

Bibliotheken und Mehrsprachigkeit aus der Perspektive von Automatisierung

Felix Sasaki

DFKI / FH Potsdam

felix.sasaki@dfki.de

Übersicht

- Was ist das mehrsprachige Web?
- Nutzen durch Bibliotheken
- Beiträge durch Bibliotheken

Inhalte

W3C 대한민국 사무국

W3C는 미국의 메이저세션과 대회(MIT), 유럽의 정보수수료법연구소(ERCIM), 그리고 일본의 개인과 대학의 세국의 대회에 흑스트를 가지고 있는 무국인 지역 조직과의 파트너십 프로그램을 가지고 있습니다. W3C 사무국은 W3C 회의 연락을 위한 저작 거점 역할을 수행하며, 또한 W3C의 기술 명세를 확보하고 다음과 같이 요청할 수 있습니다.

개발자, 설계자, 그리고 표준 전문가들에게 W3C 외연을 색색을 출전하게 하고, 향후 외연을 만들 수 있는 지역 기관의 기합을 촉려합니다.

▶ Video on the Web 2008 개최 (사무국)

2008년 9월 08일 ~ W3C 대한민국 사무국과 IPTV 포럼 코리아는 2008년 9월 25일과 26일 서울 노현동의 건설회관 대강당에서 Video on Demand 서비스를 활용해 IPTV 콘텐츠를 제공하는 웹2.0 시대에 진정으로 사용자가 원하는 IPTV비디오 서비스를 무엇인가를 사용자와 공유 비포장과 행사 등록은 9월 22일까지이며, 온라인으로 등록 확인을 가능하다 ([\[온라인등록\]](#))

▶ 2011 월드 웹 컨퍼런스 개최 (사무국)

2008년 9월 08일 ~ W3C-Magyar Iroda

Nézet: német mobil nyomtatás Nyelv: Magyar English

SZOLGÁLTATÁSAINK | W3C-TAGSÁG | RÖLINK | KAPCSOLAT

Felhasználói Program Akadálymentességi Útmutató (UAAG) 2.0 frissítve
2010. április 6.

A Felhasználói Program Akadálymentességi Útmutató Munkacsoport (User Agent Accessibility Guidelines Working Group) meghirdette az általában Felhasználói Program Akadálymentességi Útmutató 2.0 (UAAG 2.0) munkatervet. Az UAAG meghatározza, hogyan kell a böngészőknek, médiás lejátszóknak és más "felhasználói programoknak" támogatni az akadálymentességet a fogatkozó emberek számára és hogyan kell egysülműenük a kiegészítő technológiákkal. A munkaterv négy részre tagolódik: a részletekkel a beszélgetésről támogatásáról, a videoelőzetés ránkénti részről és egy új fejezet a megfelelőségről. A munkacsoport egy új segítséget is kozredit, az UAAG 2.0 alkalmanként (Implementing UAAG 2.0) címen, ezt nyilvános munkatervként. Elszerezték az UAAG munkatervről valamint többi információt a Web Akadálymentességi Kezdeményezőről (WA).

Mobilweb beállításokat tartalom átalakítási útmutató előrehaladása
2010. április 6.

A Mobilweb Beállításokat tartalmazó Munkacsoport (Mobile Web Best Practices Working Group) szívesen látja a Mobilweb Akadálymentesítők Beállításokat (Mobile Web Application Best Practices) elızetés javaslatának alkalmazását. A dokumentum célja hogy segítse az igény gazzadás dinamikus mobilwebs alkalmazások fejlesztését. Összegyűjti a legnépszerűbb működési eljárásokat, támogatva azokat jobb felhasználói élményt tesznek lehetővé, és türelmet adnak a környezetnek. A W3C az alkalmazások egy alkalmazásai jelentési összefoglaló (implementation report template) kitöltésére kéri. A munkacsoport szintén ma több közös felhívást a Websztártan Általánosított 1.0 munkavizsgárára (Guidelines for Web Content Transformation Proxies 1.0), amely útmutatóként szolgál a tartalomalkotó profiloknál, hogy lehetősége-s, és ha igen, hogyan a webstártan átalakítására. További információ a mobilweb kezdeményezőről: Mobile Web Initiative.

A WebCGM 2.1 W3C ajánlás lett
2010. április 2.

A Web Számtípusú Grafikai Metáfájl Munkacsoport (WebCGM Working Group) kiadta a Web Számtípusú Grafikai Metáfájl (WebCGM 2.1) W3C ajánlását. A Számtípusú Grafikai Metáfájl (CGM) egy ISO szabvány, melyet az ISO/IEC 8632:1999 definiál, a 2D vektor és a kevert vektor/grafikai felcserélhetőségeire. A WebCGM a CGM egyik profja, amely weblapokat és a weblap alkalmazásokra optimalizált teknikai ábrák, elektronikai dokumentáció, geodaták meglételeseinek eléréséhez. Egyik előnye a hiperlinkekkel való közelítéssel teremteni a grafikai kifejezéstől valamint az egyszerűség és használhatóság között. Egyik előnye a hatékony, egységes metadat elemekkel történő készítés támogatja a hiperlinkelésökkel a dokumentum navigációt, valamint lehetőséget tenni a WebCGM képi tartalmaknak keresését és lekérdezését. A WebCGM 2.1 finomította és kiélezte a WebCGM 2.0 tulajdonságait. További információ: Graphics Activity.

Megjelent a Biztonságos Webs Környezet: Felhasználói Felületek Útmutatója végső munkaterv

Ελληνικό Γραφείο του W3C

Views: desktop mobile print

ΕΣΧΗΜΑ ΤΟΥ W3C | ΠΕΡΙΓΡΑΦΗ | ΣΤΟΧΟΙ ΤΟΥ W3C

ΕΛΛΗΝΙΚΟ ΓΡΑΦΕΙΟ ΤΟΥ W3C

Το Ελληνικό Γραφείο του W3C εξέδωσε τη νέα αντίρρηση στο πλαίσιο της διοικήσης των δεκάετων λειτουργίας του. Μπορείτε να τη διαβάσετε εδώ, σε μορφή pdf.

ΕΔΡΑ ΕΛΛΗΝΙΚΟΥ ΓΡΑΦΕΙΟΥ ΤΟΥ W3C

Το Ελληνικό Γραφείο του W3C εξέδωσε τη νέα αντίρρηση στο πλαίσιο της διοικήσης των δεκάετων λειτουργίας του. Μπορείτε να τη διαβάσετε εδώ, σε μορφή pdf.

FORTH Institute of Computer Science

To ΙΤΕ έχει μια καταθέματη περιόδου διεθνών συναντήσεων αριθμούς στη διεθνών γνωστή και εφαρμοσμένη εργασία στην πλατφόρμα της.

Potential..

W3C ने अन्यायी संस्करण का प्राप्ति किया है। W3C ने अन्यायी संस्करण का प्राप्ति किया है।

के लिए अल्टरेंशनल योजनाओं के बारे में विवरणित करता है।

स्थानीय काफिनो (Local Links)

schandra@mit.gov.in पर जेज रखते हैं, इसे इसे

- Website in English
- दृष्टिकोणी (W3C) विवरणित
- आगत व W3C संस्करण गतिविधियां
- दृष्टिकोणी (W3C) के संस्करण संस्करण
- वर्तमान समाचार, समाचार-प्र
- प्रतीक्षा (Feed back)
- इमेल-लिस्ट (Mailing List)
- दृष्टिकोणी (W3C) आगती कारबोल वर्षों का जारी

अंतर्राष्ट्रीय काफिनो (International Links)

- विश्व कारबोल (World Offices)

आगामी आवाज

W3C के दस्तावेजों का दिनी में अनुग्रह

- सारांश
- दिनदी
- यात्राएँ
- तेज़ी
- संकेत

Interfaces

The image displays three side-by-side screenshots of the W3C CSS Validation Service website, illustrating its multilingual capabilities and various input methods.

- Top Screenshot (English):** Shows the English version of the service. It features a "CSS Validation Service" logo and a sub-header "Validación de hojas de estilo en cascada (CSS) y documentos (X)HTML". Below this are tabs for "URI", "파일 업로드" (File Upload), and "직접 입력" (Direct Input). A search bar labeled "URI 검사" (Check URI) is present. The top navigation bar includes links for Deutsch, English, Español, Français, 한국어, Italiano, Nederlands, 日本語, Polski, Português, Русский, Svenska, Български, Українська, Čeština, Romanian, and 简体中文.
- Middle Screenshot (Korean):** Shows the Korean version of the service. The main title is "CSS Validation Service" and the sub-header is "캐스케이딩 스타일 시트 (CSS) 와 스타일 시트를 포함한 (X)HTML 문서 검사". The tabs and search bar are identical to the English version. The top navigation bar includes links for Deutsch, English, Español, Français, 한국어, Italiano, Nederlands, 日本語, Polski, Português, Русский, Svenska, Български, Українська, Čeština, Romanian, and 简体中文.
- Bottom Screenshot (Spanish):** Shows the Spanish version of the service. The main title is "CSS Validation Service" and the sub-header is "Verifica Hojas de Estilo en Cascada (CSS) y documentos (X)HTML con hojas de estilo". The tabs and search bar are identical to the English version. The top navigation bar includes links for Deutsch, English, Español, Français, 한국어, Italiano, Nederlands, 日本語, Polski, Português, Русский, Svenska, Български, Українська, Čeština, Romanian, and 简体中文.

Validation Options: Each screenshot shows a "Validación por URI" section with a "Check" button. Below this, there are tabs for "mediante URI", "mediante Carga de Archivo" (File Upload), and "mediante Entrada directa" (Direct Input). A note at the bottom of the page encourages users to "Donate" to support the validators.

Footers: The bottom of each screenshot includes a "Nota" (Note) about validating embedded CSS, links to "Acerca de este servicio" (About this service), "Documentación" (Documentation), "Download", "Comentarios" (Comments), and "Créditos" (Credits), and a copyright notice from 1994-2009 W3C.

Zugriff auf Inhalte – Services zur Überbrückung von Sprachgrenzen



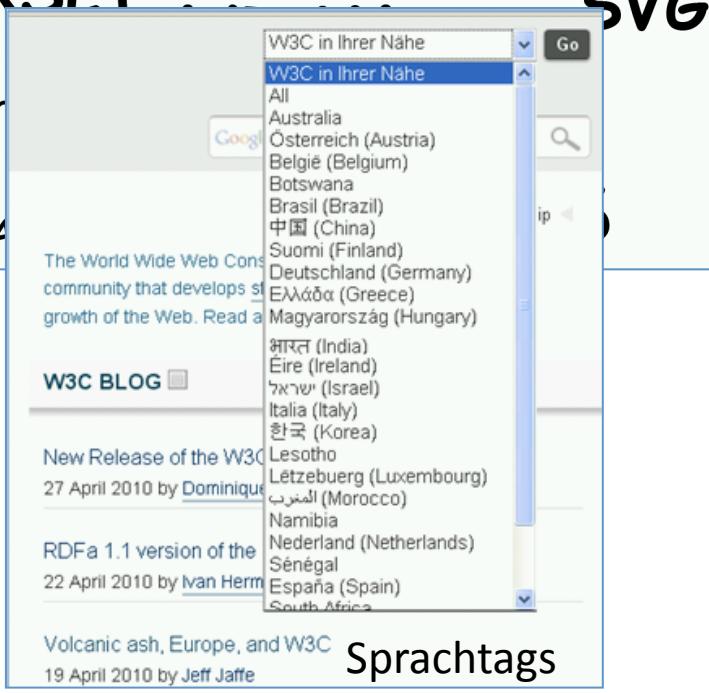
The screenshot shows the Google Translate homepage. At the top, the "Google übersetzer" logo is displayed. Below it, there are two input fields: "Von: Englisch" and "Nach: Deutsch", separated by a blue double-headed arrow icon. To the right of these fields is a large "Übersetzen" button. Below these controls is a large, empty rectangular text area for pasting text. At the bottom of the page, there is a instruction text: "Geben Sie Text oder eine Website-Adresse ein oder lassen Sie ein Dokument übersetzen."

Übersicht

- Was ist das mehrsprachige Web?
- Nutzen durch Bibliotheken
- Beiträge durch Bibliotheken

Nutzen I: Technische Infrastruktur

Unicode in Webtechnologien



OWL RDF SOAP
XPath VoiceXML

XSLT

Math

WA

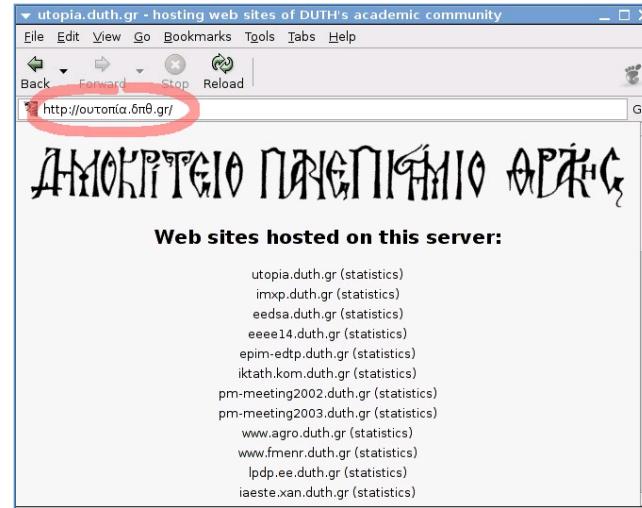
The World Wide Web Consortium is the community that develops standards to ensure the continued growth of the Web. Read about our work.

W3C BLOG

- New Release of the W3C... 27 April 2010 by Dominique
- RDFa 1.1 version of the... 22 April 2010 by Ivan Herman
- Volcanic ash, Europe, and W3C 19 April 2010 by Jeff Jaffe

Sprachtags
wie "en", "en-us", "de", ...

Internationalisierte Domännamen (IDN)



Internationalisierte Identifikatoren (IRI)



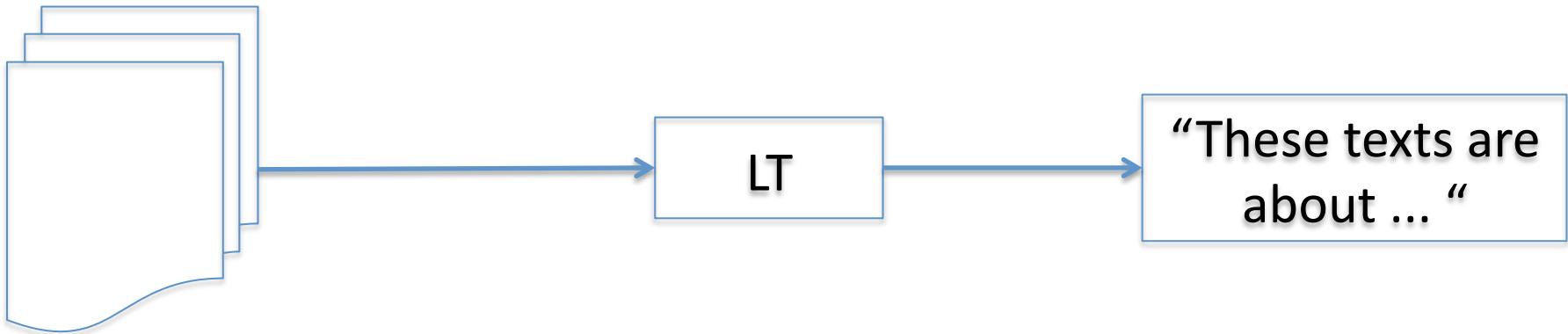
Nutzen II: automatisierte Informationsverarbeitung

- Suche über Sprachgrenzen hinweg
 - Automatische Übersetzung, Cross Language Information Retrieval
- Anreicherung mit (mehrsprachigen) Informationen
 - Verlinkung von Normdateien über Sprachgrenzen hinweg (DNB > VIAF > LC)

Nutzung von Sprachtechnologie („Language Technology“)

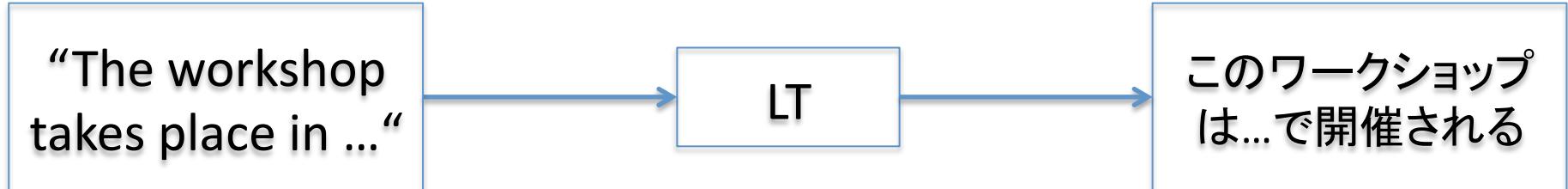
Sprachtechnologie: exemplarische Anwendungen

- Automatische Zusammenfassung



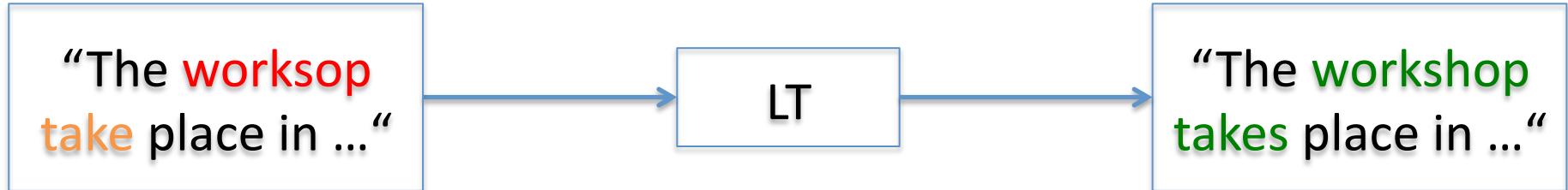
Sprachtechnologie: exemplarische Anwendungen

- Maschinelle Übersetzung



Sprachtechnologie: exemplarische Anwendungen

- Rechtschreib- / Grammatikkorrektur



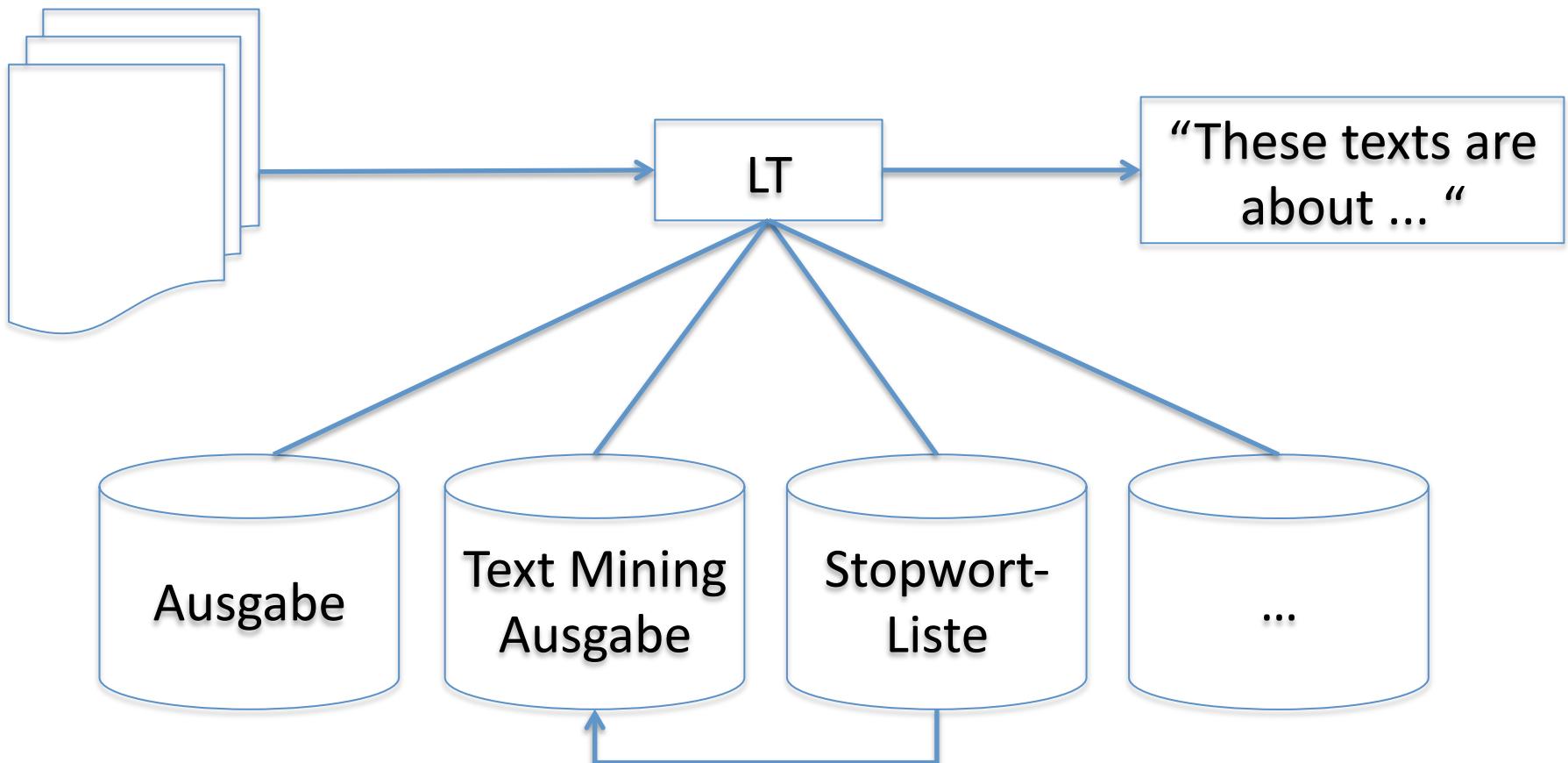
- Und viele weitere Anwendungen
 - Koreferenzauflösung, Diskursanalyse, Erkennung von Eigennamen, Generierung natürlicher Sprache, ...

Grundlagen:
Was haben alle diese
Anwendungen von
Sprachtechnologie gemeinsam?

Sie nutzen Ressourcen

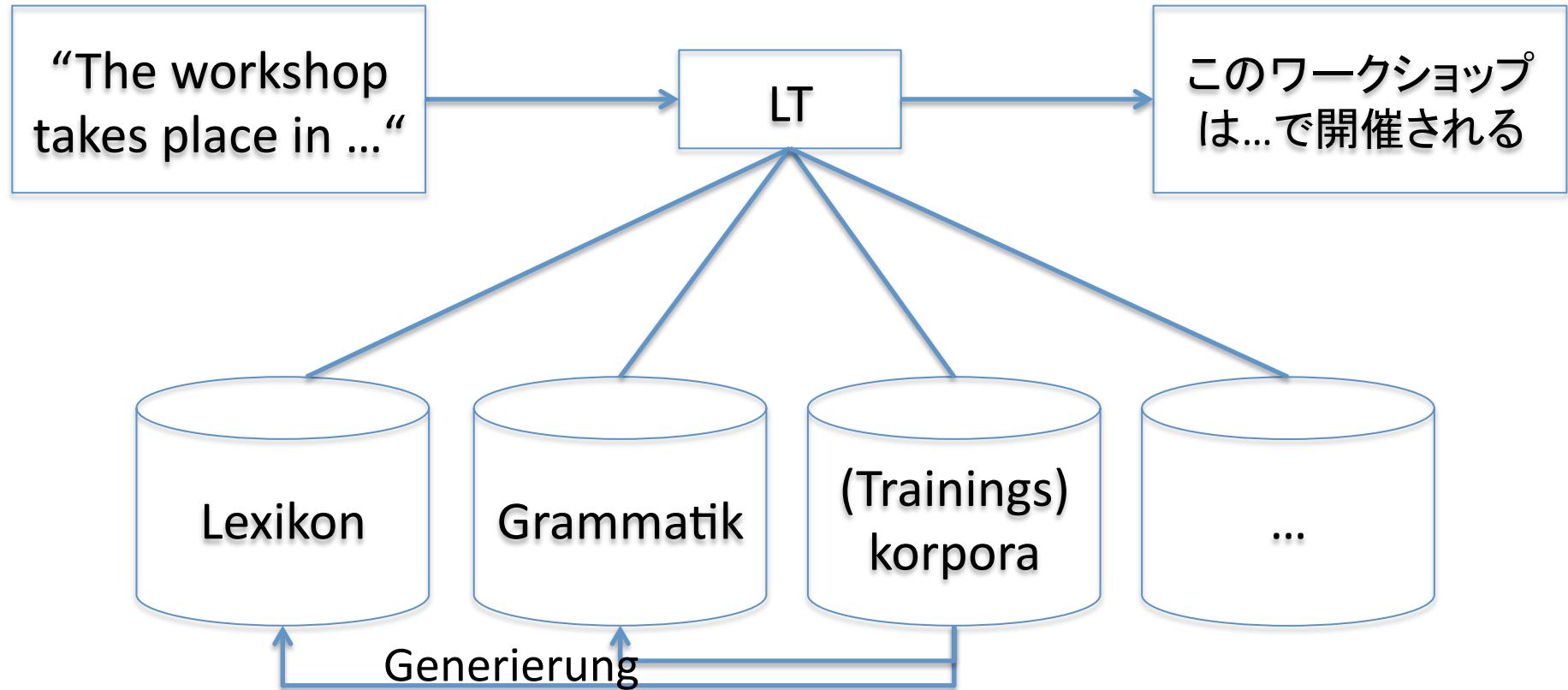
Ressourcen in der Sprachtechnologie

- Ressourcen für die Zusammenfassung



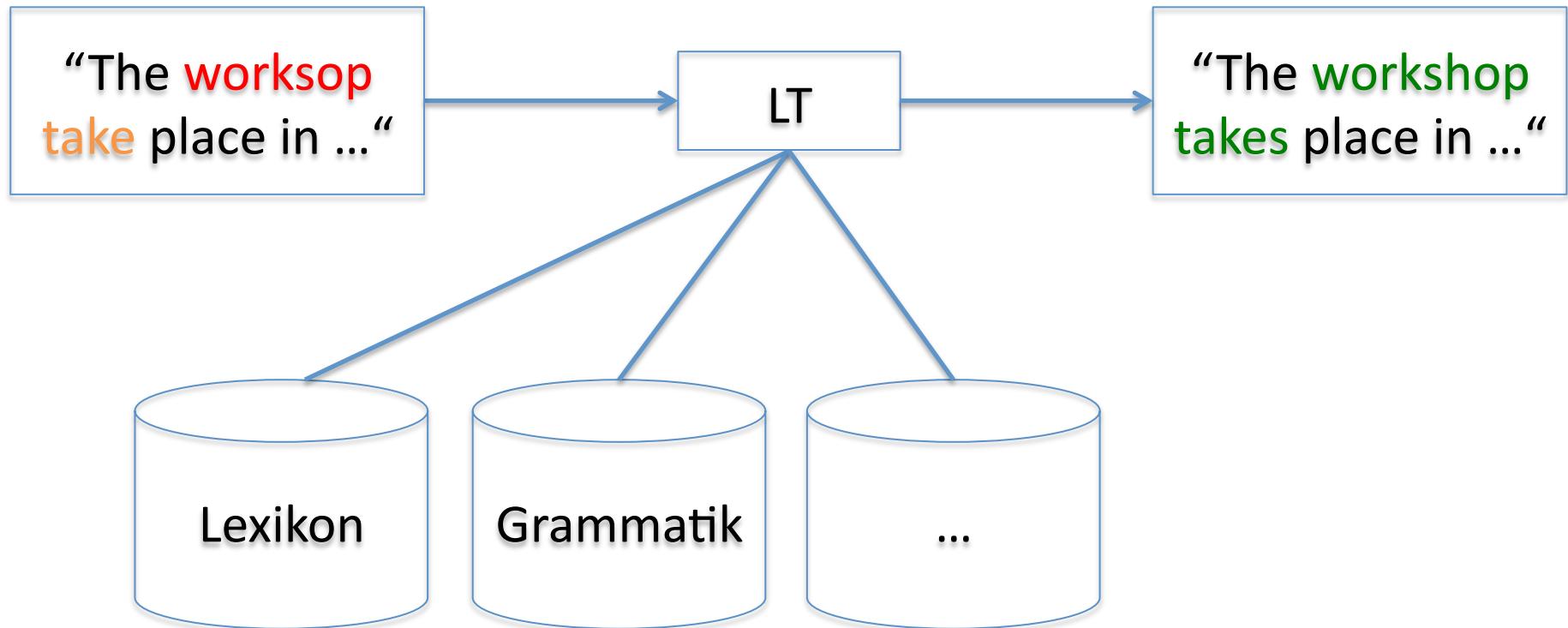
Ressourcen in der Sprachtechnologie

- Ressourcen für maschinelle Übersetzung



Ressourcen in der Sprachtechnologie

- Ressourcen für (Rechtschreib) Kontrolle



Vorläufige Zusammenfassung

- Bibliotheken nutzen im mehrsprachigen Web
 - Technische Infrastruktur
 - Automatische Methoden der mehrsprachigen Informationsverarbeitung
- Automatische mehrsprachige Informationsverarbeitung = Sprachtechnologie
 - Verschiedene Anwendungen (automat. Übersetzung, ...)
 - Immer: Nutzung von Ressourcen

Übersicht

- Was ist das mehrsprachige Web?
- Nutzen durch Bibliotheken
- Beiträge durch Bibliotheken

Vorläufige Zusammenfassung (Wiederholung und Frage)

- Bibliotheken nutzen im mehrsprachigen Web
 - Technische Infrastruktur
 - Automatische Methoden der mehrsprachigen Informationsverarbeitung
- Automatische mehrsprachige Informationsverarbeitung = Sprachtechnologie
 - Verschiedene Anwendungen (automat. Übersetzung, ...)
 - Immer: Nutzung von Ressourcen
 - Woher kommen die Ressourcen her?

Linguistische Ressourcen ...

- Linguistische Korpora
- Lexika
- „Ontologien“
- Einfache Tools
 - Morphologische Tagger
 - Syntaktische Parser
- Komplexe Tools
 - Erkennung von Eigennamen

... brauchen / umfassen Entitäten

- Linguistische Korpora
- Lexika
- „Ontologien“
- Einfache Tools
 - Morphologische Tagger
 - Syntaktische Parser
- Komplexe Tools
 - Erkennung von Eigennamen

Beispielnutzung von Entitäten (I)

- Anreicherung von Korpora mit Informationen aus Normdateien
 1. Eingabe (Ausschnitt): "Das von J. W. v. Goethe verfasste Buch ..."
 2. Semantische Verarbeitung: automatische Erkennung von Eigennamen
 3. Ergebnis: "Das von <pers pnd-link="118540238" viaf-link="24602065">J. W. v. Goethe</pers> verfasste Buch ..."

Beispielnutzung von Entitäten (II)

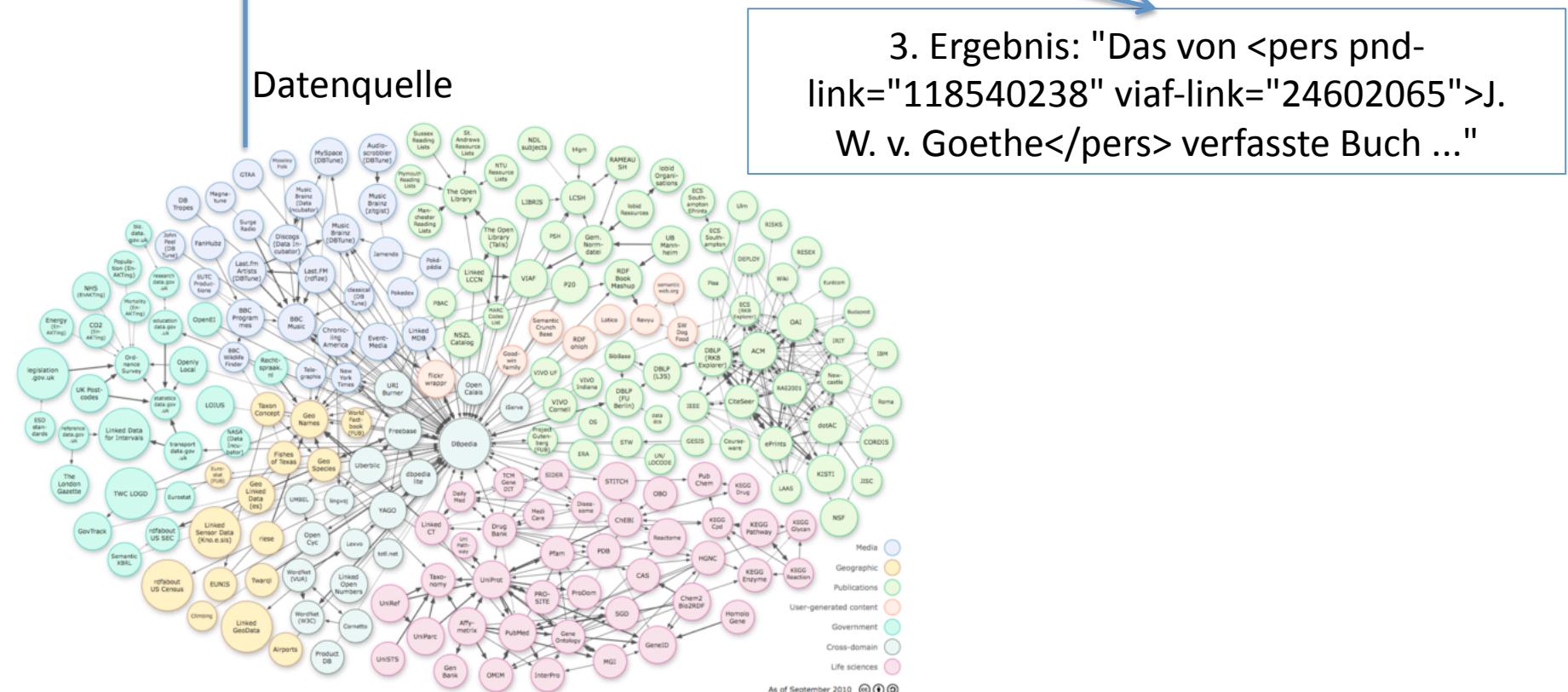
- Training und Evaluierung von Textkategorisierung, mit „getagten“ digitalisierten Büchern
 - Eingabe: Thematisch manuell klassifizierte digitale Objekte
 - Verarbeitung z.B.
 - Training von LT-Tools die Texte kategorisieren
 - Evaluierung entsprechender Tools

Sprachtechnologische Tools
verlassen sich in zunehmendem
Maße auf „externe Ressourcen“
wie „Linked Open Data“

1. Eingabe (Ausschnitt): "Das von J. W. v.
Goethe verfasste Buch ..."

2. Semantische Verarbeitung: automatische Erkennung von Eigennamen

Datenquelle



Probleme

- Heterogene Qualität der Datenquellen
- Nachhaltigkeit
- Rechtliche Verlässlichkeit
- ...

Lösungen

- Bibliotheken haben das Potential ein Rückgrat für sprachtechnologische Verarbeitung im Web zu sein
- Einige Schwierigkeiten der müssen zur Einlösung dieses Potentials gelöst werden sein
- Stichwort „Datenheterogenität“
 - Syntaktisch
 - Semantisch

Syntaktische Heterogenität

Ausschnitt aus der Entitätsdefinition

<http://viaf.org/viaf/24602065/rdf.xml>

- Goethe,+J.+W.+v.
- Goethe,+Johann+Wolfgang+von
- VonGoethe,+Johann+Wolfgang
- Heterogenität für menschlichen Leser kein Problem
- Nutzung durch automatische Verarbeitung problematisch

Semantische Heterogenität

- Semantische Beziehungen für Nomen in WordNet
 - Hyperonym („Pudel“ ist Hyperonym von „Hund“), Hyponym („Hund“ ist Hyponym von „Pudel“), Holonym („Haus“ ist Holonym von „Fenster“), Meronym („Fenster“ ist Meronym von „Haus“), ...
- Nicht immer abbildbar auf Beziehungen in bibliothekarischen Ressourcen
 - Was ist OB / UB / VB (Thesaurus), was ist Oberklasse / Unterklasse (Klassifikation)

(Forschungs)agenda

- Bei der Bereitstellung von bibliothekarischen Ressourcen im (Semantic) Web – evaluieren:
 - Prinzipielle Eignung für ausgewählte Sprachtechnologische Anwendungen
 - Konkrete Probleme (mit dem Potential der Wiederverwendbarkeit) syntaktischer und semantischer Heterogenität
- Nachhaltigkeit der Bereitstellung
- Rechtliche Verlässlichkeit / freie Verfügbarkeit

(Forschungs)agenda

- Den kleinsten gemeinsamen Nenner finden
 - Syntaktische Heterogenität: schwierig
 - Semantische Heterogenität: möglich, vgl. „Europeana Data Model“
- ... und seine Anwendungen über Community-Grenzen hinweg sichern
 - Bibliothekar trifft Sprachtechnologe!

Wurde das nicht schon getan?

- Konkordanzen zwischen kontrollierten Vokabularen
- Mapping von Ontologien
 - Direkt
 - Via Upper-Level Ontologies (SUMO etc.)
- ...

Neu:

- Ziel z.B. für Themen: sehr simple Kategorisierung
 - < 100 Klassen
- Berücksichtigung von existierenden Kategorisierungen in der Sprachtechnologie
- Ziel für Entitätsdefinitionen: wenige, leicht wiederverwendbare Entitätsdefinitionen

Hintergrund: EU-Projekt „META-NET“

- Hauptziel: Langfristige Allianz META für Sprachtechnologie in Europa bauen
 - <http://www.meta-net.eu/join-forms/organisation-form>
- META-NET umfasst mehr als 40 teilnehmende Organisationen aus 30+ Ländern
- META hat bereits über 200 Mitglieder
- Wichtig: Nutzer von und Beiträger zu Sprachtechnologie involvieren – auch Bibliotheken!

META-NET

- Event: META-FORUM 2011
- Budapest, 27.-29. Juni 2011
- Ziel: Nutzer / Sprachtechnologieentwickler / „Entscheider“ zusammenbringen
- Ziele für die Sprachtechnologie in den nächsten 10 Jahren diskutieren
- Details und bald Registrierung unter

<http://www.meta-net.eu/events>

Übersicht

- Was ist das mehrsprachige Web?
- Nutzen durch Bibliotheken
- Beiträge durch Bibliotheken

Vielen Dank für Ihre Aufmerksamkeit

