WHO FAMILY OF INTERNATIONAL CLASSIFICATIONS NETWORK MEETING

WHOFIC/04.040

Reykjavik, Iceland
24-30 October, 2004

Title: **Nomenclatures and Thesauri – Alternatives to the ICD-10 Alphabetical Index?**

Authors: **Michael Schopen**
Birgit Krause

Purpose: for discussion

Recommendations:

Abstract:
The current format of the Alphabetical Index to ICD-10 is a major obstacle to its use in computer systems for automated coding. For the ICD-10-GM (German Modification) a thesaurus was developed as an alternative to the classical index. It is based on a collection of current medical terminology and available in a database. A traditional index can be produced from the database. This paper will discuss the advantages and disadvantages of this approach and will also compare it with the current developments in SNOMED.

Introduction

The Alphabetical Index to ICD is a useful tool for book readers and for manual coding. The indented format is a condensed way of presenting longer textual descriptions of diagnoses. It makes extensive use of *see* and *see also* references to save space and to direct the reader to relevant entries in the index.

However, this format is not very useful for integration into computer based coding systems. Even if self-containing titles are created automatically from the book version, these texts are not very helpful in a database. The table of neoplams and the table of chemical substances create further problems.

The alternative to the traditional alphabetical index is a dictionary with self-containing texts in natural word order that can be used as a database in computerized environments. However, such a dictionary is not very helpful for manual coding or for book readers.

Traditional Index Format	Computer Dictionary
Diabetes, diabetic - adult-onset E11.- - arising in pregnancy O24.4 - - affecting fetus or newborn P70.0 - bone change – code to E10-E14 with fourth character .6	 Adult-onset diabetes E11.9 Diabetes arising in pregnancy O24.4 Diabetes arising in pregnancy, affecting fetus or newborn P70.0 Diabetic bone change E14.6

The German Thesaurus of Diagnostic Terms

As a solution to that problem we have tried to generate a book version automatically from a computer dictionary. That means that the self-containing texts in the computer dictionary have to be permuted automatically to create those entries that are needed in a traditional alphabetical index and that software is needed to condense these texts into the indented format.

Computer Dictionary	Converted Index Format
 Adult-onset diabetes E11.9 Diabetes arising in pregnancy O24.4 Diabetes arising in pregnancy, affecting fetus or newborn P70.0 Diabetic bone change E14.6	Diabetes - adult-onset E11.9 - arising in pregnancy O24.4 - - affecting fetus or newborn P70.0 Diabetic - bone change E14.6

This project started in 1995 by the collection of physicians' current terminology. It was decided not to start with terms from the Tabular List or from the Alphabetical Index, but to consult medical records and to collect those diagnostic phrases actually in use.

In the database it is possible to store several codes for each diagnostic phrase. At present codes for the German translation of the WHO edition and codes for the German

Modification ICD-10-GM are stored so that both an Alphabetical Index for the German Modification and for mortality coding can be created.

In 2000 the database had reached a size where it seamed no longer feasible to maintain several text permutations for each phrase. Based on the experiences with markup languages at DIMDI, it was decided to add XML markup to the phrases that allows to create the permutations automatically. This was necessary as the German language is a challenge for computerized processing due to its intricate grammatical structure.

> Akuter transmuraler anterolateraler Myokardinfarkt

from the computer dictionary should produce

> Anterolateral, Infarkt, Myokard, akut, transmural
> Infarkt, Myokard, akut, transmural, anterolateral
> Myokard, Infarkt, akut, transmural, anterolateral
> Transmural, Infarkt, Myokard, akut, anterolateral

In the German language the following markup is added to the phrase:

```
<NP>
  <A> Akuter</A>
  <A>transmuraler</A>
  <A>anterolateraler</A>
  <NP><S>Myokard</S><S>infarkt</S></NP>
</NP>
```

NP indicates a noun phrase, A an adjective and S a substantive. This markup allows for automated stemming of adjectives and analysis of composite substantives.

After this work was completed for the collection of terms in current use, a new stage of the project was opened by adding the inclusion notes from the Tabular List and relevant terms from the traditional Alphabetical Index. Meanwhile, more than 120,000 permuted entries are generated automatically from about 64,000 phrases in the database. The next steps comprise integration of the table of neoplasms and the table of chemical substances and inclusion of the *see* and *see also* references.

Already now the Thesaurus is the most important tool for everyday coding in German hospitals and surgeries, both in printed form and also in computer systems. The Alphabetical Index is still in use for mortality coding. When the integration of the Alphabetical Index has been finished, the Thesaurus will also replace the Index in mortality coding. This is possible as both codes for morbidity coding (ICD-10-GM) and for mortality coding (ICD-10-WHO) are available for each phrase.

Problems

The most important disadvantage of the Thesaurus is its size. If it is to become almost complete, the database will grow exponentially.

For a phrase like

Superficial injury of head

the following table gives an impression of the possible combinations to construct phrases for the dictionary:

Kind of superficial injury	Topography
abrasion blister contusion bruise haematoma insect bite nonthermal blister nonvenomous insect bite splinter without major open wound superficial foreign body without major open wound	ear eye face gum jaw oral cavity palate periocular area scalp temporomandibular joint area tongue tooth

In this example, the combination of phrases leads to 10 * 12 = 120 entries in the dictionary. After consulting the Tabular List it becomes obvious that for fractures the situation is even worse.

The current solution is to distinguish between phrases that are only stored in the computer dictionary and phrases that are also printed in the paper version. Furthermore, a long list of stop words suppresses frequent words as lead in terms for the paper version.

Perspective for the Future

In the future it is desirable to put a semantic structure underneath the current database so that information on the meaning of the phrases or – even better – on the meaning of words in the phrases becomes available. This approach is very much comparable to the semantic representation of knowledge in the Systematized Nomenclature of Medicine (SNOMED).

Thus, the question is raised whether a nomenclature like SNOMED with high-quality mappings to the ICD can replace an Alphabetical Index if it is possible to generate a book version in indented format automatically from the nomenclature. We assume that – for the English language – that is even possible without the additional XML markup that is necessary for proper automatization in the German language.