



Content Analytics und Linked Open Data*

Neue Chancen auch für Ihre Inhalte

Felix Sasaki (DFKI / W3C Fellow)

Christian Lieske (SAP SE)

Der Text aus dem Tekom Jahrestagung
Tagungsband findet sich am Ende dieser Präsentation
Diese Präsentation wurde gefördert durch LIDER
(EU FP7 Projekt Nr. 610782)



Überblick

Beispielhafte Möglichkeiten für Sie

- Live und in Farbe

Theorie hinter der Praxis

- Aus 3000 Metern Höhe

Annäherung

- Auf Augenhöhe

Weitere Möglichkeiten

- Für Sie, Ihre Kunden oder ... Konkurrenten

Einstieg

- Lernen und mitgestalten

Über uns

Felix Sasaki

DFKI//W3C



- Senior Researcher im DFKI (LT-Lab)
- W3C Büro Deutschland-Österreich
- Davor W3C Mitarbeit in Japan
- Themengebiete: Anwendung von Webtechnologien für Repräsentation und Verarbeitung mehrsprachiger Daten
- Studium der Japanologie, Linguistik, Webtechnologien in Deutschland und Japan

Christian Lieske

SAP SE



- Knowledge Architect (SAP Language Services, Globalization Services)
- Themengebiete: Multilinguale Produktion, Maschinelle Sprachverarbeitung, Internationalisierung
- Mitarbeit in World Wide Web Consortium (W3C), OASIS, Unicode Consortium
- Studium: Informatik und Maschinelle Sprachverarbeitung (einschließlich maschinelle Übersetzung und gesprochene Sprache)

Der Inhalt der vorliegenden Veröffentlichung ist persönlich und spiegelt nicht unbedingt die Meinung oder die Position unserer Arbeitgeber wieder



Beispielhafte Möglichkeiten für Sie – Demos

Die [SAP SE](#) mit Sitz im [baden-württembergischen Walldorf](#) ist der nach Umsatz größte europäische (und außeramerikanische) Softwarehersteller sowie der weltweit viertgrößte [4]. Tätigkeitsschwerpunkt ist die E [Geschäftsprozesse](#) eines Unter Produktion, [Lagerhaltung](#) und

Der Name stand zeitweise [syn](#) ein integriertes [betriebswirtsch](#) [SAP R/2](#) für [Großrechner](#) und [S](#) Finanzbereich ([SAP BankAnaly](#) [Warehouse](#) - [SAP BW](#)) und biet HANA).

html head title

Mein Reiseführer mit angereicherten Linked Data Inhalten :)

head

body p b Vor Anreicherung: b p

p Willkommen in Stuttgart! p

p Ein berühmter Schriftsteller in span Stuttgart span ist Gustav Schwab. p

p Ein anderer Schriftsteller neuerer Zeit ist Robert Steiner. p

p Ein Buch was nicht in Stuttgart spielt: q Berlin Alexanderplatz q ". p

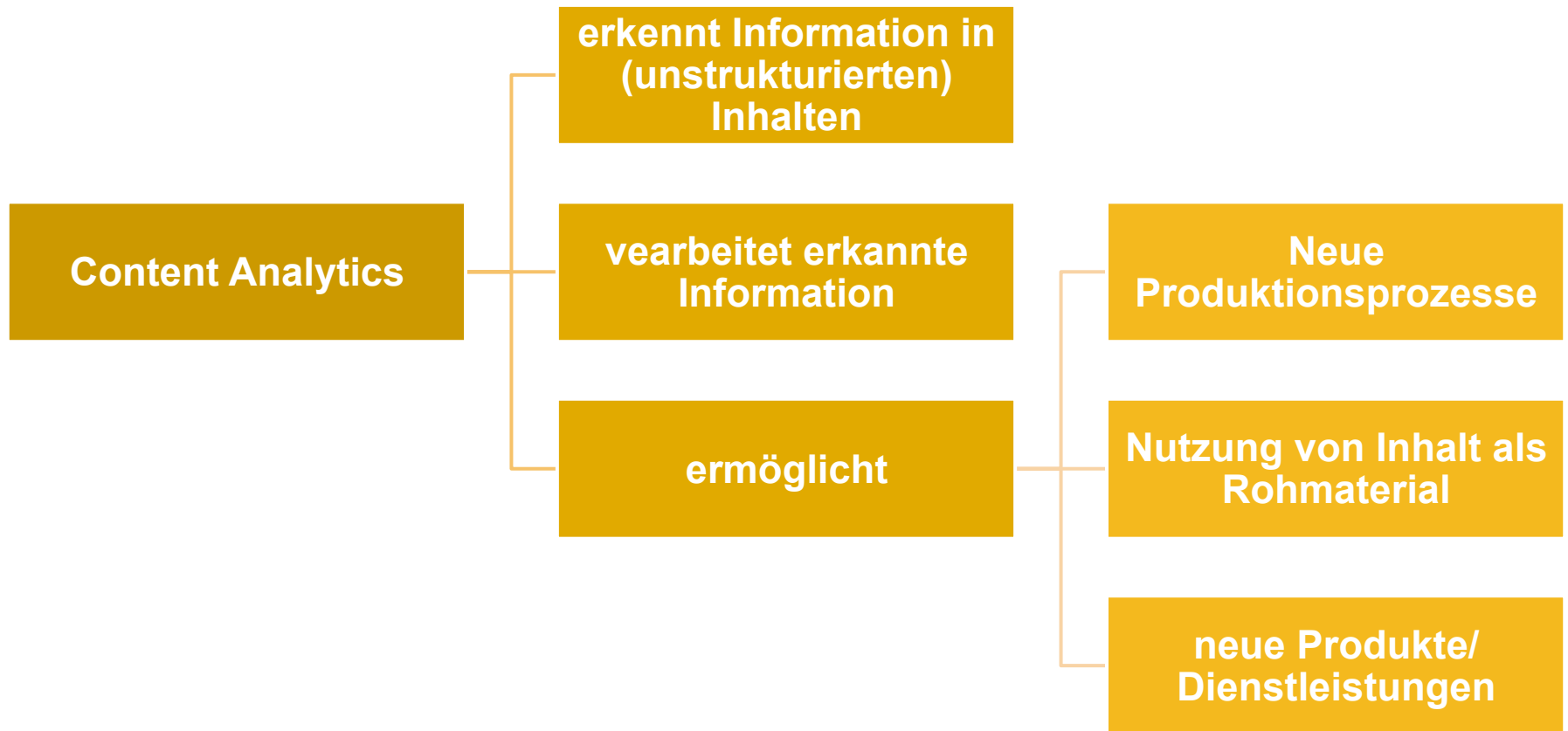
p b Nach Anreicherung: b p

p Willkommen in span a span Stuttgart span a (Einwohner: 587655) span! p

p Ein berühmter Schriftsteller in Stuttgart ist span a span Gustav Schwab span a (geboren span 1792 span) span. p

p Ein anderer Schriftsteller neuerer Zeit ist span a span Robert Steiner span a (geboren span 1976 span) span. p

Beispielhafte Möglichkeiten für Sie – Erläuterung



Theorie hinter der Praxis – Aus 3000 Metern Höhe

(Adaptierte) Sprachverarbeitung zum Erkennen von Entitäten (und Aussagen)

Prozessierung wie beispielsweise Verknüpfung

Nutzung moderner Inhaltsrepräsentation

Daten als Basis (für Sprachverarbeitung und Prozessierung)

- Proprietär: Listen mit Produktnamen, Adjektive mit positiver Konnotation, ...
- Standardisiert: „Linked Data



Annäherung – Auf Augenhöhe (Sprachverarbeitung) 1/3

Entitäten erkennen

Input

- Welcome to Stuttgart

Output

- Stuttgart ist eine Entität (Klasse „Stadt“)

Aussagen erkennen

Input

- Stuttgart is the capital of Baden-Württemberg

Output

- Stuttgart und Baden-Württemberg sind verknüpft. Stuttgart ist die Hauptstadt von Baden-Württemberg.

Annäherung – Auf Augenhöhe (Sprachverarbeitung) 2/3

Entitäten/Aussagen verknüpfen

Input

- Stuttgart is the capital of Baden-Württemberg

Output

- „Stuttgart“ ist verlinkt zum Wikipedia-Artikel über Stuttgart

Entitäten eindeutig zuordnen

Input:

- Armstrong erreicht Paris

Output

- „Armstrong“ wird als der Fahrradfahrer identifiziert

Annäherung – Auf Augenhöhe (Sprachverarbeitung) 3/3

rttembergischen Walldorf ist der (sche) <u>Softwarehersteller</u> sowie die Entwicklung von <u>Software</u> zur Abwicklung von <u>Buchführung</u>, <u>Controlling</u> und <u>Personalwesen</u>. um für das ehemals zentrale Softwarepaket, das als <u>R/3</u> für <u>Client-Server-Systeme</u> (z.B. - <u>SAP BA</u>, <u>SAP AFI</u> u. a.), im Data Warehouse eigene Datenbanklösungen an (	Disambiguation Method: ☐ prior ☐ prior+sim ☒ prior+sim+coherence Parameters Prior-Similarity-Coherence balancing ratio: prior VS. sim. balance = 0.4 (prior+sim.) VS. coh. balance 0.6 Ambiguity degree 5 Coherence robustness test threshold: 0.9 Coherence Measure: MilneWitten Entities Type Filters: Mention Extraction:	"Word sense disambiguation and entity linking together" Babelfy Powered by BabelNet Text to babelfy:
DBpedia Spotlight	AIDA	Babelfy

Annäherung – Auf Augenhöhe (Linked Data) 1/3

About: Waldorf (Baden)

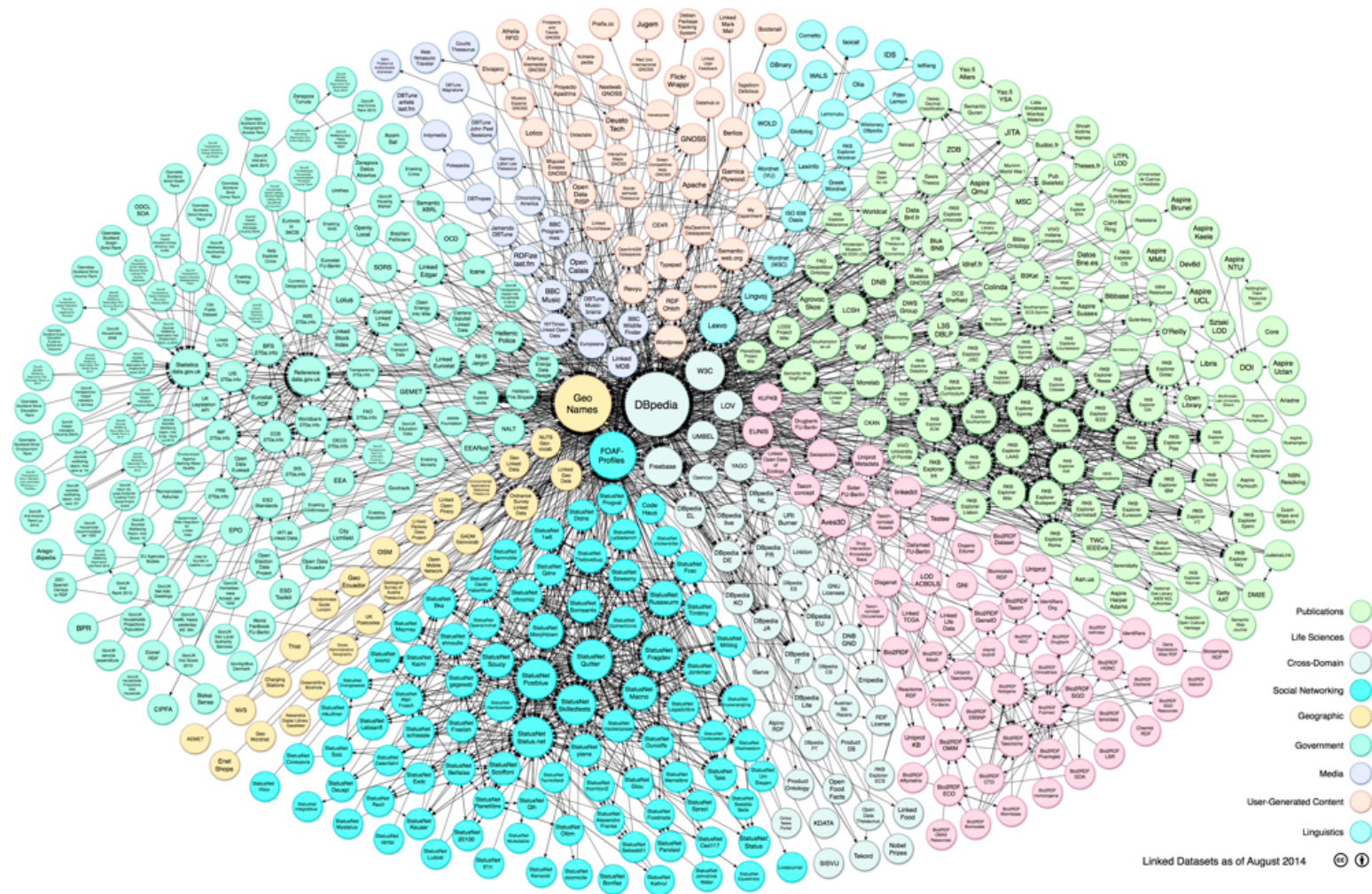
An Entity of Type : [municipality](#), from Named Graph : <http://dbpedia.org>, within Data Space : [dbpedia.org](#)



Waldorf ist eine Stadt im Süden des Rhein-Neckar-Kreises in Baden-Württemberg. Überregional bekannt ist sie durch den Hauptsitz des Softwareunternehmens SAP und das Autobahnkreuz Waldorf. Mit der Nachbarstadt Wiesloch bildet sie ein Mittelzentrum in der Metropolregion Rhein-Neckar mit einem Einzugsbereich von etwa 60.000 Menschen.

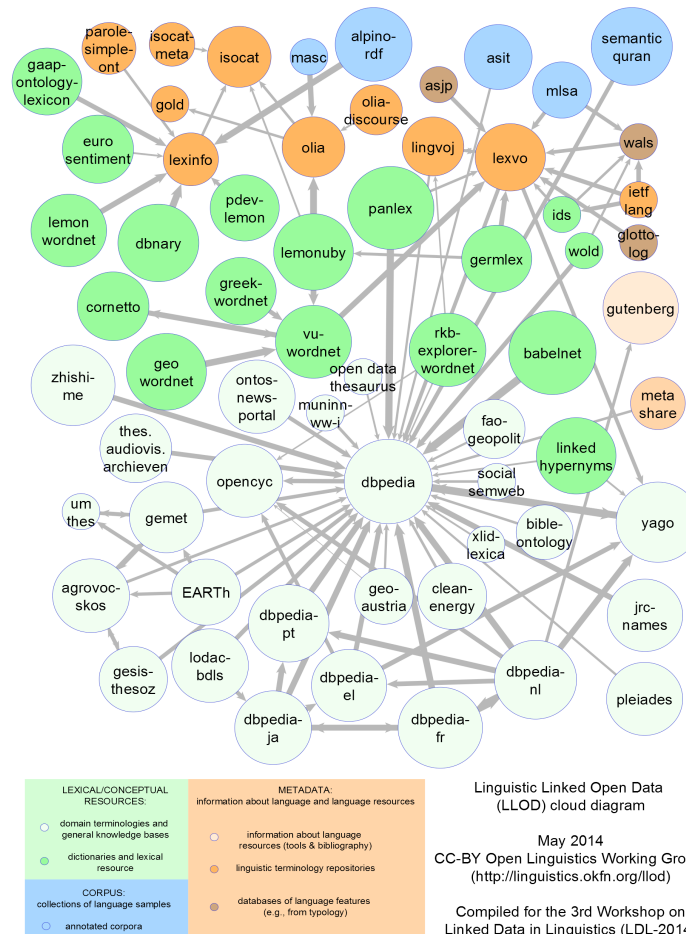
Property	Value
dbpedia-owl:PopulatedPlace/areaTotal	■ 19.91
dbpedia-owl:abstract	■ Waldorf is a town in the Rhein-Neckar-Kreis of Baden-Württemberg in Germany.Waldorf is currently probably best known as the city that headquarters the world's third largest software company SAP, but it is also the birthplace of the millionaire John Jacob Astor, at the time of his death the wealthiest man in the United States. Astor's descendants of the Astor family perpetuated the North-Americanized version Waldorf in the names of The Waldorf-Astoria Hotel and Waldorf salad. It is also the origin of the name of the Waldorf-Astoria-Zigarettenfabrik and through it Waldorf education.Waldorf is considered to be one of the richest towns in Europe and has often been awarded with the prize most economically attractive community in Germany.The town is referred to, though not actually named, in John le Carré's novel Absolute Friends (2003), much of which is set in Heidelberg. ■ Waldorf ist eine Stadt im Süden des Rhein-Neckar-Kreises in Baden-Württemberg. Überregional bekannt ist sie durch den Hauptsitz des Softwareunternehmens SAP und das Autobahnkreuz Waldorf. Mit der Nachbarstadt Wiesloch bildet sie ein Mittelzentrum in der Metropolregion Rhein-Neckar mit einem Einzugsbereich von etwa 60.000 Menschen. Waldorf gilt – gemessen an Einnahmen und Rücklagen pro Einwohner – als eine der reichsten Städte Europas und wurde schon häufig als wirtschaftlich attraktivste Gemeinde Deutschlands ausgezeichnet.Über den Waldorfer Auswanderer Johann Jacob Astor ist die Stadt indirekter Namensgeber der Hotels Waldorf-Astoria, des Waldorfsalats, der Waldorf-Astoria-Zigarettenfabrik und somit auch der Waldorfschulen.
dbpedia-owl:areaCode	■ 06227
dbpedia-owl:areaTotal	■ 19910000.000000 (xsd:double)

Annäherung – Auf Augenhöhe (Linked Data) 2/3



Linked Open Data Cloud

Annäherung – Auf Augenhöhe (Linked Data) 3/3



Linguistic Linked Open Data Cloud

Linked Open Data Clouds – Datensätze und Themen

Linked Open Data			Linguistic Linked Open Data
Öffentliche Daten	183	18.05%	Lexika
Publikations(Meta)daten	96	9.47%	Thesauri
Medizin, Biologie	83	8.19%	Einzelsprachlich
User-generated Content	48	4.73%	Mehrsprachig
Domänenübergreifend	41	4.04%	...
Multimedia	22	2.17%	
Geoinformationen	21	2.07%	
Soziales Web	520	51.28%	
Total	1014		

Annäherung – Auf Augenhöhe (Standardisierung)

Basis: Webtechnologien
(HTTP, URIs RDF,
SPARQL)

Standardisiert vom W3C

Mehr und mehr frei
verfügbar: „Linked Open
Data“

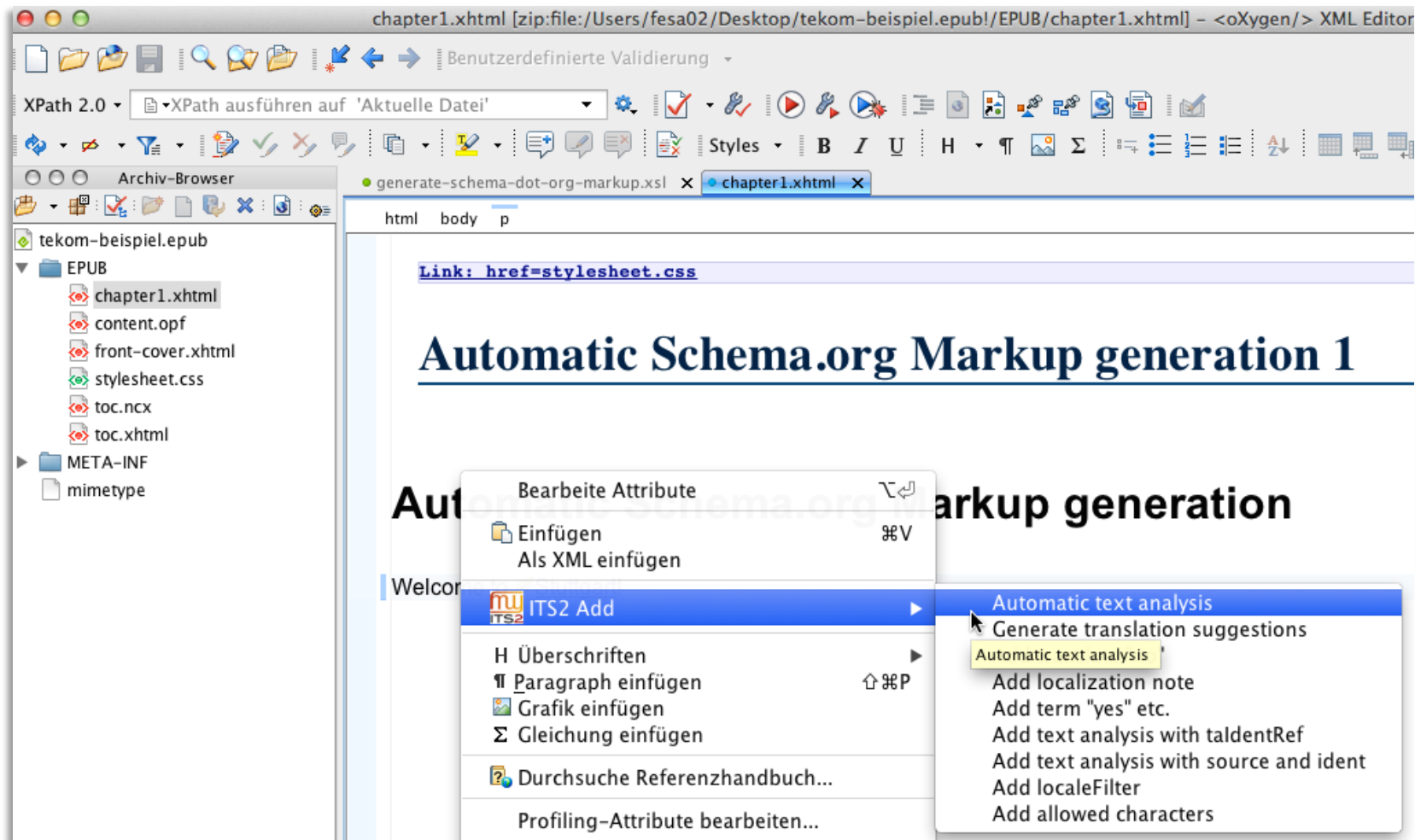
Strukturierte,
erweiterbare,
semantische
Inhaltsrepräsentationen
(XML, HTML5, ...)

Standardisierte
Annotation: W3C
Internationalization Tag
Set (ITS)



Weitere Möglichkeiten – Neue Produktionsprozesse

Autorenhilfe: automatische Textanalyse in oXygen



Weitere Möglichkeiten – Rohmaterial

Sprachverarbeitung (ULI)

smarted. This particular charge was passed by. The second charge was of a wrong done towards his father, Sir Ferdinand Gorges, and to the State. (William Bradford, public domain. <http://catalog.hathitrust.org/Record/008651224>)

Segment List

1. Of these, and also his commission, it pleased him to let the Governor here take a copy.
2. He gave them notice of his arrival by letter, but before they could visit him, he went eastward with the ship he came in; but a storm got up, and as they lacked a good pilot to harbour them there, they bore up for this harbour.
3. He and his men were kindly entertained here, and stayed fourteen days.
4. In the meantime Mr.
5. Weston arrived with his small ship, which he had now recovered.
6. Captain Gorges, who in- formed the Governor here that one purpose of his going east was to meet with Mr.
7. Weston, took this opportunity to call him to account for some abuses he had to lay to his charge.
8. Whereupon he called Weston before him and some other of his assistants, with the Governor of this place; and charged him first with the ill carriage of his men in Massachusetts, by which means the peace of the country had been disturbed, and he himself and the people he had brought over to settle there were much prejudiced.
9. To this Mr.
10. Weston easily answered, that what had been done, was in his absence, and might have befallen anyone; he left them sufficiently provided, and supposed they would be well governed; for any error committed he had sufficiently smarted.
11. This particular charge was passed by.
12. The second charge was of a wrong done towards his father, Sir Ferdinand Gorges, and to the State.
- 13.
14. (William Bradford, public domain. <http://catalog.hathitrust.org/Record/008651224>)

smarted. This particular charge was passed by. The second charge was of a wrong done towards his father, Sir Ferdinand Gorges, and to the State. (William Bradford, public domain. <http://catalog.hathitrust.org/Record/008651224>)

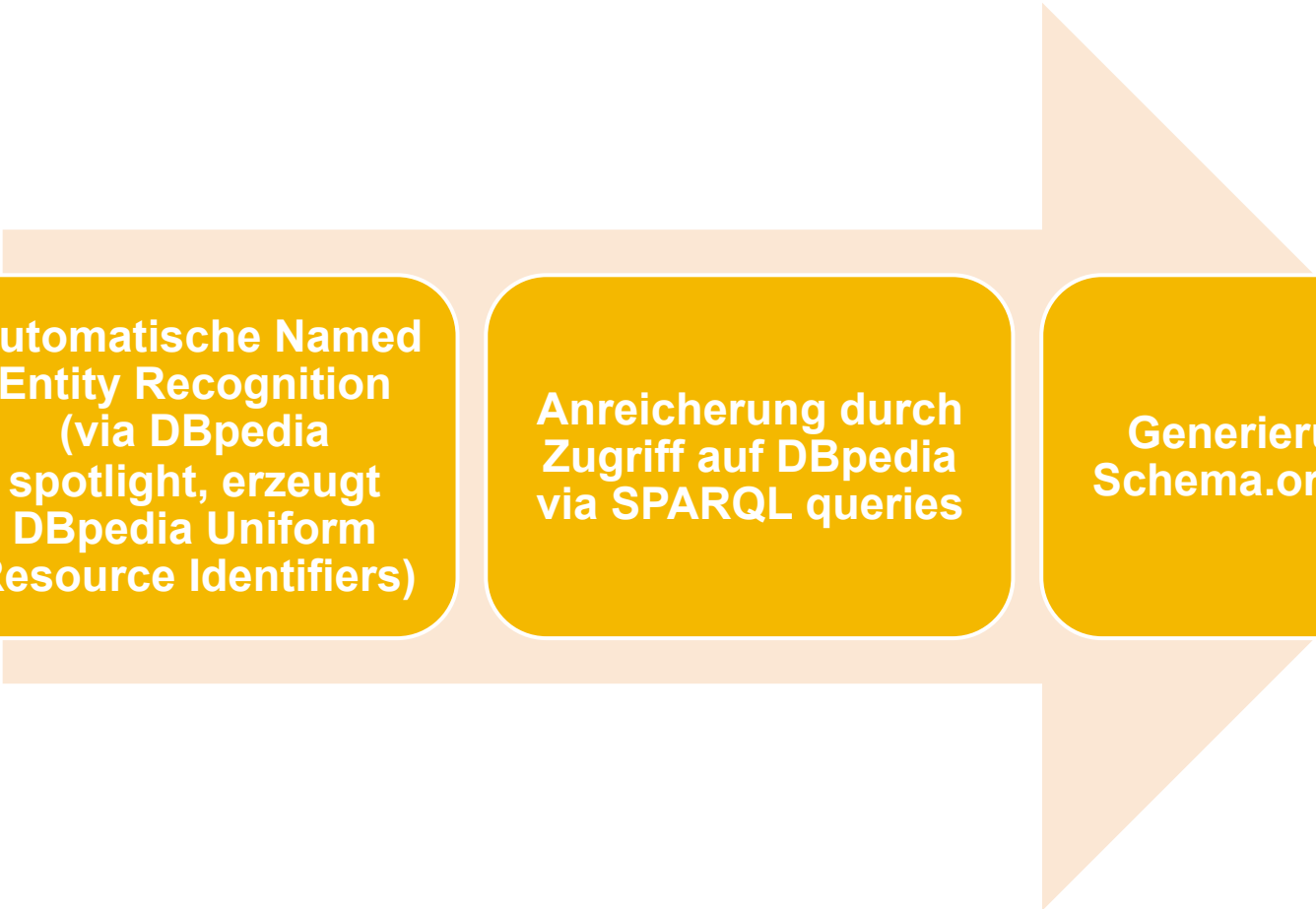
Segment List

1. Of these, and also his commission, it pleased him to let the Governor here take a copy.
2. He gave them notice of his arrival by letter, but before they could visit him, he went eastward with the ship he came in; but a storm got up, and as they lacked a good pilot to harbour them there, they bore up for this harbour.
3. He and his men were kindly entertained here, and stayed fourteen days.
4. In the meantime Mr. Weston arrived with his small ship, which he had now recovered.
5. Captain Gorges, who in- formed the Governor here that one purpose of his going east was to meet with Mr. Weston, took this opportunity to call him to account for some abuses he had to lay to his charge.
6. Whereupon he called Weston before him and some other of his assistants, with the Governor of this place; and charged him first with the ill carriage of his men in Massachusetts, by which means the peace of the country had been disturbed, and he himself and the people he had brought over to settle there were much prejudiced.
7. To this Mr. Weston easily answered, that what had been done, was in his absence, and might have befallen anyone; he left them sufficiently provided, and supposed they would be well governed; for any error committed he had sufficiently smarted.
8. This particular charge was passed by.
9. The second charge was of a wrong done towards his father, Sir Ferdinand Gorges, and to the State.
- 10.
11. (William Bradford, public domain. <http://catalog.hathitrust.org/Record/008651224>)

Status

Weitere Möglichkeiten – Neue Produkte/Dienstleistungen

Search Engine Optimization (via schema.org markup)



**Automatische Named
Entity Recognition
(via DBpedia
spotlight, erzeugt
DBpedia Uniform
Resource Identifiers)**

**Anreicherung durch
Zugriff auf DBpedia
via SPARQL queries**

**Generierung von
Schema.org Markup**

Schritt 1: Automatische Named Entity Recognition

Input

- `<p>Welcome to Stuttgart.</p>`

Output (als ITS 2.0 “Text Analysis” Markup)

- `<p>Welcome to <span its-ta-ident-ref="http://dbpedia.org/resource/Stuttgart"...>Stuttgart</span>!.</p>`

Schritt 2: Anreicherung durch Zugriff auf DBpedia via SPARQL queries

Nutzt DBpedia URIs aus Schritt 1.

Beispielabfrage, prüft ob Entität ein Ort (place) ist

- **SELECT ?wikiURI ...**
WHERE{ <<http://dbpedia.org/resource/Stuttgart>>
rdf:type <<http://schema.org/Place>>. ... }

Schritt 3: Generierung von Schema.org Markup

Input

- Teil 1: SPARQL Query Ausgabe und
- Teil 2: Dokument
- `<p>Welcome to <span its-ta-ident-ref="http://dbpedia.org/resource/Stuttgart"...>Stuttgart</span>!</p>`

Output: Dokument mit Schema.org Markup

- `<p>Welcome to
<span ... itemscope=""
itemtype="http://schema.org/Place">
... Stuttgart...!</p>`

Outputtest mit „Structured Data Testing Tool“

<http://www.google.com/webmasters/tools/richsnippets>

Extrahierte strukturierte Daten

Item

type: <http://schema.org/place>

property:

url: [Stuttgart](#)

name: Stuttgart



W3C LD4LT (Linked Data for Language Technology) Community Group

- <http://www.w3.org/community/ld4lt/>
- <https://www.w3.org/community/ld4lt/join>
- https://www.w3.org/community/ld4lt/wiki/Lider_roadmapping_activities
- Teilnahme ist frei 😊

LIDER: EU Projekt, November 2013 – Oktober 2015

- Basisplanung („Roadmap“) für eine umfangreiche „Linguistic Linked Data Cloud“ und Anwendungen im Bereich Content Analytics zu schaffen
- Anwendungsszenarien und Anforderungen von ... Ihnen



Zusammenfassung und Dank

Sprachverarbeitung, große Faktenbestände und Standardisierte Technologien erlauben veränderte Produktion, neue Verwertung und innovative Dienstleistungen

Die Europäische Gemeinschaft hat das Potential erkannt und kooperiert mit dem W3C, um Interessenten zusammenzuführen

Fragen & Antworten

Kontakt:

Felix Sasaki (fsasaki@w3.org), Christian Lieske (christian.lieske@sap.com)



Content Analytics und Linked Open Data

Neue Chancen auch für Ihre Inhalte

Felix Sasaki (DFKI / W3C Fellow)

Christian Lieske (SAP SE)



Ihre Meinung ist uns wichtig! Sagen Sie uns bitte, wie Ihnen der Vortrag gefallen hat. Wir freuen uns auf Ihr Feedback per Smartphone oder Tablet unter

<http://LT3.honestly.de>

oder scannen Sie den QR-Code



Das Bewertungstool steht Ihnen auch noch nach der Tagung zur Verfügung!

Disclaimer

All product and service names mentioned and associated logos displayed are the trademarks of their respective companies. Data contained in this document serves informational purposes only. National product specifications may vary.

This document may contain only intended strategies, developments, and is not intended to be binding upon the authors or their employers to any particular course of business, product strategy, and/or development. The authors or their employers assume no responsibility for errors or omissions in this document. The authors or their employers do not warrant the accuracy or completeness of the information, text, graphics, links, or other items contained within this material. This document is provided without a warranty of any kind, either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or non-infringement.

The authors or their employers shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials. This limitation shall not apply in cases of intent or gross negligence.

The authors have no control over the information that you may access through the use of hot links contained in these materials and does not endorse your use of third-party Web pages nor provide any warranty whatsoever relating to third-party Web pages.

Content Analytics und Linked Open Data – Neue Chancen auch für Ihre Inhalte

Dr. Felix Sasaki, DFKI GmbH/W3C, Berlin
Christian Lieske, SAP AG, St. Leon-Rot

Die maschinelle Auswertung großer Datenbestände in Echtzeit (real-time big data content analytics) ist Realität. Sie ermöglicht Anwendungen wie beispielsweise das Erkennen von Produktfehlern durch die Analyse von Social Media Beiträgen wie beispielsweise twitter tweets.

Derzeit beginnt die nächste Stufe der Entwicklung. Sie fügt den existierenden Ansätzen zweierlei hinzu:

- Inhalte werden vermehrt durch standardisierte, erweiterbare Repräsentationen erfasst. In der technischen Redaktion werden dafür beispielsweise XML-basierte Repräsentationen wie die Darwin Information Typing Architecture (DITA) eingesetzt. Durch die Kombination von Standardisierung und Flexibilität können Inhalte leicht unter anderem mit zusätzlichen Informationen angereichert bzw. verknüpft werden.
- Im Web werden frei zugängliche Wissensquellen als Basis von Informationsanreicherung oder –verknüpfung bereitgestellt. Eine Technologiebasis, welche Repräsentation und Verarbeitung dieser Wissensquellen in standardisierter Form erlaubt, wird als „Linked Open Data“ bezeichnet

„Content Analytics“ umfasst Ansätze für die Analyse, Generierung und Verknüpfung von Inhalten.

Im Folgenden wird Content Analytics insbesondere im Web-Kontext betrachtet. Dazu wird zunächst das Thema umrissen. Anschließend wird auf momentane Schwachstellen bzw. neue Anwendungsbereiche des Feldes eingegangen. Danach werden die Grundzüge von „Linked Open Data“ vorgestellt und beschrieben wie damit Schwachstellen beseitigt werden und neue Anwendungsbereiche entstehen. Beispiele für die Anwendungsbereiche zeigen wie insbesondere Inhalte der technischen Dokumentation mit „Linked Open Data“ und Content Analytics einen Mehrwert erhalten bzw. erzeugen.

Einführung: Was ist Content Analytics?

Produkt-bezogene twitter tweets wie "X schon nach 2 Monaten kaputt" und "Knopf an X funktioniert nicht" können sowohl für Käufer als auch für Verkäufer wichtige Information enthalten. "Content Analytics" bezeichnet eine Familie von Technologien, die aus solchen unstrukturierten Texten Informationen extrahieren oder gar Schlussfolgerungen ziehen. Beispiele:

- "Sentiment Analysis" generiert Information wie "15 Prozent aller tweets zu Produkt X enthalten eine negative Bewertung"
- "Business Intelligence" setzt auf landesspezifischem "Sentiment Analysis" auf um herauszufinden, ob es signifikante Bewertungsunterschiede zwischen Ländern/Märkten gibt

Technologische Grundlagen und Beispiele für Content Analytics

Basis von "Content Analytics" ist eine Reihe von Technologien aus der maschinellen Sprachverarbeitung. Beispiele:

- Erkennen/Extraktion von benannten Entitäten - oft einschließlich Zuordnung zu einer Entitätsklasse (named entity recognition/extraction)

Willkommen in Stuttgart
Stuttgart=city

- Erkennen/Extraktion von Beziehungen/Prädikaten (relationship extraction)

Stuttgart ist die Landeshauptstadt von Baden-Württemberg
(Stuttgart, ist-Landeshauptstadt-von, Baden-Württemberg)

Ein beispielhaftes Werkzeug für „Content Analytics“ ist DBpedia Spotlight“. Es bietet unter anderem named entity recognition an. Details dazu:

- Datengrundlage ist DBpedia, eine riesige, aus Wikipedia generierte Sammlung von Fakten
Verschiedene Nutzungsmöglichkeiten wie beispielsweise ein User Interface (siehe <http://dbpedia-spotlight.github.io/demo/> sowie Abbildungen 1 und 2) und eine Programmierschnittstelle (siehe <http://9nyurl.com/dbspotlight-demo>)

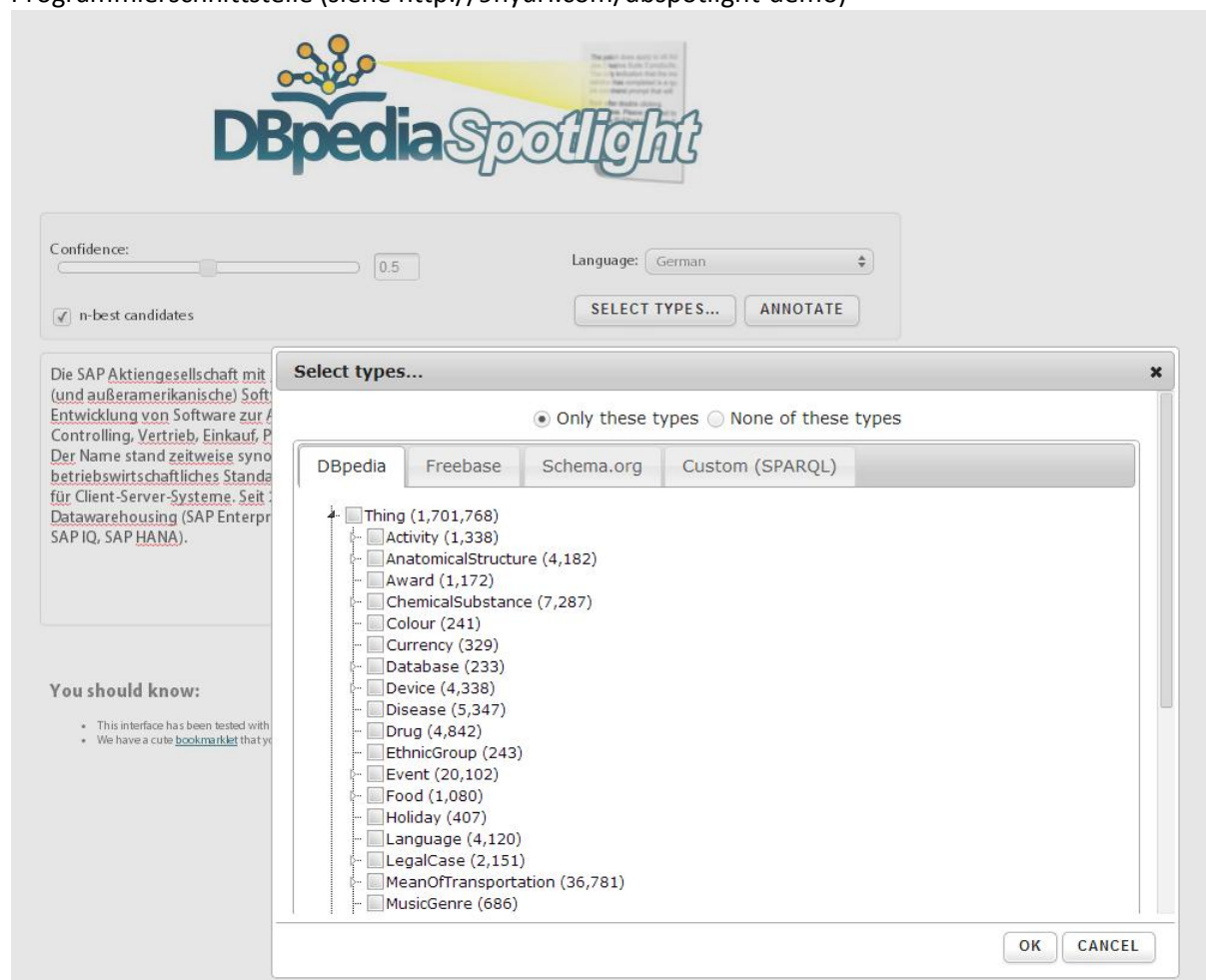


Abbildung 1: User Interface von DBpedia Spotlight – Eingabe



Abbildung 2: User Interface von DBpedia Spotlight – Ausgabe

Programmierschnittstellen wie die von DBpedia sind oft der Schlüssel zu mächtigen und nutzerfreundlichen Einsatzszenarien. <http://www.youtube.com/watch?v=F6zlW6blF5k> zeigt die Kopplung von DBpedia Spotlight mit dem oXygen Editor. Durch die Kopplung können Nutzer des Editors Texte automatisch mit Informationen aus Wikipedia annotieren.

Grundlagen von (Linguistic Linked) Open Data

Die Leistungsfähigkeit von Werkzeugen für „Content Analytics“ hängt unter anderem von der genutzten Datengrundlage ab. Beispiel DBpedia: Da die englische Wikipedia wesentlich größer ist als beispielsweise die deutsche, ist auch die deutsche DBpedia kleiner. Dementsprechend liefert die erwähnte Kopplung von DBpedia Spotlight mit dem oXygen Editor für das Deutsche andere Annotationen als für das Englische.

DBpedia ist ein prototypisches Beispiel für Datenquellen, die unter dem Begriff „Open Data“ in den letzten Jahren im Web verfügbar sind. Die für „Content Analytics“ relevanten dieser Datenquellen beinhalten Entitäten, Entitätsklassen, Entitätsrelationen, (Vorzugs-)Benennungen von Entitäten in verschiedenen Sprachen, Schreibvarianten etc.

DBpedia und viele andere Open Data Datenquellen nutzen einen standardisierten Technologiestack: „Linked Data“. Im Mittelpunkt stehen neben generellen Internettechnologien (wie beispielsweise dem http-Protokoll) zwei Grundprinzipien:

- Standardisierte Daten-Repräsentation (via RDF = Resource Description Framework)
- Standardisierter Zugriff (via SPARQL)

Durch diese Grundprinzipien können Datensätze ohne spezielle proprietäre Verarbeitung (Konvertierung, proprietäre Abfrage) genutzt werden – unter anderem für Werkzeuge des „Content Analytics“. Die Grundprinzipien ermöglichen auch eine einfache Verknüpfung der Datenquellen zur sogenannten „Linked Data Cloud“. Die für „Content Analytics“ besonders relevanten Datenquellen firmieren unter dem Begriff „Linguistic Linked Data Cloud“. Sie repräsentieren unter anderem Lexika, Textsammlungen zum Training von Tools, konzeptuelle Quellen wie Wortnetze etc. als Linked Data.

Technische Dokumentation und Content Analytics – Nutzungsszenarien

Inhalte der technischen Dokumentation können mit der Welt der „Content Analytics“ auf verschieden Art zusammen wachsen – insbesondere wenn die Inhalte auf standardisierten, strukturierten Formaten (z.B. DITA, DocBook, ePub, HTML5) basieren. Beispiele:

- Unterstützung beim Erstellen (vgl. die erwähnte Kopplung von DBpedia Spotlight mit dem oXygen Editor)
- Überarbeiten oder Verifizieren von Informationen (z.B. Nutzung von named entity recognition um die Zuordnung zu Entitätsklassen zu verifizieren)
- Automatische Anreicherung von Information bzw. Verknüpfung
- Einbeziehung von Dokumentstruktur in die maschinelle Sprachverarbeitung um beispielsweise Segmentierungsfehler zu vermeiden

Technische Dokumentation und Content Analytics – Marktperspektiven

Aus Sicht von Inhaltserstellern ergibt sich aus der Kombination mit „Content Analytics“ ein Marktpotential. Inhalte werden zu Rohstoffen für die gesamte insbesondere automatisierte Wertschöpfungskette für Content.

Ein Beispiel ist die automatische Generierung von strukturierten, sogenannten schema.org Markup Informationen (siehe https://www.w3.org/International/its/wiki/Its2-in-oXygen#For_oXygen_16.0 und Abbildung 3). Diese Informationen werden von Suchmaschinen interpretiert und erhöhen so die Sichtbarkeit von Inhalten im Web.

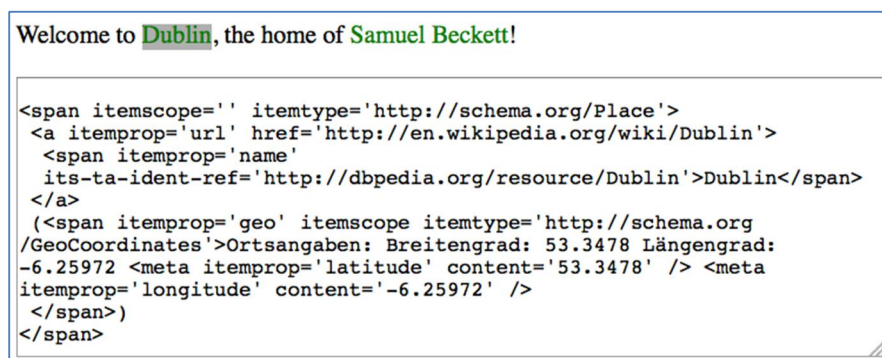


Abbildung 3: Generierung von schema.org Markup Informationen im Browser unter Nutzung von Open Data. Ausgangsbasis ist die Open Data Ressource <http://dbpedia.org/resource/Ireland>.

Inhaltsersteller haben also mit „Content Analytics“ Werkzeugen die Möglichkeit, Kunden zusätzlich zu den eigentlichen Inhalten deren Optimierung für Suchmaschinen anzubieten.

Einstieg in „Content Analytics“ durch Mitarbeit

Der „Content Analytics“ Zug nimmt derzeit Fahrt auf. Eine gute Möglichkeit, den Anschluss nicht zu verpassen bzw. zu steuern, ist die aktive Mitarbeit an relevanten Themen. Eine gute Möglichkeit dafür ist das EU Projekt LIDER „Linked Data as an enabler of cross-media and multilingual content analytics for enterprises across Europe“ (siehe <http://www.lider-project.eu/>). LIDER geht unter anderem folgenden Fragen nach:

- Welche Anwendungsszenarien für Linked Data gibt es?
- Welche Eigenschaften sind für die industrielle Nutzung von Open Data Inhalten relevant (z.B. hinsichtlich Lizenzen)?
- Welche Monetarisierungsmodelle sind für Open Data existent oder denkbar?

LIDER diskutiert diese und andere Fragen in unterschiedlichen Dimensionen. Die Erstellung mehrsprachiger Inhalte (mit den Teilbereichen Übersetzung und Lokalisierung) hat dabei einen hohen Stellenwert. Gleiches gilt für die Frage, wie Berufsfelder durch Open Data und „Content Analytics“ verändert werden. Wird aus dem Content Architect ein Content Curator?

Als Rahmen für die Diskussion nutzt LIDER die W3C „Linked Data for Language Technology“ Community Group (siehe <https://www.w3.org/community/ld4lt/>). An dieser Gruppe kann jeder Interessent/jede Interessentin mitwirken. Zu den Aktivitäten zählen Telefonkonferenzen, Workshops, Umfragen etc. Die (Zwischen)Ergebnisse werden der Öffentlichkeit im Gruppenwiki zur Verfügung gestellt.

Für Rückfragen:
fsasaki@w3.org; christian.lieske@sap.com