

ICD-10 Kodierung mit dem MorphoSaurus-System

Markó K, Daumke P, Pätzold J, Zaiss A
 Institut für Medizinische Biometrie und Informatik, Universitätsklinikum Freiburg, Deutschland

Einleitung Mit der Einführung der DRGs (Diagnosis Related Groups) als leistungsorientiertes und pauschalierendes Vergütungssystem in Deutschland hat die Verschlüsselung von Krankheiten und Behandlungsverfahren besondere Bedeutung erlangt und stellt für den behandelnden Arzt einen nicht unerheblichen Teil seiner ärztlichen Tätigkeit dar. Als Ordnungssysteme werden vor allem die Internationale Klassifikation der Krankheiten (derzeit ICD-10-GM Version 2006) und der Operationenschlüssel nach § 301 Sozialgesetzbuch V (OPS-301) eingesetzt, welche in der deutschen Ausgabe vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) veröffentlicht werden.

Für eine effiziente Zuordnung ärztlicher Informationen zu diesen Ordnungssystemen sind intelligente Kodiersysteme erforderlich. Ingener et al. [1] definieren hierzu spezifische Kriterien derartiger Kodiersysteme, die die Korrektheit, Qualität und Performanz der Kodierung entscheidend beeinflussen. Hierzu zählen neben den konkret unterstützten Datenbeständen vor allem die Möglichkeit, auf verschiedene Arten Zugang zu den Klassifikationen zu erhalten, sowie die Qualität des textorientierten Zugangs (Schlagwortsuche) zu den Daten. Eine effiziente Schlagwortsuche zeichnet sich durch folgende Kriterien aus: Unterstützung von Synonymie und Homonymie, die Zerlegung von Wortkomposita in ihre Wortbestandteile, die Berücksichtigung von Umlauten, orthografischen Varianten sowie von sprachlichen Phänomenen wie Derivation und Flexion. Weiterhin sollten Abkürzungen, Akronyme und arztindividuelle Kürzellisten aufgelöst, und die Treffer der Suche nach Relevanz sortiert dargestellt werden.

In dieser Arbeit wird ein neuartiger ICD-Kodierbrowser (SUITSEARCH ICD) vorgestellt, der die oben genannten Kriterien berücksichtigt. Er basiert auf dem MorphoSaurus-System, welches ursprünglich zur sprachübergreifenden Dokumentenrecherche entwickelt wurde [2]. MorphoSaurus verwendet das Prinzip der morpho-semantischen Indexierung, durch das medizinische Wörter in morphologisch sinnvolle Einheiten (Subwörter) zerlegt werden. Der Transfer zwischen Benutzeranfrage und Ordnungssystem geschieht auf Ebene der Subwörter. Die Ergebnisse werden nach Relevanz sortiert und automatisch werden entsprechende Verweise im systematischen Verzeichnis dargestellt sowie zusätzliche Informationen zu dem entsprechenden Eintrag angezeigt. In den folgenden Abschnitten werden das MorphoSaurus-System und die Funktionen des Kodierbrowsers genauer erläutert.

Das MorphoSaurus-System Die Morpho-Semantische Indexierung (MSI) ist ein von den Autoren entwickeltes Verfahren, welches verschiedene intra- und interlinguale morphologische Phänomene wie Flexion, Derivation und Komposition berücksichtigt. Ihr Nutzen für die sprachübergreifende Dokumentenrecherche wurde bereits gezeigt [2]. Das Verfahren beinhaltet ein spezielles Subwort-Lexikon, deren Einträge die kleinsten bedeutungstragenden Einheiten biomedizinischer Begriffe darstellen. Subwörter gleicher Bedeutung werden in sprachunabhängige Äquivalenzklassen als sogenannte Morphem-Identifiers (MIDs) zusammengefasst. Die Morpho-Semantische Indexierung biomedizinischer Wörter besteht prinzipiell aus drei Schritten:

1. orthografische Normalisierung: Zunächst werden sprachspezifische Ersetzungsregeln durchgeführt. So wird das deutsche ‚ß‘ in ‚ss‘ sowie die Umlaute ‚ä‘, ‚ö‘, und ‚ü‘ in ‚ae‘, ‚oe‘ und ‚ue‘ umgewandelt.

2. Morphologische Segmentierung: Auf Basis des Subwort-Lexikons werden nun die orthografisch normalisierten Dokumente in semantisch sinnvolle Einheiten segmentiert. Mittels eines endlichen Automaten werden dabei ungültige Zerlegungen ausgeschlossen.

3. Semantische Normalisierung: Anhand der Äquivalenzklassen werden nun alle relevanten lexikalischen Einheiten durch ihre entsprechenden MIDs ersetzt. Das Ergebnis ist eine morpho-semantisch normalisierte, sprachunabhängige Darstellung dieser Wörter (siehe Abbildung 1).

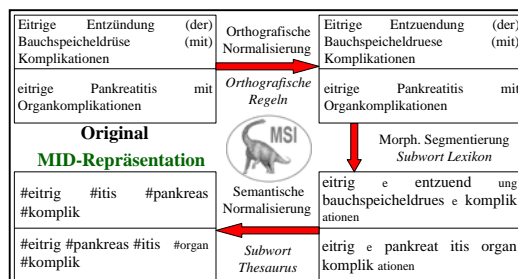


Abbildung 1: Morpho-Semantische Indexierung: Jeweils in der oberen Hälfte ist eine mögliche Suchanfrage eines Arztes abgebildet, darunter der passende ICD Eintrag. Auf Ebene der Subwörter ist eine deutliche Übereinstimmung der zwei Phrasen zu erkennen.

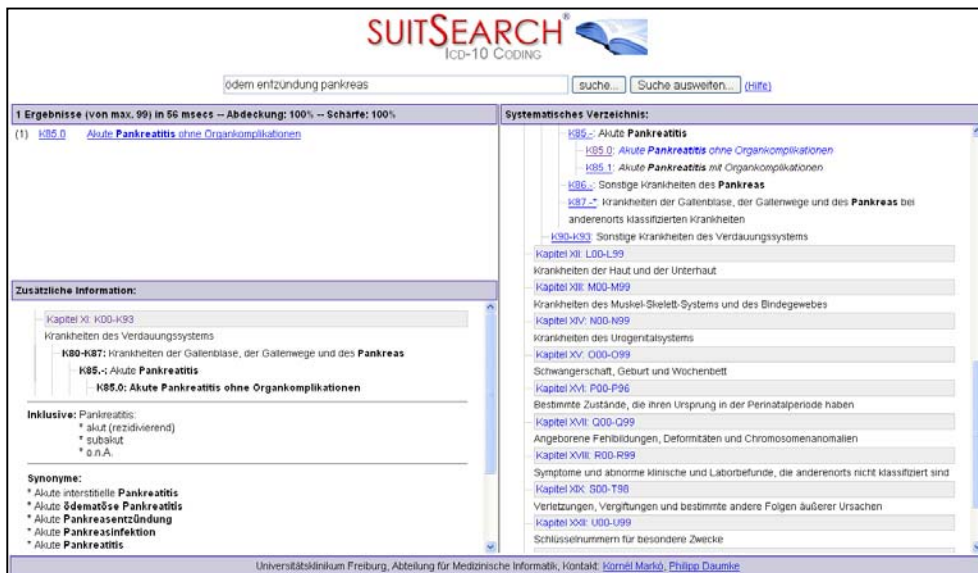


Abbildung 2: Screenshot von SUITSEARCH ICD. Oben befindet sich die Eingabefeld für zur Suche in der ICD-Klassifikation. Darunter befinden sich drei Fenster. Das linke obere Fenster stellt die Ergebnisse nach Relevanz sortiert dar. Im Fenster darunter erscheinen zusätzliche Informationen des entsprechenden ICD-Codes. Das rechte Fenster ermöglicht das Navigieren im systematischen Verzeichnis der ICD-Klassifikation.

Informationen wie Synonyme aus dem alphabetischen Verzeichnis (Diagnosesynthesaurus) sowie Inklusiva, Exklusiva und Hinweistexte des systematischen Verzeichnisses angezeigt. Das rechte Fenster stellt das systematische Verzeichnis der ICD-Klassifikation dar. Bei Mausklick auf eines der Ergebnisse wird in diesem Fenster automatisch die entsprechende Baumstruktur aufgeklappt sowie das linke untere Fenster mit den jeweiligen zusätzlichen Informationen aktualisiert. Dem Benutzer stehen somit alle relevanten Daten auf einen Blick zur Verfügung.

Die Suchfunktion von SUITSEARCH ICD erstreckt sich über das systematische und alphabetische Verzeichnis der ICD Klassifikation. Die Suche berücksichtigt Synonymie auf Basis von Subwörtern, so dass z.B. bei der Suche nach „Entzündung der Bauchspeicheldrüse“ der richtige ICD-Code K.85.0 „Akute Pankreatitis ohne Organkomplikationen“ gefunden wird. Diese Form der Synonymerkennung bietet eine sinnvolle Erweiterung zu den

SUITSEARCH ICD verfügt über eine übersichtliche Eingabemaske und drei verschiedene Fenster zur Darstellung der Ergebnisse einer Suchanfrage. Das linke obere Fenster listet die Treffer nach Relevanz sortiert auf. *Abdeckung* und *Schärfe* stellen zwei verschiedene Gütemaße für die Trefferrelevanz dar. *Abdeckung* gibt an, wie viele Suchterme auf Ebene der Subwörter in den ICD-Codes gefunden wurden. *Schärfe* ist ein intern berechneter Relevanzfaktor, der sich aus verschiedenen Aspekten wie Subwörterhäufigkeiten und deren Gewichtung innerhalb der ICD Codes zusammensetzt. Über den Button *Suche ausweiten...* kann bei Bedarf die Relevanz der Ergebnisse reduziert und dadurch gleichzeitig die Anzahl der Treffer ausgeweitet werden.

Im linken unteren Fenster werden zusätzliche

im Diagnosesaurus bereitgestellten Synonymen. Ferner werden Wortkomposita wie „Pankreatitiskomplikationen“ ebenso unterstützt wie orthografische Varianten und Flexions- und Derivationsvariationen („komplizierte Pankreatitis“). Darüber hinaus erlaubt SUITSEARCH ICD die Eingabe von Wortkürzeln wie z.B. „ak pank“ (für „akute Pankreatitis mit/ohne Organkomplikationen“) oder „pso gu“ (für „Psoriasis guttata“). Auch Kombinationen aus Synonymie und Wortkürzeln (Bsp. „Schuppenflechte gu“ für „Psoriasis guttata“) oder die Eingabe von ICD-Nummern (z.B. „L40.4“) führen zum richtigen ICD-Code.

Literatur

- [1] Ingenerf, J., Thurmayr, R., Thurmayr, G.R., Graubner, B., Zaiß, A. (1997). Diagnosen- und Prozedurenkodierung sowie Entgeltzuordnung: Ein Kriterienkatalog zur Beschreibung und Auswahl von Softwaresystemen. PMD (Praxis Medizinischer Dokumentation) 17 (4), 69-73.
- [2] Markó K., Schulz S., Hahn U.: MorphoSaurus - Design and Evaluation of an Interlingua-based, Cross-language Document Retrieval Engine for the Medical Domain. Methods of Information in Medicine. 4/2005(44): 537-545