the Swiss National Research Foundation: grant number 2100-066898.01/1
[2] University of Neuchâtel, Switzerland, email: thorsten.kurz@unine.ch
[3] University of Neuchâtel, Switzerland, email: kilian.stoffel@unine.ch

is not exactly corresponding to the terminologies used in the ontologies.
4. Some hospitals do already have code-based accounting systems in place that have to be mapped to the new TARMED ontology.

This paper is about methods that can make codifying more reliable and more efficient and about the tools, which implement such methods. Section 2 presents morphological operations on query terms, Section 3 discusses how to personalise ontologies, Section 4 describes the resulting tools, and Section 5 evaluates the performance of a current prototype.

2 FINDING CODES AND CONCEPTS

A strictly hierarchical access to the concepts through chapters and sub-chapters can only be efficient, if the chapter titles are sufficiently explicit about their content, and even in this case it is possible that a code belongs semantically to more than only one chapter. To find a concept in an ontology in this way is far from obvious, especially if the user is not familiar with this ontology. In health care services, where a large part of the coding is done not by coding experts, but by young assistant doctors and nurses, this is one aspect that leads to bad coding.

Therefore it is a promising idea to allow for an alternative access to the concepts, based on keywords and other methods from the information retrieval (IR) domain [9]. Each concept has a name consisting of at least one, but most times several words that can be searched for. Like in a search engine a query field is provided, which is used to search concepts by their keywords.

For clustering the results of a query the hierarchical structure of ontologies turns out to be very useful. Instead of displaying the resulting concepts as a flat list, the concepts of the result set are presented in the hierarchical structure that is obtained by using the original ontology, and keeping only those branches of the hierarchy that contain a class of the result set. This is especially effective for large result sets with more than 10 results [4].

2.1 Morphological operations with medical terms

A well-known morphological operation in IR systems is stemming, which is applied to words in order to reduce variant word forms to common roots. This operation improves a systems ability to match query and document vocabulary. Basically, it consists in removing the plural, verb or attribute endings of words, e.g., the words *hopeful*, *hopefulness*, *hopes*, and *hoping*, are reduced to their stem *hope*.

Medical terms differ from common language words in such as that they are often combinations of several Latin or Greek words. Each of these words carries meaning and contributes to the meaning of the term. Sometimes such word parts can stand alone as well as being part of a more complex word. That means that for one medical concept there can be many more or less similar synonymous expressions. For example the following lines are synonymous in ICD 10:

1. Kératoconjonctivite
2. Kérato-conjonctivite
3. Kératite superficielle avec conjonctivite

Here it would not be enough to simply cut off the endings. To be able to do a successful matching of the expressions above, it is necessary to cut the complex words into their meaningful parts. Consequently a stemming algorithm has been developed that reduces a medical terms into the smallest possible meaningful parts and cuts off meaningless pieces. Basically the algorithm works similar to the Porter algorithm [7] with the difference that its structure is more modular, which permits to adapt it progressively to the characteristics of the endings of medical expressions. Additionally there is a component recognition module that uses a list of known word components of medical terms [3].

2.2 Proposing alternative query terms

The current version of our search engine returns all concepts in which all words in the query are present (logical conjunction, AND). Being too specific in a query can therefore be inhibiting and prevent any concepts from appearing in the results, whereas a less specific query might eventually return a result set with concepts that satisfy the requirements of a user.

Trying out different modified versions of a query manually is annoying and time consuming. A better approach is to let the system propose less specific, potentially successful alternative queries with less keywords in case that a query is unsuccessful and returns zero results.

3 PERSONALISING ONTOLOGIES

The hierarchical structure and the vocabulary used in an ontology reflect necessarily the application domain for which it was conceived and the world view of its editors. This is not much of a problem, if the application domain stays the same over the time and all users share the same world view. But there are different points of view for an health insurance company, for the accounting department in an hospital and for a medical doctor or a nurse describing their therapeutic acts. And such differences exist not only between different types of organisations, but also between organisations of the same type, e.g., there can be differences in the vocabularies that are used in different hospitals.

There are hospitals, which have established an internal extension to the ICD 10 concepts that provide a finer distinction of diseases combined with their own vocabulary. Most of the Swiss hospitals are also already using their own creations or modifications of classification systems for therapeutic acts for internal accounting. And of course those classification systems are far from compatible with the logic imposed by the TARMED ontology.

With all these differences it is clear that there is a need for a common ontology to provide a common ground and a shared language, but there is also need to maintain a localised, even personalised access to the concepts in an ontology.

3.1 Adding synonyms

If codes are retrieved through an interface that supports keyword queries, the easiest way to provide a personalised vocabulary is to allow personal expressions and keywords for each concept to be used.

This can be accomplished by either changing the concept names to user-specific names or by attaching user-specific synonyms to each concept. The second solution was implemented in the here presented retrieval systems. The integration of a medical synonym dictionary was considered too, but this approach turned out to be too unspecific, resulting in a considerably decreased precision for the queries.

3.2 Adding access paths

In TARMED we find parallel to the hierarchical organisation of concepts in chapters also service groups and service blocks that allow to access the same concepts based on different criteria like different branches and qualifications.

Swiss hospitals most times have already local classification systems for therapeutic acts in use and contrary to TARMED the doctors and nurses are familiar with these systems. So the most obvious step is trying to translate the existing concepts to new concepts in TARMED. Due to differences in the definition of the concepts e.g. TARMED is quite specific about the qualification required to carry out a medical act, duration of the act and inclusions or exclusions of other concepts - there cannot be an automatic direct one to one translation from the existing local concepts.

However it is possible to reuse the structure of existing local classification systems and to add to each local concept the potentially applicable TARMED concepts. Or otherwise to define a new hierarchy in order to structure the TARMED concepts according to departments, services or other personal criteria.

To support this, we have created an web-based ontology editor that allows to add new concepts and relations between concepts to an existing ontology, to assign to concepts different labels and to modify them.

4 ONTOLOGY TOOLS

Prototypes for ICD 10 and TARMED code retrieval systems were build with high performance knowledge bases as backed [2, 1]. The systems implement the methods presented in Section 2 and Section 3 to various degrees and are accessible through standard web browsers. The ICD 10 Explorer (c.f. Fig. 1) was our first prototype system and its query processing is not as advanced as that of the TARMED Explorer, i.e. the ICD 10 Explorer does a matching of the query with the ontology and returns the highest ranking results, even if they do not match a 100% of the given keywords.

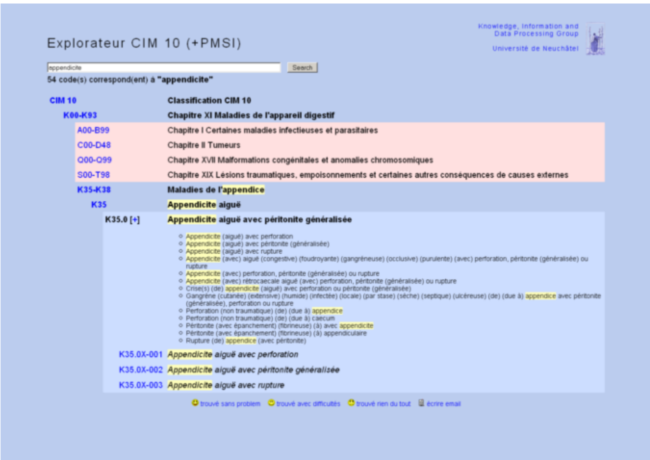

Figure 1. The ICD 10 Explorer. Appearances of the query terms in the ontology are highlighted yellow. The concepts in the red box in the upper half of the screen are references to other chapters that contain also matching concepts, but are mutually exclusive with concepts from the current chapter.

The TARMED Explorer on the other hand returns only results that match 100% of the given keywords, but it proposes less specific queries that would lead to results, if no results are found for the original query. Neither the ICD 10 nor the TARMED Explorer allow online editing and enhancing of ontologies so far, this is the domain of another system, the TARMED Editor (c.f. Fig. 2), which offers only ontology editing so far, but no keyword search.

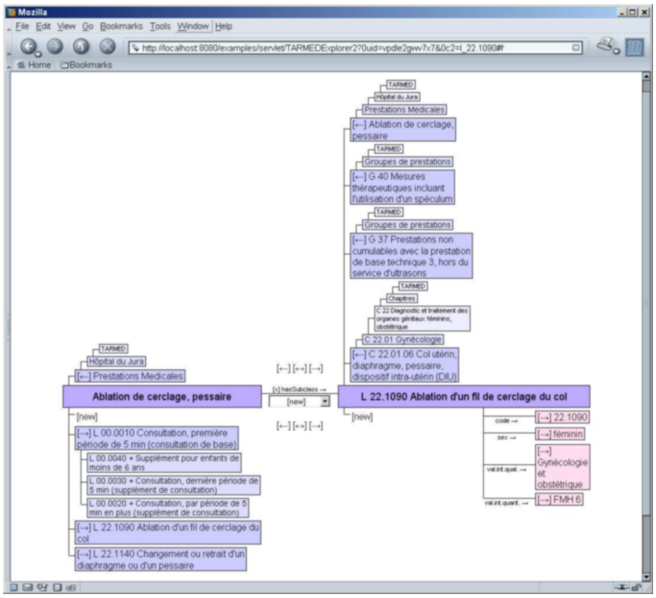

Figure 2. Mapping a local classification system for accounting to the TARMED ontology with the TARMED Editor.

5 EVALUATION

5.1 Code retrieval processes

The online prototypes of the ICD 10 and TARMED code retrieval systems have found many regular users in different hospitals in the French-speaking part of Switzerland and France. Among them is a French hospital that has made the ICD 10 system part of its coding procedure, so that we have currently the data of more than 40000 code retrieval processes to evaluate. An analysis of the log file shows a wide user base ranging from one-time users to users that come back frequently on a very regular basis (c.f. Table 1).

Table 1. Distribution of code retrieval processes per IP address (user).

Code Retrieval Processes per IP Address (User)	IP Addresses		Retrieval Processes	
1-4	6811	(90.9%)	11211	(26.5%)
5-49	603	(8.1%)	8983	(21.2%)
50-499	68	(0.9%)	10526	(24.8%)
≥500	10	(0.1%)	11635	(27.5%)
total	7492	(100.0%)	42355	(100.0%)

A typical code retrieval process starts with a query from the user. The system uses the chapter hierarchy of an ontology to build a result tree and displays consequently only these chapters that contain codes that match the query. Then the user can start to browse through the result tree and eventually reformulate his query if necessary. All user actions are registered in a log file, which contains IP addresses, access times and dates, query expressions and the chapters and concepts that were visited. Thus we define a code retrieval process as a continuous sequence of queries and consequent browsing actions that are generated by one IP address, in which the queries have a similarity either in regard to their keywords or in regard to their results. *Continuous sequence* means that the sequence of actions is not interrupted for longer than a specific time interval. A visual log file analyser gives a semi-structured view on those code retrieval processes (c.f. Fig. 3).

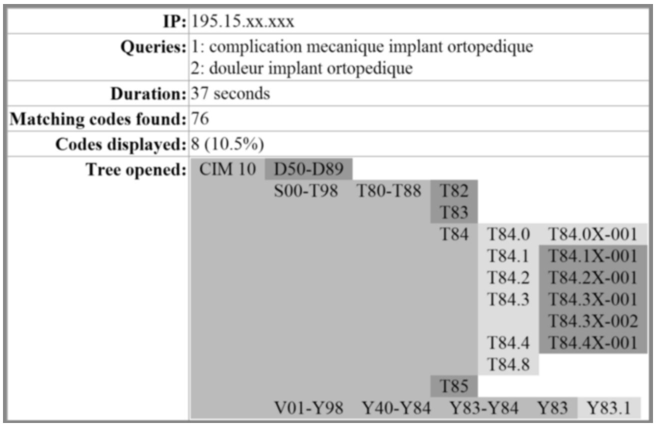

Figure 3. A semi-structured view of a code retrieval process with two queries that shows which parts of the result tree have been visited or ignored. Medium grey: Chapters of the result set that have been opened by the user. Dark grey: Chapters and codes of the result set that have been ignored by the user. Light grey: Codes that have been viewed by the user.

5.2 Efficiency, recall and precision

The average number of queries in a code retrieval process is 1.4, actually for 76% of the code retrieval processes only one single query was sufficient. The causes for code retrieval processes with larger query numbers are spelling errors, attempts to find a concept through a synonym that is unknown to the system and last but not least too unspecific queries that lead to too many results. Since the original report or bill that a user tries to codify remains unknown, it is never possible to make an absolute judgement if a returned code is correct purely based on the log data. Here it turns out to be helpful that results are not presented as a flat list, but as a tree structure. So instead of browsing through all results, the user decides which chapters seem interesting and worth opening. In fact usually the majority of chapters of the result tree is not visited, and this user behaviour can be used to distinguish between potentially correct and presumably incorrect results of query. An evaluation can be done based on the observation of which part of the result tree was visited, which one was ignored and which additional concepts were visited that were not part of the original result tree. Thus we define the *precision* of a result set as the percentage of visited concepts in comparison to the overall size of the result set. The precision is a measure for the relevance of a result set, the higher the value, the more relevant is the returned result set. Further we define the *recall* of a result set as the percentage of the visited concepts that were in the result set in comparison to the overall number of visited concepts during a code retrieval process. The recall is a measure for the completeness of a result set, the higher the value, the more complete is the returned result set. Recall and precision are here no absolute measures, but measures relative to a given system context, which allow to judge the impact of modifications of this system. Comparing those measures immediately after the first query and after the last query for all queries of code retrieval process together shows whether the additional queries really led to new information, i.e. increased recall and at what cost, i.e. decreased precision.

ICD 10 prototype uses a variant of the stemmer for French medical term which was discussed in Section 2. Based on the data from the log file the code retrieval processes were reconstructed and sent to another system, which uses no stemming at all. Table 2 shows the comparison of the two systems. The usefulness of the stemming system becomes evident by the fact that the recall of the system with the stemmer reaches already at 79% after the first query, which is 14% better than a system without stemmer achieves with all the queries of a code retrieval process together. In each of the two systems can be observed that the further queries lead to increased recall (6%-9%), at the cost of some precision (4%-5%), but the recall of the system with the stemmer is 20% higher after the first query and 23% higher for the overall code retrieval process. The differences in terms of precision are far smaller, the system with the stemmer returns results that contain 1% to 2% less of relevant entries than the system without a stemmer. That means for accepting 2% more of not relevant codes in the result set, it contains 23% more of the codes that can be considered relevant for the request.

Table 2. Performance of code retrieval systems with and without a stemmer module for medical terms.

	No Stemmer	Stemmer
Precision of first Query	40%	39%
Precision all Queries together	36%	34%
Recall of first Query	59%	79%
Recall all Queries together	65%	88%

6 CONCLUSION AND FUTURE WORK

The development and testing of these tools is being carried out in close contact with medical professionals that are providing us with valuable feedback. Apart from positive feedback from those users, it was possible to quantify the benefits of a domain specific stemming algorithm. While most search engines are reluctant with the use of stemming because of fear of a considerable loss of precision, it could be demonstrated in case of the ICD 10 ontology that the loss of precision is smaller by factor ten compared to the gain of recall of potentially relevant codes. Considering the fact, that the user rarely knows the exact name of a code that he is looking for, there will be always several potentially correct answers to his query. Instead of forcing the user to browse through long lists of results, the hierarchical structure of ontologies allows to structure the results in a way that in average only 34% of the results are visited and taken into consideration by the user. This percentage becomes as small as 9% for large result set with more than 100 entries. The system doesn't try to be more intelligent than its users and therefore does not impose presumably ”correct” answers on them, but presents information successful in a way that supports their decision making.

Both, the ICD 10 and the TARMED Explorer are currently running as stand-alone applications that can also be integrated in bigger frameworks as lookup-tools. The TARMED Editor will be evolving towards a multi-user editing environment and there is research going on to use the TARMED Explorer in combination with a voice recognition system for formulating the queries offline in a dictaphone.

ACKNOWLEDGEMENTS

We would like to thank Mr. Jean-Claude Vergriete for providing us with extensive lists of medical word components and synonyms and we also would like to thank the users of our system for their valuable feedback.

REFERENCES

[1] C. Ciorascu, I. Ciorascu, and K. Stoffel, ‘knOWLer - ontological support for information retrieval systems’, in *Proc. of 26th Annual International ACM SIGIR Conference, Workshop on Semantic Web*, Toronto, (2003).

[2] I. Ciorascu, C. Ciorascu, and K. Stoffel, ‘Scalable ontology implementation based on knOWLer’, in *Proc. of 2nd International Semantic Web Conference (ISWC2003), Workshop on Practical and Scalable Semantic Systems*, Florida, (2003).

[3] T. Kurz and K. Stoffel, ‘An advanced stemming algorithm for creating concept signatures of medical terms’, in *Proc. of the ES 2001, Research and Development in Intelligent Systems XVIII*, pp. 273–281, Cambridge, (2001).

[4] T. Kurz and K. Stoffel, ‘Enriching ontologies for improved access to web documents’, in *Proc. of IADIS WWW/Internet 2002*, pp. 815–818, Lisbon, (2002).

[5] Organisation mondial de la santé, *Classification statistique Internationale des maladies et des problèmes de santé connexes*, OMS, Genève, 1995.

[6] N. A. Najda, J. P. Vader, and A. Marazzi, ‘Cohérence du codage diagnostique: étude de contrôle de qualité dans les hôpitaux vaudois et valaisans’, *Médecine sociale et préventive*, (2306), (2000).

[7] M. Porter, ‘An algorithm for suffix stripping’, *Program*, **Vol. 14**(3), 130–137, (1980).

[8] TARMED Suisse. http://www.tarmedsuisse.ch.

[9] R. B. Yates and B. R. Neto, *Modern Information Retrieval*, Addison-Wesley-Longman, 1999.