

Andreas Holzinger, Klaus-Martin Simonic, Johannes Steyrer

## Information Overload – Stößt die Medizin an ihre Grenzen?

### Kurz gefasst:

- Medizinische Dokumentation liegt – obwohl immer von einem „Multimedia-Zeitalter“ gesprochen wird – meistens in Form von textueller Information vor.
- Dies wird umgangssprachlich und fälschlicherweise Freitext genannt, formal handelt es sich um eine nicht standardisierte Informationsdarstellung.
- Eine Unterstützung durch Expertensysteme ohne vorhergehende systematische Aufarbeitung dieser Dokumente ist praktisch nicht möglich.

**Moderne Informationstechnologie ermöglicht raschen Zugriff auf immer größere Datenmengen. Mehr Daten heißt aber nicht mehr Information und schon gar nicht mehr Wissen. Der heutige Informationsbegriff ist geprägt von den allgegenwärtigen und stets verfügbaren Massenmedien und wandelt sich mehr und mehr zum Synonym für die theoretische Möglichkeit einer allumfassenden Informiertheit. Während die technische Performanz rapide steigt, stößt die kognitive „Performance“ der End-Benutzer an ihre Grenzen. Beispielhaft sei dies an Hand der medizinischen Dokumentation verdeutlicht: Eine elektronische Patientenakte kann mehrere hundert Einzeldokumente enthalten. Informationssysteme bringen diese Informationen auf Knopfdruck an den medizinischen Arbeitsplatz. Doch dort bleibt für die Entscheidungsfindung nur wenig Zeit. Rund fünf Minuten sind es im Durchschnitt [1]. Unter diesen engen zeitlichen Rahmenbedingungen wird das Erfassen der relevanten Information selbst zum kritischen Faktor.**

### **Einige (sprachliche) Charakteristika medizinischer Dokumentation**

Die übliche Praxis medizinischer Dokumentation weist einige zentrale Merkmale auf:

1. Mediziner kombinieren grundsätzlich verschiedene Sprachen innerhalb eines Schriftstücks, insbesondere Latein, Deutsch und Englisch. Rein formal muss daher von einem multilingualen Ansatz ausgegangen werden.
2. Die Anzahl der computerlinguistisch verwendbaren größeren, deutschsprachigen Korpora der medizinischen Sprache ist beschränkt [2, 3] und eine entsprechende Entwicklung ist unter einigen Tausend Personenjahren nicht realisierbar.
3. Der Großteil der Krankenakten wird nie auf sprachliche Schwächen hin überprüft. Orthografische Mängel sind in der klinischen Routine daher an der Tagesordnung, grammatikalische nicht selten und sinnzerstörende „Verschreiber“ kommen durchaus vor. Dies ist nicht verwunderlich, da – aufgrund der mehrheitlich internen Verwendung der Aufzeichnungen – der Dokumentcharakter in der täglichen Routine in den Hintergrund tritt. Solange der Text noch für die Kollegin/den Kollegen verständlich ist, gibt es keinen zwingenden Grund sich mit der Sprache

auseinanderzusetzen.

4. Abkürzungen finden häufig Verwendung: Die medizinische Sprache ist eine Fachsprache und verfügt über ein eigenes Vokabular. Doch diese sprachlichen Vereinbarungen respektive Gepflogenheiten unterliegen starken lokalen Variationen und die Dialekte werden zudem, vor allem in der internen Dokumentation, gerne verkürzt geschrieben. So können ganze Befunde wie schwer lesbare Telegramme aussehen. Eine typische Formulierung aus der Routinebefundung lautet etwa: „Ca. 2x1cm große, ovaläre Verschattung im UF li. Rez. re. lat. frei, li. teiladhärent.“ Solche Schreibweisen sind für geschulte Mediziner fraglos interpretierbar, die maschinelle Handhabung derartiger Texte erweist sich jedoch als extrem problematisch. Während die synonyme Verwendung von Abkürzungen noch als relativ unkritisch klassiert werden kann, sind Homonyme (die identische Bezeichnung unterschiedlicher Begriffe) bei medizinischen und/oder allgemein gebräuchlichen Worten und Abkürzungen nur im kontextuellen Zusammenhang halbwegs sicher auflösbar. In Kombination mit einer inkonsistenten Verwendung von Satzzeichen oder alternierenden Schreibweisen kann die Situation schnell zum Klassifikationsalpträum werden. Dazu ein Beispiel: ca. (circa), Ca. (Karzinom) und Ca (Kalzium) sind homonyme Abkürzungen zwischen Alltagssprache und medizinischer Fachsprache. Durch einen simplen Tippfehler kann aus dem Satz: „Geschwulst, ca. 3 cm groß“ der Wortlaut: „Geschwulst, ca, 3 cm groß“ und inhaltlich aus einer kleinen Schwellung (z.B. durch einen Bienenstich) ein potentiell tödliches Karzinom werden. Dementsprechend versagen Analysewerkzeuge, die auf lexikalischen und einfachen semantischen Vergleichen basieren, bei klinischen Informationsobjekten.
5. In der Medizin werden von automatisierten Prozessen sehr hohe und stabile Erkennungsraten verlangt (>95%). Das obige Beispiel verständlicht, warum Mediziner ein gehöriges Maß an Skepsis gegenüber informationstechnischen Verfahren haben, insbesondere wenn diese für den Benutzer wenig kontrollierbar sind.

### **Standardmethoden der Informationsextraktion**

Der entscheidende Schritt, um medizinische Befundtexte inhaltlich zu erschließen, ist die Extraktion der darin enthaltenen medizinischen Aussagen. Diese Fakten können auf ein geeignetes Begriffssystem abgebildet und mit medizinischen Codes kombiniert werden. Nach dieser semantischen Informationsextraktion wird in einem weiteren informationstechnischen Schritt eine entsprechende Präsentationsform (eine Informations-Präsentation) für die Interaktion zwischen Ärzten und System entwickelt, also zwischen den End-Benutzern und der Benutzerschnittstelle (Informations-Interaktion).

Ein möglicher Lösungsansatz dafür ist die Codierung beziehungsweise Beschlagwortung der in den nicht standardisierten Texten (Freitexten) enthaltenen Information. Im Gegensatz zur Alltagssprache existieren für die medizinische Fachsprache eine Vielzahl von kontrollierten Vokabularen, Klassifikationen und Nomenklaturen. In diesem Bereich hat die Medizin eine lange Tradition, teilweise mit weltweit akkordierten Schemata [4]. Probleme ergeben sich aus relativierenden und mehrdeutigen Ausdrucksformen sowie der Verwendung von nicht vereinheitlichten Abkürzungen [5].

Eine Analyse medizinischer Texte kann folglich, auch bei syntaktisch und orthographisch korrekten Inhalten, nicht auf das einfache klassische Informationsretrieval mittels Suche nach Worten oder Wortbestandteilen beschränkt bleiben. Neben der medizinischen Bedeutung des jeweiligen Begriffs selbst ist der Kontext des „umgebenden“ Informationsobjekts, eventuell des ganzen Netzwerks von Informationsobjekten, mit einzubeziehen. Erst dadurch kann eine semantische Suche optimal unterstützt werden [6].

Im Wesentlichen kommen Algorithmen aus den Gebieten der Verarbeitung natürlicher Sprache (Natural Language Processing als Teil der Computerlinguistik), des Informationsretrieval und des maschinellen Lernens zur Anwendung. Die informationstechnischen Strukturen reichen von einfachen Listen wie CSV-Dateien (beispielsweise als Input für maschinelles Lernen) über relationale Datenbanken bis hin zu Topic Maps für abstrakte Modelle zur Formulierung von Wissensstrukturen.

Bei der Analyse von standardisierten Daten kommen die klassischen Methoden des Datamining zum Einsatz. Als Datamining werden Prozesse zur „Knowledge Discovery in Databases“ bezeichnet, die eng verwandt sind mit maschinellem Lernen. Eine Überprüfung der Datenqualität und Datenplausibilität ist sinnvoll, da Datenbestände – auch wenn sie bereits in strukturierter Form vorliegen – unvollständig bzw. nicht nachvollziehbar sein können. Ist die Qualität der Daten überprüft, plausibel und auf das gewünschte Schema abgebildet, werden die nicht standardisierten Daten verarbeitet. Zur Informationsextraktion werden die üblichen Schritte des Textminings vollzogen. Im Gegensatz zum Datamining handelt es sich nun um ein „Knowledge Discovery in Textual Information“, oder kurz: Textual Data Mining, welche die Entdeckung von Bedeutungsstrukturen aus nichtstrukturierten oder schwachstrukturierten Texten mit Hilfe von statistischen Methoden ermöglicht [7].

### **Fazit:**

Die Qualität des ärztlichen Handelns hängt wesentlich von der verfügbaren relevanten Information ab. Diese zu finden wird zur Schlüsselaktivität in medizinischen Entscheidungsprozessen. Moderne Informationssysteme können helfen, Recall und Precision der Befund- und Informationsrecherche zu verbessern [8].

### **Literatur:**

- [1] Gigerenzer, G. (2008) Gut Feelings: Short Cuts to Better Decision Making London, Penguin.
- [2] Hahn, U. & Wermter, J. (2004) Tagging medical documents with high accuracy. Pricai 2004: Trends in Artificial Intelligence, Proceedings. Berlin, Springer-Verlag Berlin, 852-861.
- [3] Wermter, J. & Hahn, U. (2005) Massive biomedical term discovery. Discovery Science, Proceedings. Berlin, Springer-Verlag Berlin, 281-293.
- [4] Klar, R. & Graubner, B. (1997) Medizinische Dokumentation. In: Seelos, H.-J. (Ed.) Medizinische Informatik, Biometrie und Epidemiologie. Berlin, New York, de Gruyter, 13-42.
- [5] Holzinger, A., Geierhofer, R. & Errath, M. (2007) Semantische Informationsextraktion in medizinischen Informationssystemen. Informatik Spektrum, 30, 2, 69-78.
- [6] Holzinger, A., Geierhofer, R., Modritscher, F. & Tatzl, R. (2008) Semantic Information in Medical Information Systems: Utilization of Text Mining Techniques to Analyze Medical Diagnoses. Journal of Universal Computer Science, 14, 22, 3781-3795.
- [7] Hippner, H. & Rentzmann, R. (2006) Text Mining. Informatik-Spektrum, 29, 4, 287-290.
- [8] Lenz, R., Beyer, M., Meiler, C., Jablonski, S. & Kuhn, K. (2005) Informationsintegration in

### Die Autoren:

Andreas Holzinger begann seine Karriere als Lehrling in der Informationstechnik und studierte erst nach zweitem Bildungsweg und Industrietätigkeit Nachrichtentechnik, Physik/Psychologie sowie Medienpädagogik/Soziologie. Er promovierte mit einem Thema aus der Kognitionswissenschaft und habilitierte sich in Angewandter Informatik an der Technischen Universität Graz. Er war Gastprofessor in Berlin, Innsbruck, Wien und London. Derzeit leitet er die Forschungseinheit Human-Computer Interaction am Institut für Medizinische Informatik der Medizinischen Universität Graz und ist als Experte für die EU tätig. Seine Arbeitsgebiete umfassen u.a. Informations- und Wissensmanagement und Multimediale Informationssysteme.

[holzinger@wissensmanagement.net](mailto:holzinger@wissensmanagement.net)

Klaus-Martin Simonic studierte Elektrotechnik und Technische Mathematik/Informatik in Graz und promovierte in Medizinischer Informatik an der TU Graz. Er ist stellvertretender Institutsvorstand des Instituts für Medizinische Informatik, Statistik und Dokumentation und Datenschutzbeauftragter der Medizinischen Universität Graz, sowie technischer Direktor des International Forum for Telesurgical Research and Education (IFTRE). Seine Arbeitsschwerpunkte umfassen Medizinische Informationssysteme, Qualitäts- und Wissensmanagement in der Medizin und Telemedizin.

[simonic@wissensmanagement.net](mailto:simonic@wissensmanagement.net)

Johannes Steyrer studierte Soziologie und Betriebswirtschaftslehre in Wien. Nach dem Studium war er mehrere Jahre in der Unternehmensberatung in den Bereichen Strategie- und Organisationsentwicklung tätig. Er ist Psychotherapeut (individualpsychologischer Analytiker), Management-Coach und -Trainer in Kommunikation, Personalführung und Teamentwicklung. Seit 1997 ist Herr Steyrer ao. Univ. Prof. an der Wirtschaftsuniversität Wien mit den Schwerpunkten Personalführung und Karriereentwicklung von Managern. Von 2002 bis 2004 leitete er den „Interdisziplinären Universitätslehrgang für Sozialwirtschaft, Management und Organisation Sozialer Dienste“ (ISMOS). Seit 2005 ist er Leiter des „MBA Studiums Healthcare Management“ an der WU Wien.

[steyrer@wissensmanagement.net](mailto:steyrer@wissensmanagement.net)