

Tags: records management big data, informatique,

Mehr Freiheit für die Daten!

Neue Technologien wie beispielsweise LTE (Long Term Evolution, ein Mobilfunknetz mit höheren Übertragungsgeschwindigkeiten für Daten) oder multifunktionale Tablets (leichte und flache Rechner mit berührungsempfindlichem Bildschirm, aber ohne Tastatur wie beispielsweise das iPad) eröffnen Möglichkeiten zur Entwicklung neuartiger Anwendungen, von denen wir heute noch gar nicht wissen, dass wir dafür einmal einen dringenden Bedarf verspüren werden. Aber noch ist das Potenzial der älteren (aber noch nicht veralteten) Technologien zur Anreicherung bestehender und Entwicklung neuer Anwendungen bei Weitem nicht ausgeschöpft.

Im Fokus stehen dabei die Daten, nicht die Technologien. Der Artikel stellt Potenziale in den folgenden drei Bereichen zur Diskussion:

- Sichtbarkeit und Datenqualität von Webseiten
- Mobiler Zugang zu Daten im Web
- Linked (Open) Data

Sichtbarkeit und Datenqualität von Webseiten

Milliarden von Webseiten halten Informationen für uns bereit, und Millionen von Datenbanken warten darauf, dass wir im Web auf sie stossen und sie nach nützlichen Daten durchsuchen. Viele der bereitgehaltenen Informationen finden wir aber nicht, weil sie im Web nicht sichtbar sind: Nutzer und Nutzerinnen von Websuchmaschinen begnügen sich bekanntlich sehr oft mit den Treffern auf der ersten Seite der Resultatliste (Bar-Ilan et al., 2007). Die Sichtbarkeit bestehender Webseiten (web visibility) zu überprüfen und nötigenfalls zu erhöhen, ist ein erster Schritt, die angesprochenen Zielgruppen zu erreichen. Hinweise darauf, wie die Sichtbarkeit von Webseiten erhöht werden kann, gibt zum Beispiel die Google-Webmaster-Tools-Hilfe (<http://www.google.com/support/webmasters/>). RSS-Newsfeeds (Really Simple Syndication) erhöhen die Sichtbarkeit von Neuigkeiten (z.B. Veranstaltungshinweise).

Allerdings bedeutet die gute Sichtbarkeit einer Webseite nicht zwingend auch, dass ihr Inhalt eine gute Qualität aufweist. Die Datenqualität einer Webseite (data quality) ist ein Mass dafür, inwieweit die dargebotenen Daten das Informationsbedürfnis einer suchenden Person in einer bestimmten Situation (context) befriedigen (Cappiello et al., 2004). Stellen wir uns eine Familie vor, die in der Nähe von Steinach^{Steinach wurde rein zufällig als Beispiel gewählt.} in den Ferien weilt und einen regnerischen Nachmittag gerne in einer Bibliothek verbringen möchte. Wird sie im Web herausfinden, wo sich die nächstgelegene Bibliothek befindet (visibility)? Und wenn sie die Webseite der Freihandbibliothek Steinach gefunden hat: Werden ihr die dargebotenen Informationen Antwort auf ihre Fragen geben, z.B. «Wann ist die Bibliothek geöffnet?», «Können die Kinder Bücher aus den Gestellen nehmen und in einer ruhigen Ecke lesen?», «Welche Zeitschriften liegen auf?», «Gibt es in der Bibliothek auch Kaffee und Kuchen?». Vielleicht hat die Bibliothek gar kein Angebot für Gäste? Ist das auf der Webseite vermerkt, damit unsere Familie die Bibliothek nicht vergebens aufsucht (data quality)? Die Datenqualität ist ein Konzept mit mehreren Dimensionen, wobei die Bedürfnisse der Zielgruppen im Zentrum stehen. Zur Ermittlung der Bedürfnisse eignen sich Befragungen nur bedingt. Als nützlich erweist sich dafür beispielsweise die Persona-Technik (siehe z.B. Chur Evaluation Laboratory: Personas, <http://www.cheval-lab.ch/cheval-wissensbasis/gestaltungsunterstuetzung/personas/>).

Anstatt einer Suchmaschine wünschen sich Personen oft eine Übersichts- oder Einstiegsseite, um im Web rasch Informationen zu einem bestimmten Thema finden zu können. Webportale erfüllen diese Aufgabe. Sie zeichnen sich durch eine eingängige Webadresse (URL) aus. Beispielsweise bietet <http://www.museums.ch> einen Überblick über alle Museen der Schweiz und den Zugang zu den Webseiten der einzelnen Museen. Hingegen ist <http://www.bibliothek.ch> eine (von Privatpersonen gestaltete) ungepflegte Webseite, deren Datenqualität sehr zu wünschen übrig lässt. Gibt es kein anderes Portal für die Schweizer Bibliotheken? Doch: das Internet Clearinghouse Schweiz (<http://www.ichschweiz.ch>), die von der Schweizerischen Nationalbibliothek und der Universitätsbibliothek Bern unterhaltene virtuelle Informationsplattform der Schweizer Bibliotheken. Ist diese Webadresse eingängig? Welche Daten stellt das Portal bereit? Welche Zielgruppen spricht es an? Gehört die Familie in unserem Beispiel dazu? Die Freihandbibliothek Steinach würde sie auf dieser Plattform jedenfalls nicht finden.

Mobiler Zugang zu Daten im Web

Immer mehr Personen nutzen für das Surfen im Web ihr Handy bzw. Smartphone [1](#). Unsere Zielgruppen können damit (fast) überall und jederzeit auf die Angebote im Web zugreifen. Aber passen die vielen Informationen und bunten Bilder, wie sie auf aktuellen Webseiten zu finden sind, auch auf kleine Bildschirme? Eher weniger. Trotzdem müssen wir für den mobilen Zugang auf unsere Inhalte nicht unbedingt mit viel Aufwand eine App entwickeln und pflegen. Mit «Apps» bezeichnet man Programme für mobile Endgeräte eines bestimmten Typs, z.B. für iPhones, Android Handys oder Geräte, die mit dem Betriebssystem Windows Phone 7 ausgerüstet sind. Es gibt sowohl kostenpflichtige als auch kostenlose Apps. Beide Arten von Apps müssen jeweils aus einem Webportal (z.B. dem App Store für iPhone Apps, dem Market für Android Apps) über das Internet auf das eigene Gerät übertragen und installiert werden. Schon heute liegt Webseiten das Konzept der konsequenten Trennung von Inhalt und Form zugrunde. Mit der Hypertext Markup Language HTML wird der Inhalt beschrieben (die Daten) und mit Cascading Style Sheets CSS die Form der Darstellung auf dem Bildschirm festgelegt. Für die Darstellung der Daten wird lediglich ein Browser benötigt. Unterschiedliche CSS-Definitionen genügen, um die Inhalte auf den verschiedenen Endgeräten mit dem passenden «Look and Feel» zu präsentieren. Während für die unterschiedlichen mobilen Endgeräte (iPhone, iPad, Android Handy usw.) je eine eigene App entwickelt werden muss, lässt sich eine entsprechend gestaltete Webseite auf allen Geräten nutzen. Und wenn wir die Inhalte aktualisieren, ist das für alle Besucherinnen und Besucher der betreffenden Webseite sofort sichtbar. Auch die speziellen Funktionen der neuen Smartphones, beispielsweise die Positionsbestimmung mit GPS [2](#), werden von Webseiten unterstützt, wenn sie in der Version 5 von HTML (HTML5) beschrieben und mit «eingebetteten», in JavaScript formulierten Programmteilen ergänzt sind. Die Standardisierung von HTML5 ist noch nicht ganz abgeschlossen. Wäre es trotzdem sinnvoll, HTML5 für neue Webprojekte in Betracht zu ziehen? Ja, sicher: Die aktuellen Versionen der gängigen Browser auf allen Endgeräten unterstützen HTML5. Und nicht zuletzt ist HTML5 die Grundlage des (wirklich) neuen E-Book Standards EPUB [3](#).

Linked (Open) Data

Auf Webseiten sind die Daten in Texte, Tabellen oder Grafiken «eingearbeitet». Wenn wir beispielsweise eine Bibliotheks-Webseite betrachten, können wir den Kasten mit den Öffnungszeiten und der Adresse, die Tabelle mit den Neuzugängen und die Liste mit den Veranstaltungshinweisen erkennen. Ein Computerprogramm kann das nicht. Wäre das denn nötig?

Machen wir ein Gedankenspiel: Die mit einem International Standard Identifier for Libraries and Related Organizations (ISIL) gekennzeichneten Gedächtnisinstitutionen (Archive, Bibliotheken und Museen) liefern der Schweizerischen Nationalbibliothek (NB) Datensätze mit der Kontaktadresse, Angaben zum Bestand und weiteren Informationen. Diese Daten sind in der Archivdatenbank HelveticArchives (<https://www.helveticaarchives.ch/suchinfo.aspx>) im Web zugänglich. Sie sind aber teilweise nicht mehr aktuell (für die Freihandbibliothek Steinach verzeichnet der betreffende Datensatz³ beispielsweise 2184 Sachbücher und 191 DVD, während im Online-Katalog der Freihandbibliothek Steinach heute (Mai 2011) 2302 Sachbücher und 443 DVD erfasst sind). Das Problem träte nicht auf, wenn die einzelnen Institutionen die Datensätze in standardisierter Form auf ihren Webseiten publizieren (und aktualisieren) würden, sodass diese bei Bedarf von einer Webapplikation der NB automatisch «gelesen» und für die Generierung der ISIL-Datensätze genutzt werden könnten. Eine weitere Webanwendung könnte die Datensätze nutzen, um daraus die Inhalte für ein Bibliotheksportal herauszufiltern, dessen Pflege nur geringen Aufwand mit sich brächte. Solche Überlegungen stecken hinter dem Konzept der Linked Open Data. Linked Data sind Datensätze, die «Dinge» in strukturierter Form mit standardisierten Auszeichnungselementen beschreiben und Verweise auf verwandte Datensätze enthalten. Der Zugriff auf Linked Data Sets kann durch eine entsprechende Verwaltung der Zugriffsrechte eingeschränkt werden. Bei Linked Open Data werden die interessierenden Daten von den Dateninhabern (data owners) im Web frei zur Verfügung gestellt und gepflegt. (Heath, 2011), das aktuell grosse Beachtung findet.

Eine weitere Anwendung von Linked Open Data stellte die Veröffentlichung der bibliografischen Daten der Schweizer Bibliotheken dar. Seit 2009 publiziert beispielsweise die Bibliothek der Europäischen Organisation für Nuklearforschung CERN in Genf ihre bibliografischen Daten im MARCXML-Format auf der Webseite www.cern.ch/bookdata. Die Bereitstellung der Daten in dem für Linked Data vorgesehenen Format RDF (Resource Description Format) ist in Arbeit. Vergleichbare Projekte sind auch in anderen Ländern bereits abgeschlossen oder im Gang. Beispiele dafür sind der Linked-Open-Data-Dienst des Hochschulbibliothekszenentrums des Landes Nordrhein-Westfalen (<http://lobid.org/de/>) und der nationale Suchdienst LIBRI für Bibliothekstitel Schwedischer Universitäts- und Forschungsbibliotheken sowie ungefähr 20 öffentlicher Bibliotheken in Schweden (<http://libris.kb.se/>).

Die Bereitstellung der Datensätze ist das eine. Das andere ist deren Nutzung. Dafür sind passende Webanwendungen nötig. Die Offenheit der Daten erlaubt Experimente. Mit den heutigen Webtechnologien ist es beispielsweise auch einer Mittelschülerin möglich, in ihrer Maturaarbeit eine Anwendung für natürlichsprachige Anfragen an Linked Open Data zu programmieren. Auf die Anfrage: «Welche Bücher von Andrea Camilleri sind in deutscher Sprache erhältlich, und wie lauten die Titel der entsprechenden Originalausgaben?», würde die Webanwendung die verfügbaren Datensätze «durchforsten» und relevante Einträge miteinander so verknüpfen, dass auch gerade sichtbar würde, bei welchen Bibliotheken die betreffenden Bücher ausgeliehen werden könnten. Einen ersten Eindruck vom Potenzial der Linked Open Data kann die Webseite von dbpedia (<http://dbpedia.org>) vermitteln. In diesem Projekt werden aus den Beiträgen von Wikipedia strukturierte Informationen gewonnen und in Linked Open Data Sets abgelegt, an die differenzierte Anfragen gestellt werden können.

«The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.» (Berners-Lee, 2006).

Wie wäre es, mit eigenen Projekten an der Umsetzung der Visionen von Tim Berners-Lee mitzuwirken?

Quellen

– Bar-Ilan, J. et al., «User Ranking of Search Engine Results», J. Am. Soc. Information Science and Technology, vol. 58, no. 9, 2007, pp. 1256–1266.

– Berners-Lee, T., «Linked Data», 2006. (<http://www.w3.org/DesignIssues/LinkedData.html>)

– Cappiello, C., Francalanci, C., and Pernici, B., «Data Quality Assessment from the User's Perspective», Proc. Int'l Workshop Information Quality in Information Systems, ACM Press, 2004, pp. 68–73.

– Heath, T., and Bizer, Ch., «Linked Data: Evolving the Web into a Global data Space», 2011, <http://linkeddatabook.com/edit...> 1.0/.

1 Accenture: Mobile Web Watch 2010 (<http://www.accenture.com/de-de/Pages/insight-mobile-web-2010-summary.aspx>).

2 Global Positioning System.

3 <https://www.helveticaarchives.ch/detail.aspx?ID=253476>.



Gabi Schneider

HTW Chur



Bruno Wenk

HTW Chur

Résumé

Français

Les nouvelles technologies comme le LTE (Long Term Evolution) ou les tablettes multifonctions ouvrent des possibilités pour développer des applications d'un nouveau genre, dont nous ne soupçonnons même pas aujourd'hui la future utilité. Cela dit, le potentiel des technologies plus anciennes (mais pas encore dépassées) pour enrichir les applications existantes et en développer de nouvelles n'est de loin pas épuisé.

Les auteurs se focalisent dans cet article sur les données et non pas sur les technologies, et discutent les potentiels qui existent dans les trois domaines suivants:

- Visibilité et qualité des données des sites web: il s'agit dans un premier temps d'examiner et le cas échéant d'augmenter la visibilité des sites web existants (web visibility) pour toucher les publics cibles auxquels on entend s'adresser.
- Accès mobile aux données du web: de plus en plus de personnes utilisent leur téléphone portable, resp. leur smartphone pour surfer sur le web. Les groupes cibles peuvent accéder ainsi (presque) partout et en tout temps aux offres proposées sur le web. Mais les nombreuses informations et les images couleurs que l'on y trouve conviennent-elles aux petits écrans?
- Linked (Open) Data: dans les pages web, les données sont intégrées dans des textes, des tableaux ou des graphiques. Il s'agit de données qui décrivent les «choses» de manière structurée avec des éléments de marquage standardisés et qui contiennent des renvois vers des données apparentées. L'accès aux Linked Data Sets (ensembles de données corrélés) peut être restreint par une gestion ad hoc des droits. Les données sont gérées et mises à disposition sur le web par leurs propriétaires (data owners).

Les auteurs soulignent que la mise à disposition des données est une chose et que leur utilisation en est une autre. Il faut pour se faire une application web ad hoc. Et de citer Berners-Lee en guise de résumé: «The Semantic Web isn't just about putting data on the web. It is about making links, so that a person or machine can explore the web of data. With linked data, when you have some of it, you can find other, related, data.» (2006).