

Selbstorganisierende Netzwerke fürs Katastrophenmanagement

Self-organizing networks for disaster management

**Informatiker und Seismologen entwickeln ein
Erdbeben-Frühwarnsystem und testen es in Istanbul**

*Computer scientists and seismologists have developed
an earthquake early-warning system and are testing it in Istanbul*

Text: Ljiljana Nikolic





Die Box auf dem Tisch sieht unscheinbar aus. Ein grauer Kasten, etwa 20 mal 10 Zentimeter groß. »Das war ein ganz gewöhnlicher WLAN-Router, wir haben eine neue Software, einen Sensor, GPS und einen Seismometer eingebaut«, erklärt Informatik-Professor Joachim Fischer, auf dessen Schreibtisch das Gerät liegt. 20 dieser am Institut für Informatik anfangs selbst zusammen gebauten Kästen befinden sich seit zwei Jahren auf den Dächern Istanbuls, genauer gesagt im Stadtteil Ataköy. Sie sind die erste Generation eines neuartigen, auf selbstorganisierenden Sensornetzwerken basierenden Erdbebenfrühwarnsystems. Es wird im Graduiertenkolleg »Metrik« am Standort Adlershof in Zusammenarbeit mit anderen Partnern wie dem Deutschen Geoforschungszentrum Potsdam entwickelt und erforscht.

Warum Istanbul? Die türkische Metropole muss mit einer erschreckenden Prognose leben: Seismologen rechnen in den nächsten Jahren und Jahrzehnten mit einem verheerenden Erdbeben in der elf Millionen Einwohner zählenden Stadt. »Hauptsächliche Ursache für alle schweren Erdbeben sind Spannungen, die sich an den Rändern tektonischer Platten aufbauen«, verdeutlicht Fischer. Im Falle von Istanbul, historisch

betrachtet Konstantinopel, reibt die Eurasische an der Anatolischen Platte entlang der Nordanatolischen Verwerfung. »Diese Bewegung wird durch Reibungskräfte gebremst, die dazu führen, dass sich Spannungen aufbauen. Wird diese Spannung stärker als die Reibung, wird ein Erdbeben ausgelöst.« Entlang der Verwerfung, die sich vom Nordosten bis zum Nordwesten der Türkei bis in die Region Istanbul zieht, haben sich in den vergangenen 70 Jahren mehr als 30 große Beben ereignet, das letzte in Izmet. 80 Kilometer von Istanbul bebte die Erde mit einer Stärke von 7,4 auf der Richterskala. 24.000 Tote, 40.000 Verletzte und 20.000 zerstörte Häuser waren die Folge dieses Naturereignisses in einem relativ dünn besiedelten Areal.

Aus diesen Gründen haben die Informatiker und Geowissenschaftler, die in Metrik zusammenarbeiten, die Mega-City als Testort gewählt für die von ihnen erforschten Technologien für Frühwarn- und Katastrophenmanagement-Systeme. »Wir arbeiten an drahtlosen Netzarchitekturen, die ohne aufwändige zentrale Verwaltung und Leitungstechnik auskommen, kostengünstig im Bau wie im Stromverbrauch sind und auf dem Prinzip der Selbstorganisation funktionieren«, erklärt Fischer. Die Informatiker sprechen von ad-hoc-Netzwerken oder Selbst-

Auf den Dächern des Wissenschaftsstandorts Adlershof: Professor Fischer (r.) und seine Mitarbeiter installieren Sensorknoten.

On the roofs of the Adlershof Science and Technology Park: Professor Fischer and his colleagues are installing sensor nodes.

Joachim Fischer hat von 1973 bis 1978 Mathematik an der Humboldt-Universität studiert. 1982 promovierte er zur Modellierung und simulativen Verhaltensnachbildung kooperativer Systeme. Sechs Jahre später erfolgte die Habilitation, ebenfalls an der HU, mit einer Arbeit zum Thema »Rapid Prototyping verteilter Systeme«. Dazwischen verbrachte er ein Forschungsjahr am Informatik-Institut der Universität Warschau.

1994 wurde er zum Professor für Systemanalyse, Modellierung und Computersimulation an der Humboldt-Universität berufen. Im Mittelpunkt seiner Forschung steht die Modellierung von Rechnersystemen in ihrer Einbettung in reale zu steuernde und ebenfalls zu modellierende komplexe Umgebungen. IT-gestützte industrielle und wissenschaftliche Workflows spielen dabei eine besondere Rolle. Neben der Systemmodellierung geht es ihm auch um die Bereitstellung geeigneter Modellbeschreibungssprachen inklusive begleitender domänenspezifischer Software-Werkzeuge wie Editoren, Compiler, Interpreter, Simulatoren, Debugger und Animatoren. Fischer ist Sprecher des interdisziplinären Graduiertenkollegs Metrik.

fischer@informatik.hu-berlin.de
Tel 030 · 2093-3109

Joachim Fischer studied mathematics at Humboldt-Universität from 1973 to 1978. He was awarded his doctorate on the modelling and behaviour simulation of cooperative systems in 1982. Six years later he qualified as a professor, also at HU, with a thesis on the »Rapid Prototyping of Distributed Systems.« In-between he spent a research year at the University of Warsaw's Institute of Computer Science.

He was appointed Professor of Systems Analysis, Modelling and Computer Simulation at Humboldt-Universität in 1994. His research focuses on the modelling of computer systems and their integration into real, complex environments that need to be controlled and modelled. IT-based industrial and scientific workflows play an important role in this context. In addition to systems modelling, he also studies the provision of suitable model-description languages, including accompanying domain-specific software tools such as editors, compilers, interpreters, simulators, debuggers, and animators. Fischer is the spokesperson of the METRIK interdisciplinary research training group.

organisierenden Kommunikationssystemen, Technologien, die sich im Anfangsstadium ihrer Entwicklung befinden. Das in Metrik entwickelte, trägt den Namen Sosewin, »Self-organizing Seismic Early Warning Information Network«.

Eines muss der Wissenschaftler und Sprecher des Kollegs dabei klarstellen: »Frühwarnsystem bedeutet nicht, dass wir in der Lage wären, ein Erdbeben vorherzusagen, das kann niemand, denn Erdbeben kündigen sich nicht an.« Es

geht vielmehr darum, mit solchen Systemen, weiter entfernte Gebiete vor den nahenden zerstörerischen Wellen zu warnen – nachdem die Erde schon im jeweiligen Epizentrum gebebt hat. Dazu muss man wissen, dass Erdbeben sich in unterschiedlichen Wellenformen in allen Richtungen ausbreiten. Seismologen unterscheiden unter anderem zwischen P(rimär)- und S(ekundär)-Wellen. Die beiden Wellenarten unterscheiden sich grundsätzlich. Während die P-Welle mit über fünf Kilometer pro Sekunde im Durchschnitt doppelt so schnell wie die S-Wel-

Sie sind die erste Generation eines neuartigen, auf selbstorganisierenden Sensornetzwerken basierenden Erdbebenfrühwarnsystems

They are the first generation of a novel earthquake early-warning system based on self-organizing sensor networks

le ist, die sich mit drei Kilometer pro Sekunde ausbreitet, überträgt sie nur geringe Energiemengen. Das bedeutet, dass P-Wellen als Ankündigungen der zerstörerischen S-Wellen interpretiert werden können.

Die Zeitspanne zwischen dem Eintreffen von Primärwelle und Sekundärwelle für einen Ort ist die verbleibende Vorwarnzeit. Je weiter Epizentrum und Besiedlungsgebiet voneinander entfernt sind, desto größer die Chancen, bei einem starken Beben Menschen zu mobilisieren.



Metrik

www.gk-metrik.de

Das Institut für Informatik ist bundesweit das einzige, das Sprecherschaften in zwei Graduiertenkollegs der Deutschen Forschungsgemeinschaft hält. Das interdisziplinär ausgerichtete Graduiertenkolleg Metrik »Modellbasierte Entwicklung von Technologien für selbstorganisierende dezentrale Informationssysteme – zur Anwendung im Katastrophenmanagement« verbindet Forschungsinitiativen von zwölf Professorinnen und Professoren aus der Informatik und Geographie der Humboldt-Universität, dem Geoforschungszentrum Potsdam, dem Zuse-Institut Berlin sowie dem Fraunhofer First.

Erforscht und erprobt werden neuartige Kommunikationsinfrastrukturen für Katastrophenmanagementsysteme, die sich von traditionellen drahtgebundenen und drahtlosen

IT-Infrastrukturen unterscheiden. Sie benötigen im Kern keine zentrale Verwaltung und passen sich in selbstorganisierender Art und Weise an sich ändernde Belastungssituationen ihrer Umgebung an. 14 Doktoranden und zwei Post-Doktoranden forschen zurzeit am Kolleg und werden in technischen Arbeiten von verschiedenen Studentengruppen unterstützt. Das von der Deutschen Forschungsgemeinschaft geförderte Kolleg befindet sich in der zweiten Förderphase.

Humboldt-Universität's Department of Computer Science is the only such institute in the country that hosts two research training groups of the German Research Foundation. The interdisciplinary research training group Metrik (a German acronym standing for »Model-based Development of Technologies for Self-Organizing Distributed Information Systems for Use in Disaster Management«) links the research initiatives of twelve professors in the fields of computer science and geography

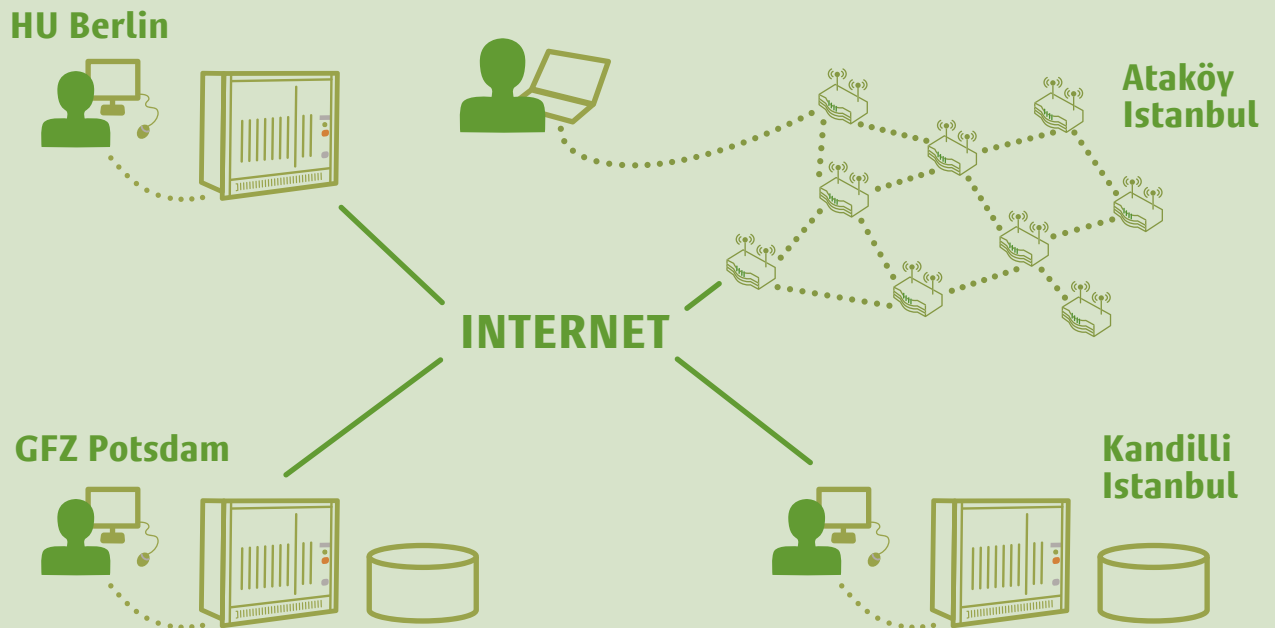
at Humboldt-Universität, the German Research Centre for Geosciences, Potsdam, the Zuse Institute, Berlin, and the Fraunhofer First.

The group focuses on researching and testing novel communication infrastructures for disaster management systems which differ from traditional wired and wireless IT infrastructures. They essentially need no central management, but self-organize to adapt to changing demands from their environment. 14 PhD students and two post-docs are currently engaged in research at Metrik, supported in technical work by various student groups. The group is in its second phase of funding by the German Research Foundation.

Schnelle Kommunikation

Instant communication

Handelt es sich um ein Beben, werden innerhalb von zwei Sekunden das Istanbuler Erdbebenzentrum Kandilli und die Forscher am Institut für Informatik und am Deutschen Geoforschungszentrum Potsdam informiert. *If it is a quake, the Kandilli Earthquake Centre in Istanbul, the researchers at the HU's Department of Computer Science and the German Research Centre for Geosciences, Potsdam, are informed within two seconds.*



Zurück nach Istanbul, den Dächern von Ataköy und dem Netzwerk, in dem die in 100 bis 200 Meter Abstand installierten 20 Knoten drahtlos und selbstorganisierend miteinander kommunizieren und durch Sonnenenergie am Tage und von einer Batterie in der Nacht gespeist werden.

»Die Kommunikation der einzelnen Knoten, die per Software seismische Wellen identifizieren, lässt sich mit der Teamarbeit im Unternehmen vergleichen«, meint Fischer. Ein Knoten allein kann eine Erschütterung nicht sicher als P-Welle identifizieren, es könnte sich auch um eine Erdbewegung anderen Ursprungs handeln. »Deshalb nutzen wir die örtliche Verteilung unserer Sensoren aus, bei der wir die Ausbreitung der Welle in Raum und Zeit berücksichtigen. Wenn ein mit GPS- und Sensor erfasstes Signal in einem Knotenpunkt als eine P-Welle gehalten wird, kommuniziert der von uns entwickelte Knoten eine Verdachtsnachricht an seinen Gruppenleiter.« Dieser informiert nach Eintreffen von Verdachtsmeldungen weiterer Mitglieder seiner Gruppe benachbarte Gruppenleiterknoten. Der erste Gruppenleiter, bei dem nun eine gewisse Anzahl von Meldungen eingetroffen ist, kann nun einen Alarm auslösen. »Dabei wird die Funktionalität des Systems weder durch ausfallende

Knoten, noch durch neu dazugekommene behindert, sie werden vom Netz selbst erkannt und integriert«, unterstreicht Fischer. »Ein beliebiger Knoten im Netz übernimmt zeitweilig die Rolle eines Senders, Übermittlers oder Empfängers von Meldungen und muss diese bei Bedarf auch wechseln können.« Hinzukommt, dass jeder Knoten über einen Seismometer die lokalen Erschütterungen seines Aufstellungsortes zu erfassen hat und in der Lage sein muss, auch gegebenenfalls den Job eines Gruppenleiters zu übernehmen. Erst durch die Kooperation der Knoten im Team kann ein vorbeifahrender Lastwagen oder andere Erschütterungsursachen von einer P-Welle seismischen Ursprungs unterschieden werden. Handelt es sich um ein Beben, werden innerhalb von zwei Sekunden das Istanbuler Erdbebenzentrum Kandilli und die Forscher am Institut für Informatik und am Deutschen Geoforschungszentrum Potsdam informiert.

Doch was, wenn das zu erwartende Epizentrum vielleicht nur 40 Kilometer von der bedrohten Metropole entfernt ist, und die Zeit zwischen P- und S-Welle und damit die Vorwarnzeit gerade mal fünf Sekunden beträgt? »Möglich sind dann immer noch eine Reihe von Maßnahmen, die in erster Linie nachfol-



Das drahtlose Maschennetzwerk für Testzwecke in Adlershof soll aus 100 Knotenpunkten bestehen. *The wireless mesh network for testing purposes in Adlershof shall consist of 100 nodes.*

gende, manchmal noch schlimmere Katastrophen wie Gasexplosionen, Brände und Ähnliches stark reduzieren könnten, indem beispielsweise Gasleitungen automatisch abgesperrt werden, Strom ausgeschaltet und chemische Produktionen gedrosselt werden«, verdeutlicht der Forscher. Mit Hilfe des Systems soll außerdem der Grad entstandener Zerstörungen nachfolgender S-Wellen und Oberflächenwellen schnell ermittelt werden. Eine Information, die für die zügige Durchführung von Bergungs- und Rettungsaufgaben wichtig ist, wäre die Erfassung maximaler Erdbeschleunigungen in einem engen Raster mit weniger als einem Kilometer Abstand.

Damit stehen die Wissenschaftler rund um Metrik vor einer der größeren Herausforderungen bei der Erforschung der Ad-hoc-Netzwerke. Es muss gesichert sein, dass das Netzwerk auch nach einem schweren Beben funktionsfähig gemacht werden kann. Sicherheitsprobleme gibt es bereits schon jetzt, denn die installierten Knotenpunkte dienen manchem Istanbul als Ersatzteillager. Außerdem müssen auch große Netzwerke, wie sie eine Stadt wie Istanbul braucht, sicher und zuverlässig funktionieren.

Um die Forschung voranzutreiben, möchte Professor Fischer mit seinen Kollegen ein drahtloses Maschennetzwerk mit

100 Knotenpunkten auf den Dächern des Wissenschaftsstandorts Adlershof für Testzwecke installieren. Mit einem Erdbeben ist hier sicherlich nicht zu rechnen. P- und S-Wellen werden die Informatiker deshalb simulieren. Die Signalanalysekomponenten in den Knoten haben dann nicht auf ihre Seismometer zu hören, sondern werden Datenkonserven auswerten. Dabei soll das Netz auch für den Einsatz auf neue Anwendungsfelder durch alternative Sensorik, wie Kameras, Lärm- und Feinstaubsensoren, vorbereitet werden.

The box object

on the table doesn't look like anything special. A grey box about 20 cm by 10 cm. »It used to be a perfectly ordinary WiFi router. We installed some new software, a sensor, GPS and a seismometer,« explains computer science professor Joachim Fischer, on whose desk the device sits. For two years now, 20 of these boxes, the first of which were self-assembled at the Department of Computer Science, have been installed on the roofs of Istanbul – in the district of Ataköy to be more precise. They are the first generation of a novel earthquake early-warning system based on self-organizing sensor networks. It is being developed and studied at the »Metrik« research training group on the Adlershof campus in collaboration with other partners such as the German Research Centre for Geosciences, Potsdam (see box).

Why Istanbul? Well, the Turkish capital has to live with a frightening prediction: seismologists think a devastating earthquake will hit this city of eleven million inhabitants in the coming years and decades. »The main cause of all severe earthquakes is the tensions that build up on the edges of tectonic plates,« explains Fischer. In the case of Istanbul (or Constantinople as it used to be called), the Eurasian Plate is rubbing against the Anatolian Plate along the North Anatolian Fault. »This movement is being slowed down by frictional forces, which builds up tensions; if the tension becomes greater than the friction, this triggers an earthquake.« Along this fault, which runs from northeast to northwest Turkey as far as the Istanbul region, more than 30 major earthquakes have occurred in the past 70 years, the last one in Izmet. This quake – 80 kilometres from Istanbul – measured 7.4 on the Richter scale. 24,000 people were killed, 40,000 injured and 20,000 houses destroyed as a result of this natural disaster in a relatively sparsely populated area.

For these reasons, the computer scientists and geoscientists who work together in Metrik have chosen the mega-city of Istanbul as a testing site for the early-warning and disaster-management-system technologies they are researching. »We are working on wireless network architectures that don't require an expensive centralized administration and cable technology, are cheap to make, low on power consumption and work

on the principle of self-organization,« explains Fischer. The computer scientists speak of ad-hoc networks or self-organizing communication systems – technologies that are in the early stages of development. The system developed in Metrik is called Sosewin, »Self-Organizing Seismic Early Warning Information Network«.

Fischer, spokesman for the research training group, has to make one thing clear, though: »An early-warning system doesn't mean we can predict an earthquake; nobody can do that because earthquakes don't give any prior warning.« Rather, what such systems aim to do is to warn more distant regions of the approaching destructive waves – after the earth has already shaken at the respective epicentre. We need some background information here. Earthquakes spread out in all directions in different wave forms. Seismologists distinguish, for example, between P(rietary) and S(econdary) waves. The two types are fundamentally different. While the P-wave, at more than five kilometres a second, is twice as fast on average as the S-wave, which spreads at three kilometres a second, it only transfers small amounts of energy. This means that P-waves can be interpreted as a warning of the imminent, destructive S-waves.

So the time interval between the arrival at a site of the primary wave and the secondary wave represents the remaining warning time. The further the epicentre is from a residential area, the better the chances of mobilizing people before a strong quake strikes.

Returning to Istanbul, to the roofs of Ataköy: the 20 nodes of the network have been installed 100 to 200 metres apart; they communicate with each other wirelessly and in a self-organizing way and are powered by solar energy during the day and by batteries at night.

Die Kommunikation der einzelnen Knoten lässt sich mit der Teamarbeit im Unternehmen vergleichen

The communication between the individual nodes can be compared with teamwork within a company



»**The communication between the individual nodes**, which identify seismic waves using software, can be compared with teamwork within a company,« says Fischer. One node alone cannot reliably identify a vibration as a P-wave; it could always be some other earth movement from another source. »That's why we exploit the local distribution of our sensors and look at the spread of the wave in space and time. If a signal recorded in a node by GPS and a sensor is regarded as a P-wave, the node we have developed sends a ›suspicion message‹ to its group leader.« If suspicion messages also arrive from other members of its group, it informs neighbouring group-leader nodes. The first group leader that receives a certain number of group-leader messages can now trigger an alarm. »The system even continues to function when nodes cease functioning or when a new one is added, since these are recognized and integrated by the network itself,« Fischer underlines. »At times any node in the network can temporarily assume the role of an emitter, transmitter or receiver of messages; it must also be able to change roles when necessary.« Furthermore, every node has to sense the local vibrations at its installation site using a seismometer and must be able, when necessary, to take on the job of group leader. Only through cooperation between the nodes in the team can a passing truck or other causes of vibration be distin-

guished from a P-wave of seismic origin. If it is a quake, the Kandilli Earthquake Centre in Istanbul, the researchers at the HU's Department of Computer Science and the German Research Centre for Geosciences, Potsdam, are informed within two seconds.

But what if the expected epicentre is perhaps only 40 kilometres from the threatened city, and the time between the P- and S-waves (i.e. the warning time) is only 5 seconds? »Even then, a number of measures are still possible primarily geared to reducing the effects of the following disasters, which are sometimes even worse – such as gas explosions, fires and the like – e.g. by automatically closing gas mains, switching off electricity and cutting back chemical production,« explains the researcher. Another aim is that the system can help to quickly determine the degree of damage that will be caused by the following S-waves and surface waves. One piece of information that is important for the swift implementation of search and rescue operations would be to record the peak ground accelerations within a narrow grid with points less than 1 kilometre apart.

This is one of the bigger challenges for the Metrik scientists in their research into ad hoc networks. It must be ensured that the network continues to work well, or can be quickly repaired even after a major quake. There are already security problems, because



the installed nodes serve some Istanbul residents as a source of spare parts. Furthermore, large networks, like the ones needed by a city the size of Istanbul, must function securely and reliable.

To advance the research, Professor Fischer and his colleagues are installing a wireless mesh network with 100 nodes on the roofs of the Adlershof Science and Technology Park for testing purposes. Since an earthquake is hardly likely here, the computer scientists have to simulate P- and S-waves. The signal analysis components in the nodes are therefore not listening to their own seismometers, but will evaluate stored data. In this context, the network is also being prepared for use in new applications using alternative sensory systems such as cameras, noise or particulate-matter sensors.

Es muss gesichert sein, dass das Netzwerk auch nach einem schweren Beben funktioniert
It must be ensured that the network continues to work well, or can be quickly repaired even after a major quake





L'Aquila nach dem Beben
L'Aquila after the earthquake

Per Equator ins Erdbebengebiet *Getting into the earthquake zone via Equator*

Wenn irgendwo auf der Welt die Erde bebzt, sammeln die Wissenschaftler des Geoforschungszentrums Potsdam (GFZ) schnell Informationen: Die Mitglieder der Deutschen Task Force Erdbeben müssen beispielsweise schnell entscheiden, ob sie ins Erdbebengebiet fahren oder nicht. Dabei sind Daten über Stärke und Tiefe des Bebens, frühere Beben, die Struktur der Region oder auch Hinweise auf geologische Besonderheiten in der Gegend von entscheidender Bedeutung. Die Wissenschaftler arbeiten dabei hauptsächlich mit Informationen verschiedener Erdbebeninformationsagenturen sowie anderen Quellen, die ihnen bis vor kurzem nur über eine Vielzahl von Webseiten, Veröffentlichungen und Büchern vorlagen und recherchiert wurden mussten.

Nun setzen die Wissenschaftler die Software Equator ein. Sie wurde von dem Studierenden der Informatik, Lars Döhling, als Studienabschlussarbeit im Rahmen der Kooperation des GFZ mit der Humboldt-Universität im Graduiertenkolleg

Metrik entwickelt. Equator steht für »Earthquake data collector« und ermöglicht es, statt mit vielen Quellen mit einem automatisch abrufbaren, strukturierten Dokument zu arbeiten, das die wichtigsten Informationen anderer Quellen halbautomatisch integriert. »Das System ist problemlos zu bedienen. Geologen können aus der Liste der aktuellen Meldungen aus der GFZ-Erdbebenagentur Geofon einen Eintrag anlegen. Basierend auf diesen Daten sucht das System dann aus anderen Erdbeben-Agenturen die dazu passenden Meldungen und teilt sie in thematische Abschnitte«, berichtet Döhling.

Eine Arbeitserleichterung für das GFZ. »Equator erfreut sich wachsender Popularität, es spart viel Zeit, nicht per Hand recherchieren zu müssen und trotzdem schnell einen Überblick über die verschiedenen, sich oft auch widersprechenden Daten zu bekommen«, erklärt GFZ-Mitarbeiter Heiko Woith. Das System werde auch eingesetzt, um die vielen Journalistenanfragen, die das Forschungszentrum bei Erdbeben erreichen, schnell zu bewältigen.

Die Wissenschaftler arbeiten über eine Update-Funktion immer mit der aktuellsten Version, können aber auch auf vorherige Versionen zurückgreifen. »Alle automatisch erkannten Koordinaten der einzelnen Abschnitte werden außerdem auf einer integrierten Karte dargestellt«, sagt Döhling. Natürlich können alle Daten bei Bedarf in übersichtlicher Form ausgedruckt werden. »Ein weiterer Vorteil des Systems liegt darin, dass die

Nutzer den ganzen Datenbestand zu einem Beben auch mit Stichwörtern automatisch durchsuchen und widersprüchliche Informationen so leichter erkennen können.«

Döhling, der mittlerweile Doktorand bei Metrik ist, forscht weiter an der Frage, wie Daten über Erdbeben schnell gewonnen werden können und bedient sich diesmal des Text Minings, einer Methode, mit deren Hilfe Informationen, die nur in textueller Form vorliegen, schnell erfasst werden können. »Beim Text Mining geht es darum, eine Fülle von elektronisch vorliegenden Dokumenten automatisch mit maschinellen Lernverfahren auf bestimmte Themen zu durchforsten und dabei relevante von irrelevanten Informationen zu trennen«, erklärt Informatik-Professor Ulf Leser, Doktorvater von Döhling und Text Mining-Spezialist (siehe auch Text Seite 174). In der Phase unmittelbar nach einem Erdbeben geht es vor allem darum, verlässliche Zahlen über Verletzte oder Tote sowie deren genaue Position aus einer Fülle von Nachrichten zu extrahieren. »Mitunter werden zu einem Beben in zehn unterschiedlichen Texten zehn verschiedene Zahlen geliefert. Es wird von Verletzten, Verschütteten, Toten berichtet«, unterstreicht Leser. Mit Text Mining wird an einem Bündel von automatischen Verfahren geforscht, die es ermöglichen sollen, solche Zahlen automatisch in einer Fülle von Dokumenten zu finden und so zu verlässlicheren Schätzungen zu gelangen.



Whenever the earth shakes anywhere in the world, speed is of the essence when the scientists at the »GFZ German Research Centre for Geosciences« collect the information. The members of the German Task Force on Earthquakes, for example, need to know quickly whether to travel to the quake zone or not. Data on the strength and depth of the earthquake, earlier quakes, the structure of the region and evidence of special geological features in the area are all of critical importance. In this context the scientists work mainly with information from different earthquake information agencies, as well as various other sources, which, until recently, was only accessible via a large number of websites, publications and books and had to be researched every time.

Now the scientists use a software program called Equator. It was developed by Lars Döhling, a computer science student, as a final degree assignment in the context of cooperation between the GFZ and Humboldt-Universität at the METRIK research training group. Equator stands for »Earthquake Data Collector« and enables staff to work – not with many different sources, but – with an automatically retrievable, structured document that integrates the most important information from other sources semi-automatically. »The system is easy to operate. Geologists can create an entry from the list of current news reports from Geofon, GFZ's earthquake alert system. Based on these data, the system then

searches for matching reports from other earthquake agencies and divides them into thematic sections,« Döhling explains.

This makes life much easier for the GFZ. »Equator is becoming increasingly popular. It saves us a lot of time because we no longer have to search manually, yet it still provides a quick overview of the various, often-contradictory data,« explains GFZ staff member Heiko Woith. The system is also used to quickly manage the inquiries from journalists that flood into the Research Centre when news of an earthquake spreads.

The scientists always work with the latest version via an update function, but can also fall back on previous versions. »In addition, all the automatically detected coordinates of the individual sections are displayed on an integrated map,« says Döhling. Of course, all the data can be printed out in an easy-to-read format whenever necessary. »Another advantage of the system is that the users can also automatically search all the available data on a quake using keywords, making it easier to recognize conflicting information.«

Döhling, now a PhD student with METRIK, is continuing his research into how data on earthquakes can be obtained quickly. This time he is looking into text mining, a method by which information that is only available in text form can be quickly processed. »With text mining you can automatically trawl through a large number of electronic documents on specific subjects using

Equator erfreut sich wachsender Popularität, es spart viel Zeit, nicht per Hand recherchieren zu müssen

Equator is becoming increasingly popular. It saves us a lot of time because we no longer have to search manually

machine learning systems; you can also separate relevant from irrelevant information,« explains computer-science professor Ulf Leser, Döhling's PhD supervisor and text-mining specialist (see also page 174). In the immediate post-quake phase, the main priority is to extract reliable figures on injuries or deaths, as well as their exact position, from many news reports. »Sometimes, when there is an earthquake, ten different figures are quoted in ten different texts. There are reports on people who have been injured, buried, killed,« says Leser. Text mining involved research into a whole bundle of automatic procedures which, it is hoped, will make it possible to find such figures from a wealth of documents automatically and in this way arrive at more reliable estimates.