

Linking Open Data: Anwendungen für Endanwender

Herbert Mühlburger

(Graz University of Technology
muehlburger@student.tugraz.at)

Zusammenfassung Dieser Artikel beschäftigt sich mit Anwendungen, die das Linking Open Data Konzept verwenden. Zuerst wird auf das Open Data Movement eingegangen und anschließend auf das W3C SWEO Linking Open Data Project. Es sind zahlreiche Anwendungen entstanden, die Linked Open Data verwenden und daraus einen Mehrwert für Benutzer erzielen. Eine dieser Anwendungen ist Freebase, die von der Firma Metaweb Technologies Inc. entwickelt wurde. Weitere Anwendungen sind aus dem MIT SIMILE Project hervorgegangen und werden hier auch aufgelistet. Der Artikel geht bewusst nicht auf Details ein, sondern stellt einen Überblick über die verschiedenen Anwendungen dar.

1 Einführung

Open Data gewinnt immer mehr an Bedeutung. Ein Beispiel dafür ist die steigende Anzahl der Artikel in Wikipedia¹. Ein Artikel in Wikipedia kann wichtige Informationen enthalten, die in anderen Systemen auch von Bedeutung sind, dort aber nicht direkt verwendet werden können. Durch das W3C SWEO Linking Open Data Community Project wird nun versucht, diese Informationen strukturierter abzuspeichern und mit Metadaten anzureichern. Diese Strukturierung und Anreicherung wird mittels semantischer Technologien erzielt. Durch diese semantischen Technologien ist es nun anderen Systemen auch möglich, Daten aus Wikipedia und anderen Open Data Sets zu verwenden. Es wird somit die Interoperability verbessert. Diese Arbeit beschäftigt sich deshalb im nächsten Kapitel zuerst näher mit Linking Open Data bevor in Kapitel 3 auf verschiedene Anwendungen für Endanwender eingegangen wird.

2 Linking Open Data

Unter Linking Open Data versteht man das Verlinken von im Web frei zugänglichen und frei verfügbaren Informationsquellen durch semantische Technologien. Durch diese Verlinkung können neue Anwendungen entstehen, die auf diesen Linked Open Data aufbauen. Auf die verschiedenen Anwendungen wird jedoch erst in Kapitel 3 näher eingegangen. Im nächsten Abschnitt wird zuerst näher auf Open Data und das Open Data Movement eingegangen.

¹ Wikipedia: <http://www.wikipedia.org/>

2.1 Open Data Movement

Das Open Data Movement² zielt darauf ab, Informationen für jeden frei zugänglich und online verfügbar zu machen. Diese frei verfügbaren Informationen werden als Open Data Sets bezeichnet. In der folgenden Liste sind einige der Open Data Sets aufgezählt:

- Wikipedia³
- Wikibooks⁴
- Geonames⁵
- MusicBrainz⁶
- WordNet⁷
- DBLP Bibliography⁸

Die Informationen dieser Open Data Sets sind durch RDF [Manola and Miller, 2004] Links beschrieben. Im Unterschied zum “herkömmlichen” Web, bei dem Dokumente über HTML Links verbunden sind, wird bei den Open Data Sets nicht von Dokumenten gesprochen, stattdessen werden die verlinkten Bestandteile als Ressourcen bezeichnet. RDF Links beinhalten im Unterschied zu HTML Links auch semantische Informationen zu der verlinkten Ressource. Nach Tim Berners-Lee wird RDF so beschrieben: “RDF is to data what HTML is to documents, ...” [Berners-Lee, 2007] Ein RDF Link beschreibt, dass eine Resource eine bestimmte Beziehung zu einer anderen Resource hat. [Bizer et al., 2008] Diese Beziehungen können unterschiedlichster Natur sein. So kann z.B. ein RDF Link, der Informationen über Menschen verbindet, ausdrücken, dass sich zwei Menschen kennen. Beschreibt ein RDF Link Informationen einer bibliographischen Datenbank, so kann durch den Link ausgedrückt werden, dass eine bestimmte Person der Autor einer Publikation ist. [Bizer et al., 2008]

Das durch diese Verlinkung entstehende Web wird auch als Web of Data oder Semantic Web bezeichnet. Im Semantic Web sind die Ressourcen aus verschiedenen Quellen miteinander verlinkt. Die Verlinkung verschiedener Informationsquellen (Data Sets) wird durch das W3C SWEO Community Project Linking Open Data (LOD) [W3C, 2008b] unterstützt.

² Open Data Movement: http://en.wikipedia.org/wiki/Open_Data

³ Wikipedia: <http://www.wikipedia.org/>

⁴ Wikibooks: <http://www.wikibooks.org/>

⁵ Geonames: <http://www.geonames.org/>

⁶ MusicBrainz: <http://musicbrainz.org/>

⁷ WordNet: <http://wordnet.princeton.edu/online/>

⁸ DBLP Bibliography: <http://www.informatik.uni-trier.de/~ley/db/>

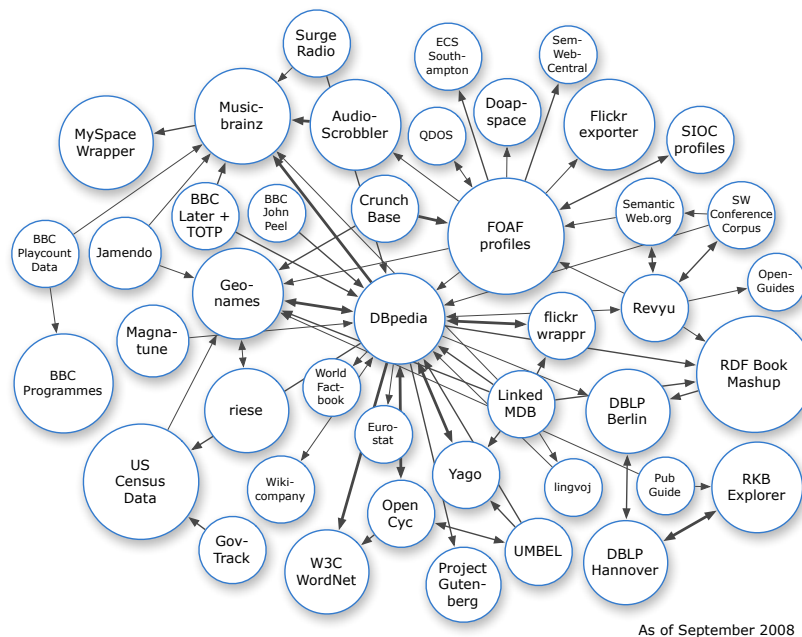


Abbildung 1: Linked Open Data Cloud [W3C, 2008a]

2.2 Linked Open Data

Sir Tim Berners-Lee beschrieb im Jahr 2006 die grundlegende Idee des Linked Data Konzeptes. [Berners-Lee, 2006] In seinem Beitrag behandelt er eine Menge von Regeln um die Akzeptanz des Semantic Web durch Linked Data zu verbessern. Das W3C SWEO Linked Open Data (LOD) Community Project ist ein kollaborativer Ansatz um dem Semantic Web durch die Veröffentlichung von (Open) Data Sets im RDF Format eine breitere Akzeptanz zu verschaffen. Weiters werden die verschiedenen Open Data Sets auch untereinander durch RDF [Manola and Miller, 2004] Links verknüpft. [Bizer et al., 2007]

In Abbildung 1 ist die Verlinkung der derzeit verfügbaren Open Data Sets ersichtlich.

Durch die Verlinkung der einzelnen Open Data Sets per RDF haben sich bis dato schon einige interessante Anwendungen für Endanwender entwickelt. Auf die verschiedenen Anwendungen wird im folgenden Kapitel näher eingegangen.

3 Anwendungen für Endanwender

Durch die Verlinkung der verschiedenen Open Data Sets per RDF und durch die Verknüpfung dieser verlinkten Data Sets konnten bis dato schon einige inter-

essante Anwendungen entwickelt werden. Im Folgenden wird speziell auf Freebase sowie auf das MIT Projekt SIMILE eingegangen.

3.1 Freebase

Freebase wurde von der Firma Metaweb Technologies Inc.⁹ entwickelt und stellt eine Semantische Datenbank dar. Freebase ist eine offene Datenbank die Informationen über die Welt enthält.[Metaweb Technologies, 2008] Offen bedeutet in diesem Zusammenhang, dass von der Community neue Informationen erstellt und für die Community verfügbar gemacht werden. Es steht auch jedem frei, neue Applikationen durch die Verwendung des Freebase API zu erstellen. Freebase ist somit wie Wikipedia ein kollaboratives System zur Wissenserstellung. Ein großer Unterschied zu Wikipedia besteht jedoch darin, dass die in Freebase enthaltenen Informationen über Semantische Technologien gespeichert werden. In Wikipedia sind die Informationen als Artikel einer bestimmten Kategorie gespeichert. Diese Artikel können in anderen Applikationen jedoch nur schwer wiederverwendet werden. In Freebase werden Informationen über sogenannte Fakten und Statistiken aufgelistet. Fakten sind durch RDF Links verlinkt, wodurch also auch semantische Informationen zu jeder Resource gespeichert werden. Mehrere Fakten ergeben somit ein Thema, das verschiedene Informationen verbindet. Im herkömmlichen Sinne findet man Informationen je nach Typ in verschiedenen Datenbanken.

So kann zum Thema Arnold Schwarzenegger in einer Filmdatenbank herausgefunden werden, dass er ein Schauspieler ist. In einer Bodybuilder Datenbank findet man heraus, dass er Bodybuilder und sogar Mister Universe war. In einer politischen Datenbank wird man herausfinden, dass Arnold Schwarzenegger Gouverneur des Bundesstaates Kalifornien ist. Da Arnold Schwarzenegger in Freebase durch RDF Links mit weiteren Informationen verbunden ist, gibt es dort nur ein Thema zu Arnold Schwarzenegger, das alle ihn betreffenden Informationen bündelt. Das Thema Arnold Schwarzenegger kann als Informationsverteiler angesehen werden, der durch jegliche neue Informationen erweitert werden kann. Dadurch ist es in Freebase sehr einfach Informationen unterschiedlichster Art über Arnold Schwarzenegger zu finden.

Es gibt derzeit schon eine Reihe an Applikationen, die das API von Freebase verwenden:

- Freebase Parallax¹⁰

- Powerset¹¹

⁹ Metaweb Technologies Inc.: <http://www.metaweb.com/>

¹⁰ Freebase Parallax: <http://mqlx.com/~david/parallax/>

¹¹ Powerset: <http://powerset.com/>

- Archiportal¹²
- History Of Sciences 2¹³
- FMDb¹⁴
- Dipity Timelines¹⁵
- MetaJam¹⁶
- History of a Day¹⁷

Durch die Verlinkung der verschiedenen Themen per RDF werden auch semantische Suchmaschinen darin unterstützt, bessere Suchergebnisse zu liefern, da RDF Links im Gegensatz zu HTML Links maschinenlesbarer und maschinenverständlicher sind.

3.2 SIMILE Project

Die folgenden Informationen wurden aus der SIMILE Project Seite¹⁸ entnommen. SIMILE steht für “Semantic Interoperability of Metadata and Information in unLike Environments” und wurde in Zusammenarbeit der MIT Libraries und dem MIT CSAIL (MIT Computer Science and Artificial Intelligence Laboratory) entwickelt. Das Ziel des MIT SIMILE Projektes ist es, die systemübergreifende Verwendung (inter-operability) von digitalen Informationen, Schemata, Vokabularen, Ontologien, Metadaten und Services zu verbessern. [Mazzocchi et al., 2006] Im Zuge des SIMILE Projektes wurden schon einige frei verfügbare Applikationen entwickelt. Im Folgenden wird kurz auf einige dieser Applikationen eingegangen, die schon einen gewissen Reifegrad erreicht haben.

3.2.1 Timeline

Timeline ist ein DHTML-basiertes AJAXy Widget, das für die Visualisierung von zeitabhängigen Events verwendet werden kann. Da Timeline ausschließlich in JavaScript und HTML implementiert wurde, lässt es sich somit sehr leicht in ein bestehendes System einbinden. [François Huyn, 2008c]

¹² Archiportal: <http://dev.mqlx.com/~zak/arch/>

¹³ History Of Sciences 2: <http://lindenb.integrangen.org/xulhistory/>

¹⁴ FMDb: <http://jameshome.com/freebase/fmdb/index.html>

¹⁵ Dipity Timelines: <http://www.dipity.com/>

¹⁶ MetaJam: <http://metajam.mobi/>

¹⁷ History of a Day: http://www.tigerlight.com/mjt/day_history.html

¹⁸ About SIMILE: <http://simile.mit.edu/wiki/SIMILE:About>

3.2.2 Timeplot

Timeplot ist ein DHTML-basiertes AJAXy Widget, das es ermöglicht, zeitlich abhängige Serien darzustellen und über diesen überlagernde Events darzustellen. [Mazzocchi, 2008b]

3.2.3 Exhibit

Exhibit ist ein in Javascript geschriebenes Web Application Framework, das es ermöglicht, strukturierte Daten durch Faceted Browsing [Hearst, 2008] darzustellen und zu erkunden. Exhibit basiert auf JavaScript und HTML und ist somit sehr einfach in bestehende Systeme einzubinden. [François Huyn, 2008a]

3.2.4 Gadget

Gadget ist ein XML Inspektor, der durch das Zusammenfassen von riesigen XML Datenmengen, das Verstehen dieser XML Daten erheblich erleichtert. [Mazzocchi, 2008a]

3.2.5 Welkin

Welkin ist ein RDF Visualizer der es ermöglicht, RDF in Form eines Graphen zu visualisieren. [Mazzocchi and Ciccarese, 2008]

3.2.6 Seek

Seek [François Huyn, 2008b] ist eine Mozilla Thunderbird Erweiterung, die effizienteres Suchen von Email Mozilla Thunderbird durch Faceted Browsing ermöglicht. Beim Faceted Browsing wird durch das Setzen verschiedener Filter das Suchen nach Informationen erleichtert. [Hearst, 2008]

4 Schlussfolgerung und Ausblick

Linking Open Data ist ein interessanter Ansatz, der sich hoffentlich noch mehr durchsetzen wird. Die hier beschriebenen Anwendungen zeigen einige der Möglichkeiten, die Linking Open Data bietet. Im Bezug auf das Semantic Web stellen die verlinkten Open Data Sets die Basis dar, auf der neue Applikationen aufbauen können, um semantische Technologien einer breiten Öffentlichkeit praxisnah näherzubringen. Bleibt zu hoffen, dass sich noch mehr Institutionen und Unternehmen dem Grundgedanken von Open Data verschreiben und ihre Informationen in strukturierter Form der Webgemeinschaft zur Verfügung stellen.

References

- [Berners-Lee, 2006] Berners-Lee, T. (2006). Linked data. <http://www.w3.org/DesignIssues/LinkedData.html>, last visited at 2009-01-14.
- [Berners-Lee, 2007] Berners-Lee, T. (2007). Testamony to the united states congress, hearing on the "digital future of the united states: Part 1 – the future of the world wide web. <http://dig.csail.mit.edu/2007/03/01-ushouse-future-of-the-web.html>, last visited at 2009-01-14.
- [Bizer et al., 2007] Bizer, C., Heath, T., Ayers, D., and Raimond, Y. (2007). Interlinking open data on the web. <http://tomheath.com/papers/bizer-heath-eswc2007-interlinking-open-data.pdf>, last visited at 2009-01-15.
- [Bizer et al., 2008] Bizer, C., Heath, T., Idehen, K., and Berners-Lee, T. (2008). Linked data on the web. <http://CEUR-WS.org/Vol-369/paper00.pdf>, last visited at 2008-09-29.
- [François Huyn, 2008a] François Huyn, D. (2008a). Mit simile project: Exhibit. <http://simile.mit.edu/exhibit/>, last visited at 2009-01-16.
- [François Huyn, 2008b] François Huyn, D. (2008b). Mit simile project: Seek. <http://simile.mit.edu/seek/>, last visited at 2009-01-16.
- [François Huyn, 2008c] François Huyn, D. (2008c). Mit simile project: Timeline. <http://simile.mit.edu/timeline/>, last visited at 2009-01-16.
- [Hearst, 2008] Hearst, M. A. (2008). Uis for faceted navigation: Recent advances and remaining open problems. flamenco.berkeley.edu/papers/hcir08.pdf, last visited at 2009-01-16.
- [Manola and Miller, 2004] Manola, F. and Miller, E., editors (2004). *RDF Primer*. W3C Recommendation. World Wide Web Consortium. <http://www.w3.org/TR/rdf-primer/>, last visited at 2009-01-16.
- [Mazzocchi, 2008a] Mazzocchi, S. (2008a). Mit simile project: Gadget. <http://simile.mit.edu/wiki/Gadget>, last visited at 2009-01-16.
- [Mazzocchi, 2008b] Mazzocchi, S. (2008b). Mit simile project: Timeplot. <http://simile.mit.edu/timeplot/>, last visited at 2009-01-16.
- [Mazzocchi and Cicarese, 2008] Mazzocchi, S. and Cicarese, P. (2008). Mit simile project: Welkin. <http://simile.mit.edu/welkin/>, last visited at 2009-01-16.
- [Mazzocchi et al., 2006] Mazzocchi, S., Garland, S., and Lee, R. (2006). Simile: Practical metadata for the semantic web. <http://www.xml.com/pub/a/2005/01/26/simile.html>, last visited at 2009-01-15.
- [Metaweb Technologies, 2008] Metaweb Technologies, I. (2008). Freebase faq. <http://www.freebase.com/view/freebase/faq>, last visited at 2009-01-16.
- [W3C, 2008a] W3C (2008a). Linked open data cloud. http://www4.wiwi.fu-berlin.de/bizer/pub/lod-datasets_2008-09-18_blank.pdf, last visited at 2009-01-16.
- [W3C, 2008b] W3C (2008b). Swox linked open data community project. <http://esw.w3.org/topic/SwoxIG/TaskForces/CommunityProjects/LinkingOpenData>, last visited at 2009-01-16.