

Steffen Lohmann, Philipp Heim, Timo Stegemann, Lena Tetzlaff und Jürgen Ziegler

# Entdecken und Explorieren von Zusammenhängen im Semantic Web

## Discovering and Exploring Relationships in the Semantic Web

Semantic Web\_Relationship Discovery\_Semantic User Interfaces\_Linked Data\_Interactive Graph Visualization

**Zusammenfassung.** Auch wenn das heutige Web eine gewaltige Menge an Informationen bereithält, unterstützt es nur unzureichend die Entdeckung und Exploration von Zusammenhängen. Hier verspricht das Semantic Web durch eine Graph-basierte Informationsrepräsentation prinzipiell bessere Unterstützung. Allerdings müssen die im Semantic Web beschriebenen Zusammenhänge erst extrahiert und in eine für den Nutzer verständliche Darstellung gebracht werden. Aus diesem Grund wurde der RelFinder entwickelt, der Zusammenhänge in Datenbeständen des Semantic Web auffindet, nachvollziehbar visualisiert und für den Nutzer erfahrbar macht. Um die interaktive Exploration zu unterstützen, wurden neuartige Visualisierungs- und Interaktionsformen umgesetzt, die erst durch die Repräsentationen des Semantic Web ermöglicht werden. Dieser Artikel zeigt anhand einer Nutzerstudie, welche Potenziale sich damit gegenüber den textbasierten Darstellungsformen des heutigen Web ergeben. Es wird deutlich, dass das Semantic Web in Kombination mit intuitiven Darstellungs- und Bedienkonzepten potenziell sehr geeignet ist, um einen schnellen Überblick über Zusammenhänge zwischen Informationsobjekten zu erhalten.

**Summary.** Although today's Web offers a massive amount of information, its support for the discovery and exploration of relationships is often not sufficient. The graph-based information representation of the Semantic Web promises better support for these kinds of tasks. However, the relationships described in the Semantic Web need first to be extracted and presented in a user-friendly way. The RelFinder has been developed exactly for this purpose: It searches for relationships in data sets of the Semantic Web and provides an interactive visualization to support user exploration. In doing so, it implements novel visual and interactive features that are only possible due to the representation forms of the Semantic Web. The article presents a user study that illustrates the potentials of this approach compared to the text-based presentation forms of the current Web. It shows the suitability of the Semantic Web in giving quick impressions of relationships between objects, if equipped with intuitive and usable interfaces.

## 1. Einleitung

Was haben Volkswagen und Toyota gemeinsam? Wie hängen Albert Einstein und Kurt Gödel zusammen? Von welchen Personen wurden sowohl Picasso als auch Matisse beeinflusst? Solche und ähnliche Fragestellungen sind mit Hilfe des heutigen Web häufig nur schwer zu beantworten. Mit Suchmaschinen wie Google oder in Wissenssammlungen wie

Wikipedia lassen sich Zusammenhänge in der Regel nur aufspüren, wenn die Suchbegriffe gemeinsam in Textpassagen vorkommen. Die Suchenden müssen die Textpassagen auswerten, was in einem hohen Leseaufwand resultieren kann. Dabei ist nicht sichergestellt, dass sich in den Textpassagen überhaupt Zusammenhänge finden lassen: die Begriffe können ebenso rein zufällig gemeinsam auftreten oder in unterschiedlicher Weise verwendet werden (z. B. bei homonymen Begriffen). Aus diesen Gründen bietet das heutige, Dokumenten-orientierte Web nur

wenig Unterstützung, um sich einen schnellen Überblick über Zusammenhänge zu verschaffen.

Die feingranulare Architektur des Semantic Web hingegen bietet hier prinzipiell mehr Unterstützung. Im Semantic Web werden Entitäten (wie „Berlin“ oder „Deutschland“) nicht innerhalb von Texten, sondern als eigenständige Informationsobjekte repräsentiert. Jedes Objekt ist dabei über einen Uniform Resource Identifier (URI) eindeutig beschrieben und verfügt über explizite Relationen zu anderen Objekten, mit denen es in Bezie-

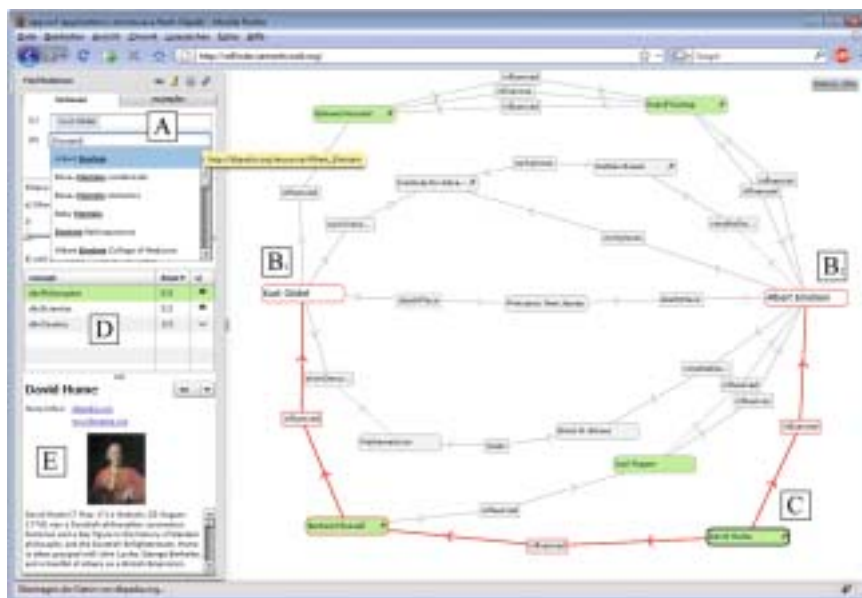


Bild 1: Die Benutzungsschnittstelle des RelFinder

hung steht (z.B. „ist Hauptstadt von“). Hierdurch entsteht eine umfangreiche Graph-Repräsentation der Informationsobjekte und ihrer Beziehungen, die auch als „Giant Global Graph“ (Berners-Lee 2007) bezeichnet wird und Suchanfragen und Analysen ermöglicht, die über die Möglichkeiten des Dokumenten-orientierten Web hinausgehen.

Mit dem RelFinder<sup>1</sup> haben wir – mit Unterstützung durch die AKSW-Gruppe der Universität Leipzig<sup>2</sup> – ein Werkzeug entwickelt, das auf die strukturierten Datenbestände des Semantic Web zugreift und Zusammenhänge zwischen beliebigen Informationsobjekten identifiziert und visualisiert. Hierbei macht sich der RelFinder die Graph-Repräsentation des Semantic Web zu Nutze, um neben direkten Beziehungen auch Zusammenhänge über mehrere Informationsobjekte hinweg zu berücksichtigen. Die gefundenen Zusammenhänge werden anschließend als Graph visualisiert und können über Filter- und Highlighting-Funktionen interaktiv exploriert werden. Im Folgenden stellen wir diesen Ansatz näher vor

und zeigen anhand einer Nutzerstudie, dass diese Form der Suche und Darstellung von Zusammenhängen gegenüber der konventionellen, textbasierten Suche und Präsentation eine Reihe von Vorteilen bietet und neue Potenziale eröffnet.

## 2. Entdecken und Explorieren von Zusammenhängen mit dem RelFinder

Grundlegende Aufgabe des RelFinder ist es, Zusammenhänge zwischen Informationsobjekten innerhalb von Datenbeständen des Semantic Web zu finden und anschaulich zu visualisieren. Darüber hinaus unterstützen Filter- und Explorationsmöglichkeiten den Nutzer bei der Analyse von gefundenen Zusammenhängen. Der RelFinder ist in Adobe Flex<sup>3</sup> implementiert und ist in jedem Webbrowser mit installiertem Flash-Player ausführbar. Er lässt sich auf beliebige Datenbestände anwenden; erforderlich ist lediglich ein Server, über den die Daten standardisiert per SPARQL<sup>4</sup> zugegriffen werden können.

Eine Beschreibung der genauen Funktionsweise und verwendeten Algorithmen des RelFinder findet sich in (Heim et al. 2009). Deshalb begnügen wir uns an dieser Stelle mit einer kurzen Erläuterung der grundlegenden Funktionsweise und wollen anschließend vielmehr auf die explorativen Funktionalitäten der Benutzungsschnittstelle eingehen, die erst

durch das Aufsetzen auf die strukturierten Datenbestände des Semantic Web ermöglicht werden.

### 2.1 Grundlegende Funktionsweise des RelFinder

Bild 1 zeigt die Benutzungsschnittstelle des RelFinder, in der nach Zusammenhängen zwischen Kurt Gödel und Albert Einstein auf Grundlage der in DBpedia<sup>5</sup> verfügbaren Relationen gesucht wurde. Wird der RelFinder gestartet, sind zunächst nur die beiden Eingabefelder in der linken oberen Ecke sichtbar (Bild 1, A), in die der Nutzer seine Suchbegriffe eingibt. Diese Suchbegriffe werden anschließend passenden Objekten der Datenbasis zugeordnet, die die Startknoten der Graph-Visualisierung bilden (Bild 1, B1 & B2). Daraufhin werden über den RelFinder-Algorithmus Informationsobjekte und Relationen gesucht, die die Startknoten miteinander verbinden. Für die Verbindungssuche können verschiedene Parameter vordefiniert und teilweise auch zur Laufzeit durch den Nutzer variiert werden, wie etwa die maximale Länge der gesuchten Verbindungen (d. h. über wie viele Informationsobjekte die Startknoten maximal miteinander verbunden sein dürfen). Die Graph-Visualisierung wird sukzessive in einer Animation aufgebaut und kann vom Nutzer interaktiv exploriert werden.

### 2.2 Disambiguierung der Sucheingaben

Eine Unterstützungsform, die erst durch die Repräsentationsformate des Semantic Web in vollem Umfang ermöglicht wird, ist die Disambiguierung von Nutzer-eingaben. Sie gewährleistet, dass Nutzer-eingaben den intendierten Objekten in der Datenbasis zugeordnet werden und beispielsweise bei homonymen Begriffen keine Fehlzusammenhänge erfolgen, die letztlich zu irrelevanten Ergebnissen und erhöhtem Nutzeraufwand bei der Auswertung führen.

Während Google keinerlei Disambiguierung anbietet – alle Verwendungen eines Suchbegriffs werden gleichbehandelt –, bietet Wikipedia hier bereits mehr Unterstützung: Nach Eingabe von mehrdeutigen Begriffen – v.a. Akronymen wie „RDF“ – erscheint vielfach eine Übersichtsseite, die die verschiedenen Begriffsverwendungen auflistet. Zusätzlich

<sup>1</sup> <http://relfinder.semanticweb.org>

<sup>2</sup> <http://aksw.org>

<sup>3</sup> <http://www.adobe.com/products/flex>

<sup>4</sup> <http://www.w3.org/TR/rdf-sparql-query/>

<sup>5</sup> Die Datenbasis von DBpedia umfasst mehr als 2,6 Millionen Informationsobjekte, zu denen in über 270 Millionen RDF-Tripeln strukturierte Informationen verfügbar sind (Bizer et al. 2009). Nahezu die komplette Datenbasis wurde automatisiert aus Wikipedia extrahiert.

unterstützen sowohl Google als auch Wikipedia den Nutzer während der Eingabe von Suchbegriffen mit einer Auto-vervollständigung, jedoch beinhaltet diese kaum Disambiguierungshilfen (bzw. sind die Hilfen im Falle von Wikipedia recht spärlich und dadurch teilweise unklar)<sup>6</sup>.

Im Gegensatz dazu findet im RelFinder jedes Mal eine Disambiguierung der Eingabe statt, was die Präzision der Suche erhöht, da irrelevante Bedeutungen eines Suchbegriffs unberücksichtigt bleiben. Dabei kommt ein Algorithmus zum Einsatz, der zunächst eine automatische Disambiguierung versucht und bei unklaren Zuweisungen eine manuelle Disambiguierung vom Nutzer verlangt. Zusätzlich erhält der Nutzer bereits während der Eingabe Disambiguierungsvorschläge, in dem passende Informationsobjekte in einer Liste unterhalb des Eingabefeldes angezeigt werden (Bild 1, A). Für die Sortierung der Vorschläge wird ein ‚count‘-Wert der Objekte – sofern verfügbar – herangezogen, der den Grad ihrer ‚Vernetztheit‘ innerhalb der Datenbasis ausdrückt (durch Auswertung der eingehenden und ausgehenden Eigenschaftsrelationen, ähnlich dem PageRank-Verfahren, vgl. Page et al. 1999). Bestenfalls sind die Vorschläge somit nach der Wahrscheinlichkeit sortiert, mit der sie auf die Nutzereingabe zutreffen. Als zusätzliche Unterstützung der Disambiguierung wird die URI jedes vorgeschlagenen Informationsobjekts als Tooltip angezeigt.

### 2.3 Typisierte Relationen

Im Gegensatz zu klassischen Hyperlinks sind Relationen zwischen Informationsobjekten im Semantic Web typisiert, d. h. mit semantisch eindeutigen Bezeichnern versehen, die die Relationen beschreiben. Da der RelFinder diese typisierten Relationen auswertet, um nach Zusammenhängen zwischen Informationsobjekten zu suchen, ist im Gegensatz zu Hyperlinks nicht nur bekannt, dass eine Beziehung besteht, sondern auch, welcher Art diese Beziehung ist. Somit können die gefundenen Verbindungen – wie in RDF-Graphen üblich – nach dem Subjekt-Prädikat-Objekt-Schema ähnlich wie normale Sätze gelesen werden (z. B.: Kurt

Gödel arbeitete am Institute for Advanced Studies; das Institute for Advanced Studies war auch Arbeitsstätte von Albert Einstein).

Darüber hinaus haben wir interaktive Funktionalitäten entwickelt, die die Lesbarkeit von gefundenen Zusammenhängen in der Graph-Visualisierung weiter steigern. Zum einen wird ein „roter Faden“ durch den Graphen angezeigt, sobald ein Informationsobjekt ausgewählt wurde, der alle hiermit verbundenen Beziehungen hervorhebt und dadurch das visuelle Nachverfolgen von Zusammenhängen erleichtert (Bild 1, C). Zum anderen können Informationsobjekte vom automatischen Graph-Layout entkoppelt werden, indem der Nutzer sie per Drag&Drop an die gewünschte Position verschiebt: Sobald ein Objektknoten im Graphen verschoben wird, wird er an der neuen Position verankert und ein Stecknadel-Symbol (Bild 1, C) erscheint (entsprechend nennen wir das Feature auch Pick&Pin). Will der Nutzer den Objektknoten zurück an das automatische Layout übergeben, braucht er nur das Stecknadel-Symbol anzuklicken, woraufhin die Verankerung gelöst wird. Somit ist es auf intuitive Weise möglich, einzelne Objektknoten aus dem Zentrum des Graphen ‚herauszuziehen‘, um sie separat zu betrachten.

Darüber hinaus können die Relationstypen zur Filterung und zum visuellen Highlighting eingesetzt werden. So ist es im RelFinder beispielsweise möglich, sich Relationen eines bestimmten Typs hervorheben zu lassen. Auch können Relationstypen, die nicht von Interesse sind, ausgeblendet werden. Alle gefundenen Relationstypen werden hierfür in einer sortierbaren Tabelle in der linken Seitenleiste angezeigt.

### 2.4 Klassenzugehörigkeit

Da Informationsobjekte im Semantic Web typischerweise bestimmten Klassen zugeordnet sind, lassen sich auch diese Klassen sinnvoll zur Analyse und Filterung von Suchergebnissen einsetzen. Im RelFinder werden die Klassen – genau wie die Relationstypen – in einer sortierbaren Tabelle in der Seitenleiste angezeigt und können hierüber in der Graph-Visualisierung interaktiv hervorgehoben sowie ein- und ausgeblendet werden. Im abgebildeten Beispiel wurden beispielsweise alle Informationsobjekte ausge-

blendet, die der Klasse „Country“ angehören (Bild 1, D).

Die Klassenzuweisungen werden im RelFinder darüber hinaus genutzt, um Objekte der gleichen Klasse im Graphen als zugehörig zu markieren: Wenn der Nutzer beispielsweise ein Informationsobjekt der Klasse „Philosoph“ im Graphen auswählt, werden auch alle anderen Objekte der Klasse „Philosoph“ hervorgehoben (vgl. Bild 1), so dass möglicherweise ebenfalls relevante Informationsobjekte leichter entdeckt werden. Auf diese Weise unterstützt der RelFinder neben der Visualisierung von Beziehungen auf der Objektebene auch eine indirekte Darstellungsform für Beziehungen auf der Klassenebene durch interaktives Highlighting.

### 2.5 Zusatzinformationen

Die Datenbestände des Semantic Web verfügen häufig über Zusatzinformationen zu Klassen, Objekten und Relationen, auf die ebenfalls strukturiert zugegriffen werden kann. Im RelFinder werden verfügbare Basisinformationen zum ausgewählten Objekt in der Seitenleiste angezeigt. Im Falle von DBpedia sind dies beispielsweise eine Kurzzusammenfassung, eine exemplarische Grafik sowie Hyperlinks zum zugehörigen DBpedia-Eintrag und Wikipedia-Artikel (Bild 1, E).

## 3. Evaluation: RelFinder, Google und Wikipedia im Vergleich

Aus den angeführten Unterstützungsformen ergeben sich zahlreiche Evaluationsfragen, von denen wir einige in einer vergleichenden Nutzerstudie genauer untersucht haben. Als Vergleichsanwendungen wählten wir Google und Wikipedia als derzeit populärste Vertreter der Anwendungsklassen Suchmaschine und Wiki im Web. Zudem nahmen wir eine gewisse Vertrautheit der Studienteilnehmer mit diesen Werkzeugen an, was eine Erläuterung ihrer Funktionsweise überflüssig machte.

Da alle drei Anwendungen sehr unterschiedliche Interaktionskonzepte aufweisen und die Benutzungsschnittstellen von Google und Wikipedia im Gegensatz zum RelFinder nicht für die Suche von Zusammenhängen optimiert wurden, sind die folgenden Ausführungen eher quali-

<sup>6</sup> Alle Vergleiche mit Google und Wikipedia in diesem Artikel betreffen Versionen dieser Anwendungen, wie sie zwischen Juni und September 2009 im Web verfügbar waren.

tativer Natur. Ziel der Evaluation war es nicht, die Anwendungen als solche miteinander zu vergleichen, sondern vielmehr, die umgesetzten Darstellungs- und Interaktionsformen mit Blick auf die Suche und Exploration von Zusammenhängen zu bewerten. Insbesondere interessierte uns, in welchen Fällen eine Graph-basierte Visualisierung von Beziehungen der textbasierten Darstellung vorzuziehen ist und welche qualitativen Schlussfolgerungen sich daraus ableiten lassen.

### 3.1 Studiendesign

An der Nutzerstudie nahmen insgesamt zwölf Probanden im Alter von 21 bis 47 Jahren teil, die aus Studenten und Mitarbeitern der Universität rekrutiert wurden. Die allgemeine Vertrautheit der Probanden mit Google und Wikipedia war hoch (Google:  $M = 9$ ,  $SD = 1,2$ ; Wikipedia:  $M = 8,8$ ,  $SD = 1,0$ ; auf einer Skala von 1–10 mit Ausprägungen von 1 = „gar nicht vertraut“ bis 10 = „sehr vertraut“).<sup>7</sup> Der RelFinder war den meisten Probanden unbekannt; lediglich vier Probanden hatten vor der Studie bereits erste Erfahrungen mit dem RelFinder gesammelt und äußerten eine gewisse Vertrautheit ( $M = 4,7$ ;  $SD = 2,2$  für die vier Probanden).

Aufgabe der Probanden war es, mit allen drei Anwendungen Zusammenhänge zwischen jeweils vorgegebenen Informationsobjekten zu suchen. Bei der Auswahl der Informationsobjekte haben wir uns auf solche beschränkt, zwischen denen sich in einem Pre-Test mit vergleichsweise geringem Aufwand Verbindungen in allen drei Anwendungen finden ließen (z. B. achteten wir im Falle von Google darauf, dass sich nach Eingabe der Objektbezeichnungen möglichst auf der ersten, höchstens auf der zweiten Ergebnis-seite Lösungen finden ließen). Außerdem wählten wir Informationsobjekte aus drei allgemeinen Kategorien (Personen, Orte und Kultur), wobei jeder Studienteilnehmer Objekte aus jeder Kategorie präsentiert bekam.

Die Nutzerstudie umfasste drei Typen von Aufgaben, die mit allen drei Anwendungen bearbeitet werden sollten. In der

ersten Aufgabe sollte nach einer, in der zweiten nach drei Verbindungen gesucht werden. Auch in der dritten Aufgabe galt es, drei Verbindungen zu suchen, doch wurde hier zusätzlich die Klasse vorgegeben, der die Informationsobjekte in den gefundenen Verbindungen angehören sollten. Im Folgenden sind konkrete Beispiele für drei Nutzeraufgaben angeführt, wie sie in der Studie verwendet wurden:

- Aufgabe 1 (Kategorie: Orte): Welche Verbindung besteht zwischen Heidelberg und Karlsruhe? Bitte nennen Sie eine Verbindung sowie die Art der Verbindung.
- Aufgabe 2 (Kategorie: Personen): Welche Verbindungen bestehen zwischen Johann Wolfgang von Goethe und Friedrich Schiller? Bitte nennen Sie 3 Verbindungen sowie die Art der Verbindungen.
- Aufgabe 3 (Kategorie: Kultur): Nennen Sie 3 Filme, die Verbindungen zwischen Quentin Tarantino und Samuel L. Jackson herstellen. Geben Sie auch hier die Art der Verbindungen an.

Die konkreten Objekte, zwischen denen Verbindungen gesucht werden sollten, wurden über die Werkzeuge hinweg variiert, genau wie die Klasse, nach der in der dritten Aufgabe gefragt wurde (im obigen Beispiel „Film“). Die Bearbeitungszeit pro Aufgabe war auf max. drei Minuten festgesetzt. Konnte ein Studienteilnehmer eine Aufgabe innerhalb dieser Zeit nicht lösen, wurde die Bearbeitung abgebrochen und mit der nächsten Aufgabe fortgesetzt.

Als Datenbasis für den RelFinder verwendeten wir in der Nutzerstudie DBpedia (s. Abschnitt 2.1). Da es sich hierbei um eine englischsprachige Datenbasis handelt, beschränkten wir auch die Suche in Google und Wikipedia auf englischsprachige Inhalte. Ferner stellten wir sicher, dass die Probanden über Google nicht auf Wikipedia-Artikel zugriffen. Ansonsten nahmen wir keine Einschränkungen vor, so dass die Probanden beliebige Strategien anwenden konnten, um die Aufgaben mit den jeweiligen Anwendungen zu lösen. Die Unterschiedlichkeit der durchsuchten Datenbestände ließ sich nicht vermeiden, erachteten wir jedoch als nicht weiter problematisch, da durch die Pre-Tests

sichergestellt war, dass sich alle Aufgaben in allen Werkzeugen prinzipiell mit einer Bearbeitungszeit unterhalb der vorgegebenen drei Minuten beantworten ließen.

### 3.2 Studiendurchführung

Die Studie begann mit einer kurzen Einführung und Erläuterung der drei Aufgabentypen, die es von den Teilnehmern zu bearbeiten galt. Anschließend wurden die drei Anwendungen in systematisch variierender Reihenfolge präsentiert, wobei mit jeder Anwendung alle drei Aufgabentypen beantwortet werden sollten. Die Reihenfolge der vorgegebenen Informationsobjekte blieb über alle Durchgänge gleich, so dass die in den Anwendungen einzugebenden Suchbegriffe ebenfalls systematisch variierten (s. o.). Zusätzlich sollte die Zufriedenheit mit der gefundenen Lösung nach jeder Aufgabe auf einer Skala von 1–10 (Ausprägungen: 1 = „gar nicht zufrieden“, 10 = „voll zufrieden“) kurz bewertet werden.

Nachdem alle Aufgaben mit einer Anwendung bearbeitet wurden, folgte ein Fragebogen, in dem die Unterstützung der jeweiligen Anwendung bei der Aufgabenbearbeitung noch einmal zusammenfassend hinsichtlich der Dimensionen Zufriedenheit, Effizienz und Sicherheit bewertet werden sollte.<sup>8</sup> Hatten die Probanden die Aufgaben mit allen drei Anwendungen bearbeitet, sollten sie abschließend auch die verschiedenen Darstellungs- und Interaktionsformen hinsichtlich ihrer Eignung zur Unterstützung der Aufgabenbearbeitung miteinander vergleichen und bewerten. Dies geschah sowohl schriftlich über einen Fragebogen als auch mündlich in einem offenen Interview.

Die Probanden waren angehalten, ihr eigenes Hintergrundwissen zu den befragten Informationsobjekten bei der Aufgabenbearbeitung soweit möglich auszublenden. In jedem Fall durften sie nur Verbindungen nennen, die sie nachweislich über die Anwendungen gefunden haben. Entsprechend war es nicht erlaubt, weitere Begriffe bei der Suche zu verwenden, die auf eigenem Vorwissen beruhen (z. B. „Weimarer Klassik“, wenn nach Verbindungen zwischen Goethe und Schiller gefragt wurde).

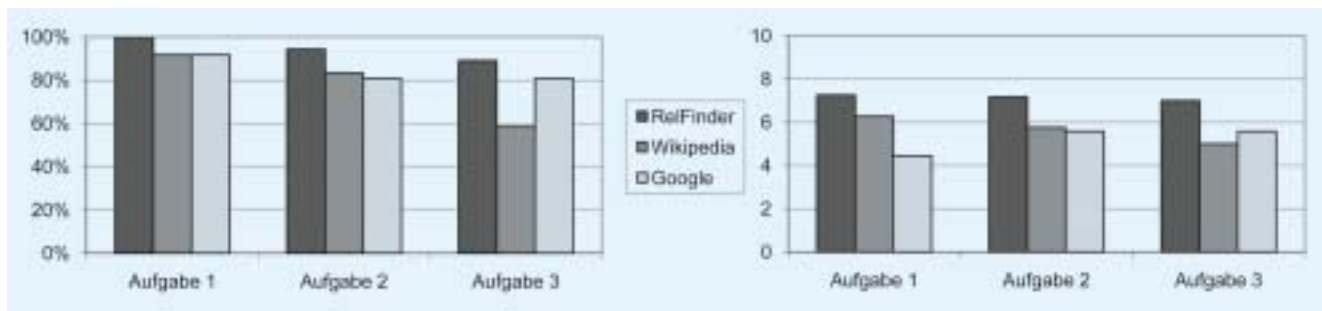
### 3.3 Ergebnisse

Bild 2a gibt die durchschnittliche Anzahl an gelösten Aufgaben in Prozent je

<sup>7</sup>  $M$  = Mittelwert;  $SD$  = Standardabweichung

<sup>8</sup> Die Bewertungen auf den verschiedenen Dimensionen wurden durch mehrere Aussagen auf einer 5-fach abgestuften Likert-Skala von 1 = „stimme nicht zu“ bis 5 = „stimme voll zu“ abgeprüft.





**Bild 2:** a) Durchschnittliche Anzahl an erfolgreich gelösten Aufgaben (in Prozent je Aufgabentyp); b) Durchschnittliche Zufriedenheit mit der Aufgabenlösung (auf einer Skala von 1–10 je Aufgabentyp)

Anwendung und Aufgabentyp wieder.<sup>9</sup> Im Vergleich der Anwendungen liefert der RelFinder augenscheinlich über alle Aufgabentypen hinweg die beste Unterstützung. Ein ähnliches Ergebnis zeigen die Zufriedenheitsbewertungen, die direkt im Anschluss an die Lösung der Aufgaben abgefragt wurden und in Bild 2b dargestellt sind. Diese Ergebnisse deuten darauf hin, dass eine Graph-basierte Suche und Darstellung von Zusammenhängen wie im RelFinder gegenüber den textbasierten Alternativen vorzuziehen ist.

Ein differenzierteres Bild erlauben die Nutzerbewertungen und -kommentare, die wir über die Fragebögen und Interviews erfasst haben. Bild 3 zeigt die Bewertungen der Probanden auf den drei Dimensionen, die wir direkt im Anschluss an die Präsentation jeder Anwendung über die Likert-Skala abgefragt haben. Auch hier waren die Probanden mit den Unterstützungsformen des RelFinder insgesamt zufriedener als mit denen von Google und Wikipedia. Die Sicherheit im Umgang mit den Anwendungen wurde etwa gleich bewertet. Dies spricht für ein gelungenes Interaktionsdesign des RelFinder, da die Vertrautheit der Probanden mit Google und Wikipedia – wie oben angegeben – sehr viel höher war. Am deutlichsten ist der Unterschied jedoch

bei der Effizienz – hier verzeichnet der RelFinder die besten Werte. Dieses Ergebnis deutet darauf hin, dass die Graph-basierte Darstellung im RelFinder Zusammenhänge sehr viel schneller erkennen lässt als die textbasierten Darstellungen in Google und Wikipedia.

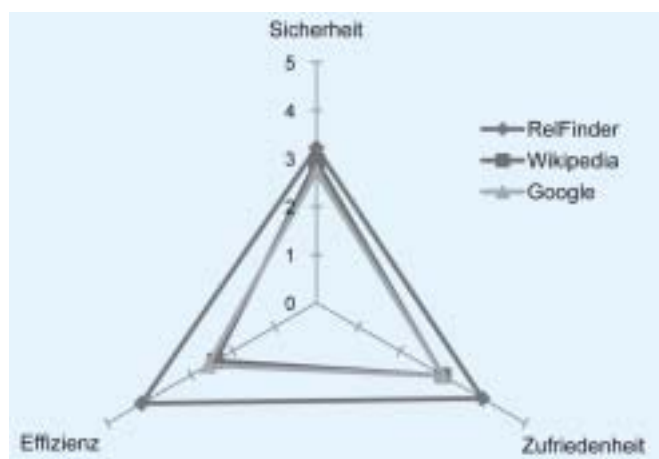
Die Auswertung der freien Antworten der Studienteilnehmer in den Interviews führte zu teilweise konträren Ergebnissen: Während sieben Teilnehmer die Graph-Darstellung als besonders übersichtlich hervorhoben, bewerteten sie drei als unübersichtlich. Diese unterschiedlichen Bewertungen sind vermutlich darauf zurückzuführen, dass die Graph-Dichte in Abhängigkeit der vorgegebenen Informationsobjekte variierte (auch wenn sehr hohe Graph-Dichten durch die automatische Aktivierung von initialen Filtern im RelFinder vermieden werden, vgl. Heim et al. 2009). Vier Teilnehmer hoben die Zusatzinformationen in der Seitenleiste des RelFinder als wesentlich hervor, um Weiteres zu den Informationsobjekten und deren Zusammenhängen zu erfahren. Drei Teilnehmer wussten die gerichteten Kanten und die Bezeichnungen der Relationen in der

Graph-Darstellung nicht recht zu interpretieren.

Auf die Frage „Wenn ich nach Verbindungen zwischen zwei Begriffen suchen würde, dann würde ich folgendes Tool verwenden“ entschieden sich acht Teilnehmer für den RelFinder und je zwei für Google und Wikipedia. Begründet wurde die Entscheidung beim RelFinder insbesondere mit der übersichtlichen Darstellung der Zusammenhänge, der Reduzierung auf das Wesentliche und der Zuweisung von Nutzereingaben zu Informationsobjekten, die sich gesondert handhaben und anschaulich gegenüberstellen lassen. Die teilweise recht langen Antwortzeiten des Servers bis zur Anzeige der Disambiguierungsvorschläge und Suchergebnisse wurden beim RelFinder als verbesserungswürdig angemerkt. Ebenso wurde kritisiert, dass die im RelFinder gefundenen Beziehungen in vielen Fällen nicht zufriedenstellend sind, was jedoch der verwendeten Datenbasis (in diesem Fall also DBpedia) zuzuschreiben ist. Bei Wikipedia und Google wurden die Fülle und Tiefe der dargebotenen Informationen als Vorteil genannt. Die Entscheidung für Google oder Wikipedia wurde

**Bild 3:**

Bewertungen der drei Anwendungen bezüglich Sicherheit, Zufriedenheit und Effizienz im Zusammenhang mit der Aufgabenbearbeitung (aggregiert über die Zustimmung zu Aussagen auf einer 5-stufigen Likert-Skala)



<sup>9</sup> Die zur Aufgabenbearbeitung benötigte Zeit ließ sich als objektives Maß nicht verwenden, da die Probanden zum einen nicht angehalten waren, die Aufgaben möglichst schnell zu lösen – Vorgabe war lediglich, unterhalb der drei Minuten zu bleiben – und die Bearbeitungszeiten zum anderen starken, interindividuellen Schwankungen unterlagen, die sich unter anderem aus den variierenden Antwortzeiten der Anwendungen ergaben (v.a. beim RelFinder, dessen Performanz in der Studie wesentlich von den Antwortzeiten des DBpedia-Servers abhängig war).

in erster Linie mit der höheren Vertrautheit bei der Bedienung dieser Anwendungen begründet.

Die Frage nach der grundsätzlich hilfreichen Form der Darstellung ergab eine klare Präferenz für die Graph-Visualisierung: Die Probanden sagten aus, dass sich die Graph-Darstellung sowohl für die schnelle Suche nach Verbindungen ( $M = 4,1$ ;  $SD = 0,7$ ) als auch für das Verständnis der Art von Verbindungen ( $M = 4,0$ ;  $SD = 0,7$ ) und das Nachvollziehen von Zusammenhängen ( $M = 3,9$ ;  $SD = 0,8$ ) besser eignen würde als eine Textdarstellung (Differential von 1–5; Pole Text- und Graph-Darstellung); wobei dieses Ergebnis sicherlich stark durch die Art der Aufgaben in der Studie beeinflusst ist. Für eine andere Art von Aufgaben, die ein tieferes Textverständnis verlangen, würden die Bewertungen vermutlich weniger eindeutig ausfallen.

## 4. Diskussion und Ausblick

Der RelFinder schnitt in der vergleichenden Nutzerstudie insgesamt sehr positiv ab. Dies liegt natürlich unter anderem daran, dass er für die Art von Aufgabenstellungen, wie sie von den Studienteilnehmern zu bearbeiten waren, entwickelt und optimiert wurde. Auch aufgrund der geringen Probandenzahl erlauben die Ergebnisse nur limitierte Aussagen, Interpretationen und Schlussfolgerungen. Nichtsdestotrotz geben der RelFinder und die Ergebnisse der Studie einen Einblick in die Möglichkeiten und Potenziale, die durch die Datenbestände des Semantic Web in Kombination mit einfach zu bedienenden Benutzungsschnittstellen entstehen können.

Das Semantic Web verspricht Antworten auf neue Arten von Fragestellungen sowie effiziente Darstellungs- und Zugriffsformen basierend auf strukturiert vorliegenden Informationen. Der RelFinder ist eine Beispielanwendung, die konkret zeigt, welche Vorteile eine Graph-basierte Repräsentation und Visualisierung gegenüber textuell vorliegenden Informationen haben kann. Insbesondere beim Zugriff auf Faktenwissen und für ein schnelles Entdecken von Zusammenhängen scheint diese Form der Repräsentation und Visualisierung vorteilhaft. Zweifelsohne sind für ein tieferes Ver-

ständnis von Zusammenhängen und betroffenen Informationsobjekten jedoch zusätzliche, textbasierte Inhalte notwendig. Deshalb ist es wichtig, dass auch diese Zusatzinhalte bzw. Verlinkungen zu entsprechenden Texten in den Datenbanken des Semantic Web hinterlegt sind, so dass der Nutzer bei Bedarf auf diese zugreifen kann.

In der angeführten Nutzerstudie haben wir eine ältere Version des RelFinder eingesetzt, die Zusammenhänge zwischen lediglich zwei Informationsobjekten sucht. In der aktuellen Version können auch zwischen mehr als zwei Informationsobjekten Zusammenhänge ermittelt werden, was neue Anwendungsfälle ermöglicht. Zudem lässt sich die aktuelle RelFinder-Version über eine XML-Datei auf einfache Weise für beliebige SPARQL-Schnittstellen konfigurieren. Aktuell testen wir den RelFinder beispielsweise im medizinischen Bereich, indem wir auf die Datenbasis des Linking Open Drug Data (LODD)-Projekts<sup>10</sup> aufsetzen. Dieser Anwendungsbereich verspricht interessante Einsatzmöglichkeiten, wie beispielsweise die Suche nach nicht-offensichtlichen Zusammenhängen beim kombinierten Einsatz von Medikamenten.

Der RelFinder integriert aktuell unter anderem die Suche von Zusammenhängen in der LOD-Cloud,<sup>11</sup> in der verschiedene Datenbestände des Semantic Web miteinander verknüpft sind (vgl. Bizer, Heath und Berners-Lee 2009). Was beim Zugriff auf die LOD-Cloud – aber auch auf andere Datenbestände wie DBpedia oder LODD (s. o.) – immer wieder offenbar wird, ist die noch unzureichende Qualität der dort repräsentierten Relationen. Vielfach lassen sich keine sinnvollen und relevanten Zusammenhänge finden. Entsprechend schwierig war es, im Rahmen der Pre-Tests für die Studie Beispiele zu finden, die sowohl im RelFinder als auch in Google und Wikipedia zu einigermaßen sinnvollen Zusammenhängen führten. Zudem sind die verwendeten Relationstypen in vielen Datenbanken bisher wenig differenziert und kaum weiterführend beschrieben. Da die Nutzbarkeit von Anwendungen wie dem RelFinder jedoch wesentlich von der Qualität der

zugrundeliegenden Datenbanken abhängt, ist deren Bedeutung für den Erfolg des Semantic Web nicht zu unterschätzen. Auch die teilweise sehr langen Antwortzeiten auf SPARQL-Anfragen stellen immer noch ein Problem dar, da sie einer flüssigen Nutzerinteraktion entgegenstehen. Hier zeichnen sich jedoch bereits vielversprechende Lösungsansätze ab, die die Geschwindigkeit der Verarbeitung von SPARQL-Anfragen verbessern (z. B. Stocker et al. 2008), so dass dieses Problem nur eine Frage der Zeit zu sein scheint.

## Danksagung

Wir danken Jens Lehmann und Sebastian Hellmann von der Forschungsgruppe AKSW der Universität Leipzig für ihre Unterstützung bei der Umsetzung des RelFinder sowie Jörg Schüppel, der zusammen mit Jens Lehmann und Sören Auer die grundlegende Idee und Basisfunktionalität für den RelFinder entwickelt hat (Lehmann, Schüppel und Auer 2007).

## Literatur

- Berners-Lee, T.: Giant Global Graph. Decentralized Information Group (DIG) Breadcrumbs, 2007; <http://dig.csail.mit.edu/breadcrumbs/node/215>
- Bizer, C.; Heath, T.; Berners-Lee, T.: Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems* **5** (3) (2009) 1–22; im Druck
- Bizer, C.; Lehmann, J.; Kobilarov, G.; Auer, S.; Becker, C.; Cyganiak, R.; Hellmann, S.: DBpedia – A Crystallization Point for the Web of Data. *Journal of Web Semantics* **7** (2009) 154–165; im Druck
- Heim, P.; Hellmann, S.; Lehmann, J.; Lohmann, S.; Stegemann, T.: RelFinder: Revealing Relationships in RDF Knowledge Bases. *Proceedings of the 4th International Conference on Semantic and Digital Media Technologies*. Heidelberg, Berlin: Springer, 2009; im Druck
- Lehmann, J.; Schüppel, J.; Auer, S.: Discovering Unknown Connections – The DBpedia Relationship Finder. *Proceedings of the 1st Conference on Social Semantic Web*. Bonn: Köllen, 2007; S. 99–110
- Page, L.; Brin, S.; Motwani, R.; Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. *Stanford InfoLab*, 1999.
- Stocker, M.; Seaborne, A.; Bernstein, A.; Kiefer, C.; Reynolds, D.: SPARQL Basic Graph Pattern Optimization Using Selectivity Estimation. *Proceeding of the 17th International World Wide Web Conference*. New York, NY: ACM, 2008; S. 595–604,

<sup>10</sup> <http://esw.w3.org/topic/HCLSIG/LODD>

<sup>11</sup> <http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData>



1



2



3



4



5

**1 Steffen Lohmann** ist seit Herbst 2009 am DEI Laboratory der Universidad Carlos III in Madrid tätig. Zuvor war er wissenschaftlicher Mitarbeiter am Lehrstuhl für Interaktive Systeme und Interaktionsdesign der Universität Duisburg-Essen. Seine Forschungsinteressen liegen in den Bereichen Mensch-Computer-Interaktion und webbasierte Systeme. Insbesondere beschäftigt er sich mit Nutzer- und Community-orientierten Entwicklungsprozessen, Wissensmanagement und -visualisierung sowie neuartigen Interaktionsformen im Social und Semantic Web.

E-Mail: [slohmam@inf.uc3m.es](mailto:slohmam@inf.uc3m.es)

**2 Philipp Heim** studierte Computervisualistik an der Universität Koblenz-Landau und schrieb 2006 seine Diplomarbeit extern am DFKI (Deutsches Forschungszentrum für künstliche Intelligenz) in Saarbrücken. Nach seinem Studium war er zwei

Jahre als wissenschaftlicher Mitarbeiter am Lehrstuhl für Interaktive Systeme und Interaktionsdesign an der Universität Duisburg-Essen beschäftigt und wechselte danach nach Stuttgart. Seitdem ist er als wissenschaftlicher Mitarbeiter am Lehrstuhl für Visualisierung und interaktive Systeme an der Universität Stuttgart angestellt und beschäftigt sich mit interaktiver Graph-Visualisierung in Kombination mit Semantic-Web-Technologien.

E-Mail: [philipp.heim@vis.uni-stuttgart.de](mailto:philipp.heim@vis.uni-stuttgart.de)  
<http://www.vis.uni-stuttgart.de/~heimpp/>

**3 Timo Stegemann** studiert Angewandte Informatik an der Universität Duisburg-Essen und ist seit April 2008 studentische Hilfskraft am Lehrstuhl für Interaktive Systeme und Interaktionsdesign. Dort beschäftigt er sich mit der Gestaltung und Umsetzung von innovativen User Interfaces und interaktiven Visualisierungen im Bereich webbasierter Systeme.

E-Mail: [timo.stegemann@uni-due.de](mailto:timo.stegemann@uni-due.de)  
[www.interactivesystems.info](http://www.interactivesystems.info)

**4 Lena Tetzlaff** erhielt ihren Bachelor of Science in Angewandte Kommunikations- und Medienwissenschaften an der Universität Duisburg-Essen im Jahr 2008 und führt dort aktuell ihr Studium zur Erreichung des Mastergrades fort. Sie arbeitet seit einem Jahr als studentische Hilfskraft am Lehrstuhl Interaktive Systeme und Interaktionsdesign der Universität Duisburg-Essen und unterstützt dort die Durchführung von Evaluationen.

E-Mail: [lana.tetzlaff@uni-due.de](mailto:lana.tetzlaff@uni-due.de)  
[www.interactivesystems.info](http://www.interactivesystems.info)

**5 Prof. Dr.-Ing. Jürgen Ziegler** ist seit 2003 Inhaber des Lehrstuhls Interaktive Systeme und Interaktionsdesign der Universität Duisburg-Essen. Seine Forschungsschwerpunkte liegen in den Feldern Mensch-Computer-Interaktion, Informationsvisualisierung und kontextadaptive Webanwendungen.

E-Mail: [juergen.ziegler@uni-due.de](mailto:juergen.ziegler@uni-due.de)  
[www.interactivesystems.info](http://www.interactivesystems.info)