

INNOVATIVE NUTZUNG UND BEREITSTELLUNG MEDIZINISCHER INTERNETINHALTE

Denecke K¹

Kurzfassung

Im Internet finden sich eine Vielzahl von Informationsangeboten zu den Themen Gesundheit, Medizin und Wellness. Dazu gehören neben Internetseiten medizinische Weblogs, Frageportale, Foren etc. In diesem Beitrag geben wir einen Überblick über Methoden, die entwickelt wurden, um medizinische Internetinhalte automatisch zu verarbeiten und deren innovative Nutzung zu ermöglichen. Mit diesen Methoden können u.a. Meinungstrends oder inhaltliche Trends festgestellt werden. Weiterhin werden potentielle Anwendungen dieser Methoden beschrieben.

1. Einleitung

In Weblogs, Frageportalen, Foren (sog. Social Media Daten) und auf Internetseiten finden sich eine Vielzahl von Informationen zu den Themen Gesundheit, Medizin und Wellness. Diese Inhalte und die entsprechenden Web-Portale gewinnen als Informationsquelle nicht nur für den einzelnen Nutzer (Patienten, Ärzte, usw.), sondern auch für Krankenkassen, Pharmakonzerne u.a.m. zunehmend an Gewicht. Heutzutage sind Konsumenten längst keine passiven Rezipienten mehr. Konsumentenmeinungen oder Gedanken zu bestimmten Themen werden schnell und unkontrolliert ausgetauscht und oftmals im Internet öffentlich verfügbar gemacht. Für Unternehmen, Krankenkassen, Krankenhäuser und andere Einrichtungen ist es relevant, diese Meinungen zu analysieren, sie in den Ursprüngen zu erkennen, um ggf. durch Marketingstrategien oder gezielter Information darauf zu reagieren. Auf diese Weise können Markenbotschaften zielgerichteter, entsprechend der aktuellen "Gemütslage" von Konsumenten platziert werden. Die Marketingkommunikation durch das Internet als Medium lässt sich nicht nur wirkungsvoller und planbarer gestalten, sondern es lässt sich zugleich auch eine schnellere Durchdringung von Marketingbotschaften erzielen. Trends in Meinungen und Themen werden aufgespürt und Kampagnen gezielt entwickelt. Außerdem können Informationen bereitgestellt werden, die dem aktuellen Informationsbedarf angepasst sind.

Die enorme und ständig zunehmende Datenmenge, die im Internet allgemein und in medizinischen Social Media Plattformen im Speziellen zu finden ist, macht eine manuelle Verarbeitung und Analyse der geäußerten Meinungen und bereitgestellten Inhalte unrealisierbar. Da automatische Verarbeitungsmethoden fehlen, bleiben diese Daten bisher ungenutzt. Bestehende Methoden zur Analyse medizinischer Dokumente wurden für die Verarbeitung klinischer Dokumente entwickelt und werden überwiegend zur Unterstützung der automatischen Kodierung eingesetzt (z.B. die Systeme MedLEE [5] und MedIE [6]). Klinische Dokumente und medizinische Social Media Daten unterscheiden sich allerdings sprachlich und inhaltlich. Während in klinischen Dokumenten vor allem

¹ Forschungszentrum L3S, Hannover

die Behandlung von Patienten in Fachsprache dokumentiert wird, finden sich in Social Media Dokumenten Erfahrungen und Informationen zu Krankheiten und Behandlungen, die in Alltagssprache beschrieben werden.

In diesem Beitrag werden Methoden vorgestellt, die konkret zur Verarbeitung von medizinischen Social Media Daten entwickelt wurden und die dazu genutzt werden können, relevante medizinische Inhalte in diesen Dokumenten zu analysieren und aufzubereiten. Dazu gehört die quantitative Analyse von Meinungen (z.B. wie viele Einträge enthalten negative Meinungen zu dem Arzneimittel?), die Ermittlung von Meinungstrends (z.B. Wie haben sich die Meinungen im Laufe des Jahres verändert?), die Identifizierung von Meinungsführern und Experten oder die Bestimmung von aktuellen thematischen Schwerpunkten in diesen Plattformen. Weiterhin werden wir potentielle Anwendungsszenarien für die Nutzung von Social Media Daten beschreiben.

2. Medizinische Social Media Daten

Verschiedene Quellen zu medizinischen bzw. gesundheitsrelevanten Themen sind im Internet verfügbar. Zum einen werden von Krankenkassen, Krankenhäusern, Gesundheitseinrichtungen und anderen Anbietern auf Webseiten Informationen zu medizinisch- bzw. gesundheitsrelevanten Themen bereitgestellt. Zum anderen sind im Rahmen der Entwicklungen von Web 2.0 Technologien Werkzeuge entstanden, die es ermöglichen, Meinungen und Erfahrungen auszutauschen. Dies geschieht über sog. Social Media Plattformen wie Foren, Frageportale, Expertenforen oder Weblogs. Im Bereich der Medizin äußern sich Mediziner, Pflegepersonal, aber auch Patienten in diesen Plattformen zu den unterschiedlichsten medizinischen Themen. Dabei werden Meinungen geäußert und zahlreiche Informationen bereitgestellt. Weblogs sind Onlinetagebücher, deren Autoren ihre Gedanken, Meinungen und zu bestimmten Themen, aber auch sehr allgemeine Informationen im Internet bereitstellen (z.B. <http://www.mein-diabetes-blog.de>). Institutionen nutzen Weblogs inzwischen auch, um aktuelle Informationen zu bestimmten Themen bereitzustellen (z.B. <http://www.gesundheits-weblog.de>). In Foren finden Diskussionen zu bestimmten Themen statt. Frageportale wie <http://www.rund-ums-baby.de> ermöglichen es Nutzern (Patienten), ihre Fragen zu Krankheiten, Behandlungen, Symptomen oder anderen (medizinischen) Themen zu stellen. In einigen Portalen dürfen ausschließlich Experten die Fragen beantworten. Andere Portale erlauben allen Nutzern das Beantworten von Fragen (z.B. Yahoo! Answers). Da Meinungen vorwiegend in Foren, Blogs und Frageportalen ausgetauscht werden, liegt der Schwerpunkt unserer Arbeit auf der Analyse von medizinischen Online-Daten aus solchen Plattformen.

3. Methoden

Um die Inhalte von Social Media Daten und Internetseiten zu analysieren und relevante Informationen zu Themen und Meinungen identifizieren zu können, sind Methoden des Textmining notwendig. Das Textmining zielt darauf ab, bisher unbekannt Zusammenhänge aus Inhalten von Textdokumenten zu identifizieren [4]. Dazu werden Methoden aus dem Information Retrieval, computerlinguistische Verfahren wie die Wortartenerkennung oder der gezielten Extraktion von Informationen sowie statistische Verfahren zur Auswertung gewonnener Daten genutzt. In diesem Kapitel werden die Methoden vorgestellt, die wir bisher entwickelt haben, um Internetinhalte zu analysieren. Entsprechende Anwendungen, die die Methoden nutzen, werden im folgenden Kapitel beschrieben.

Abbildung 1 skizziert einen möglichen Ablauf einer Analyse von Social Media Daten. Da gerade in Weblog-Einträgen sehr verschiedene, nicht zwingend medizinische Themen zu finden sind, ist es nötig, zunächst solche Einträge zu ermitteln, die tatsächlich medizinische bzw. gesundheitsrelevante Themen diskutieren. Die relevanten Dokumente können dann analysiert werden, z.B. kann das Thema eines Beitrages bestimmt werden oder die zu einzelnen Themen geäußerten Meinungen. Abschließend werden die Ergebnisse so aufbereitet, dass sie durch Menschen interpretiert oder in anderen Anwendungen genutzt werden können.

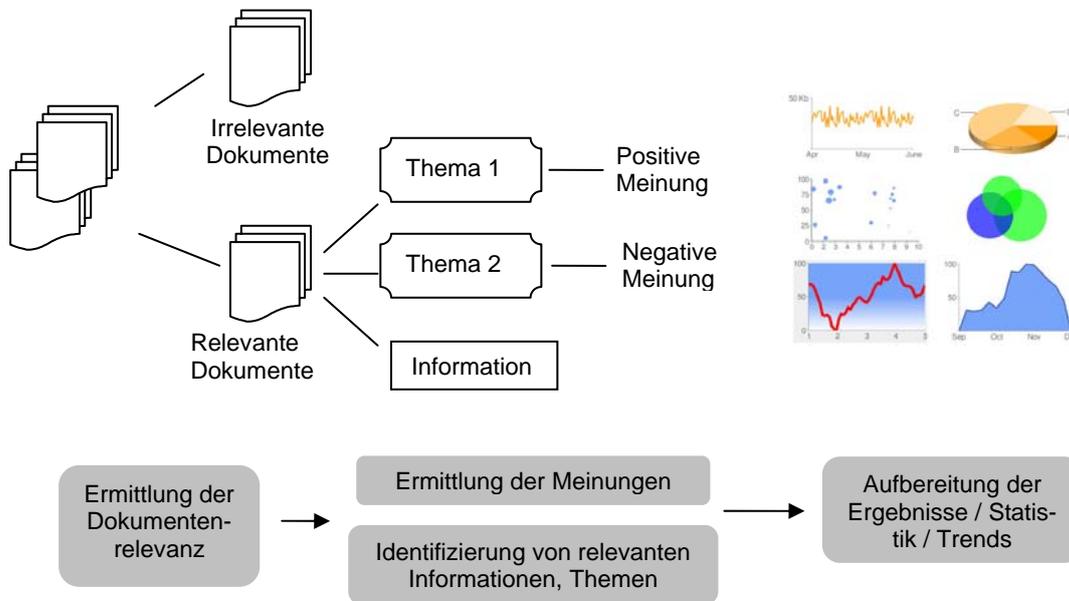


Abbildung 1: Möglicher Ablauf der Analyse und Auswertung von Webinhalten

Eine Voraussetzung für die maschinelle Auswertung und Verarbeitung natürlichsprachiger Dokumente, sind Methoden zur semantischen Repräsentation. Die Autorin entwickelte ein solches Verfahren speziell für die Verarbeitung von klinischen Dokumenten [1]. Dabei werden wesentliche Inhalte eines Dokumentes in eine maschinenverständliche, aber auch für Menschen lesbare Form überführt. Diese semantische Repräsentation von Dokumenten ermöglicht erst eine automatische Weiterverarbeitung, z.B. eine gezielte Extraktion von Inhalten wie Angaben zu Medikamenten oder zu Diagnosen. Damit ist es beispielsweise möglich, automatisch das Thema eines Dokuments zu charakterisieren und thematisch verwandte Dokumente bzw. Webseiteninhalte zu bestimmen.

Weiterhin steht ein Verfahren zur Klassifizierung von Webinhalten nach deren Informationsgehalt bereit [3]. Dabei wird ermittelt, ob ein Beitrag eher informativ ist oder ob eher eine Meinung geäußert wird. Das Verfahren wurde bisher erfolgreich, mit einer Korrektheit von 87%, für die Klassifizierung von Weblog-Einträgen getestet. Diese Zusatzinformation zum Informationsgehalt kann zum Ranking von Suchergebnissen genutzt werden. Wir konnten zeigen, dass beim Blog Post Retrieval die Nutzerzufriedenheit erhöht werden kann, indem informative Texte an höheren Positionen in einem Suchergebnis aufgeführt werden als affektive Texte.

Darüber hinaus entwickelten wir ein Verfahren zur Ermittlung von in Textdokumenten ausgedrückten Meinungen [2]. Auf Basis von linguistischen und lexikalischen Ressourcen, die für die Analyse englischsprachiger Dokumente bereits zur Verfügung stehen, wurde ein Verfahren entwickelt, mit dem Meinungen aus Dokumenten in verschiedenen Sprachen extrahiert werden können. Ursprünglich wurde das Verfahren für die Identifikation von Meinungen aus Filmkritiken in verschiedenen

Sprachen entwickelt. Es konnte aber auch erfolgreich dazu eingesetzt werden, Kommentare zu Medikamenten hinsichtlich der ausgedrückten Meinung zu klassifizieren.

4. Anwendungsszenarien

4.1. Anwendungen für Krankenkassen

Krankenkassen sind aufgrund des großen Konkurrenzkampfes daran interessiert, ihr Leistungsangebot an die Kunden anzupassen. Dies kann durch eine Analyse von medizinischen Social Media Daten unterstützt werden. Aus diesen Daten kann die Außenwirkung der Krankenversicherung ermittelt werden (z.B. Wer spricht wie über die Versicherung und ihre Angebote im Internet?). Meinungsführer und Experten sowie Behandlungstrends können identifiziert werden. Weiterhin wird ermöglicht, Meinungsäußerungen zur Krankenkasse allgemein, zu einzelnen Produkten oder zu Konkurrenten zu untersuchen. Auf Basis dieser Informationen kann ggf. die Marketingstrategie angepasst werden oder gezielt Information in Umlauf gebracht werden.

Identifizierte Internettrends zu Gesundheitsthemen können genutzt werden, um den „eigenen“ Versicherten qualitativ hochwertige Informationen zu diesen Themen bereit zu stellen. Dabei kann gezielt auf besondere Leistungen der Krankenkasse in diesen Bereichen eingegangen werden. Stellt man beispielsweise fest, dass aktuell viele Diskussionen und Fragen zum Thema Brustkrebs auftauchen, kann verstärkt auf den Internetseiten der Krankenkasse oder über andere Medien über dieses Thema informiert und auf Leistungen der Krankenkasse in diesem Bereich hingewiesen werden. Informationswünsche und (nicht erfüllte) Informationsbedürfnisse von Patienten bzw. Laien können damit erkannt werden und ggf. Abhilfe geschaffen werden.

4.2. Anwendungen für Pharmakonzerne

Textmining-Technologien, wie sie in diesem Beitrag vorgestellt werden, können durch Pharmaunternehmen auf verschiedene Weise genutzt werden. Dazu gehört

- die frühzeitige Erkennung von Streitfälle,
- die Identifizierung unerwünschter Nebenwirkungen,
- die Anpassung von Marketingstrategien,
- die Identifizierung von Meinungsführern und Meinungstrends in Foren / Blogs, sowie
- die Ermittlung von Experten bzw. Personen mit hoher Expertise bzgl. eines bestimmten Themas.

Streitfälle lassen sich frühzeitig erkennen, indem in Weblogs oder Foren beschriebene Nebenwirkungen und Begleiterscheinungen identifiziert und geprüft werden. Pharmaunternehmen prüfen in regelmäßigen Abständen, ob ihre Medikamente unerwünschte Nebenwirkungen zeigen. Im Internet berichten Kunden über ihre Erfahrungen mit Arzneimitteln und gerade auch über Nebenwirkungen. Diese Prüfungen können mit Ergebnissen einer Webanalyse angereichert und damit unterstützt werden. Außerdem können Trendanalysen zur Entwicklung und Anpassung eigener Marketingstrategien in Pharmaunternehmen genutzt werden.

Eine systematische Analyse der Meinungen und Erfahrungen, die von Patienten in Weblogs oder Foren hinsichtlich bestimmter Medikamente ausgedrückt werden, erlaubt Marktanalysen und Darstellungen von Trends in Hinblick auf bestimmte Arzneimittel. Kundenmeinungen zu eigenen Produkten werden transparenter. Gleichzeitig können Meinungen über Arzneimittel konkurrierender Unternehmen zum Vergleich identifiziert werden. Experten oder Meinungsführer können beobach-

tet werden und ihre Meinungen und Aussagen entsprechend beim Marketing berücksichtigt werden. Auf Basis des gewonnenen Wissens werden relevante Markenbotschaften kreiert. Ein permanentes Monitoring sowie die Just-in-time-Abfrage sich daraus entwickelnder Meinungsbilder erlauben dabei eine gezielte Steuerung des fortlaufenden Kommunikationsprozesses innerhalb von Web 2.0.

4.3. Nutzen für Informationsanbieter

Anbieter von Informationen zu Gesundheit und Medizin, wie z.B. Vitamet.de stellen auf ihren Webseiten Informationen zu verschiedenen gesundheitsrelevanten Themen bereit. Aber nicht alle Inhalte sind für jeden Informationssuchenden geeignet, da die Nutzerabsicht und sein Wissenstand sehr stark variieren können. Beispielsweise ist eine Person, die eher allgemeines Interesse an einer Krankheit hat an anderen Informationen interessiert, als ein Betroffener. Ein Arzt sucht andere Informationen als ein Patient, etc. Hier sind Methoden nötig, die es ermöglichen, den Zugriff auf die bereitgestellten Informationen zu personalisieren, d.h. einem Nutzer sollen nur für ihn passende Informationen angeboten werden.

Das folgende Beispiel illustrierte eine mögliche Nutzung der in Abschnitt 3 vorgestellten Methoden. Vitamet stellt auf der Webseite Informationen zu verschiedenen medizinischen Themen bereit. Dabei sind die Texte inhaltlich unterschiedlich stark detailliert und nicht immer in vollem Umfang für jeden Informationssuchenden geeignet. Zusätzlich werden über DocmedTV, einem Teil von Vitamet, weitere Informationen in freitextlicher Form zu verschiedenen Themen bereitgestellt, die oftmals sehr umfassend sind.

Bei der momentanen Suchfunktion auf der Vitamet-Webseite handelt es sich um eine Standard-suchmaschine, die die Inhalte der Domain durchsucht. Kenntnisstand eines Nutzers oder weitere Informationen zu dem Nutzer bleiben unberücksichtigt. Besucher dieser Webseiten bzw. Informationssuchende könnten besser von deren Inhalten profitieren, wenn die Informationsbereitstellung den jeweiligen Bedürfnissen angepasst werden könnte.

Um eine personalisierte Suche auf der Vitamet-Webseite realisieren zu können und damit das Suchergebnis individuell auf den einzelnen Nutzer abstimmen zu können, ist es nötig, die Inhalte der Webseite im Vorfeld zu analysieren und mit semantischen Informationen anzureichern (siehe *Abbildung 2*). Diese Zusatzinformationen beinhalten u.a.

- den Grad der Detailliertheit der Beschreibung (z.B. sehr detaillierte oder sehr allgemeine Beschreibung),
- den Grad des vom Informationssuchenden benötigten Interesses bzw. Hintergrundes der Suche (z.B. Betroffener, Arzt, allgemeines Interesse, Angehöriger),
- den zum Verstehen nötigen Erkenntnisstand (z.B. kein Fachwissen nötig), oder
- das allgemeine Thema des Beitrages (Krankheit, Behandlung, Arzneimittel, etc).

Die in Abschnitt 3 vorgestellten Methoden liefern bereits solche Zusatzinformationen zum Inhalt eines Dokumentes oder Webeintrages. Die mit diesen Methoden identifizierten zusätzlichen Informationen können dann von der auf Vitamet.de bereitgestellten Suchmaschine genutzt werden, um personalisiert, d.h. auf die jeweiligen Nutzer und deren Bedürfnisse angepasst, Informationen bereitzustellen. Dem Nutzer muss auf der Webseite allerdings die Möglichkeit eingeräumt werden, einige Informationen zu seinem Informationsbedürfnis bzw. seinem Kenntnisstand bzgl. des gesuch-

ten Themas einzugeben. Diese Angabe wird dann genutzt, um ein entsprechendes Suchergebnis zusammenzustellen.

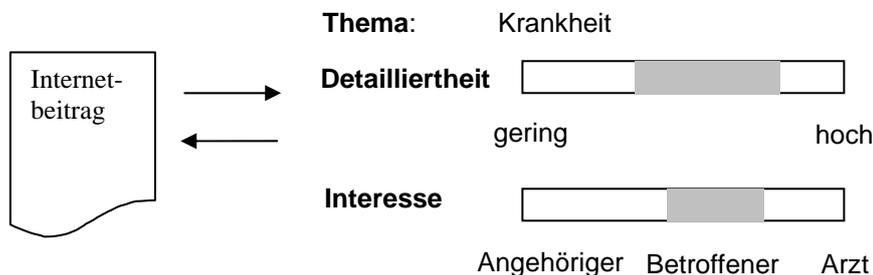


Abbildung 2: Beispiel für identifizierte Zusatzinformationen

5. Schlussfolgerungen

Die Vielzahl der vorgestellten Anwendungen zeigt, dass die Analyse medizinischer Social Media Daten durch Textmining-Technologien Informationsprozesse im Gesundheitswesen sinnvoll unterstützen kann. In diesem Beitrag wurden Methoden vorgestellt, mit denen solche Daten analysiert und aufbereitet werden können. Sie müssen noch erweitert werden, um alle beschriebenen Anwendungsszenarien realisieren zu können. Dazu gehören die automatische Klassifizierung von Dokumenten nach erforderlichem Kenntnisstand und eine stufenweise Einordnung nach Grad der Detailliertheit der Beschreibung. Insgesamt bieten Social Media Daten im Kontext Medizin eine wesentliche Informationsquelle, die unter Ausnutzung von Textmining-Technologien erschlossen werden kann und damit den Weg für eine Reihe von Anwendungen eröffnet.

6. Literaturangaben

- [1] DENECKE K: Semantic Structuring of and Information Extraction from Medical Documents Using the UMLS. *Methods of Information in Medicine*, 2008; 5(47)
- [2] DENECKE K: How to Assess Customer Opinions Beyond Language Barriers? *Proceedings of the Third IEEE International Conference on Digital Information Management*, 2008.
- [3] DENECKE K, NEJDL W: How valuable is medical social media data? *Content Analysis of the Medical Web*, Accepted for: *Journal of Information Science. Special Issue on Web Search*, 2009.
- [4] FELDMAN R, SANGER J: *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2006
- [5] FRIEDMAN, C: A broad-coverage natural language processing system. *Proc AMIA Symp 2000*; 270-274
- [6] ZHOU X, HAN H, CHANKAI I, PRESTRUD A, BROOKS A.: Approaches to text mining for clinical medical records. In: *Proceedings of the 2006 ACM Symposium on Applied Computing (Dijon, France)*. SAC '06, ACM Press, New York, April 2006. S. 235-239