

Wer sucht, der soll finden! - Ein Überblick über Verschlagwortung und Suche medizinischer Lerninhalte

Seek, and you shall find! - An overview on indexing and searching medical learning content

• Matthias Holzer¹ • Michael Pfähler² • Inga Hege¹ • Martin Fischer¹

Die rasche Entwicklung computerbasierter Lernprogramme in den letzten Jahren hat zu einer Vielzahl von Angeboten geführt, die für die Lernenden schwierig zu überblicken sind. Daher ist eine gute Suchfunktion ein Schlüssel zur curricularen Integration von e-Learning. Es werden die beiden Hauptstrategien beschrieben, um eine einfache und effiziente Suche in medizinischen Lehrinhalten zu ermöglichen: Indexierung anhand ausgewählter Terminologien und die freie Suche im Text. Eine Indexierung bietet den Vorteil der Abstraktion über verschiedene Synonyma eines Begriffs wohingegen die freie Suche auch Feinheiten des Inhalts erschließen kann, die der Autor nicht explizit mit Metadaten beschrieben hat. Es werden zudem konkrete Einsatz Erfahrungen mit dem fallbasierten CASEPORT-Lernsystem und dem Prüfungsfragenmanager IMSm aufgezeigt.

Over the past years the rapid development of computer-based training in medicine has created a multitude of learning units, that are difficult to overview for the learners. We therefore suggest that good search facilities play key role in the curricular integration of e-learning. We will describe the two main strategies for a simple and efficient search in medical learning content: Indexing with selected terminologies and a free-text search. Indexing can overcome the problems with unexpected synonyms while the free text search also exposes detail that are not explicitly annotated by the author. Further we will report about our experiences with the case-based portal CASEPORT and the test item management system IMSm.

Keywords: full text search, teaching, patient simulation, problem-based learning, documentation, abstracting and indexing, vocabulary, controlled

¹ Schwerpunkt Medizindidaktik, Med. Klinik Innenstadt, Klinikum der Ludwig-Maximilian-Universität München, München, Deutschland

² FH Ulm, Ulm, Deutschland

Einleitung

Das Lernen am Computer erweitert seit vielen Jahren das Spektrum der Lernmedien in der Aus-, Fort- und Weiterbildung der Medizin. In diesem Zusammenhang stehen elektronische Fassungen von Lehrbüchern, Bild- und Audiodatenbanken, Lehrvideos, Lexika, Suchmaschinen, fallbasierte Lernsysteme und Patientensimulationen sowie weitere Lerninhalte zur Verfügung, die das lineare Lernen aus Büchern und das Lernen in Lehrveranstaltungen ergänzen.

Um diese Fülle von elektronischen Lernangeboten sinnvoll zu nutzen, ist es wichtig, dass Lehrende und Lernende einen einfachen und systematischen Zugang zu den Lerninhalten haben. Eine Verbesserung der Auffindbarkeit von Lern- und Prüfungsinhalten erhöht wesentlich die Effizienz der Erstellung und Nutzung solcher Inhalte und vereinfacht ihre curriculare Integration. Zu diesem Zweck stehen generell zwei verschiedene Ansätze zur Verfügung: Zum einen kann die Suche in einem Lernsystem oder Portal anhand einer systematischen Terminologie erfolgen, wobei den einzelnen Lerninhalten vorab manuell vom Autor oder automatisch passende Einträge dieser Terminologie zugeordnet wurden. Oder es kann eine Volltextsuche erfolgen, wobei die Lern- oder auch Prüfungsinhalte nach bestimmten Wörtern, Wortbestandteilen oder Phrasen durchsucht werden können.

Im Rahmen dieses Beitrags soll ein Überblick über Verschlagwortungs- und Suchtechnologien mit einem speziellen Fokus auf die medizinische Lehre gegeben werden und Beispiele für den Einsatz dieser Technologien in konkreten Projekten angeführt werden. Die Möglichkeiten der einzelnen Technologien werden beleuchtet, ihre Probleme aufgezeigt und, falls vorhanden, geeignete Lösungen für die Probleme sowie weiterführende Literatur vorgestellt.

Methoden

• Verschlagwortung

Definition der Begriffe Klassifikation und Thesaurus

- Eine Klassifikation ist eine Zusammenfassung von Objekten zu Klassen, die ihrerseits hierarchisch zueinander in Relation gesetzt werden können. Die in der Medizin am häufigsten verwendete Klassifikation, die International Classification of Diseases (ICD), unterteilt die medizinischen Diagnosen zu Abrechnungszwecken in hierarchische Gruppen, wobei jede Diagnose nur einer Gruppe angehören kann.

- Ein Thesaurus, wie auch eine Ontologie, ist ein nach Wortfeldern (Synonyme, Unter- und Oberbegriffe) geordnetes Modell eines Themengebietes. Durch die Zugehörigkeit vieler Begriffe zu mehreren Wortfeldern sind Thesauri im Gegensatz zu Klassifikationen meist polyhierarchisch organisiert.

Medizinische Standard-Klassifikationen und -Thesauren

In der folgenden Aufstellung werden nur Systeme genannt, die in deutscher Sprache verfügbar sind, die nicht auf ein bestimmtes medizinisches Fachgebiet beschränkt sind und die bereits in Lehr- und Prüfungssystemen zum Einsatz gekommen sind.

MeSH

Die Medical Subject Headings (MeSH) sind ein von der National Library of Medicine (NLM) der USA erstellter und gepflegter Thesaurus. Sie bestehen aus Termen, die in eine hierarchische Struktur mit 11 Ebenen eingeordnet sind und so die Beschreibung eines Textes auf verschiedenen Detailebenen erlauben.

Aktuell beinhalten die MeSH 22.997 Terme. Zusätzlich steht ein Lexikon mit weiteren 151.000 Termen zur Verfügung, um das Auffinden der kanonischen Schreibweise von MeSH-Termen zu ermöglichen („Entry terms“). Die wichtigste Anwendung des MeSH [1] ist die Indexierung der medizinischen Literaturdatenbank MEDLINE [2]. Vorteile der MeSH sind die gute Abdeckung des medizinischen Fachwortschatzes, ihre Aktualität und der breite Einsatz. Die MeSH sind in deutscher Sprache gegen einen Betrag zwischen 500 und 5000 Euro je nach Verwendungszweck beim Deutschen Institut für Medizinische Dokumentation und Information (DIMDI) [3] zu beziehen.

ICD10

Die "International Classification of Diseases" (ICD-10) wurde ursprünglich von der Weltgesundheitsorganisation entwickelt und ist in deutscher Sprache kostenfrei über die Webseiten des DIMDI erhältlich. Die ICD ermöglicht über 12.000 verschiedene Kodierungen im Bereich Krankheiten und Krankheitssymptome. In Deutschland wird die ICD-10 zur Verschlüsselung von Todesursachen und zur Verschlüsselung von Diagnosen in der ambulanten und stationären Versorgung verwendet und hat somit eine weite Verbreitung in Kliniken und Praxen. Der Vorteil der ICD-10 ist die genaue Abdeckung aller Krankheitsbilder und Symptome, die Aktualität sowie die Verfügbarkeit eines Diagnosenthesaurus, der beim Auffinden der kanonischen Einträge hilft. Demgegenüber stehen als Nachteile die ausschließliche Beschränkung auf Diagnosen, die oft noch die Verwendung einer weiteren Terminologie

erforderlich macht und die inhärente Ausrichtung auf die Verschlüsselung von Diagnosen zu Abrechnungszwecken, die sich in der strikt monohierarchischen Form niederschlägt.

Im Zusammenhang mit der ICD-10 sind auch die weiteren über das DIMDI zu beziehenden internationalen Klassifikationen zu erwähnen, die aber jeweils nur einen sehr speziellen Fokus besitzen: Der Operationen- und Prozedurenschlüssel (OPS), die Anatomisch-Therapeutisch-Chemische Klassifikation mit definierten Tagesdosen (ATC/DDD) sowie die Medizinprodukteklassifikation (UMDNS).

SNOMED

Die aktuelle deutsche Ausgabe von SNOMED CT (Systematic Nomenclature of Medicine – Clinical Terms) [4] enthält 273.778 Terme und Synonyma, die in verschiedenen Achsen (Topographie, Prozedur, Morphologie, etc.) jeweils hierarchisch organisiert sind. Zudem sind über 600.000 Begriffsbeschreibungen und etwa 1,47 Mio. semantische Links enthalten. Durch den multiaxialen Aufbau von SNOMED ergibt sich die Kodierung eines medizinischen Konzeptes als Abbildung in einem mehrdimensionalen Raum. Eine Beschreibung der SNOMED II-Thesaurus findet sich in [5] und [6].

Aussagemodell von SNOMED (basierend auf SNOMED II [6]):

- Eine Prozedur P
- für eine morphologische Veränderung M
- in einer Topographie T
- verursacht durch ein Agens E
- verbunden mit einer Funktionsstörung F

Zusätzlich zu den Achsen P, M, T, E und F existiert die Diagnosenachse D.

Die hohe Komplexität von SNOMED ermöglicht einerseits eine hohe Ausdruckskraft, erfordert aber auch eine umfangreiche Einarbeitung um die Möglichkeiten dieses Thesaurus zu nutzen. Eine pauschale Angabe zu den Kosten von SNOMED ist nicht möglich, kann aber unter Angabe der Einrichtung und des Einsatzzwecks auf der Homepage angefragt werden.

UMLS

Das Unified Medical Language System (UMLS) [7], wie MeSH von der US National Library of Medicine (NLM) entwickelt, enthält ebenso wie SNOMED medizinische Begriffe (Metathesaurus) und ihre semantischen Beziehungen (Semantic Network). Das UMLS führt als Metathesaurus die Bezeichnungen aus etwa 100 heterogenen Einzelvokabularen, darunter auch

SNOMED, MeSH und die ICD-10, in 15 Sprachen zusammen und verknüpft sie jeweils mit eigenen Konzepten. Diese wiederum stehen durch ein semantisches Netzwerk miteinander in Beziehung. Da außer den erwähnten Quellen kaum deutsche Vokabulare verwendet wurden, ist es fraglich, ob die zusätzliche Komplexität des Systems die Ergebnisse einer Suche verbessern kann.

Spezifische Klassifikationen für die Medizinische Lehre

Da oft durch eine spezifisch zugeschnittene Klassifikation eine Verschlagwortung für Autor und Lerner vereinfacht werden kann wurden vielfach auch eigene weniger umfangreiche Vokabulare für die Verschlagwortung medizinischer Lern- und Prüfungsinhalte entwickelt.

Leitsymptome (CASEPORT)

Im fallbasierten Lernportal CASEPORT [8] sollen Lernfälle so beschrieben werden, dass das Problem des Patienten beim Arztbesuch dargestellt wird, ohne die schlussendliche Diagnose zu verraten. Dafür wurde aufbauend auf der Glasgow Master list of Clinical Presentations eine Liste von etwa 200 Leitsymptomen entwickelt und bei der Indexierung der CASEPORT-Lernfälle weiter adaptiert. Die Leitsymptome werden auf einem Terminologieserver der FH Ulm als Web-Service zur Verfügung gestellt.

Lernziele (Ludwig-Maximilians-Universität München (LMU), Fernuniversität in Hagen)

An der medizinischen Fakultät der LMU wurde in Zusammenarbeit mit der Fernuniversität in Hagen (Lehrstuhl Prof. Schlageter) ein Lernzielsystem entwickelt [9], das die Verknüpfung spezieller Lernziele mit Lehrveranstaltungen einerseits und Katalogeinträgen andererseits ermöglicht.

Die zugrunde gelegten Kataloge sind der Gegenstandskatalog für den schriftlichen Teil des Zweiten Abschnitts der Ärztlichen Prüfung (GK2) des Instituts für medizinische und pharmazeutische Prüfungsfragen (IMPP) und dem Hamburger Lernzielkatalog [10] basiert, Letzterer basiert auf dem Swiss Catalogue of Learning Objectives, der wiederum in Anlehnung an den niederländischen Blueprint 2001 entwickelt wurden. Die speziellen Lernziele werden in einem standardisierten Satz erfasst, der eine Fähigkeit des Studenten beschreibt.

Fächerlisten (CASEPORT/IMSm)

Eine ebenso einfache wie wirkungsvolle Einteilung von Lernangeboten ist, trotz des von der neuen Approbationsordnung geforderten fächerübergreifenden

Lernens, die Einteilung nach medizinischen Fachgebieten.

Funktionssysteme (Charité)

Am Berliner Universitätsklinikum Charité wird zudem, um dem fachübergreifenden Lernen Rechnung zu tragen, eine Einteilung nach Funktionssystemen (Bewegung, Herz/Kreislauf, Verdauung etc.) vorgenommen, die sich trotz der Kürze der Liste mit nur 11 Einträgen für die Verschlagwortung von Prüfungsfragen bewährt hat.

Navigation in kontrollierten Terminologien

Unabhängig von der verwendeten Terminologie kann die Navigation in Form einer Suche über allen Einträgen und Synonyma erfolgen oder durch einen hierarchischen Zugang. Die erste Möglichkeit bietet sich insbesondere für unerfahrene Nutzer oder bei aufwändigen Hierarchien an. Im Idealfall können beide Möglichkeiten miteinander kombiniert werden.

• Volltextsuche am Beispiel des Item Management System für die Medizin - IMSm

Die zweite Möglichkeit gewünschte Lerninhalte in einem großen Archiv aufzuspüren, ist die Verwendung einer Freitextsuche. Da die Suche in einem großen Datenbestand naturgemäß sehr lange dauern kann, bietet sich hier der Aufbau von Indizes (Konkordanz, Inverted index) an, um die Suche zu beschleunigen.

Eine Erhöhung der Retrievalleistung eines solchen Systems auf Kosten der Genauigkeit (Spezifität) der Ergebnisse kann durch eine fehlertolerante Suche erreicht werden.

Während in einer einfachen Freitextsuche nur exakt nach einem spezifischen Text gesucht werden kann, erlauben Suchoperatoren eine Festlegung der Suchmodalitäten durch den Benutzer. Eine weitere Methode, die speziell zur Zerlegung von medizinischen Komposita geeignet ist, ist die Morphosaurus-Methode [11], die Wörter zur Indexierung in ihre kleinsten sinntragenden Einheiten zerlegt.

Im Rahmen der Programmierung des „Item Management System für die Medizin“ [12] wurde am „Schwerpunkt Medizindidaktik“ der LMU München eine Volltextsuche für Prüfungsfragen mit Hilfe der Open-Source-Bibliothek Lucene [13] programmiert. Das webbasierte IMSm-System ist ein Kooperationsprojekt der Universitäten Berlin, Heidelberg und München und dient dazu, Prüfungsfragen der Medizin-Studiengänge zu verwalten und daraus Prüfungen zu generieren.

Da verschiedene Fakultäten und Universitäten ihre Prüfungsfragen in das IMSm einspielen und gegensei-

tig zur Verfügung stellen, kommt dem Auffinden der vorhandenen Items eine zentrale Rolle zu. Neben einer Einordnung in Klassifikationssysteme (Fachgebiet, Funktionssystem, Lernziele (geplant)) kommt hier eine Volltextsuche zur Anwendung. Die dabei aufgetretenen Probleme und ihre Lösungen werden im folgenden Abschnitt behandelt.“

Ergebnisse

• Benutzererfahrungen mit der terminologie-basierten Suche im CASEPORT-System

In einer in diesem Jahr durchgeführten Benutzerumfrage zum CASEPORT-Portal haben von 778 eingeladenen Nutzern des Portals 131 u.a. ihre Erfahrungen mit der strukturierten CASEPORT-Profisuche bewertet. Die drei prominentesten Suchfelder im oberen Bereich der Suchmaske greifen auf kontrollierte Terminologien zu (Fachgebietsliste, Leitsymptomliste, ICD-10) (Abbildung 1).

The screenshot shows a search interface with several sections:

- Fachgebiete:** A dropdown menu labeled 'Fachgebiet auswählen'.
- Leitprobleme:** A dropdown menu labeled 'Leitproblem nachschlagen'.
- Diagnosen:** A dropdown menu labeled 'ICD 10 nachschlagen' with a checked box for 'inkl. Verdachtsdiagnosen'.
- Zielgruppen:** A dropdown menu.
- Falltitel:** A text input field.
- Lernziele:** A text input field.
- Autoren:** A text input field.
- Kurzinformation:** A text input field.
- Institution:** A text input field.
- Bearbeitungsdauer:** A dropdown menu.
- Lernsystem:** A dropdown menu.
- Geändert nach:** A dropdown menu showing '1', 'Januar', and '19€'.
- Medien:** A dropdown menu.
- Buttons: 'RESET', 'Hilfe und Suchtipps', 'Systemvoraussetzungen', and 'SUCHEN'.

Abbildung 1: Die Profisuche in CASEPORT

Die Benutzer konnten ihre Zustimmung zu folgenden Aussagen auf der Skala von 1 (=trifft voll zu) bis 5(=trifft nicht zu) angeben (Abbildung 2, Tabelle 1).

Tabelle 1: Bewertung der CASEPORT-Suche

| Aussage | Mittelwert |
|--|------------|
| Die Ergebnisanzeige der Fallsuche ist übersichtlich. | 2,5 |
| Die Profi-Fallsuche ist zu komplex. | 3,2 |
| Die Suchresultate erscheinen ausreichend schnell. | 2,7 |

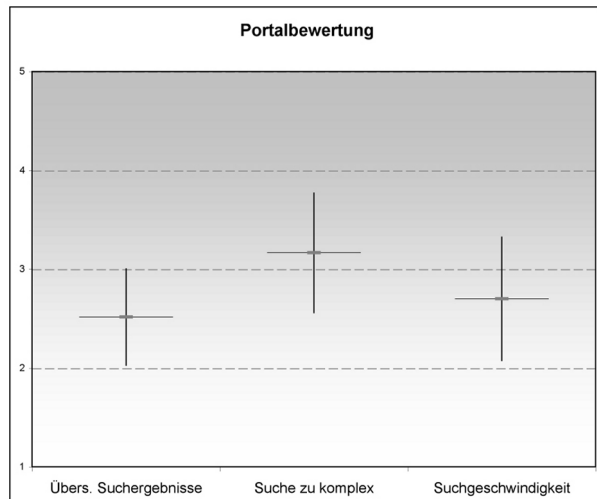


Abbildung 2: Bewertung der CASEPORT-Suche (1=trifft voll zu, 5=trifft nicht zu; vertikale Balken geben die Standardabweichung an)

Hier wird deutlich dass allein die Anzahl der möglichen Suchparameter, trotz übersichtlicher Anordnung, zu einer Überforderung bei vielen Suchenden geführt hat. Als mögliche Lösung könnte man einzelne Suchfelder standardmäßig ausblenden und nur auf explizite Anforderung des Benutzers einblenden. Eine andere Lösung wäre es, die Fachgebiete direkt in einem Menu aufzuführen. In CASEPORT wurde das Problem dadurch entschärft, dass neben der Profisuche noch die Standardsuche zur Verfügung steht mit der die Fälle nach Fachgebieten durchsucht werden können.

• Volltextsuche: Probleme und Lösungsansätze

In diesem Abschnitt sollen anhand der Erfahrungen bei der Implementierung der Volltextsuche im IMSm-System generelle Probleme bei der Volltextsuche behandelt und mögliche Lösungen aufgezeigt werden. Mit Hilfe der Volltextsuche sollen Fragen anhand des Vorkommens bestimmter Wörter oder Phrasen an definierten Stellen der Frage gefunden werden. Eine Volltextsuche ermöglicht oft ein schnelles Auffinden der gewünschten Information, kann jedoch auch ungewünschte Treffer anzeigen (zu geringe Spezifität) bzw. Informationen nicht anzeigen, die für den Suchenden von Interesse wären (zu geringe Sensitivität).

Da die Volltextsuche im Gegensatz zur Suche in einem Thesaurus wort- und nicht begriffsbasiert ist, entstehen eine Reihe von Problemen, die im Folgenden genannt und für die Lösungsansätze vorgestellt werden sollen.

So werden z.B. synonyme Bezeichnungen nicht adäquat in die Suche einbezogen und unterschiedliche Flexionsformen oder Schreibweisen führen zu unterschiedlichen Suchresultaten.

Um dennoch möglichst viele, für den Suchenden relevante Treffer zu erhalten, können folgende Verfahren verwendet werden.

Synonyme und Synonymlisten

Problem: Eventuelle synonyme Bezeichnungen eines Suchwortes werden nicht gefunden.

Lösung: Erstellung von Synonymlisten: Jeder Bezeichnung wird eine Vorzugsbenennung zugewiesen. Diese Beziehung kann für die Suche so aufgelöst werden, dass automatisch auch nach synonymen Bezeichnungen gesucht wird. Diese Funktion wurde für das IMSm bislang nicht implementiert. Mit der Implementierung dieser Funktion würde sich die Volltextsuche ein Stück auf die Suche in kontrollierten Terminologien zubewegen. Wegen des Aufwands bei der Erstellung von Synonymlisten wäre dann jedoch die direkte Verwendung des MeSH und seiner Entry-Terms vorzuziehen.

Flexionsformen und Stemming-Algorithmen

Problem: Andere Flexionsformen (Singular/Plural, Deklinationsformen eines Verbs) des Suchbegriffs werden nicht gefunden.

Lösung: Durch die Verwendung geeigneter „Stemming-Algorithmen“ wird ein Wort, sowohl bei der Indexierung als auch bei der Suche, auf seinen Wortstamm reduziert. Stemming-Algorithmen sind sprachspezifisch. In Lucene kommt eine leicht abgewandelte Fassung des von Jörg Caumanns in [14] vorgestellten Algorithmus zum Einsatz. In einer von uns durchgeführten Studie konnte gezeigt werden, dass das Stemming die Retrievalrate um mindestens einen halben Prozentpunkt erhöht. Als einfacher Ersatz für das Stemming bietet sich für die Abstraktion von Flexionsformen auch die Wildcardsuche an (siehe Abschnitt fehlertolerante Suche), die jedoch im Deutschen nicht immer zum Ziel führt (Haus <-> Häuser).

Unterschiedliche Schreibweisen und fehlertolerante Suche

Problem: Unterschiedliche oder falsche Schreibweisen des Suchbegriffs werden nicht gefunden (z.B. Appendizitis oder Appendicitis)

Lösung: Einen guten Überblick zum Thema gibt unter anderem die Webseite [15]. Verschiedene Verfahren, wie z.B. der Levenshtein-Algorithmus, auch als Edit-Distanz bezeichnet, erlauben es, auch ähnlich geschriebene Bezeichnungen mit in die Suche aufzunehmen. Bei diesem Verfahren wird die Ähnlichkeit des Suchwortes mit allen Wörtern im Index berechnet, so dass ähnliche Wörter ab einem gewissen Grad an Übereinstimmung ebenfalls als Treffer im Suchergebnis mit

ausgegeben werden können. Allerdings ist die Berechnung des Levenshtein-Abstandes aufwändig, da das Suchwort mit allen anderen Wörtern im Corpus verglichen werden muss. Um in Lucene nach ähnlichen Schreibweisen zu suchen, kann der Suchende eine Tilde (~) hinter das Suchwort setzen. Damit wird z.B. bei der Suche nach Appendizitis~ auch Appendicitis gefunden.

Zudem kann der Benutzer in Lucene durch die Verwendung eines Sterns (*) das System anweisen, alle Wörter, die mit dem Suchstring beginnen als Ergebnis anzuzeigen (z.B. Sinusk* findet alle Wörter, die mit 'Sinusk' beginnen: Sinuskurve, Sinusknoten, etc.).

Nicht in Lucene implementiert ist hingegen die Suche nach phonetisch ähnlichen Begriffen mittels des Soundex- bzw. Double-Metaphone-Verfahrens [16]. Ebenso wird keine orthographische Normierung vorgenommen, wie dies beispielsweise von [17] beschrieben wurde (z.B. y und ie werden zu ie, c und k werden zu c). Diese Art der Normierung ist besonders in der medizinischen Fachsprache erfolgreich.

Stoppwörter und Stoppwortlisten

Problem: Es gibt in jeder Sprache einige Wörter, die sehr häufig auftreten, aber wenig bzw. keine relevante Information enthalten, die so genannten Stoppwörter, z.B. Artikel und Konjunktionen. Die Suche nach einem Stoppwort würde eine sehr große Treffermenge generieren.

Lösung: Bei der Indizierung wird eine Stoppwortliste benutzt, so dass häufig vorkommende Wörter nicht in den Index aufgenommen werden. Für den medizinischen Kontext empfiehlt sich eine spezifische Anpassung der Stoppwortliste: Hier sind beispielsweise die Begriffe Krankheit und Syndrom häufig und unspezifisch.

Benutzerdefinierte Einstellungen der Suche für fortgeschrittene Benutzer

Da die Lucene-Suche per Default ähnlich wie Google nach allen eingegebenen Wörtern unabhängig von ihrem Auftreten im Text sucht, wird es dem Benutzer ermöglicht eine Phrase zu suchen, indem er sie in doppelte Hochkommata (") einschließt. So findet z.B. der Suchtext "malignes Melanom" nur Items in denen beide Wörter unmittelbar hintereinander stehen.

Um nach Wörtern zu suchen, die innerhalb eines Items nur nahe beieinander stehen, setzt man diese in doppelte Hochkommata und ergänzt eine Tilde und die Anzahl der Wörter, die maximal zwischen ihnen stehen darf.

So findet z.B. "Dosis lethal"~10 nur Items in denen maximal 10 Wörter zwischen "Dosis" und "lethal" stehen.

Im Falle strukturierter Daten, wie sie beispielsweise bei den Prüfungsfragen im IMSm vorliegen, ist es sinnvoll, den Suchbereich z.B. auf den Fragentext oder die Antworten einschränken zu können. Um nur in gewissen Feldern zu suchen, kann in Lucene der Feldname gefolgt von einem Doppelpunkt vor das Suchwort gesetzt werden (z.B. Fragentext:Laryngoskopie). Gültige Feldnamen im IMSm sind: Kontext, Kurzttext, Vignette, Fragentext, Kommentar, RichtigeAntworten, FalscheAntworten.

Diskussion und Fazit

Bezüglich des Einsatzes kontrollierter Vokabulare ist festzustellen, dass diese von den Benutzern nur akzeptiert werden, wenn sie einfach zu bedienen sind. Die CASEPORT-Profisuche zeigt, dass auch eine gut strukturierte Suchmaske die Benutzer überfordern kann.

Es gibt keine Suchstrategie, die für jede Form von Lerninhalten exakt passt. Während es in Lernfällen, die sich der Student selbst erschließen soll, nicht immer sinnvoll ist eine Volltextsuche anzubieten, um eine „zufällige“ Preisgabe der Diagnose zu vermeiden, ist diese Methode zum Beispiel für ein Lehrbuch dringend angezeigt. Ebenso kann auch eine Berücksichtigung der Rolle des Suchenden im System sinnvoll sein: So werden in CASEPORT für den suchenden Dozenten Differenzialdiagnosen von der finalen Diagnose unterschieden, wohingegen für den Studenten alle Diagnosen gleich dargestellt werden.

Letztlich ist eine genaue Analyse des Problems für die Suchstrategie entscheidend und außerdem sollte die Akzeptanz durch die Suchenden bezüglich der Benutzerschnittstelle, der Suchdauer und der Adjustierung von Spezifität und Sensitivität frühzeitig getestet werden.

Korrespondenzadresse:

• Matthias Holzer, Schwerpunkt Medizindidaktik, Med. Klinik Innenstadt, Klinikum der Ludwig-Maximilian-Universität München, Ziemssenstr. 1, 80336 München
Matthias.Holzer@lrz.uni-muenchen.de

Literatur:

[1] MeSH Thesaurus [homepage on the Internet]. National Library of Medicine [updated 2006 Apr 10; cited 2006 Jun 30]. Available from: <http://www.nlm.nih.gov/mesh/>.

- [2] MEDLINE Database [homepage on the Internet]. National Library of Medicine [updated 2006 Apr 10; cited 2006 Jun 30] Available from: <http://www.ncbi.nlm.nih.gov/pubmed/>
- [3] DIMDI [homepage on the Internet]. Deutsches Institut für Medizinische Dokumentation und Information [updated 2006 Jun 29; cited 2006 Jun 30] .Available from: <http://www.dimdi.de/>.
- [4] SNOMED [homepage on the Internet]. Systematic Nomenclature of Medicine [updated 2006 Apr; cited 2006 Jun 30]. Available from: <http://www.snomed.org/>.
- [5] Rothwell DJ, Hause LL. SNOMED and microcomputers in anatomic pathology. *Med Inf (Lond)*. 1983;8:23-31.
- [6] Wingert F. SNOMED- Systematische Nomenklatur der Medizin. Berlin: Springer Verlag; 1984.
- [7] UMLS [homepage on the Internet]. Unified Medical Language System [updated 2006 May 01; cited 2006 Jun 30]. Available from: <http://www.nlm.nih.gov/research/umls/>.
- [8] Fischer MR, CASEPORT-Konsortium. CASEPORT: Systemintegrierendes Portal für die fallbasierte Lehre in der Medizin. Proceedings des 6. Workshop der GMS-Arbeitsgruppe Computerunterstützte Lehr- und Lernsysteme in der Medizin.
- [9] Lernziele.net [homepage on the Internet]. Lernzieldatenbank der LMU [updated 2006 May 14; cited 2006 Jun 30]. Available from: <http://www.lernziele.net/>.
- [10] Hamburger Lernzielkatalog [homepage on the Internet]. Hamburger Lernzielkatalog [updated 2003 Nov 18; cited 2006 Jun 30]. Available from: <http://www.uke.uni-hamburg.de/studierende/.../downloads/zg-studierende/.../Hamburger-Lernzielkatalog.pdf>.
- [11] Morphosaurus [homepage on the Internet] Morphosaurus Projekt [updated 2005 Jun; cited 2006 Sep 14]. Available from: <http://www.morphosaurus.de/>.
- [12] IMS-m.org [homepage on the Internet]. Item Management System für die Medizin [updated 2003 Jun 30; cited 2006 Jun 30]. Available from: <http://ims-m.org/>.
- [13] Apache Lucene [homepage on the Internet]. Apache Software Foundation [updated 2006 Jun 19; cited 2006 Jun 30]. Available from: <http://lucene.apache.org>.
- [14] Caumanns J. A Fast and Simple Stemming Algorithm for German Words [homepage on the Internet] [updated 1999 Oct; cited 2006 Sep 14]. Available from: <http://www.inf.fu-berlin.de/inst/pubs/.../tr-b-99-16.abstract.html>.
- [15] Phonetic Searches with Double Metaphone and Alternatives [homepage on the Internet]. The Code Project [updated 2003 Jun 30; cited 2006 Jun 30]. Available from: <http://www.codeproject.com/string/dmetaphone6.asp>.
- [16] The Double Metaphone Search Algorithm [homepage on the Internet]. Lawrence Philips [updated 2003 Apr 15; cited 2006 Jun 30]. Available from: <http://www.ddj.com/184401251?pgno=2>.
- [17] Briegl B. Eine Methode zur lexikabasierten Indexierung von Diagnosen (LBI-Methode). Bericht Nr. 1/1992 des Instituts für Med. Biometrie und Informatik der Universität Heidelberg.