

Semantisch angereicherte Präsentationsschichten für geisteswissenschaftliche Webanwendungen

Methodenvergleich und Referenzimplementierung

Toschka, Patrick

patrick.toschka@adwmainz.de

Akademie der Wissenschaften und der Literatur | Mainz, Deutschland

Hintergrund

Ausgangspunkt der folgenden Betrachtungen ist eine Master-Thesis, die sich mit einem Vergleich von Methoden zur Anreicherung von Präsentationsschichten geisteswissenschaftlicher Webanwendungen mit semantisch strukturierten Metadaten befasst. Viele aktuelle Forschungsapplikationen lagern Metadaten in eine eigene Datenschicht aus. Meist können diese Daten über spezifische Schnittstellen bezogen werden. In diesem Szenario bietet die Präsentationsschicht meist nur eine optische Darstellung dieser Daten ohne eine tiefergehende semantische Strukturierung. Alternativ können Metadaten aber auch direkt im HTML-Quelltext eingebettet werden. Eine externe, maschinelle Verarbeitung der Metadaten dort scheint allerdings noch aufwendig, da angeblich wenige bis keine Standards für die Auswertung von Daten direkt aus der Präsentationsschicht vorhanden sind.

Das Web hat sich weiterentwickelt. Große Konzerne wie Suchmaschinenbetreiber haben Methoden und Standards vorangetrieben, die im privatwirtschaftlichen Sektor neue Ansätze ermöglichen. Über Linked (Open) Data Technologien können Geräte selbstständig Inhalte fremder Quellen anfragen, interpretieren und aufbereiten. Hierfür sind Technologien erforderlich, die semantische Metadaten auf einer Vielzahl von Plattformen standardisiert zur Verfügung stellen und Inhalte miteinander verknüpfen.

Potentiale von JSON-LD und Schema.org

JSON-LD hat sich durch die Verwendung der JSON-Syntax, der Übersichtlichkeit des Quellcodes und der Auslagerung aus dem eigentlichen Dokument heraus als de facto Standard herausgestellt. Bereits bei der

Evaluation des Schema.org-Vokabulars wird klar, dass dieses gut geeignet ist, grundsätzliche semantische Aussagen abzubilden. Beispielsweise ist die Auszeichnung von Beziehungen zwischen Personen wie *x kennt y*, *x ist dieselbe Person wie y* oder *x ist verwandt mit y* out-of-the box mit dem Schema.org-Vokabular möglich. Bei komplexeren semantischen Relationen wie *x hat schon von y gehört*, *x empfiehlt y* oder *x spottet über y* fehlen in Schema.org zuweilen noch bestimmte Aussagemuster. Dieser Situation kann aber durch die Einbindung zusätzlicher LOD-Vokabulare gut begegnet werden.

Die modulare Ausbaufähigkeit von JSON-LD birgt somit für eine semantische Anreicherung der Präsentationsschichten von Digital Humanities Anwendungen viele Möglichkeiten. Im Bereich Digitaler Editionen beispielsweise fügt sich eine Einbeziehung der Metadaten in die Präsentationsschicht gut in das Konzept der Edition als Interface ein.¹ Auch für die Interoperabilität webbasierter geisteswissenschaftlicher Ressourcen in Einklang mit den FAIR-Prinzipien spielt eine semantische Strukturierung der Präsentationsschicht eine wichtige, momentan aber noch unterschätzte Rolle.² So könnten Browser-Plugins zukünftig automatisch Vorschläge für verwandte Einträge auf anderen Webressourcen anzeigen, ohne dass dabei ein Mehraufwand bei der Datenpflege betrieben oder gar eine spezifische Schnittstelle implementiert werden muss.

Methodik

Es gilt also herauszufinden, inwiefern das Schema.org Vokabular für den geisteswissenschaftlichen Anwendungsfall zum einen qualitativ exakt genug ist und ob, zum anderen, alle nötigen Aussagen mit diesem Vokabular und eventuellen Erweiterungen getroffen werden können. Mittels Referenzimplementierungen auf Basis bestehender Webapplikationen, die bisher keine der genannten Technologien in der Präsentationsschicht nutzen, können die Potentiale genauer eingeschätzt werden. Dazu müssen Sets projektspezifischer Metadaten zunächst aufgearbeitet und dann unter Verwendung verschiedener Einbettungsmethoden in HTML verankert werden. Ziel ist es, möglichst viele verschiedene Anwendungsfälle auf diese Art und Weise vergleichend zu erfassen und dabei die möglichen Probleme mit den Technologien und dem Schema.org Vokabular herauszuarbeiten. Daraus leitet sich ab, ob sich einzelne Einbettungsverfahren gut oder weniger gut für geisteswissenschaftliche Webanwendungen eignen, welche Probleme auftreten und worauf bei zukünftigen Implementierungen geachtet werden sollte.

Fußnoten

1. Siehe zum Begriff der Edition als Interface auch <https://www.i-d-e.de/publikationen/schriften/bd-12-interfaces/> (letzter Zugriff: 12. Januar 2019).
2. Siehe zu den FAIR-Prinzipen (Findable, Accessible, Interoperable, und Re-usable) auch http://www.forschungsdaten.org/index.php/FAIR_data_principles (letzter Zugriff: 12. Januar 2019).

Bibliographie

Carroll, Jeremy J. / Bizer, Christian / Hayes, Patrick / Stickler, Patrick (2004): *“Named Graphs, Provenance and Trust”*, pdf Format, URL: <http://wifo5-03.informatik.uni-mannheim.de/bizer/SWTSGuide/carroll-ISWC2004.pdf> (letzter Zugriff: 08. Januar 2019).

Google Inc. (2018): *“Understand how structured data works”*, URL: <https://developers.google.com/search/docs/guides/intro-structured-data>, text/html Format, (letzter Zugriff: 04. Januar 2019).

Guha, R.V. / Brickley, Dan / MacBeth, Steve (2015): *“Schema.org: Evolution of Structured Data on the Web”* in: Queue - Structured Data, Vol. 13 No. 9, text/html Format, URL: <https://queue.acm.org/detail.cfm?id=2857276> (letzter Zugriff: 08. Januar 2019).

Halpin, Harry / Herman, Ivan / Hayes, Patrick J. (2009): *“When owl:sameAs isn’t the Same: An Analysis of Identity Links on the Semantic Web”*, pdf Format, URL: <https://www.w3.org/2009/12/rdf-ws/papers/ws21> (letzter Zugriff: 08. Januar 2019).

Lahntaler, Markus (2012): *“Third Generation Web APIs. Bridging the Gap Between REST and Linked Data”*, pdf Format, URL: <http://www.markus-lanthaler.com/research/third-generation-web-apis-bridging-the-gap-between-rest-and-linked-data.pdf> (letzter Zugriff: 04. Januar 2019).

W3C World Wide Web Consortium (2014): *“JSON-LD 1.0. A JSON-based Serialization for Linked Data”*, text/html Format, URL: <http://www.w3.org/TR/2014/REC-json-ld-20140116/> (letzter Zugriff: 04. Januar 2019).

W3C World Wide Web Consortium (2015): *“RDFa Core 1.1 – Third Edition. Syntax and processing rules for embedding RDF through attributes”*, text/html Format, URL: <http://www.w3.org/TR/2015/REC-rdfa-core-20150317/> (letzter Zugriff: 04. Januar 2019).

Wettlaufer, Jörg (2018): *“Der nächste Schritt? Semantic Web und digitale Editionen”* in: Digitale Metamorphose: Digital Humanities und Editionswissenschaft, Hg. von Roland S. Kamzelak / Timo Steyer. 2018 (= Sonderband der Zeitschrift für digitale Geisteswissenschaften, 2), text/html Format, URL: http://dx.doi.org/10.17175/sb002_007 (letzter Zugriff: 08. Januar 2019).