

Interaktive Exploration und Visualisierung von semantischem Wissen mit SemaPlorer

Interactive Exploration and Visualization of Semantic Knowledge with SemaPlorer

Keywords: Exploration, Visualisierung, Facetted Browsing, Semantisches Wissen, Web 2.0, SemaPlorer

Zusammenfassung. SemaPlorer ist eine einfach zu bedienende Anwendung, die es Endanwendern erlaubt, einen verteilten, sehr großen Datensatz gemischter Qualität und von heterogener Semantik in Echtzeit zu explorieren und zu visualisieren. Benutzer können sich damit über eine interessante Region wie eine Stadt oder Ferienregion informieren. Die Visualisierung erfolgt mit Hilfe einer Karte, Medienansicht und verschiedenen kontextuellen Sichten auf die Daten, die es dem Benutzer erlauben, interaktiv mit den Datensätzen zu interagieren. Für SemaPlorer verwenden wir verschiedene semantische Datenquellen wie DBpedia, GeoNames, WordNet und persönliche FOAF-Dateien. Zudem ist ein großer, nach RDF konvertierter Datensatz von Flickr integriert worden. Weitere Datenquellen können sehr einfach in SemaPlorer hinzugefügt werden. Wir haben eine formative Evaluierung der SemaPlorer-Anwendung mit 20 Testpersonen durchgeführt. Die Ergebnisse dieser Evaluation werden analysiert und deren Auswirkung auf zukünftige Arbeiten skizziert.

Summary. SemaPlorer is an easy to use application that allows end users to interactively explore and visualize a very large, mixed-quality and semantically heterogeneous distributed semantic data set in real-time. Its purpose is to acquaint oneself about a city, touristic area, or any other area the user is interested in. The data is visualized using a map, media, and different context views. By this, the users can interact with the large data set. SemaPlorer leverages different semantic data sources like DBpedia, GeoNames, WordNet, and personal FOAF files. It connects with a large Flickr data set converted to RDF. Additional data sources can easily be added to SemaPlorer. We conducted a formative evaluation of SemaPlorer with 20 test subjects. The results of this evaluation are analyzed and their implication to future work is outlined.

1 Einleitung

Informationen über Reiseziele werden heutzutage selbstverständlich im Internet gesucht. Dazu werden zahlreiche Wikis, Portale und Webseiten aufgesucht, die eine unüberschaubare Anzahl von Texten, Bildern und Metainformationen enthalten, die von Internetnutzern online gestellt werden. Diese für den Benutzer schnell, sinnvoll und optisch ansprechend nutzbar zu machen, ist eine Herausforderung, der mit der von den Autoren entwickelten SemaPlorer-Anwendung Rechnung getragen wird. Die SemaPlorer-Anwendung verknüpft verschiedene, sehr große Datenquellen unterschiedlicher Herkunft und Qualität auf intelligente Art und Weise und stellt sie dem Benutzer dar. Im Gegensatz zu bestehenden Kartenbasierten Anwendungen, wie zum Beispiel Google Maps, erlaubt die semantische Verknüpfung der Informationsquellen in SemaPlorer die Suchanfragen über verschiedene Datenquellen wie zum Beispiel GeoNames oder Flickr. So kann zum Beispiel die Suchanfrage nach Fotos auf Flickr zu einem bestimmten Thema wie „streetart“ über Orte in GeoNames erfolgen. Damit kann der Benutzer Anfragen wie „Finde mir alle Streetart-Bilder um den Fernmeldeturm in Berlin“ stellen. Die Ergebnismenge kann dabei hinsichtlich weiterer Kriterien eingeschränkt werden, wie zum Beispiel entlang des Erstellungszeitpunktes der Bilder oder entlang der Benutzer, die diese Bilder hochgeladen haben. Zudem hat die SemaPlorer-Anwendung im Gegensatz zu bestehenden Kartenbasierten Anwendungen explizite Kenntnis über die Art der Anfragen. Wird zum Beispiel die Suche nach Fotos weiter eingeschränkt auf „space invaders“, eine spezielle Art von Streetart, so kann dies in der SemaPlorer-Anwendung durch einfaches Auswählen aus den weiteren Tags zu den aktuell dargestellten Bildern erfolgen. Bei existierenden, Kartenbasierten Anwendungen ist dem System dieser Suchkontext nicht

bekannt. So werden beispielsweise in Google Maps die Kategorien Restaurants und Unterkünfte als relevant angezeigt. Analog kann, beispielsweise um Straßenkunst in Paris mit Straßenkunst in Berlin zu vergleichen, der Suchkontext in der SemaPlover-Anwendung durch Austausch der Ortsanfrage auf Berlin einfach durchgeführt werden. Ein solcher Kontextwechsel ist beispielsweise mit Google Maps nicht möglich.

2 Verteilte Infrastruktur für semantische Daten

Mit der SemaPlover-Anwendung greifen wir auf ein Netz verbundener Datenbestände zu. Diese sind in einer einzigen verteilten Infrastruktur integriert, um generischen Zugang zu den semantischen Multimedia-Daten zu erhalten. Die verschiedenen Datenbestände werden über SPARQL-Endpunkte zur Verfügung gestellt. Über solche Endpunkte können semantische Datenbanken, die Wissen in der Sprache RDF (Resource Description Framework) vorhalten, über die Anfragesprache SPARQL angesprochen werden. Damit können nahezu beliebige Datenquellen ad hoc zur Dateninfrastruktur von SemaPlover hinzugefügt werden. Um Informationen aus dieser verteilten Infrastruktur abzurufen und zu visualisieren, bedienen wir uns mit der SemaPlover-Anwendung dem sogenannten „Blended Browsing and Querying“-Ansatz (Munroe, Ludscher und Papakanstantinou 2000). Die Nutzer können sich durch nahezu beliebige Datensätze unter Verwendung verschiedener Ansichten (Facetten) wie Ort, Zeit, Personen und Tags navigieren (Hearst 2006). Wenn der Benutzer mit der Anwendung interagiert, werden dabei gleichzeitig mehrere Anfragen an die zugrunde liegende Speicher-Infrastruktur gesendet, um die entsprechenden Ergebnisse zu berechnen. Die Ergebnisse werden mittels einer Karte, Medien- und Kontextansichten, die die verschiedenen Facetten repräsentieren, dargestellt.

Für SemaPlover haben wir verschiedene semantische Datenquellen wie DBpedia (<http://dbpedia.org>), eine semantische Version von Wikipedia, GeoNames (<http://geonames.org>), eine umfangreiche Datenbank mit geo-referenzierten Orten, WordNet (<http://wordnet.princeton.edu>) mit einer Abbildung des englischen Sprachvokabulars und persönliche FOAF-Dateien aus der semantischen Suchmaschine Swoogle (<http://swoogle.umbc.edu>) integriert. Darüber hinaus haben wir einen partiellen Crawl, also eine partielle lokale Kopie von Flickr (<http://flickr.com>) erstellt und als einen sehr großen, nicht-semantischen Datensatz eingebunden, dessen RDF Version ca. 700 Millionen Triples umfasst. Der Datensatz umfasst alle Annotationen von Fotos auf Flickr von ca. Mai 2005 bis April 2006. Zusammen bilden diese Datenbestände einen sehr großen, semantisch heterogenen Datensatz von gemischter Qualität, die zusammen über eine Milliarde Triples ergeben. Die Verknüpfung dieser Daten erfordert eine flexible und skalierbare Speicherstruktur. Die SemaPlover-Infrastruktur besteht aus 25 RDF-Datenbanken. Die Datenbanken werden in virtuellen Maschinen auf Amazons Elastic Computing Cloud (<http://aws.amazon.com/ec2>) gehostet. Sie können wie ein einziger, virtueller RDF Speicher über einen Federator angesprochen werden. Der Federator verwendet die von den Autoren entwickelte Technologie NetworkedGraphs (Schenk und Staab 2008), einen SPARQL-basierten, verteilten View-Mechanismus für RDF und verteilte Auswertung von SPARQL-Anfragen. NetworkedGraphs erlaubt einfaches, regelbasiertes Schließen zur Laufzeit, zum Beispiel für die Integration semantisch heterogener Daten. Das Verteilen von Anfragen innerhalb der Infrastruktur wird durch eine – ebenfalls RDF-basierte – Konfiguration gesteuert und kann zur Laufzeit angepasst werden. Daher wird das Hinzufügen neuer Datenquellen sehr einfach und für die SemaPlover-Anwendung vollkommen transparent.

3 SemaPlover-Anwendung

Je komplexer die Anfragen nach Informationen über eine interessante Region wie eine Stadt oder eine Ferienregion sind desto schwerer können heutzutage Suchmaschinen und Plattformen nützliche Informationen liefern. So lassen sich beispielsweise Webseiten über Städte wie Berlin sehr einfach über Standard-Suchmaschinen wie Google finden. Andererseits ist es z.B. fast unmöglich, Orte mit Straßenkunst in Berlin zu finden. Diese Anfrage auf eine andere Stadt wie z.B. Paris zu übertragen, stellt eine zusätzliche Herausforderung für die die Anwendung dar, die die traditionellen Ansätze nicht lösen können. Mit der SemaPlover-Anwendung unterstützen wir die Anwender bei der Durchführung solch komplexer Datenexplorationen über verschiedene Datenquellen hinweg. Dabei integrieren wir das Navigieren mit Hilfe von Facetten und die traditionelle Volltextsuche und erlauben dem Nutzer somit eine frühzeitige Auflösung von möglicherweise mehrdeutigen Suchtermen. SemaPlover unterstützt verschiedene generische Facetten, wie zum Beispiel Ort, Personen und Tags. Andere Facetten können einfach konfiguriert und hinzugefügt werden.

Eine Facette kann verstanden werden als ein Filter für große Datenmengen. Zum Beispiel kann SemaPlover die Sehenswürdigkeiten einer bestimmten Stadt oder Gegend unter der Verwendung der Ort-Facette filtern und darstellen und dabei

ausschließlich Fotos von bestimmten Benutzern zeigen. Während der Benutzer mit SemaPlorer interagiert, werden unmittelbar verschiedene Anfragen im Hintergrund erstellt. Die Ergebnisse der Anfragen werden sofort in der visuellen Ansicht in der Anwendung hinzugefügt und dargestellt. Dieser Ansatz ermöglicht eine vom Nutzer gesteuerte Darstellung und interaktive Exploration der verwendeten semantischen Daten. In der SemaPlorer-Anwendung formuliert der Benutzer zunächst eine einfache Anfrage in Textform, die in der oberen linken Ecke von Bild 1 am Beispiel von „berlin“ dargestellt ist. Die Suchergebnisse werden darunter dargestellt und sind in die drei Kategorien Orte, Personen und Tags aufgeteilt. So finden sich unter Orte die Stadt Berlin (in Deutschland), der Fernmeldeturm in Berlin, aber auch die Stadt Berlin in Pennsylvania. Zudem werden verschiedene Personen mit dem Nachnamen Berlin angezeigt und Tags bzw. Worte aus Flickr und WordNet. Klickt der Benutzer auf einen Eintrag in der Ergebnisliste wie beispielsweise auf die Stadt Berlin, wechselt der Benutzer von der initialen, textuellen Suche hinüber zum facettierten Browsing. Dazu aktualisiert SemaPlorer die Ansicht in der Mitte von Bild 1, welche eine Stadtkarte von Berlin zeigt. Gleichzeitig werden Anfragen ausgeführt und die Ergebnisse als Pins in der Karte dargestellt. Wiederum gleichzeitig werden Anfragen ausgeführt, die den rechten Teil von Bild 1 mit Kontextinformationen füllen. Die Medienansicht ergänzt die Kartensicht durch eine einfache Möglichkeit in den Flickr-Bildern zu browsen (in Bild 1 nicht dargestellt).



Bild 1: Screenshot der SemaPlorer-Anwendung mit Straßenkunst in Berlin

Für jede Facette in SemaPlorer ist eine Kontextansicht definiert. Die in Bild 1 dargestellte Ort-Facette bietet z.B. Informationen aus DBpedia, beispielsweise Bevölkerung und Land. Es werden Sehenswürdigkeiten und Orte in der Nähe gezeigt („nearby places“). Die Personen-Facette enthält Persönlichkeiten, die mit diesem Ort in Verbindung stehen, Flickr-Benutzer, die geo-referenzierte Bilder aus dieser Region hochgeladen haben und Internet-Nutzer, die in dieser Region leben – identifiziert anhand ihrer FOAF-Dateien. In der Tag-Facette werden Schlagworte von Flickr (Englisch: tags) dargestellt. Da alle Facetten, wie Sehenswürdigkeiten, nahe gelegene Orte, Persönlichkeiten und Tags, interaktiv sind, kann der Benutzer in der weiteren Verwendung der SemaPlorer-Anwendung über diese Facetten in den Kontextansichten navigieren. Zum Beispiel können die Benutzer, wenn die Karte in SemaPlorer die Stadt Berlin zeigt, auf den Tag „street art“ (Straßenkunst) klicken. Sofort wird die Kartenansicht aktualisiert und die Standorte der Flickr Fotos, welche mit „streetart“ annotiert sind, angezeigt. Durch die Eingabe einer weiteren textuellen Suche nach „paris“ kann der Nutzer zwischen dem aktuellen Kontext, nämlich Straßenkunst in Berlin, zu Straßenkunst in Paris wechseln und die Straßenkunst in den beiden Städten vergleichen. Dazu klickt der Benutzer in der Ergebnisliste auf den Eintrag Paris unter der Kategorie Orte. Die SemaPlorer-

Anwendung tauscht darauf hin den aktuellen Ortskontext Berlin durch Paris aus. Dies ist möglich, da SemaPlover - im Gegensatz zu zum Beispiel Google Maps - explizites Wissen über das Konzept Ort hat.

In der aktuellen Version der SemaPlover-Anwendung werden bestimmte Fragestellungen, die sich bei der Interaktion mit der Karte ergeben können, noch nicht betrachtet. Dies betrifft zum Beispiel die Frage, wie mit einer großen Ergebnismenge wie zum Beispiel eine große Anzahl an gefundenen Photos umgegangen wird. Dies ist Gegenstand aktueller Untersuchungen und soll in zukünftigen Versionen von SemaPlover berücksichtigt werden.

4 Semantische Vernetzung der Daten

Um die facettierte, interaktive Suche und Visualisierung in SemaPlover zu unterstützen, werden die semantischen Daten aus DBpedia, GeoNames, WordNet, FOAF-Dateien und Flickr kombiniert und verschiedene Kontextansichten für SemaPlover definiert, wie im vorangegangenen Abschnitt beschrieben. Diese Kontextansichten ergeben sich aus den Eigenschaften beziehungsweise den zur Verfügung gestellten Informationen der verwendeten Daten. Im Folgenden beschreiben wir die verwendeten Daten entlang der in SemaPlover definierten Facetten und erläutern, wie sie miteinander verbunden sind.

4.1 Ort

Elemente dieser Facette beziehen sich auf die geographischen Koordinaten. Wir setzen GeoNames für Orte aller Art ein wie beispielsweise Städte und Länder. Für Sehenswürdigkeiten verwenden wir eine Kombination von GeoNames mit der Volltext-Suche auf Artikelbeschreibungen aus DBpedia und deren Kategoriebeschreibungen. Artikel in DBpedia sind mit Hilfe des SKOS Vokabulars klassifiziert (Simple Knowledge Organization System, <http://www.w3.org/2004/02/skos>), z.B. der Fernsehturm als „Sehenswürdigkeit in Berlin“. Mit SKOS können Systeme zur Wissensorganisation beschrieben werden wie Thesauri, Klassifikationsschemas und Taxonomien. Zur Erkennung von Sehenswürdigkeiten betrachten wir die SKOS-Kategorien in DBpedia, insbesondere SKOS:broader, welches hierarchische Beziehungen zwischen SKOS-Konzepten beschreibt, und berechnen die transitive Hülle aller SKOS:broader Beziehungen. Außerdem nutzen wir eine Volltext-Suche auf den Kategorienamen und schränken die Ergebnisse auf Einträge ein, die in der Kategorie dbpedia:Visitor_attractions einsortiert sind. Für die Anzeige der nahegelegenen Orte und Sehenswürdigkeiten wählen wir alle Geschwister eines gewählten Standortelements und sortieren sie auf der Grundlage der geografischen Distanz. Wenn z.B. der Arc de Triomphe in Paris ausgewählt wurde, werden als nahe Orte der Eiffelturm und Notre Dame angezeigt. Zusätzlich werden Bilder von Flickr dargestellt, die mit Geoinformationen versehen sind und sich im relevanten Kartenausschnitt befinden.

4.2 Person

In den von SemaPlover verwendeten Datensätzen haben wir drei Arten von Personen identifiziert: Diese sind Persönlichkeiten aus DBpedia, Flickr-Benutzer, die Bilder eingestellt haben, und Internet-Nutzer, die ihre FOAF-Dateien veröffentlicht haben und über Swoogle zugreifbar sind. Für jede dieser Kategorien von Personen verwenden wir eine andere Kombination der Daten. Für Persönlichkeiten wählen wir Bilder, die die ausgewählten Persönlichkeiten zeigen, basierend auf einer Volltext-Suche auf den Flickr-Tags. In Bezug auf einen Flickr-Nutzer suchen wir nach Inhalten, die durch den Benutzer veröffentlicht wurden. Für Internet-Nutzer betrachten wir den Geostandort in der FOAF-Datei (falls vorhanden) und verbinden sie mit Bildern von diesem Ort aus Flickr.

4.3 Tags

Tags stehen im direkten Zusammenhang mit den Flickr-Inhalten und können über Volltextsuche recherchiert werden. Wenn ein Tag von einem Nutzer ausgewählt wurde, zeigen wir verwandte Tags von Flickr sowie WordNet. Hinsichtlich Flickr sind dies alle Tags der aktuell dargestellten Photos. Verwandte Tags in WordNet sind die Synonyme des aktuellen Tag.

4.4 Erfolge und Erfahrungen

Bei der Erstellung des Datensatzes für unsere SemaPlover-Anwendung haben wir bemerkt, dass die Datensätze oft nicht vollständig und manchmal auch bezüglich der Semantik nicht eindeutig genug sind: Zum Beispiel fehlen in GeoNames zu einem beliebigen Eintrag Informationen über Sehenswürdigkeiten und Orte in der Nähe – Informationen, die in der HTML-

Version vorhanden sind. Trotzdem konnten wir diese Informationen durch die Verbindung der einzelnen Datensätze, wie oben beschrieben, gewinnen. Des Weiteren haben wir beobachtet, dass die Daten auch innerhalb eines einzelnen Datensatzes heterogen sind. Zum Beispiel gibt es keinen klaren Lösungsansatz für die Angabe des Geburtsortes einer Person in DBpedia. Manchmal ist es `dbpedia:cityofbirth` und manchmal `dbpedia:placeofbirth`. In SemaPlorer lösen wir diese Unklarheiten durch die Zusammenfassung der beiden Eigenschaften in einem View. Ein View erlaubt eine bestimmte Sicht auf einen Datensatz zu legen und ermöglicht somit die Vereinheitlichung der beiden Modellierungsvarianten der DBpedia durchzuführen. Während Linked Open Data, also die Verknüpfung von semantischen Datenbeständen, fortschreitet, ist es immer noch eine offene Frage, wie es für die Verwaltung von Ressourcen wie Flickr-Bilder zu nutzen ist. Wie SemaPlorer zeigt, ist eine Kartierung der Linked Open Data und die semantische Beschreibung der Flickr-Daten in RDF möglich und funktioniert z.B. mit GeoNames gut. Doch statt der Kennzeichnung von Bildern mit Tags und anschließender Kartierung dieser Tags in Zusammenhang mit Open Data wäre es gewinnbringender, direkt die semantisch reichen Linked Open Data zur Annotation, also Beschreibung der Bilder zu verwenden. Zum Beispiel könnte ein Bild, das den Eiffelturm zeigt, direkt mit dem entsprechenden Konzept für den Eiffelturm aus der DBpedia annotiert werden.

5 Planung und Durchführung der Evaluation von SemaPlorer

Die SemaPlorer-Anwendung ist entwickelt worden, um die Skalierbarkeit von semantischen Webtechnologien zu demonstrieren. Als solches wurde die SemaPlorer-Anwendung und ihre zu Grunde liegende Infrastruktur als ein technischer Demonstrator, aber nicht als eine Endbenutzer-Anwendung, die in einer echten produktiven Umgebung läuft, entworfen. Um in einer solch frühen Phase eine Rückmeldung über die Benutzbarkeit und Nützlichkeit der Anwendung und Verbesserungsvorschläge zu den Features der Anwendung zu erhalten, wurde eine formative Evaluation durchgeführt. Wir baten 20 Personen aus dem Institut für Informatik der Universität Koblenz-Landau (11 Doktoranden, 9 Studierende), SemaPlorer auszuprobieren. Die Personen sind zwischen 21 und 26 Jahren alt und haben gute bis sehr gute Computer-Kenntnisse. 18 Teilnehmer haben bereits Erfahrung mit der Nutzung von Kartenbasierten Anwendungen zur Informationsbeschaffung und Visualisierung und sind daher gute Kandidaten zur Ermittlung relevanter Rückmeldungen.

Die Evaluation der SemaPlorer-Anwendung wurde in drei Phasen unterteilt, nämlich Einführung, Test und Rückmeldung. In der Einführungsphase wurden die Teilnehmer mit SemaPlorer und seinen Features vertraut gemacht. Den Teilnehmern wurde erklärt, dass es nicht um die Messung ihrer Leistungen bei der Abarbeitung der Evaluationsaufgaben geht, sondern um die Gewinnung von Erkenntnissen zur Verbesserung von SemaPlorer. In der Test-Phase hatte jeder Teilnehmer eine festgelegte Zahl von Aufgaben auszuführen. Die Aufgaben, die ausgeführt werden sollten, sind die initiale, textuelle Suche nach der Stadt Berlin und die Verwendung des „Sights“-Features gewesen. Dann sollte das Ergebnis durch Hinzufügen des „streetart“-Tags auf die Anzeige von Bildern zu Straßenkunst eingeschränkt werden und Bilder mit Straßenkunst rund um den Berliner Sendeturm unter Benutzung des „nearby places“-Features erkundet werden. Eine spezielle Form der Straßenkunst sind „Space Invaders“-Piktogramme, die durch Hinzufügen des „spaceinvaders“-Tag gefunden werden. Die Teilnehmer wurden gebeten, „Space Invaders“ in Berlin zu finden. Anschließend sollte der Ortsbezug auf Paris abgeändert werden, um dort „Space Invaders“ zu suchen. Um Paris weiter zu erkunden, wurden die Testpersonen gebeten, nach bestimmten Flickr-Usern und interessanten Bildern, die diese aufgenommen haben, zu suchen. Zusätzlich sollten die Testpersonen nach Persönlichkeiten in Paris suchen. Schließlich sollten die Benutzer entlang semantischer Relation zu dem Wort Paris in WordNet navigieren.

Eine solch einheitliche Aufgabenstellung ist wichtig, um eine Vergleichbarkeit zwischen den einzelnen Testpersonen herstellen zu können und eine valide Rückmeldung zu erhalten. Die Teilnehmer sollten die Aufgaben soweit sie konnten selbstständig durchführen. Falls eine Aufgabe von einem Teilnehmer nicht erledigt werden konnte oder es Rückfragen gab, wurden den Teilnehmern zusätzliche Hilfestellungen und Erklärungen gegeben.

In der nachfolgenden Feedback-Phase füllten die Teilnehmer einen Fragebogen aus. Dieser wurde in Anlehnung an IsoMetrics-L erstellt (Gediga und Hamborg 1999). Es wurde jedoch keine explizite Gewichtung der einzelnen Fragen vorgenommen, sondern den Benutzern die Möglichkeit gegeben, punktuell subjektive Rückmeldungen zu geben. Der Fragebogen hat zum Ziel die Benutzbarkeit der SemaPlorer-Anwendung und die Akzeptanz der implementierten Features zu beurteilen. Es wurden Fragen zur Suche in den semantischen Daten, zur Darstellung der Ergebnisse in der Karten- und Medienansicht, zur Interaktion entlang der verschiedenen Facetten sowie zur Performanz der Ergebnisse gestellt. Der Fragebogen ist in Tabelle 1 dargestellt. Die Analyse der Ergebnisse wird im Folgenden beschrieben.

6 Analyse der Evaluationsergebnisse

In der letzten Phase, der Rückmeldung, wurden die Testpersonen gebeten den Fragebogen auszufüllen, um damit Feedback über die bereits implementierten Features in SemaPlorer und die Anwendung als Ganzes zu bekommen. Die Fragen konnten gemäß IsoMetrics auf einer Skala von 1 bis 5 bewertet werden, bei der 1 „absolut keine Zustimmung“ und 5 „absolute Zustimmung“ bedeutet. Die einzelnen Features der SemaPlorer-Anwendung sind im Fragebogen im Durchschnitt mit Werten zwischen 0.9 bis 3.3 beurteilt worden.

Wir erklären uns diese Bewertung in den unteren beiden Dritteln der Skala durch die Heterogenität der Daten, die für die SemaPlorer-Anwendung genutzt wurden, die Performanz der Anwendung und die Benutzbarkeit. Die für die SemaPlorer-Anwendung verwendeten Daten stammen aus unterschiedlichen Quellen und sind von unterschiedlicher Qualität. GeoNames, DBpedia und der Flickr-Datensatz sind durch die Mitwirkung einer großen Anzahl an Benutzern entstanden. Für solche Datensätze kann die Qualität der Anfrageergebnisse nicht garantiert werden und ist stark von der jeweiligen individuellen Anfrage abhängig. Dies spiegelt sich in unserer Evaluation durch die Beurteilung der Qualität der Suchergebnisse als mittelmäßig wider (S1). Die Aufteilung der Suchergebnisse in Orte, Tags und Personen wurde ähnlich bewertet (S2).

Hinsichtlich der Bewertung des Kontextwechsels durch die Nutzung der Suchfunktion ist eine hohe Standardabweichung festzustellen (S3). Hier zeigen die Testpersonen zwei unterschiedliche Präferenzen. Die einen mögen eine facetiierte Suche und Navigation und andere bevorzugen die klassische Interaktion. Diese Streuung ergibt die vergleichsweise schlechte Beurteilung. Hier muss eine Lösung zur Unterstützung der verschiedenen Benutzerpräferenzen erarbeitet werden.

Die Benutzbarkeit von Karten- und Medienansicht wurde als durchschnittlich beurteilt (A1 und A2). Bei der Beurteilung der einzelnen Features der Facetten wurde die Auswahl der Sehenswürdigkeiten in der Ort-Facette am besten bewertet (F1). Es wurden auch interessante Ansichten gefunden (F2). Das „nearby places“-Feature wurde ähnlich gut wie die Auswahl der Sehenswürdigkeiten bewertet (F3). Allerdings sollte die Qualität der gefundenen „nearby places“ verbessert werden (F4), die als eher schlecht beurteilt wurde. Um die Qualität zu erhöhen benötigen wir bessere Daten über benachbarte Orte als uns bisher zur Verfügung stehen, wie im Abschnitt über die semantische Vernetzung der Daten kurz diskutiert wurde. Die Navigation entlang von WordNet (F5) und die Auswahl von Prominenten aus DBpedia (F6 und F7) wurden beide als "teils-teils" bewertet. Wir denken, dass hier insbesondere das Feature der Navigation entlang WordNet zu hinterfragen ist und eventuell entfernt werden sollte. Nur die Funktion, über Flickr-User zu navigieren, wurde von den Teilnehmern abgelehnt. Anscheinend lieferte diese Suchfunktion nur sehr wenige oder uninteressante Bilder von Persönlichkeiten oder Flickr-Benutzern (F8 und F9). Hier finden wir ebenfalls eine sehr hohe Standardabweichung, die eine strikte Ablehnung dieser Features bei einem Teil der Testpersonen widerspiegelt. Bei den Testpersonen, die eine „teils-teils“-Bewertung vorgenommen haben (eine 3 auf der Skala), gehen wir davon aus, dass die zu Grunde liegenden Daten für eine eindeutige Bewertung dieser Features nicht ausreichend waren.

In der letzten Phase konnten die Testpersonen außerdem zusätzliches Feedback zu den in den Fragen genannten Funktionen geben sowie Vorschläge für weitere Funktionen machen, die sie gerne in SemaPlorer hinzufügen würden. So wurden grundsätzlich alle existierenden Funktionen zur Suche, Karten- und Medienansicht und den Facetten begrüßt. Lediglich das Browsen über WordNet, die Suche nach Persönlichkeiten in DBpedia sowie nach Flickr-Benutzern wurde von vielen Testpersonen als nicht sinnvoll erachtet, da keine passenden Ergebnisse gefunden werden konnten.

Tabelle 1. Feedback zur Suchfunktion (S1-S3), Karten- und Medienansicht (A1-A2), sowie den Facetten (F1-F9) und der Performanz (P1) der SemaPlorer-Anwendung

Frage	Mittelwert	Standardabweichung
S1: Die Suchergebnisse entsprechen meinen Erwartungen.	3.3	0.9
S2: Die Aufteilung in Orte, Tags und Personen ist intuitiv.	2.8	0.7
S3: Der Kontextwechsel mittels der Suchfunktion ist intuitiv.	1.8	1.0
A1: Die Kartenansicht ist intuitiv und einfach zu benutzen.	3.0	0.6
A2: Die Medienansicht ist eine gute Ergänzung zur Kartenansicht.	3.2	0.8
F1: Ist die Funktion zur Auswahl von Sehenswürdigkeiten sinnvoll?	3.4	0.5
F2: Haben Sie interessante Sehenswürdigkeiten gefunden?	2.8	0.7
F3: Ist die Funktion "nearby places" sinnvoll?	3.1	0.6
F4: Haben Sie interessante "nearby places" gefunden?	2.2	0.9
F5: Ist die Navigation mittels WordNet sinnvoll?	2.1	1.0
F6: Haben Sie interessante Persönlichkeiten in DBpedia gefunden?	2.4	1.0
F7: Ist diese Funktion sinnvoll?	2.4	1.0

F8: Haben Sie interessante Flickr-Benutzer gefunden?	0.9	0.8
F9: Ist diese Funktion sinnvoll?	1.7	1.0
P1: Die Antwortzeiten der Anwendung entsprechen meinen Erwartungen.	2.5	1.2

Fünf von 20 Personen schrieben, dass sie eine Erhöhung der Performanz von SemaPloer begrüßen würden. Obwohl die Antwortzeiten im Allgemeinen gut waren, haben komplexere Anfragen mehr Zeit gebraucht, als sich die Tester wünschten. Im Fragebogenteil wurde die Antwortzeit von gut bis teils-teils bewertet (P1). Diese Einstufung mag zunächst überraschen, aber wir gehen davon aus, dass den Testpersonen zum Teil kommerzielle Produkte wie Google Maps als Vergleich dienten. Die sehr hohe Varianz ist ein starker Indikator dafür. Es ist daher wichtig zu betonen, dass SemaPloer keine Anwendung ist, die auf einem Produktiv-Server läuft wie z.B. Google Maps, sondern eine technische Demonstration ist, die die Skalierbarkeit von Semantic Web Technologien zeigt. Außerdem wurden einige Vorschläge für Verbesserungen zur Benutzbarkeit der Anwendung gemacht, wie zum Beispiel den Ortswechsel über das Facetten-Menü intuitiver zu gestalten. Hinsichtlich weiterer Funktionalitäten wurden bspw. eine Verlaufsfunktion genannt, welche das Vor- und Zurückspringen in den Navigationsschritten ermöglicht, und die Auswahl mehrerer Orte, um eine Reise zu planen. Eine Person fügte als Anmerkung hinzu, dass bereits zu viele Features vorhanden sind. Sehr interessant war der Vorschlag, ein Bewertungssystem für die Vertrauenswürdigkeit der gelieferten Informationen einzubauen.

7 Verwandte Arbeiten

Der Grundgedanke der facettierten, interaktiven Suche ist die Exploration von großen Datenmengen und ist seit längerem bekannt (Yee et al. 2003). Der Gewinner der Semantic Web Challenge 2006, /facet (Schraefel et al. 2005), hat diese Idee in den Bereich von semantischen Daten eingebracht. Vor kurzem ist die Anwendung Freebase Parallax (<http://mqix.com/~david/parallax>) veröffentlicht worden, ein facettierter Browser für Exploration und Visualisierung der strukturierten Daten von Freebase (<http://www.freebase.com>). Der größte Nachteil von /facet und Freebase Parallax ist, dass sie auf zentralisierten Infrastrukturen basieren, die keinen skalierbaren Einsatz von einer großen Anzahl von Daten aus vielen verschiedenen Datenquellen erlauben. Mit SemaPloer haben wir dies erreicht und sorgen für eine facettierte, interaktive Suche und Visualisierung über einen sehr großen Satz von semantisch heterogenen und verteilten Daten von unterschiedlicher Qualität. Zwar existieren verschiedene Systeme, die hoch skalierbares Management von RDF-Daten ermöglichen, z.B. YARS2 (Harth et al. 2007). Diese Systeme zielen jedoch auf die Steuerung eines großen Volumens von RDF-Daten in einem einzigen, wenn auch möglicherweise hardwaremäßig verteilten Repository ab und nicht auf die Verknüpfung mehrerer verteilter Repositories, wie die für SemaPloer verwendete Infrastruktur. Hinsichtlich der Benutzungsschnittstelle verfolgen /facet, Freebase Parallax und SemaPloer den Ansatz einer facettierten Suche und Navigation und unterscheiden sich hier abgesehen vom konkreten Design der Benutzungsoberfläche nicht. Neben den genannten Systemen zur facettierten Suche und Navigation in semantischen Datenbeständen existieren außerdem noch klassische, Kartenbasierte Anwendungen wie zum Beispiel DBpedia Mobile (Becker und Bizer, 2008).

Kartenbasierte Anwendungen wie SemaPloer sollen interaktiv sein und den Benutzer in der Durchführung einfacher Analyseaufgaben unterstützen (Wisniewski et al. 2009). Existierende Evaluationen haben sich dabei auf unterschiedliche Aspekte konzentriert, wie z. B. die Interaktion mit einer Karte auf dem mobilen Endgerät (Wilson et al. 2006), die Navigation in einer Kartenbasierten 3D-Umgebung (Swan et al. 2003) oder der Vergleich zwischen einer 2D- und 3D-Kartennavigation (Porathe und Prison 2008). Zur facettierten, interaktiven Suche und Visualisierung existieren umfangreiche Designempfehlungen basierend auf langjährigen Erfahrungen und Evaluationen (Hearst 2006; Wilson et al. 2009). Die Evaluation einer facettierten, Kartenbasierten Anwendung wie SemaPloer, die sich der Verknüpfung sehr großer, semantischer Datenquellen unterschiedlicher Herkunft und Qualität bedient, ist bisher nicht untersucht worden.

8 Zusammenfassung

In diesem Artikel haben wir die SemaPloer-Anwendung präsentiert. Wie gezeigt wurde, ist SemaPloer ein einfach zu bedienendes Werkzeug, das dem Endnutzer erlaubt, interaktiv sehr große, verteilte, semantische Datenmengen von unterschiedlicher Qualität interaktiv zu explorieren und zu visualisieren. Die Evaluation von SemaPloer hat gezeigt, dass die interaktive Exploration und Navigation in dem großen semantischen Datenbestand mit Hilfe von Facetten für die Benutzer anwendbar ist. Bei der Weiterentwicklung von SemaPloer werden die Evaluationsergebnisse berücksichtigt. Eine zentrale

Herausforderung der Zukunft wird der sinnvolle Umgang mit der heterogenen Qualität der Datensätze sein. Des Weiteren können im Semantic Web keine Annahmen über die Art der zu visualisierenden Daten gemacht werden. Dies erfordert Benutzungsschnittstellen, die sich flexibel an die Anforderungen und das Schema der Daten anpassen können.

Danksagung: Diese Forschung wurde co-finanziert von der EU im 6. RP im NoE K-Space (027026) und dem Neon-Projekt (027595) sowie im RP7 im WeKnowIt Projekt (215453).

Literatur

Arndt, R.; Troncy, R.; Staab, S.; Hardman, L.; Vacura, M.: COMM: Designing a Well-Founded Multimedia Ontology for the Web. In: *International Semantic Web Conference*. Springer, 2007.

Becker, C.; Bizer, C.: DBpedia Mobile: A Location-Enabled Linked Data Browser. In: *1st Workshop about Linked Data on the Web*. Beijing, China: CEUR-WS, 2008.

Gediga, G.; Hamborg, K.-C.: IsoMetrics: An usability inventory supporting summative and formative evaluation of software systems. In: *HCI International*. Lawrence Erlbaum Associates, 1999.

Harth, A.; Umbrich, J.; Hogan, A.; Decker, S.: YARS2: A Federated Repository for Querying Graph Structured Data from the Web. In: *International Semantic Web Conference*. Springer, 2007.

Hearst, M.A.: Design recommendations for hierarchical faceted search interfaces. In: *SIGIR Workshop on Faceted Search*. ACM, 2006.

Hildebrand, M.; van Ossenbruggen, J.; Hardman, L.: /facet: A Browser for Heterogeneous Semantic Web Repository. In: *International Semantic Web Conference*. Springer, 2006.

Munroe, K.D.; Ludäscher, B.; Papakonstantinou, Y.: Blending Browsing and Querying of XML in a Lazy Mediator System. In: *Extending Database Technology*. Springer, 2000.

Porathe, T.; Prison, J.: Design of human-map system interaction. In: *Extended abstracts on Human factors in computing systems*. ACM, 2008.

Schenk, S.; Staab, S.: NetworkedGraphs: a declarative mechanism for SPARQL rules, SPARQL views and RDF data integration on the web. In: *WWW*. ACM, 2008.

schraefel, m.c.; Smith, D.A.; Owens, A.: The evolving mspace platform: leveraging the semantic web on the trail of the Mex. In: *Hypertext*. ACM, 2005.

Swan, J.E.; Gabbard, J.L.; Hix, D.: A Comparative Study of User Performance in a Map-Based Virtual Environment. In: *Virtual Reality*. IEEE, 2003.

Wilson, M.; Russel, A.; schraefel, m.c.; Smith, D.A.: mSpace mobile: a UI gestalt to support on-the-go info-interaction. In: *Extended abstracts on Human factors in computing systems*. ACM, 2006.

Wilson, M.L.; schraefel, m.c.; White, R.W.: Evaluating Advanced Search Interfaces using Established Information-Seeking Models. *American Society for Information Science and Technology* 7 (2009) 1407-1422.

Wisniewski, P.K.; Pala, O.; Lipford, H.R.: Grounding geovisualization interface design: a study of interactive map use. In: *Extended abstracts on Human factors in computing systems*. ACM, 2009.

Yee, K.P.; Swearingen, K.; Li, K.; Hearst, M.: Faceted metadata for image search and browsing. In: *Human factors in computing systems*. ACM, 2003.

1 Ansgar Scherp

Dr. Scherp leitet die Fokusgruppe Interactive Web der Arbeitsgruppe Informationssysteme und Semantic Web an der Universität Koblenz-Landau. Herr Scherp hat in Oldenburg studiert und dort mit Auszeichnung promoviert. Er war 2006-2008 als EU Marie Curie Fellow an der University of California in Irvine, USA tätig. Seine Interessensgebiete sind Multimedia, die semantische Modellierung von Multimedia-Inhalten, Event-basierte Multimedia-Systeme und Human-centered Computing. E-Mail: scherp@uni-koblenz.de

2 Simon Schenk

Simon Schenk ist Doktorand in der Arbeitsgruppe Informationssysteme und Semantic Web. Er hat an der FH NORDAKADEMIE und der Universität Karlstadt, Schweden Wirtschaftsinformatik studiert. Vor dem Start seiner Promotion war Herr Schenk als Unternehmensberater tätig. Herr Schenk forscht im Bereich Views und Regeln, Anfragesprachen und Trust Inference im Semantic Web.

E-Mail: sschenk@uni-koblenz.de

3 Carsten Saathoff

Carsten Saathoff ist als wissenschaftlicher Mitarbeiter in der Forschungsgruppe Informationssysteme und Semantic Web an der Universität Koblenz-Landau tätig und promoviert dort gerade zum Thema Semantische Annotation von Multimedia-Inhalten. Er hat an der Universität Oldenburg studiert und dort sein Diplom im Bereich Data Warehouses und Data Mining gemacht. Seine Forschungsinteressen liegen im Bereich Ontology Engineering, Multimedia Reasoning und (halb-)automatischer Annotation.

E-Mail: saathoff@uni.koblenz.de

4 Steffen Staab

Prof. Dr. Steffen Staab hat in Erlangen und Philadelphia studiert, wurde in Freiburg promoviert und hat sich in Karlsruhe habilitiert. Seit 2004 ist er Professor für Datenbanken und Informationssysteme an der Universität Koblenz-Landau. Seine Forschungsinteressen liegen im Bereich der Web Forschung und semantischer Technologien und ihrer Anwendungen für Zwecke des Wissensmanagement, des Software Engineering und der Verwaltung von Multimedia-Inhalten.

E-Mail: staab@uni-koblenz.de