Medical Informatics Europe '96
J. Brender et al. (Eds.)
IOS Press, 1996

Conversion Table between ICD-9 and ICD-10

A. Zaiss, S. Schulz, B. Graubner* and R. Klar
Department of Medical Informatics, Albert-Ludwigs-University Freiburg, Germany
*Kassenärztliche Bundesvereinigung (German Federal Association of Panel Doctors), Köln, Germany

Abstract: In January 1998 encoding of the diagnoses of all inpatient will change from ICD-9 to ICD-10. This necessitates the conversion of ICD-9 codes to ICD-10 codes and vice versa for different purposes. We present a first version of conversion tables between ICD-9 and ICD-10 for interactive use by physicians. The basic method to construct these tables was a modified vector space text retrieval method using word stems. Stemming was done by a lexicon based method on all available textes for ICD-9 and ICD-10. The first output was an automatic generated list of weighted mappings which was filled up by published mappings. Finally the list was revised manually. Doing this took us three months to generate the presented tables. The main results are: direct mapping from ICD-9 to ICD-10 can be done in 65%, from ICD-10 to ICD-9 in 85%. The next version of the tables will have an improved quality, a higher rate of direct mappings from ICD-10 to ICD-9 and more marked automatic mappings and will be released at the end of 1996. Then it should be possible to convert ICD-10 codes automatically into ICD-9 codes. It seems impossible to map from ICD-9 to ICD-10 automatically. The presented vector space text retrieval method is also very helpful to encode complete diagnosis or treatment phrases e.g. direct out of the discharge summary into a corresponding classification.

1. Introduction

The Ninth Revision of the International Classification of Diseases (ICD-9) has been used in East Germany since 1979 and in West Germany since 1986 to encode the main diagnoses of all inpatients. Starting January 1998 the diagnoses of all inpatients as well as outpatients will be encoded with ICD-10. This has to be done by all physicians or their co-workers for medical documentation working in hospitals and/or ambulatory care. The introduction of ICD-10 in 1998 necessitates the conversion of ICD-9 codes into ICD-10 codes and vice versa. Hospital physicians nowadays mostly use specialised ICD-9 tables for their own field and will need special tables to convert ICD-9 into ICD-10. These tables are used for clinical research, contract minimum basic data sets, epidemiological morbidity trends and public health statistics. They will be helpful for comparing data collected in ICD-10 with data formerly classified by ICD-9. The version of our conversion tables presented first supports only interactive mapping of corresponding ICD codes in both directions. A complete automatic mapping seems possible in the direction from ICD-10 to ICD-9 and will be released with a later version. An overview about conceptual classifications in medicine with a special emphasize on the German situation is given by Klar and Zaiss [8].

2. Method

The German version of the alphabetic part of the ICD-10 was released by DIMDI (Deutsches Institut für medizinische Dokumentation und Information) in September 1995. Since then all the files necessary for building different databases for ICD-9 and ICD-10 for all entries of the classifications are available. The following sources have been used:

1. **ICD-9**
   - Revised texts for the tabular list (all chapters and V-Classification) [7]
   - Additional files for the ICD-9 [7]
   - Alphabetic index [3]
2. **ICD-10**
   - Revised texts for the tabular list (without chapter XX) [4][7].
   - Alphabetic index [4]

All entries were processed by a modified lexicon based indexing (LBI) method [1][2] to generate stems of all the words used. The LBI-method consists of three substeps. Step 1 (pre-processing) divides the input string (ICD text entry) into words, substitutes abbreviations according to their meanings and standardises word spelling. Step 2 (morphological analysis) divides up a word into its component parts and removes suffixes (stemming) and at last irrelevant words. Step 3 was modified in the following manner: Instead of assigning a set of SNOMED indices to the given input string the generated word stems were assigned to the corresponding ICD code. The ICD codes, the texts and the generated stems were put into a vector space retrieval system [10][11]. This system uses term weights to assign relevance indications to each term as a function of the term frequency in all entries of one ICD code and as a function of the number of codes containing the term. Using this system all texts of the ICD-10 could be searched for in the database of ICD-9 and vice versa. The retrieval algorithm computes for each query (e.g. a ICD-10 text) a ranked retrieval output (e.g. ICD-9 texts) in decreasing order of query-result similarity. The similarity function is the normal inner product between vector elements (cosine) producing a value of 1 when all terms match and the vectors are identical and a value of 0 when no common terms exist between query and result [10].

This system was used to obtain an automatically generated list of mappings between ICD-9 and ICD-10 and this list was completed by several published mappings between ICD-9 and ICD-10. It is worthwhile mentioning here the work of the NHS (National Health Service of United Kingdom) [9] and the publications for the classification of mental disorders [5][6]. This list was manually revised applying the following rules:

- completion of all mappings which were not found automatically - mostly due to changes in medical terminology
- mapping of all levels of ICD codes (chapters, groups, 3 and 4 digit categories)
- mapping within the chapter and on the same level - as much as possible
- consideration of instructions, inclusion and exclusion notes
- consideration of dagger - asterisk notations

The result is a symmetric table between ICD-9 and ICD-10. This table maps all chapters of the ICD-9 and the V-Classification to the corresponding chapters of the ICD-10. Thus the ICD-10 chapter XX „External causes of morbidity and mortality“ is not part of the table as well as the morphology codes (M-Classification). No effort has been made to include the facultative 5 digit codes.

3. Results

3.1 Mapping from ICD-9 to ICD-10

The analyses of the **types of mappings from ICD-9 (all chapters and V-classification) to ICD-10** are presented in the next table. „1:N" mapping means, that one ICD-9 code corresponds with more than one ICD-10 code and includes also 3 three digit and 27 four digit categories with no corresponding ICD-10 code. Cardinality means the average figure of corresponding codes.

	Total	1:1 [#]	1:1 [%]	1:N [#]	1:N [%]	Cardinality
Chapter	18	16	88,9%	2	11,1%	2,00
Group	119	81	68,1%	38	31,9%	4,84
3 digit categories	985	588	59,7%	400	40,5%	3,28
4 digit categories	5685	3709	65,2%	2003	35,1%	4,11
Sum/Average	6837	4394	64,3%	2443	35,7%	3,98

The following table shows an example for the **mapping of ICD-9 codes to ICD-10 codes with texts** (ICD-10 notations are printed in italic style, the arrows indicate automatic preferred entities):

280-289	**IV. Krankheiten des Blutes und der blutbildenden Organe**		
	D50-D89 III. Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems		
280	**Eisenmangelanämie**		**Iron deficiency anemias**
	➢	*D50.- Eisenmangelanämie*	*Iron deficiency anemias*
		D50.0 Eisenmangelanämie nach Blutverlust (chronisch)	*Secondary to blood loss ...*
		D50.1 Sideropenische Dysphagie	*Sideropenic dysphagia*
		D50.8 Sonstige Eisenmangelanämien	*Other specified iron def...*
		D50.9 Eisenmangelanämie, nicht näher bezeichnet	*Unspecified iron def...*
281.-	**Sonstige Mangelanämien**		**Other deficiency anemias**
		D51.- Vitamin-B_{12}-Mangelanämie	***Vitamin-B_{12} deficiency***
		D52.- Folsäure-Mangelanämie	***Folate-deficiency***
	➢	***D53.- Sonstige alimentäre Anämien***	***Other nutritional anemias***

3.2 Mapping from ICD-10 to ICD-9:

The analyses of the **types of mappings from ICD-10 (without chapter XX) to ICD-9** are presented in the next table. „1:N" mapping means, that one ICD-10 code corresponds with more than one ICD-9 code. The figures also include 4 three digit and 27 four digit categories with no corresponding ICD-9 code. Cardinality means the average figure of corresponding ICD-9 codes.

	Total	1:1 [#]	1:1 [%]	1:N [#]	1:N [%]	Cardinality
Chapter	20	18	90,0%	2	10,0%	2,00
Group	226	148	65,5%	78	34,5%	3,64
3 digit categories	1660	1386	83,5%	278	16,7%	2,49
4 digit categories	8824	7432	84,2%	1419	15,8%	2,90
Sum/Average	10761	8984	83,5%	1777	16,5%	2,87

The following table shows an example for the **mapping of ICD-10 codes to ICD-9 codes with texts** (ICD-9 notations are printed in italic style, the arrows indicate automatic preferred entities):

D50-D89		**III. Krankheiten des Blutes und der blutbildenden Organe sowie bestimmte Störungen mit Beteiligung des Immunsystems**	
280-289		***IV. Krankheiten des Blutes und der blutbildenden Organe***	
D50-D53		**Alimentäre Anämien**	**Deficiency anemias**
	280	***Eisenmangelanämie***	***Iron deficiency anemias***
	281.-	***Sonstige Mangelanämien***	***Other deficiency anemias***
D50.-		**Eisenmangelanämie**	**Iron deficiency anemias**
	➢ ***280***	***Eisenmangelanämie***	***Iron deficiency anemias***
D50.0		Eisenmangelanämie nach Blutverlust (chronisch)	Secondary to blood loss (chronic)
	➢ ***280***	***Eisenmangelanämie***	***Iron deficiency anemias***
D50.1		Sideropenische Dysphagie	Sideropenic dysphagia
	➢ ***280***	***Eisenmangelanämie***	***Iron deficiency anemias***
D50.8		Sonstige Eisenmangelanämien	Other specified iron def...
	➢ ***280***	***Eisenmangelanämie***	***Iron deficiency anemias***
D50.9		*Eisenmangelanämie, nicht näher bezeichnet*	*Unspecified iron def...*
	➢ ***280***	***Eisenmangelanämie***	***Iron deficiency anemias***

The conversion table is available as ASCII-file and as a Windows computer program [12]. There are also text files in Winword-format.

4. Discussion

The World Health Organisation (WHO) has worked over fourteen years to produce the tenth version of the International Classification of Diseases (ICD-10). In the course of these years the changes in medical knowledge and medical terminology have been as significant, as were the changes in understanding and in requirements of the classification itself. There were many difficult situations in the mapping between ICD-9 and ICD-10, e.g.:

- new defined clinical concepts in ICD-10, e.g. HIV infection disease
- new classification concepts
- re-arrangements, e.g. ICD-9 chapter XVII versus ICD-10 chapter XIX: change of primary and secondary classification concept
- obsolescence
- partial absence

In spite of all these difficulties the method presented here enabled us to construct the desired tables within 3 months. This first version is to be used only interactively and is

designed for physicians. All mappings are done in a pragmatic way to propose meaningful mappings between ICD-9 and ICD-10. An automatic mapping would be of great help especially for statistical purposes, but seems very difficult between ICD-9 and ICD-10. It looks as if the opposite direction will be more successful and it is probably better collecting data in ICD-10 and mapping them to ICD-9 for comparision with older data. Further work will be done to improve the tables and to increase the part of 1:1 mappings in both directions. Furthermore explicitly defined 1:N mappings should be added to allow more and more automatic mapping in order to map the medical concept and not the code.

The method presented here with lexicon based generation of stems of words to use as indices in a vector space retrieval model is an appropriate and efficient method for mapping diagnosis texts into the desired ICD classification. The method can also be transferred to procedures and other tasks for (semi-) automatic classification of medical expressions. The method will work better in English or French than in German because automatically dividing up German composed words (Germans like to construct composita) is often difficult. It was a fascinating experience to apply this retrieval model, which Salton developed to avoid all the problems in developing and maintaining a thesaurus just on a thesaurus. We succeded in generating conversion tables between classifications in order to improve thesaurus or lexicon based methods in medical linguistics.

References:

[1] **Brigl** B. et al.: The LBI-Method for automated indexing of diagnosis by using SNOMED. Part 1: Design and realization. International Journal of Bio-Medical Computing 37 (1994) 237-247

[2] **Brigl** B. et al.: The LBI-Method for automated indexing of diagnosis by using SNOMED. Part 2: Evaluation. International Journal of Bio-Medical Computing 38 (1995) 101-108

[3] **DIMDI** [editor]: Internationale Klassifikation der Krankheiten, Verletzungen und Todesursachen (ICD) in der Fassung der vom Bundesministerium für Gesundheit herausgegebenen 9. Revision. Köln, Stuttgart, etc.: Kohlhammer (1993), Volumes IA, IB and II.

[4] **DIMDI** [editor]: Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme. 10. Revision. München, Wien, Baltimore: Urban & Schwarzenberg (1994/96), Volumes I, II and III

[5] **Freyberger** H. J. et al: Referenztabellen der WHO zum Kapitel V (F) der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD10): ICD-9 vs. ICD-10. Fortschr. Neurol. Psychiat. 61 (1993) 109-127

[6] **Freyberger** H. J. et al: Referenztabellen der WHO zum Kapitel V (F) der 10. Revision der Internationalen Klassifikation der Krankheiten (ICD10): ICD-10 vs. ICD-9. Fortschr. Neurol. Psychiat. 61 (1993) 128-143

[7] **Graubner** B.: ICD und Operationenschlüssel (ICPM): Terminologische Standardisierung und Aufbereitung für EDV-Systeme. In Trampisch H.J. and Lange S. (eds.): Medizinische Forschung - Ärztliches Handeln, München: MVV (1995), 399-404

[8] **Klar** R., **Zaiß** A.: Conceptual Classifications and Nomenclatures in Medicine. In H.H. Bock, W. Polasek: Data Analysis and Informationsystems. Springer Verlag Berlin (1996), 380-395.

[9] **NHS**: A Guide to the Use of Tables of Equivalence between ICD-9 and ICD-10. Document Version 3.0, IMG Reference F6110, April 1995

[10] **Salton** G.: Developments in Automatic Text Retrieval. Science Vol. 253 (1991), 974-978

[11] **Zaiß** A. et al.: Vor- und Nachteile verschiedener Methoden des Information Retrieval zur Klassierung von Diagnosetexten in die ICD-9. In Kunath H. et al. (eds.): Medizin und Information, München: MVV (1994), 248-251

[12] **Zaiß** A. et al: Überleitungstabelle zwischen ICD-9 und ICD-10. Programmgestaltung: dr. ruffing & partner. Köln: Deutscher Ärzte-Verlag. 1996