

Vom Gezwitscher zur umfassenden und aktuellen Situationseinschätzung

Philipp Heim, Thomas Schlegel, Thomas Ertl
Institut für Visualisierung und Interaktive Systeme (VIS)
Universität Stuttgart
Universitätsstraße 38
70569 Stuttgart

{Philipp.Heim|Thomas.Schlegel|Thomas.Ertl}@vis.uni-stuttgart.de

Kurzfassung: Dieser Beitrag beschreibt einen Ansatz, wie Informationsquellen des Web 2.0 genutzt werden können, um eine aktuelle und angemessene Einschätzung von komplexen und schwer zu überblickenden Situationen zu unterstützen. Dazu werden, mit Hilfe von Technologien des Semantic Web, für eine konkrete Situationseinschätzung relevante Informationen ausfindig gemacht und in kombinierter Form angezeigt. Neue Erkenntnisse über die Situation können zur Aktivierung weiterer Information führen und somit die Einschätzung iterativ verbessern.

1 Einleitung

In Leitständen und Krisenstäben müssen in kurzer Zeit, Gefahren erkannt und weitreichende Entscheidungen getroffen werden. Als Grundlage für die Entscheidungsfindung dienen Einschätzungen der vorherrschenden Situation. Eine solche Einschätzung kann jedoch nur dann situationsangemessen sein, wenn alle relevanten Informationen zugreifbar sind und in ein umfassendes Gesamtbild integriert wurden. So werden beispielsweise für die Einschätzung von Gefahrensituationen, Informationen wie zeitliche und örtliche Ausdehnung, Ursache der Gefahr, Anzahl der Opfer, Art und Umfang von Verletzungen und möglicherweise noch weitere Details benötigt. Da sich eine Situation laufend verändern kann, müssen sich auch die Informationen über diese laufend aktualisieren, um den Entscheidungskontext aktuell zu halten. Nur ein aktuelles und umfassendes Gesamtbild erlaubt eine situationsangemessene Einschätzung und damit eine optimale Reaktion zur Minimierung von Schaden.

In vielen Fällen sind jedoch weder alle, für eine Situationseinschätzung relevanten Informationen zugreifbar, noch befinden sich diese auf dem aktuellen Stand. Das ist der Fall, weil Einschätzungen oftmals aus der Entfernung getroffen werden müssen, d.h. die einschätzende Person nicht vor Ort ist, oder sich Situationen zeitlich und räumlich über einen größeren Bereich erstrecken und dadurch nicht alle relevanten Informationen immer vollständig einsehbar sind: Ereignisse können sich z.B. innerhalb von Gebäuden oder Fahrzeugen abspielen. Die Schwierigkeit liegt jedoch nicht nur im Bereich des Informationszugangs, sondern auch im Bereich der Informationsauswahl und der Integration unterschiedlicher Informationen in ein Gesamtbild. Vor allem bei zeitkritischen, komplexen und schwer einsehbaren Situationen, z.B. bei einem eingehenden Notruf, sind daher Fehleinschätzungen möglich.

Neue Erkenntnisse, die auf Grund der Visualisierung relevanter Informationen oder externer Quellen, z.B. offizieller Berichte, gewonnen wurden (Abb. 1, F), können zu einer veränderten, semantischen Definition des Ausgangspunkts für die Situationseinschätzung führen (Abb. 1, C). Durch einen veränderten Ausgangspunkt werden andere Informationen als relevant identifiziert (Abb. 1, D) und angezeigt, was wiederum zu neuen Erkenntnissen und damit zu einer iterativen Verbesserung der Einschätzung einer Situation führen kann.

Im Folgenden werden die Teilkomponenten des Systems näher beschrieben.

2.1 Verteilte, heterogene Informationsquellen

Vor allem bei der Einschätzung von Situationen, in denen aktuelle Ereignisse eine entscheidende Rolle spielen, diese aber durch bestehende Informationskanäle nicht aufgefangen werden können, bieten verteilte, heterogene Informationsquellen des Web 2.0 einen möglichen Ausweg. Internetseiten, Blogs, Diskussionsforen, soziale Netzwerke und Mikroblogging-Dienste ermöglichen jedem Bürger, in einer Art Bürgerjournalismus aktuelle Ereignisse, Meinungen und Kommentare über das Internet zu verbreiten. Dadurch werden viele unterschiedliche Perspektiven auf aktuelle Ereignisse zugänglich, die bei der Einschätzung von Situationen bisher keine Berücksichtigung fanden, möglicherweise aber relevante Informationen liefern können.

Ein aktuelles und sehr trauriges Beispiel ist der *Amoklauf von Winnenden*⁴. Hier wurden, zum ersten Mal in dieser Form in Deutschland, massiv Informationen zu einer aktuellen Gefahrensituation von Bürgern vor Ort über den Mikroblogging-Dienst Twitter bereitgestellt. Solche, bereits verfügbaren Informationsquellen anzuzapfen, um die darüber aufgefangenen Informationen dann zur Begrenzung und Vermeidung von Schaden in konkreten Gefahrensituationen einzusetzen, ist das Ziel des hier vorgestellten Ansatzes.

2.2 Semantische Annotation

Um auf die große Menge an aufgefangenen Informationen sinnvoll zugreifen zu können, werden diese semantisch annotiert. Durch die Annotation werden den, in den aufgefangenen Informationen verwendeten Zeichenketten eindeutige Bedeutungen zugewiesen, z.B. ein bestimmtes Datum, ein bestimmter Ort oder bestimmte Instanzen einer anderen Domain. Als Quelle für diese Referenzen dienen dabei bestehende, über das Internet via *SPARQL*⁵-Anfragen frei zugängliche, semantische Datensätze, wie z.B. die *LOD-Cloud*⁶. Diese enthält in *RDF*⁷ abgespeicherte, semantische Beschreibungen von Daten aus vielen verschiedenen Domains, z.B. geografische Daten, Musik-Daten, Bild-Daten, Personen-Daten und Wikipedia-Daten, sowie deren Verbindungen untereinander.

⁴ Amoklauf von Winnenden: http://www.stuttgarter-zeitung.de/stz/page/1972451_0_9223_zwischen-zu-schnell-und-zu-langsam-das-twitter-echo-von-winnenden-und-seine-folgen.html

⁵ SPARQL (Anfragesprache für RDF): <http://www.w3.org/TR/rdf-sparql-query/>

⁶ LOD-Cloud: <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>

⁷ RDF (Resource Description Framework): <http://www.w3.org/RDF/>

Auf Grund der großen Menge an stetig neuer, über das Web 2.0 aufgefangener Information, wäre eine automatische Annotation von großem Nutzen. Da jedoch die Semantik natürlicher Sprache noch nicht in ausreichendem Maße automatisch erkannt wird, ist die vollautomatische Annotation von Zeichenketten nicht immer eindeutig und damit fehlerfrei möglich. Einen möglichen Ausweg bieten hier semi-automatische Verfahren [He09a] in Kombination mit automatischen Lernstrategien [LHZ07]. Für unterschiedliche Domains, z.B. verschiedene Ideologien radikaler Gruppierungen, könnten unterschiedliche Experten Informationen manuell klassifizieren und dadurch schrittweise automatische Klassifikationsverfahren trainieren. Diese wären dann schnell genug, um große Mengen aktuell aufgefangener Information zeitnah zugreifbar zu machen.

2.3 Definition des Ausgangspunkts für die Situationseinschätzung

Angefangen von einem eingehenden Notruf, über den Ausbruch einer neuartigen Krankheit, bis zur Nachricht über den Beginn eines Krieges, lassen sich unterschiedliche Ereignisse aufführen, die ein entschiedenes und schnelles Handeln auf Grundlage einer umfassenden und aktuellen Einschätzung der Situation notwendig machen. Um für ein konkretes Ereignis relevante Informationen zu identifizieren, muss dieses, genau wie die aufgefangenen Informationen selbst, zuerst semantisch annotiert werden. Die Annotation des Ereignisses bildet dann den Ausgangspunkt für die Einschätzung der Situation.

2.4 Aktivierungsausbreitung im Semantic Web

Ausgehend von der semantischen Definition des Ausgangspunkts, findet eine Aktivierung entlang bereits im Semantic Web existierender, definierter Relationen statt, um möglicherweise relevante Informationen auffindig zu machen. Vergleichbar mit dem Ansatz des *Relationship Finder* [He09b], werden Verbindungen gesucht, die zwischen den im Web 2.0 aufgefangenen Informationen und einem konkreten Ausgangspunkt bestehen. Diese werden dann auf Grundlage unterschiedlicher Kriterien, wie Länge der Verbindung, zeitlicher, örtlicher und thematischer Zugehörigkeit, in unterschiedliche Wichtigkeitsklassen eingeteilt und dementsprechend angezeigt.

Die große Anzahl an Relationen im Semantic Web (z.B. die ca. zwei Milliarden Triple innerhalb der LOD-Cloud [Ha08]), bieten zum einen ein hohes Potential, zum anderen aber auch die Schwierigkeit wichtige von unwichtigen Relationen zu unterscheiden. Im Gegensatz zu vielen Ansätzen aus dem Bereich der *Spreading Activation* [CL75], können Relationen im Semantic Web nicht direkt mit Gewichten versehen werden, da in den meisten Fällen die hierfür notwendigen Schreibrechte fehlen. Daher können automatisch erlernte oder manuell gesetzte Bewertungen über die Wichtigkeit der Relationen meist nur an separater Stelle abgelegt werden, was die globale Zusammenarbeit zur Generierung dieser Information, sowie den Zugriff darauf erschweren.

2.5 Visualisierung, Erkenntnisgewinn und Aktualisierung

Graph-Visualisierungen eignen sich in besonderer Weise, um für einen konkreten Ausgangspunkt als relevant eingestufte Informationen, zusammen mit den sie verbindenden Relationen darzustellen. Im Graph können, über verschiedene Verbindungen bestehende Zusammenhänge auf einmal und in kombinierter Form angezeigt werden. Viele Knoten und Kanten führen jedoch häufig zu Überschneidungen und Überlappungen und somit zu einer unübersichtlichen Darstellung von Informationen [FTH05]. Mit dem *ChainGraph*-Ansatz [HL09] und dem Tool *gFacet* [HZ09] wurden, neben vielen weiteren Ansätzen zur Optimierung von Graph-Visualisierungen [DiB94], bereits Lösungen für dieses Problem vorgestellt.

Neben der kombinierten Darstellung unterschiedlicher Relationen, ist insbesondere die Nachvollziehbarkeit durch den Nutzer ein wichtiger Aspekt. Hierbei muss sowohl ersichtlich werden, aus welchem Grund bestimmte Informationen aktiviert wurden, als auch, aus welchen Quellen diese Informationen stammen. Ohne dieses Wissen können die angezeigten Informationen nicht verantwortungsvoll in die Gesamteinschätzung einer Situation einbezogen werden. Dies ist insbesondere deshalb der Fall, weil im Web 2.0 häufig unvollständige, redundante, inkonsistente oder sogar falsche Informationen eingestellt werden.

Dennoch bieten diese Informationen Einblick in Geschehen, die ansonsten nicht nachvollziehbar wären und können daher zu neuen Erkenntnissen über die Situation führen. Um diese umgehend zur Aktivierung weiterer relevanter Informationen einsetzen zu können, müssen sie in die Definition des Ausgangspunkts für die Einschätzung einfließen. Durch einen aktualisierten Ausgangspunkt werden möglicherweise andere, definierte Relationen im Semantic Web aktiviert und somit neue, für die Einschätzung relevante Informationen gefunden.

3 Diskussion

In diesem Beitrag wurde ein Ansatz vorgestellt, der die Einschätzung komplexer und schwer zu überblickender Situationen unterstützt. Durch die Verwendung von Informationen aus dem Web 2.0 und Technologien des Semantic Web können neue, bisher nicht berücksichtigte Perspektiven in die Entscheidungsfindung mit einfließen, was zu angemesseneren Reaktionen führen kann und dadurch Schaden minimiert.

3.1 Verwandte Arbeiten

Es existieren bereits viele Arbeiten zu Entscheidungsunterstützungssystemen (engl. Decision Support Systems, DSS) [TAL04]. *Hermes* [BLF08] nutzt beispielsweise bereits Technologien des Semantic Web, um relevante Informationen auffindig zu machen und neue Informationen zu klassifizieren. Allerdings muss der Nutzer relevante Informationen zeitaufwendig über die Definition von Filtern selbst zusammensuchen und bekommt diese nicht, wie im hier beschriebenen Ansatz, automatisch auf Grundlage einer Ausgangspunktbeschreibung angezeigt. Da-

durch geht kostbare Zeit verloren, die insbesondere in Gefahrensituationen von entscheidender Bedeutung sein kann.

Literatur

- [BLF08] Borsje, J.; Levering, L.; Frasincar, F.: *Hermes: a Semantic Web-Based News Decision Support System*. In: Proceedings of the ACM Symposium on Applied Computing, 2008; S. 2415-2420.
- [CL75] Collins, Allan M.; Loftus, Elizabeth F.: *A spreading-activation theory of semantic processing*. Psychological Review, 82(6), 1975; S. 407-428.
- [DiB94] Di Battista, G.; Eades, P.; Tamassia, R.; Tollis, I.G.: *Algorithms for Drawing Graphs: An Annotated Bibliography*. Computational Geometry, 1994; S. 235-282.
- [FTH05] Frasincar, F.; Telea, A.; Houben, G.-J.: *Adapting Graph Visualization Techniques for the Visualization of RDF Data*. Visualizing the Semantic Web, Springer, 2005; S. 154-171.
- [Ha08] Hausenblas, M.; Halb, W.; Raimond, Y.; Heath, T.: *What is the Size of the Semantic Web?* In: Proceedings of the International Conference on Semantic Systems (I-SEMANTICS 08), Graz, Austria, 2008; S. 9-16.
- [He09a] Heim, P.; Stegemann, T.; Lohmann, S.; Ziegler, J.: *Semantisch unterstützte Informationsextraktion aus Dokumentenmengen*. In: Mensch und Computer (M&C 09), 2009.
- [He09b] Heim, P.; Hellmann, S.; Lehmann, J.; Lohmann, S.; Stegemann, T.: *RelFinder: Revealing Connections in RDF Knowledge Bases*. In: Proceedings of the 4th International Conference on Semantic and Digital Media Technologies (SAMT 09), 2009.
- [HL09] Heim, P.; Lohmann, S.: *ChainGraph: A New Approach to Visualize Shared Properties in Resource Collections*. In: Proceedings of the 9th International Conference on Knowledge Management (I-KNOW 09), 2009.
- [HZ09] Heim, P.; Ziegler, J.: *Graph-basierte facettierte Erstellung von semantischen Suchanfragen*. In: Mensch und Computer (M&C 09), 2009.
- [LHZ07] Lohmann, S.; Heim, P.; Ziegler, J.: *Semantic Integrator: Semi-Automatically Enhancing Social Semantic Web Environments*. In: Proceedings of the 1st Conference on Social Semantic Web (CSSW), Bonn: Köllen, 2007; S. 167-172.
- [TAL04] Turban, E.; Aronson, J.; Liang, T.-P.: *Decision Support Systems and Intelligent Systems*. Prentice Hall, 2004.