

Linked Data

Ein kleiner Überblick

Margaux Gatrio

Margaux.gatrio@st.ovgu.de

Zusammenfassung. Während heute immer noch viele Texte im Internet über Hyperlinks verbunden werden, treiben sowohl Großunternehmen wie "Google", "Yahoo" und "Facebook", als auch „Linkend Open Data“, welches für jeden frei verwendbar ist, die Weiterentwicklung durch Linked Data voran. So werden ganze Bibliotheken durch Linked Data umstrukturiert und Suchmaschinen verbessert. Das gesamte Web dient dabei als Grundlage. Dazu werden veröffentlichte Daten so strukturiert und verlinkt, dass sie semantisch vernetzt sind. Diese Daten werden aber nicht einfach "nur" verlinkt, sondern dabei auch miteinander in Beziehung gesetzt. Daraus eröffnen sich für einen Nutzer eine Vielzahl von effektiveren Möglichkeiten Informationen zu sammeln.

Technisch ist dies durch URI-, RDF und HTTP-Technologien umsetzbar. Das Besondere daran ist, dass Maschinen verarbeiten können, was semantisch eigentlich nur von Menschen erfasst werden kann.

Tim Bernes-Lee, einer der größten Vorreiter in dieser Thematik, war überzeugt davon, das Linked Data das Web revolutionieren wird. Dieses Paper soll eine Übersicht darüber geben.

Keywords: Linked Data, Tim Bernes-Lee, URI, RDF, Netz, Generische Anwendungen, Linked Data-Browser, Linked Data-Suchmaschine, Domänenspezifische Anwendungen, Linked Open Data

1 Einleitung

In der heutigen Zeit, in der wir mit sehr großen Datenmengen in Berührung kommen, wird die Handhabung dieser immer schwieriger. Daraus entwickelt sich ein großes Forschungsgebiet, in dem auch Linked Data eine zunehmend wichtige Rolle spielt.

Um zu verstehen was Linked Data ist, muss man sich zuerst einmal mit dem semantischen Web auseinandersetzen. Dies basiert auf der Idee, das Web um ein Wissensnetz¹ zu erweitern. Das bedeutet, dass neben den Begriffen auch Informationen zu ihrem Kontext hinzugefügt werden. Dies kann sehr praktisch sein, zum Beispiel bei mehrdeutigen Suchanfragen. Die Suche nach "Maus" als Eingabegerät kann sich als schwierig herausstellen, wenn man nur Ergebnisse über das Tier bekommt. Im semantischen Web sollen diese Kontextinformationen aber zur Verfügung stehen.

¹ In der Informatik auch als semantisches Netz-Modell bekannt

Zur Umsetzung sind eine allgemeine, strukturierte Formatierung und der Zugang zu den Datensätzen notwendig. Genauso wichtig sind aber auch die Beziehungen zwischen den Daten. Eine Sammlung von so miteinander verbundenen Datensätzen im Web kann auch als “Linked Data” bezeichnet werden. Linked Data ist also als ein Werkzeug zur Erstellung semantischen Webs [1].

Zusammengefasst handelt es sich um Daten, die im Netz veröffentlicht wurden, maschinenlesbar sind, ihre Bedeutungen explizit definiert wurde und mit anderen externen Daten verlinkt sind [2]. Ein Beispiel zeigt Abbildung 1, in der veröffentlichten und miteinander vernetzten Datenbeständen in Linked Data-Format gezeigt werden.

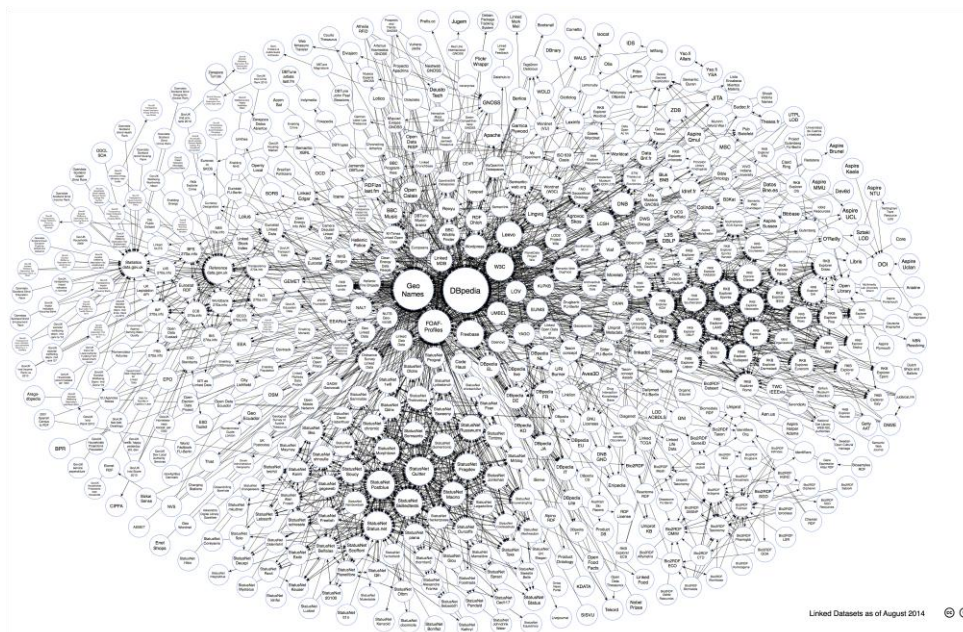


Abb 1. Verknüpfung zwischen Linked Open Data Datenbeständen (Stand April 2014) [15].

2 Prinzipien

Tim Berners-Lee stellte vier grundlegende Prinzipien und gleichzeitig technische Standards auf. Es gibt zwar keine Verpflichtung, diese einhalten zu müssen, dennoch haben sie sich durchgesetzt. Diese Prinzipien wurden von ihm 2006 in seinem Artikel Linked Data – Design Issue eingeführt [3]:

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

Um diese Prinzipien zu verstehen, ist es wichtig die Grundstruktur eines Web-Dokuments zu verstehen. Ein Web-Dokument ähnelt einer Webseite, aber erfüllt das erweiternde Kriterium von W3C², dass jedes über eine eigene URI (Uniform Resource Identifiers) verfügt. Anders als eine einfache Datei, kann es in vielen verschiedenen Sprachen oder Formaten (z.B. HTML, JPEG, RDF) aufgerufen werden [4]. Desweiteren baut es auf ein paar einfachen Standards auf: URI dient als globaler, einzigartiger Identifikationsmechanismus [5], HTML (Hypertext Markup Language) [6] als universaler Zugangsmechanismus und HTTP (Hypertext Transfer Protocol) als das gewöhnlich benutzte Inhaltsformat [7]. Außerdem bestehen Hyperlinks zwischen den Web-Dokumenten, auch wenn diese auf anderen Web-Server liegen.

Die Entwicklung und Nutzung dieser Standards ermöglicht es dem Web unterschiedliche technische Architekturen zu überwinden. Durch die Hyperlinks könne Nutzer zwischen verschiedenen Servern navigieren. Auch ist es dadurch möglich bessere Suchmechanismen zu entwickeln. Das macht Hyperlinks ausschlaggebend für die Verlinkung von Inhalten auf verschiedenen Servern zu einem einzigen globalen Informationsraum.

Linked Data baut direkt auf Web-Architekturen auf und verwendet diese Architekturen, um den Austausch von Daten auf globaler Ebene zu unterstützen [8]. Strukturierte Daten³ erleichtern dies, da Zusammenhänge zwischen Wörter oder Phrasen auch für Maschinen leichter erfassbar werden [9]. Eine Suchmaschine weiß dann z.B. um welche Art von Daten es sich handelt und kann einen sogenannten Snippet anzeigen (Abbildung 2). Übersichtlich und strukturiert werden dabei die wichtigsten Informationen dargestellt.

Strukturierte Daten für Websites, Online-Shops und Blogs ...

www.onlinemarketing-praxis.de > Suchmaschinenoptimierung ▾

★★★★★ Bewertung: 4,6 - 23 Abstimmungsergebnisse

Mehr Wissen für mehr Erfolg im Internet: "Strukturierte Daten für Websites, Online-Shops und Blogs" mit Tipps, Tricks und Know-how bei Onlinemarketing-Praxis.

Daten – Wikipedia

<https://de.wikipedia.org/wiki/Daten> ▾

Unter Daten versteht man im Allgemeinen Angaben, (Zahlen-)Werte oder formulierbare ... **strukturierte**

Daten: Die Daten (zum Beispiel in Datenbanken oder Dateien) weisen eine gleichartige Struktur auf.

Siehe auch Datenmodell ...

[Metadaten](#) · [Datenarchitektur](#) · [Paradaten](#) · [Datenstruktur](#)

Strukturiertes Daten-Markup überwachen - Search Console-Hilfe

support.google.com > Search Console-Hilfe > Optimierung ▾

Auf der Seite "Strukturierte Daten" in der Search Console werden alle strukturierten Informationen angezeigt, die Google auf Ihrer Website entdeckt hat.

Abb 2. Bei einer Google-Suche nach „strukturierten Daten“ werden verschiedene Snippets angezeigt [17].

² World Wide Web Consortium - <https://www.w3.org/>

³ Strukturierte Daten kann man sich wie eine Tabelle vorstellen. Die Datentypen sind die Spaltenbezeichnungen und die Spalteninhalte die jeweils zugeordneten Eigenschaften

3 Anwendungen

Für Linked Data gibt es bereits heute viele Nutzungsmöglichkeiten. Hier sollen einmal die Bekanntesten vorgestellt werden:

3.1 Generische Anwendungen

Generisches Linked Data-Anwendungen können Daten aus beliebigen aktuellen Domäne verarbeiten. Es gibt zwei Grundtypen von generischen Anwendungen: Linked Data-Browser und Linked Data-Suchmaschinen.

Linked Data-Browser

Auf Web-Daten kann mit Linked Data-Browser zugegriffen werden, ähnlich wie auf traditionelle Web-Dokumente mit HTML-Browser zugegriffen wird. Aber statt Hyperlinks zwischen den HTML-Seiten zu folgen, ermöglicht es Linked Data-Browser dem Nutzer zwischen verschiedenen Datenquellen durch RDF Links zu navigieren.

Dadurch kann ein Nutzer mit einer Datenquelle starten und sich dann durch ein potentiell unendliches Netz von Datenquellen, das mit RDF Links verbunden ist, bewegen. Während man zum Beispiel nach Informationen über eine Person sucht, könnte es auch interessant sein, Informationen über die Heimatstadt dieser Person zu bekommen. Durch den entsprechenden RDF-Link kann der Nutzer direkt diese Stadt aufrufen, die in einem anderen Datensatz enthalten sind [10].

Linked Data-Suchmaschinen

Es wurden zahlreiche Suchmaschinen entwickelt, die Linked Data aus dem Web mithilfe von RDF Links durchsuchen und Abfragefunktionen über aggregierte Daten zur Verfügung stellen. Sie können Daten aus tausenden Datenquellen integrieren. Dies zeigt die großen Vorteile der Linked Data-Architektur gegenüber dem Web auf, welches auf einem festen Satz von Datenquellen angewiesen sind, die proprietären Schnittstellen ausgesetzt sind.

Suchmaschinen wie Sig.ma, Falcon und SWSE stellen Schlüsselwort-basierende Suchdienste und ähneln in ihrem Interaktionsmuster den Marktführern wie Google und Yahoo. Man gibt in die Suchzeile den gesuchten Begriff ein und erhält eine Liste von möglichen Ergebnissen. Aber statt Links auszugeben, in denen der gesuchte Begriff ausgegeben wird, bieten Linked Data-Suchmaschinen mehr Interaktionsmöglichkeiten an, die die zugrunde liegende Struktur der Daten zu nutzen.

Falcons z.B. ermöglicht es, Suchergebnisse nach Klassen zu filtern und damit die Ergebnisse zu reduzieren. Beispielsweise nur Informatik relevante Resultate anzuzeigen.

Die Sig.ma Suchmaschine ermöglicht es den Nutzern, die Datenquellen zu wählen, aus denen ihre aggregierte Sicht konstruiert ist. Durch das Entfernen von geringerer Datenqualität aus ihren individuellen Ansichten, erstellen Sig.ma-Benutzer kollektiv Bewertungen für Datenquellen im Web.

Eine Suchmaschine die sich auch komplexere Suchanfragen spezialisiert ist VisiNav. Abfragen werden durch den Benutzer in einem explorativen Model formuliert und weit ausdrucksvoller Abfragen können derzeit besser als durch Google beantwortet werden [8].

Während Sig.ma, VisiNav, SWSE und Falcons Suchfunktionen für Menschen bieten, haben sich auch Suchdienste für Anwendungen entwickelt, umgesetzt durch anwendungsorientierte Indizes. Diese Indizes, wie Sindice, Swoogle und Watson bieten APIs an, durch die Linked Data-Anwendungen RDF-Dokumenten im Internet entdecken können, die eine bestimmte URI oder Schlüsselwort enthalten.

Dadurch müssen neue Linked Data-Anwendung nicht eine eigene Infrastruktur implementieren, um das gesamte Datennetz zu durchsuchen und zu indexieren. Stattdessen können Anwendungen, die diese Indizes abfragen Hinweise auf potenziell relevante RDF-Dokumente erhalten, die dann abgerufen und durch die Anwendung verarbeitet werden kann.

Es ist interessant, dass auch Suchmaschinen wie Google angefangen haben, strukturierte Daten in ihren Anwendungen zu nutzen. So nutzt Google RDF und mikroformatierte Daten um z.B. Personen, Organisationen oder Events zu beschreiben. Auch kann Google mittlerweile einfache Fragen direkt beantworten, was nur durch extrahierte Daten möglich ist [8].

3.2 Domänenspezifische Anwendungen

Es gibt auch verschiedene Linked Data-Anwendungen, die die Bedürfnisse bestimmter Nutzergruppen abdecken. Einige davon werden folgend dargestellt:

Listen von Linked Data-Anwendungen, die zur Steigerung der Transparenz von Regierung, durch die Kombination und Visualisierung von Regierungsdaten, führen, können auf den data.gov⁴ und data.gov.uk⁵ Websites gefunden werden.

Auch gibt es Anwendungen, die versuchen Linked Data in den täglichen Arbeitskontext zu bringen. Dayta.me [11] ist ein Empfehler von persönlichen Informationen. Es erweitert dazu Online-Kalender mit nützlichen Informationen im Bezug auf bevorstehende Aktivitäten. Paggr [12] bietet personalisierte Dashboards und Widgets für Daten aus dem Web.

⁴ <http://data.gov.uk/apps>

⁵ <http://data-gov.tw.rpi.edu/demo/USForeignAid/demo-1554.html><http://data-gov.tw.rpi.edu/demo/USForeignAid/demo-1554.html>

Die Linked Data-Anwendung Talis Aspire [13] hilft Pädagogen Listen für Lernressourcen zu erstellen und zu verwalten. Zehntausende von Studenten an zahlreichen Universitäten nutzen diese täglich. Dazu interagieren sie über herkömmliche Web-Schnittstelle, während die erstellten Daten hauptsächlich in RDF gespeichert werden. Eine HTTP-URI wird jeder Ressource, Ressourcenliste, Autor und Verleger zugeordnet. Durch die Verwendung von Talis Aspire wird das Angebot an Materialien bereichert.

Dbpedia Mobile⁶ hilft Touristen eine Stadt zu erkunden. Die Anwendung läuft auf einem Smartphone und bietet auf Grundlage von GPS-Daten Informationen über die umgebenden Lokationen. Nutzer können Bewertungen, Fotos und andere Informationen als Linked Data an die Anwendungen übermitteln und dadurch noch mehr Daten zur Verfügung stellen.

4 Vor- und Nachteile

Linked Data besitzt viele attraktive Vorteile gegenüber dem Web 2.0⁷. So ist es für jeden einfach Daten in das Netz zu laden. Das Einbinden von Informationen in dieser Weise kann auch als wertvoller betrachtet werden, da sie gleich mit anderen Daten in Verbindung gesetzt werden. Auch für den Endnutzer bedeutet das, dass er sehr schnell Zugang zu den neuen Informationen findet. Der größte Vorteil ist aber die standardisierte Datenrepräsentation. Bei der immer stärker steigenden Datenmenge in unserer Zeit und der besseren Vernetzung die mit der Globalisierung einhergeht, wird es auch wichtiger diese Daten integrieren zu können, was nur durch Vereinheitlichung gelingt [14].

Eine wichtige Rolle spielt auch die Offenheit von Linked Data, welche Vor- und Nachteile mit sich bringt. Das Konzept von „Linked Open Data“, ein Projekt W3Cs, welches stark von Tim Berners-Lee geprägt wurde, ermöglicht es schon seit 2007 auf freie Datensätze in RDF Format zugreifen zu können. Dass jeder offen mit Linked Data arbeiten kann, findet viel Zuspruch. Es bietet aber auch Schwachstellen. Es gestaltet sich als sehr schwierig Lizenzen oder die Vertraulichkeit von einigen Daten einzuhalten. Auch fehlt eine Instanz zur Kontrolle. Bei den Prinzipien zur Erstellung von Linked Data handelt es lediglich um Empfehlungen. Daher kann die Qualität und Relevanz von Informationen nur schwer eingeschätzt werden. Während die Idee von Linked Data Redundanzen verringern soll, indem Objekte mit den dazugehörigen Daten verknüpft werden, zeigen sich auch hier Problem auf. Wenn mehrere Nutzer

⁶ <http://wiki.dbpedia.org/projects/dbpedia-mobile>

⁷ Das frühere Web (1.0) konnte nur statische Webseiten erzeugen und darstellen. Das Konzept basierte darauf, dass nur wenige Personen als Erzeuger dienten und die vielen Benutzer lediglich in der Lage waren, sich die Inhalte anzugucken.

Mit der Weiterentwicklung bis 2005 kam es zu dynamischen Webseiten, technischen Verbesserungen und Nutzer stellten nun selbst Inhalte in das Netz. Angelehnt an Versionsnummern von Software wurde dieses nun Web 2.0 genannt.

Die Begriff Web 3.0 bezeichnet nun das Web mit der Erweiterung zum semantischen Web.

unterschiedliche Werte zu dem Objekt hinzufügen, kommt es doch zu Redundanzen. Ein besonders gutes Beispiel ist hier DBpedia. Bereits Wikipedia, woraus es abgeleitet wurde, war teilweise unvollständig und besaß Fehler. Die Daten sind also unzuverlässig [15].

Auch die Wartung der Links kann noch zu Fehler führen. Bis jetzt werden alte Links liegen gelassen. Da das System ohne Zeitbezug⁸ arbeitet, ist dies bis jetzt noch kein Problem. Mit ansteigenden Daten und damit auch mehr Aktualisierungen, werden deren Verarbeitungen zusätzlichen Rechenaufwand fordern, der irgendwann schwer zu bewältigen wird.

Der fehlende Zeitbezug führt auch dazu, dass Daten nicht versionierbar sind. Es gilt immer der zuletzt publizierte Zustand der Ressourcen. Gerade aber in unserer heutigen Zeit ist es immer wichtiger, unsere Informationsquellen nachweisen zu können [14].

Bis heute hat sich Linked Data noch immer nicht durchgesetzt. Es gibt noch zu wenige Verlinkungen im Netz und kaum Endnutzer-Anwendungen [15]. Dennoch spielt es eine große Rolle in der Entwicklung unseres Webs. So übernehmen große Konzerne wie „Google“ und „Yahoo“ viele Linked Data-Konzepte. Aber auch in der Industrie wird es mittlerweile genutzt. Allerdings eher als internes Netzwerk anstatt sich mit dem offenen Web zu verbinden. Besonders hier werden gut integrierbare Daten mit einfachem Zugang geschätzt, welche auch zu geringeren Kosten führen [15].

⁸ Daten werden zwar geupdatet, aber es ist nicht mehr nachvollziehbar wie aktuell die Quelle ist, da keine Zeitstempel gespeichert werden

5 Bibliographie

1. W3C, <https://www.w3.org/standards/semanticweb/data> [19.06.2016]
2. Linked Data – The Story so far, 2009, <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> [19.06.2016]
3. Tim Berners-Lee. Linked Data - Design Issues, 2006, <http://www.w3.org/DesignIssues/LinkedData.html> [20.06.2016]
4. W3C, Sauermann, Leo, Cyganiak, Richard, 2008, "Cool URIs for the Semantic Web", <https://www.w3.org/TR/cooluris/#oldweb> [20.06.2016]
5. Tim Berners-Lee, Fielding, Masinter. RFC 2396 - Uniform Resource Identifiers (URI): Generic Syntax, 1998, <http://www.isi.edu/in-notes/rfc2396.txt><http://www.isi.edu/in-notes/rfc2396.txt> [20.06.2016]
6. Fielding, Hypertext transfer protocol 1999, http/1.1., 1999, <http://www.w3.org/Protocols/rfc2616/rfc2616.html> [20.06.2016]
7. Raggett, Le Hors, Jacobs. Html 4.01 specification - w3c recommendation, 1999, <http://www.w3.org/TR/html401/><http://www.w3.org/TR/html401/> [20.06.2016]
8. Tom Heath, Christian Bizer, Linked Data: Evolving the Web into a Global Data Space, 2011, <http://linkeddatabook.com/editions/1.0/#htoc8> [20.06.2016]
9. Patrick Wilhelm, Strukturierte Daten – einfach erklärt, 2013, <http://seo.de/8337/strukturierte-daten-einfach-erklart/> [29.06]
10. Heath, Bizer, Cyganiak, 2007, How to Publish Linked Data on the Web <http://wifo5-03.informatik.uni-mannheim.de/bizer/pub/LinkedDataTutorial/> [20.06.2016]
11. Al-Mahrubi et al., A Personal News + Data Recommender for Your Day, 2010, http://challenge.semanticweb.org/submissions/swc2010_submission_17.pdf [20.06.2016]
12. Benjamin Nowack. Paggr: Linked data widgets and dashboards. Web Semantics: Science, Services and Agents on the World Wide Web, 7(4):272 – 277, 2009. Semantic Web challenge 2008.
13. W3C, <https://www.w3.org/2001/sw/sweo/public/UseCases/Talis/> [20.06.2016]
14. Gradmann, Hennicke, Olensky, Das Semantic Web und Linked Data, 2012, <http://edoc.hu-berlin.de/cmsj/35/gradmann-stefan-18/PDF/gradmann.pdf> [20.06.2016]
15. Pellegrini, Sack, Auer, Linked Enterprise Data, 2014, <http://link.springer.com/book/10.1007/978-3-642-30274-9> [20.06.2016]
16. Linking Open Data cloud diagram 2014, by Max Schmachtenberg, Christian Bizer, Anja Jentzsch and Richard Cyganiak, <http://lod-cloud.net/> [29.06.2016]
17. Google-Suche nach “ strukturierten Daten”, https://www.google.de/?gws_rd=ssl#q=Strukturierte+Daten [29.06.2016]