

# Frühwarnungen aus dem Netz

## ZUR ÜBERWACHUNG VON KRANKHEITSAUSBRÜCHEN MITTELS WEBDATEN

Globalisierung, Klimaveränderungen und zunehmender Flugverkehr: Infektionskrankheiten breiten sich immer schneller und weiter aus und stellen die Gesundheitsorganisationen vor große Herausforderungen.

In dem EU-Projekt M-Eco wollen Forscher des Forschungszentrums L3S Internetdaten für die Medizin und Epidemiologie nutzen, um künftige Krankheitsausbrüche frühzeitig zu erkennen.



### Bekannt, aber langsam

*Ein trüber Montagmorgen im Januar. Um 9 Uhr schaltet Andrea, Epidemiologin am Niedersächsischen Landesgesundheitsamt, den Computer an und schaut sich die aktuellen Meldedaten für Infektionskrankheiten an. Angezeigt werden einige wenige neue Grippeerkrankungen sowie eine Person, bei der Salmonellen diagnostiziert wurden. Alles in allem keine kritischen Zahlen für Erkrankungen. Kein kritischer Krankheitsausbruch in Sicht.*

Die Meldung, dass ein Patient eine hochgradig ansteckende Krankheit hat, erfolgt bislang ausschließlich über die Krankenhäuser, Labore und Arztpraxen an die jeweils lokal zuständigen Gesundheitsämter. Wird eine meldepflichtige Krankheit diagnostiziert (beispielsweise Masern), sind diese Einrichtungen aufgefordert unter Einhaltung bestimmter zeitlicher Fristen, die Diagnose zu melden. Beispielsweise werden laborbestätigte Fallzahlen, Erkrankungsraten

oder Sterberaten übermittelt und stellen die Basis für die Datenanalyse dar.

Die Mitarbeiter in den Gesundheitsämtern prüfen auf Basis der eingehenden Meldungen, ob Maßnahmen zu ergreifen sind, und leiten die Daten in regelmäßigen Abständen an das Robert-Koch-Institut weiter. Von hier aus oder von der Weltgesundheitsorganisation (WHO) kann die Warnung vor einer drohenden Pandemie ausgegeben werden. Insgesamt sind also mehrere Stationen eingeschaltet, bevor eine Warnung veröffentlicht wird und Maßnahmen ergriffen werden. Das Meldesystem ist gründlich, aber langsam.

### Bedarf an Alternativen

*13 Uhr: Nach dem Mittag meldet Andrea sich im M-Eco System an. Ein Signal in höchster Warnstufe wird ihr angezeigt. Verwundert schaut sie sich die weiteren Informationen zu dem Signal an.*



Der bisherige indikatorbasierte Surveillance-Prozess muss sich immer neuen Herausforderungen stellen. Das Projekt »Medical Ecosystem« (M-Eco) setzt an dieser Stelle an und entwickelt innovative Lösungen, um frühzeitiger auf entstehende Krankheitsausbrüche reagieren zu können. Das M-Eco System wird im Rahmen eines EU-Forschungsprojektes am Forschungszentrum L3S entwickelt. Das System nutzt anstatt offizieller Meldedaten andere Informationsquellen, um Hinweise auf Krankheitsausbrüche frühzeitig zu erkennen oder zusätzliche Informationen über die Entwicklung eines Ausbruchs zu ermitteln.

Das M-Eco System bezieht sich auf das Prinzip einer ereignisbasierten Surveillance (event-based, Ereignis-Überwachung). Danach werden unstrukturierte Informationen aus unterschiedlichen Quellen wie zum Beispiel aus TV, Radio und Internet untersucht und für die Surveillance aufbereitet und ausgewertet. Im Gegensatz zu indikatorbasierten Surveil-

lance-Systemen, können ereignisbasierte Systeme seltene oder neue Krankheitsausbrüche sofort erkennen. Grund dafür ist, dass andere Datenquellen genutzt werden, die nicht auf den klassischen Meldeweg angewiesen sind und damit früher Informationen zu Krankheitsausbrüchen zur Verfügung stellen können. In Ländern, in denen kein indikatorbasiertes Surveillance-System verfügbar ist, sind ereignis-

*Hautausschlag zu Hause bleiben mussten. Es wird spekuliert, dass sie Masern haben.*

Viele Menschen nutzen das Internet als Informations- und Kommunikationsmittel für Gesundheitsfragen und geben Auskünfte über den eigenen Gesundheitszustand. Online-medien, Weblogs, wissenschaftliche und nicht-wissenschaftliche Diskussionsforen sowie elektronische Kommu-

*benannte Informationen extrahiert, wie beispielsweise Symptome oder Örtlichkeiten. Das System erkennt eine ungewöhnliche Anzahl an Tweets mit ähnlichen Symptombeschreibungen. Es generiert ein Signal und teilt es Andrea mit.*

In regelmäßigen Abständen fragt das M-Eco System die sozialen Sensoren ab (beispielsweise RSS feeds) und sucht so nach neuen Informationen zu Krankheitsausbrüchen im Web. Dazu werden Online-Nachrichten und Twiternachrichten identifiziert, in denen Krankheiten oder deren Symptome benannt werden. Weiterhin werden zuvor festgelegte Blogs und Foren nach relevanten Nachrichten durchsucht. Relevante TV- und Radiobeiträge werden ebenfalls ermittelt.

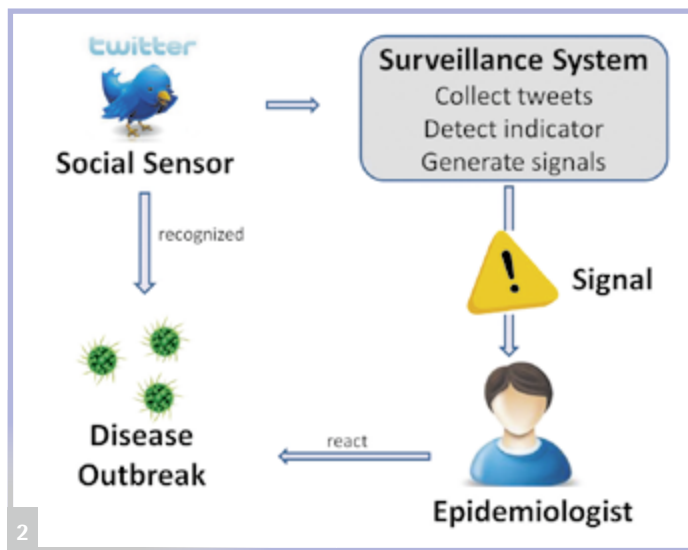


Abbildung 1  
Quelle: Fotolia

basierte Systeme oftmals die einzige Möglichkeit, um Hinweise zu erkennen. Biocaster, Medisys und ProMed Mail sind ereignisbasierte Systeme, die hauptsächlich Nachrichtenartikel und Berichte als Informationsquellen nutzen und nach Hinweisen durchsuchen. Diese Systeme werden bereits intensiv von der europäischen Gesundheitsorganisation ECDC und der WHO verwendet.

**Zusätzliche Quellen für die Frühwarnung**

13:15 Uhr: Andrea stellt fest, dass der Grund für das angezeigte Signal sehr viele Twiternachrichten sind, die Begriffe wie »Fieber« und »Hautausschlag« enthalten. Offenbar twittert eine Gruppe von Schülern über Mitschüler, die wegen Fieber und

nikation können somit ergänzende Informationen über das Auftreten von Krankheiten und deren Symptome liefern. Sie werden daher von den Gesundheitsorganisationen zunehmend als wertvolle Informationsquelle angesehen; die Informationen im Internet könnten helfen, die Möglichkeit der Früherkennung von Krankheiten zu erweitern. Ziel von M-Eco ist es, bestehende Frühwarnsysteme zu erweitern, indem diese zusätzlichen Informationsquellen genutzt werden.

**Methoden zum Ermitteln von Ereignissen in Texten**

Das M-Eco System identifiziert anhand von Schlüsselwörtern (unter anderem Symptome) relevante Tweets. Außerdem werden

Mit Methoden des maschinellen Lernens versuchen die Wissenschaftler am L3S aus den ermittelten Texten Muster zu erkennen, die als Hinweis auf Krankheitsausbrüche interpretiert werden können. Dazu werden Fakten aus den Texten extrahiert, etwa die benannten Symptome, Krankheiten oder Orte. Diese Informationen bilden die Basis für die Verfahren zur Mustererkennung. Anschließend werden statistische Verfahren angewendet, um diese Muster zu interpretieren. Sollten die Daten einen bestimmten Schwellenwert überschreiten oder Auffälligkeiten aufweisen, werden Signale generiert und dem M-Eco Benutzer mit den entsprechenden Informationen und Internetinhalten angezeigt. Dieser kann dann effektiver und frühzeitiger auf die bevorstehende Epidemie reagieren. Der Schwellenwert kann vom M-Eco Benutzer selbst, von Gesundheitsorganisationen oder durch statistische Methoden (beispielsweise Zeitreihenanalysen, Berücksichtigen von Saisonalität) berechnet und festgelegt werden.

Abbildung 2  
Im M-Eco Projekt werden soziale Sensoren genutzt, um Hinweise auf Krankheitsausbrüche zu identifizieren.

Eine große Herausforderung für die Entwicklung eines solchen Systems ist die Verarbeitung natürlicher Sprache und die geographische Ortung. Gerade in Tweets teilen Menschen ihren aktuellen Auf-

len wie zum Beispiel Youtube wird die Sprache in Text umgewandelt. Hierbei wird das System mit den Schwierigkeiten der Spracherkennung wie beispielsweise Dialekten, Sprach- und Tonqualität konfrontiert.

gen, an welchen Krankheiten und Symptomen er interessiert ist und in welcher Region. Zum anderen wird sein Nutzerverhalten berücksichtigt. Durch Ausnutzen des Feedbacks der Benutzer, beispiels-

Abbildung 3  
In Graphen auf der M-Eco Benutzeroberfläche werden Trends visualisiert, unter anderem wie viele Dokumente über die Zeit zu einem Signal ermittelt wurden.

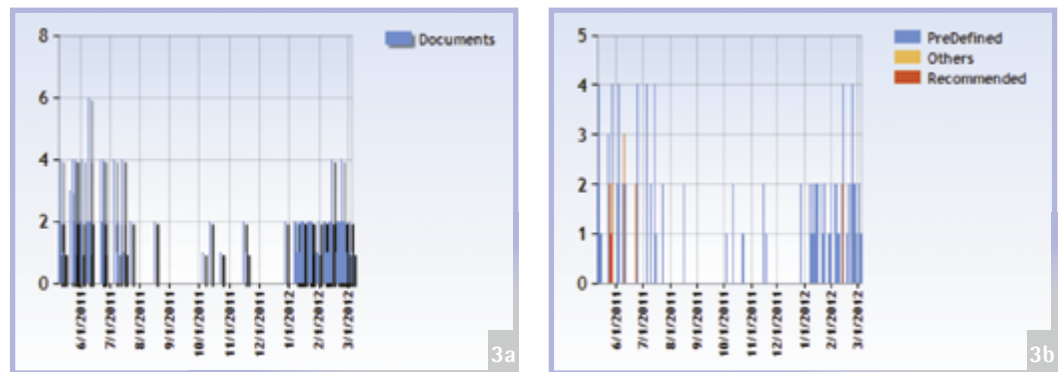


Abbildung 4  
In der Signaldefinition legt der Nutzer Krankheiten, Symptome und geographische Regionen fest.

Abbildung 5  
Erzeugte Signale werden im M-Eco System auf einer Landkarte markiert.



weise durch die Bewertung der Signale und Dokumente ist das System in der Lage, selbst die Präferenzen der Nutzer zu erlernen. Daraus werden Empfehlungen für Signale generiert. Die Informationen zu Signalen werden auf verschiedene Weise visualisiert, beispielsweise durch die Darstellung der Signale auf einer Weltkarte (siehe Abbildung 5), durch Balkendiagramme, Zeitreihen und einer Tag Cloud (siehe Abbildung 7).

### Was bringt die Zukunft?

13:30 Uhr: Andreas Nachforschungen ergeben, dass diesem Hinweis nachgegangen werden muss. Sie setzt sich daraufhin mit der regionalen Gesundheitsbehörde in Verbindung. Diese stellt fest, dass tatsächlich viele Schüler einer Schule an Masern erkrankt sind. Um eine weitere Ausbreitung zu verhindern, müssen nicht geimpfte Kinder vorerst zu Hause bleiben. Eine weitere Ausbreitung kann so verhindert werden.

enthaltensort nicht mit. Sie benutzen beim Schreiben Abkürzungen, Lautschrift, umgangssprachliche Wörter, Ironie, Sarkasmus sowie Wörter mit Doppelbedeutungen. Ferner kommt es zu Tipp- und Rechtschreibfehlern. Bei der Umwandlung von Audioquel-

### Signale und Informationen für Epidemiologen

Aus den erzeugten Signalen werden solche ausgewählt, die für den jeweiligen Benutzer relevant sind. Dazu wird zum einen seine Signaldefinition genutzt. Hier kann er festle-

Die zukünftigen Anwender des M-Eco Systems sind Epidemiologen und Mitarbeiter in Gesundheitsämtern auf kommunaler, überregionaler, nationaler und internationaler Ebe-

ne. Sie evaluieren momentan das System und versuchen herauszufinden, wann hilf-

Social Media Daten und Onlinenachrichten relevante Informationen zu dem Ausbruch



**Dr. Kerstin Denecke**

Jahrgang 1979, arbeitet am Forschungszentrum L3S und leitet seit 2010 das M-Eco Projekt. Ihr Forschungsschwerpunkt ist Textmining, vor allem mit dem Anwendungsschwerpunkt Medizin. Kontakt: [denecke@L3S.de](mailto:denecke@L3S.de)

**Literatur**

- Stewart A.; Diaz-Aviles, E.; Nanopoulos, A.; Self-Supervised Detection of Disease Reporting Events in Outbreak Reports. The 12th IEEE International Conference on Information Reuse and Integration (IEEE IRI-2011) Tuscany Suites & Casino, Las Vegas, USA, 2011
- Fisichella M.; Stewart A.; Cuzzocrea A.; Denecke K: Detecting Health Events on the Social Web to Enable Epidemic Intelligence. SPIRE 2011: the 18th edition of the International Symposium on String Processing and Information Retrieval, Pisa, Italy, 2011
- Denecke K.; Kirchner G.; Dolog P.; Smrz P.; Linge J.; Backfried G.; Dreesman J.: Event-Driven Architecture for Health Event Detection from Multiple Sources. In: Moen A.; Andersen S.K.; Aarts J.; Hurlen P. (eds.): Studies in Health Technology and Informatics. Volume 169, 2011, User Centred Networked Health Care – Proceedings of MIE 2011, 160–164

Abbildung 6  
Vertreter der WHO, INVS und anderen Gesundheitsorganisationen beraten die Wissenschaftler im M-Eco Projekt im Mai 2011 im Joint Research Centre der Europäischen Kommission im italienischen Ispra.

**M-Eco  
Medical Ecosystem – Personalized Event-based Surveillance**

Projektkoordinator: Leibniz Universität Hannover

**Kontakt**

- Dr. Kerstin Denecke  
Forschungszentrum L3S  
Appelstraße 9a, D-30167 Hannover  
Telefon: (+49) 511 76217725  
E-Mail: [denecke@L3S.de](mailto:denecke@L3S.de)  
Website: <http://www.meco-project.eu>

**Partner**

- Leibniz Universität Hannover (Deutschland)
- Aalborg University (Dänemark)
- Brno University of Technology (Tschechische Republik)
- SAIL Labs Technology (Österreich)
- Robert-Koch Institut (Deutschland)
- Niedersächsisches Landesgesundheitsamt (Deutschland)
- Joint Research Centre (Italien)

Projektlaufzeit: 01/10 –06/12  
Gesamtvolumen: EUR 2,59 Millionen  
Fördersumme: EUR 1,99 Millionen  
Instrument: STREP  
Projektidentifikationsnummer: FP7 ICT-2007.5.3

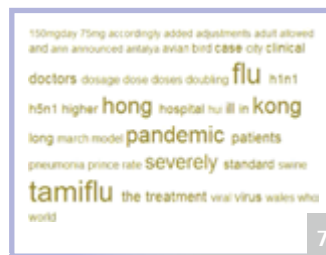


Abbildung 7  
Tag cloud, generiert aus Texten zu einem Signal

reiche Informationen in Social Media durch das System erkannt wurden.

verfügbar waren. Die Zukunft wird zeigen, ob das M-Eco System Ursachen eines künftigen Krankheitsausbruchs frühzeitiger erkennbar machen kann.

Für den Krankheitsausbruch in Deutschland, der durch E.coli Bakterien in Wasser ausgelöst wurde, konnte bereits festgestellt werden, dass in