

Jakob Voß
Verbundzentrale des GBV (VZG)

Strukturierte Metadaten in Wikipedia

Vortrag auf dem KIM-Workshop
*„Sind Metadaten teamfähig oder
was macht Metadatenformate interoperabel?“*
auf dem Bibliothekartag in Mannheim
am 04.06.2008

___TOC___

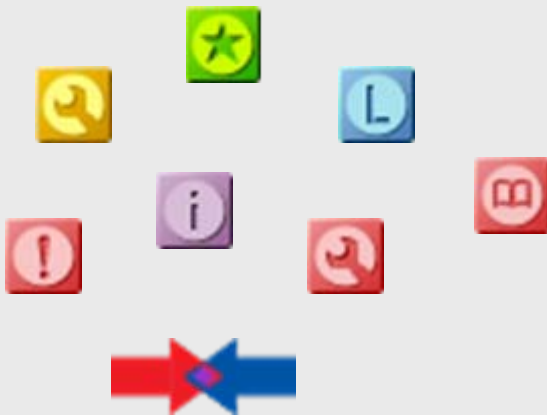
- Metadaten und Wikipedia
- Zugriff auf die Daten
- Qualität
- Anwendungsbeispiele
- Bewertung

Metadaten und Wikipedia

- Metadaten *über* Wikipedia
- Metadaten *in* Wikipedia

Metadaten und Wikipedia

- **Metadaten *über* Wikipedia**
 - Artikel, Autoren, Versionen, Verweise ...
 - Bewertung, Wartung ...



| | | | | | | | |
|---|-----------------------|---|----------------------|---------------------------------------|----------------|------------------------------|--------------------------------|
| ■ | (Aktuell) (Vorherige) | ● | 15:15, 24. Mai 2008 | Demonax (Diskussion Beiträge) | K (804 Bytes) | (Tippo) (rückgängig) | [gesichtet von Demonax] |
| ■ | (Aktuell) (Vorherige) | ● | 15:13, 24. Mai 2008 | Demonax (Diskussion Beiträge) | (804 Bytes) | (Wikilink) (rückgängig) | |
| ■ | (Aktuell) (Vorherige) | ○ | 20:50, 12. Okt. 2007 | Gu drun Meyer (Diskussion Beiträge) | K (699 Bytes) | (Korr Wikilink) (rückgängig) | |
| ■ | (Aktuell) (Vorherige) | ○ | 12:18, 24. Jan. 2007 | Janneman (Diskussion Beiträge) | K (rückgängig) | | |

Metadaten und Wikipedia

- **Metadaten *in* Wikipedia**
 - Kategorien
 - Spezielle Seiten: Listen, Chronik ...
 - Vorlagen
 - Infoboxen
 - Navigationsleisten
 - Personendaten
 - Datenbanklink (z.B. PND-Links, ISBNs ...)
 - Literaturangaben, Zitate ...
 - Interwikilinks, Begriffsklärungsseiten ...

Beispiel

```

{{Infobox Ort in Deutschland
|Art                = Stadt
|Wappen             = Wappen voelklingen.jpg
|lat_deg            = 49 |lat_min = 15
|lon_deg            = 06 |lon_min = 50
|Lageplan           =
|Bundesland         = Saarland
|Landkreis          = Regionalverband Saarbrücken
|Höhe               = 182–328
|Fläche             = 67.06
|Einwohner          = 40453<!--Bitte nicht pe
|Stand              = 2006-12-31
|PLZ                = 66301–66333
|Vorwahl            = 06898, 06802
|Kfz                = VK
|Gemeindeschlüssel = 10 0 41 519
|Gliederung         = 10 Stadtteile
|Adresse            = Rathausplatz<br />66333
|Website            = [http://www.voelklingen
|Bürgermeister     = Klaus Lorig
|Bürgermeistertitel = Oberbürgermeister
|Partei             = CDU
}}

```

| Wappen | Deutschlandkarte |
|---|--|
|  |  |
| Basisdaten | |
| Bundesland: | Saarland |
| Landkreis: | Regionalverband Saarbrücken |
| Höhe: | 182–328 m ü. NN |
| Fläche: | 67,06 km² |
| Einwohner: | 40.453 (31. Dez. 2006) |
| Bevölkerungsdichte: | 603 Einwohner je km² |
| Postleitzahlen: | 66301–66333 |
| Vorwahlen: | 06898, 06802 |
| Kfz-Kennzeichen: | VK |
| Gemeindeschlüssel: | 10 0 41 519 |
| Stadtgliederung: | 10 Stadtteile |
| Adresse der Stadtverwaltung: | Rathausplatz 66333 Völklingen |
| Webpräsenz: | www.voelklingen.de  |
| Oberbürgermeister: | Klaus Lorig (CDU) |

Zugriff auf die Daten

- Eigene Verarbeitung
 - <http://www.mediawiki.org/wiki/API>
 - Datenbank-Dumps (etwa monatlich)
Extrahierte Daten
 - DBPedia
 - Extractor
 - TemplateTiger
- Strukturierte Metadaten direkt
 - Semantic MediaWiki (geplant)



Qualität der Daten

- Spezialisierte Autorentams und Redaktionen
- Quellenangaben
- „Endredaktion“/„Releases“ geplant mit den „geprüften Versionen“
- Wie gut sind die Alternativen?!

Qualität durch Feedback

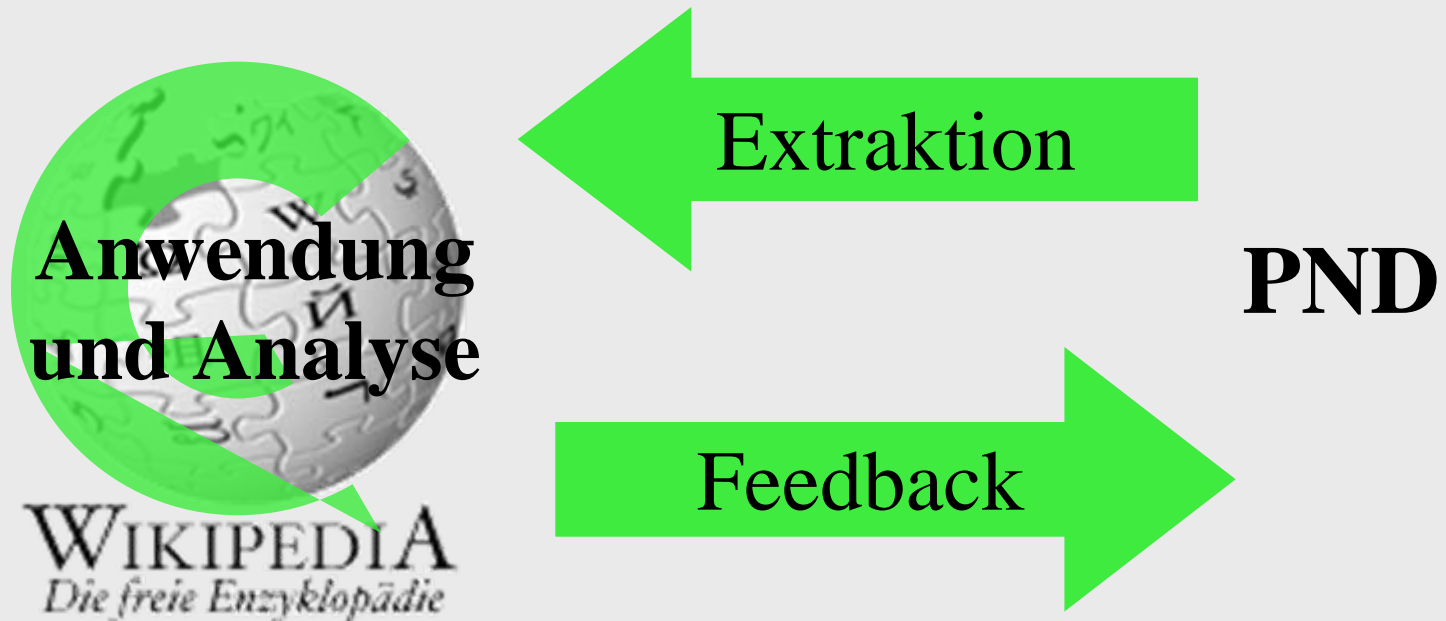


Extraktion

Anwendung
und Analyse

Feedback

Qualität durch Feedback



durchschnittlich etwa 2-3 Korrekturen/Tag
siehe [[Wikipedia:PND/Fehlermeldung]]

Anwendungsbeispiele

- RDF-Export als „Linked Data“ und Zugriff via SPARQL (DBPedia)
- Automatische Thesaurus-Erstellung
- Analyse von Literaturangaben

- Darauf aufbauende Anwendungen und Webservices („Mashups“)

http://dbpedia.org/resource/Leipzig

Einfache Ansicht



Leipzig | DBpedia.org

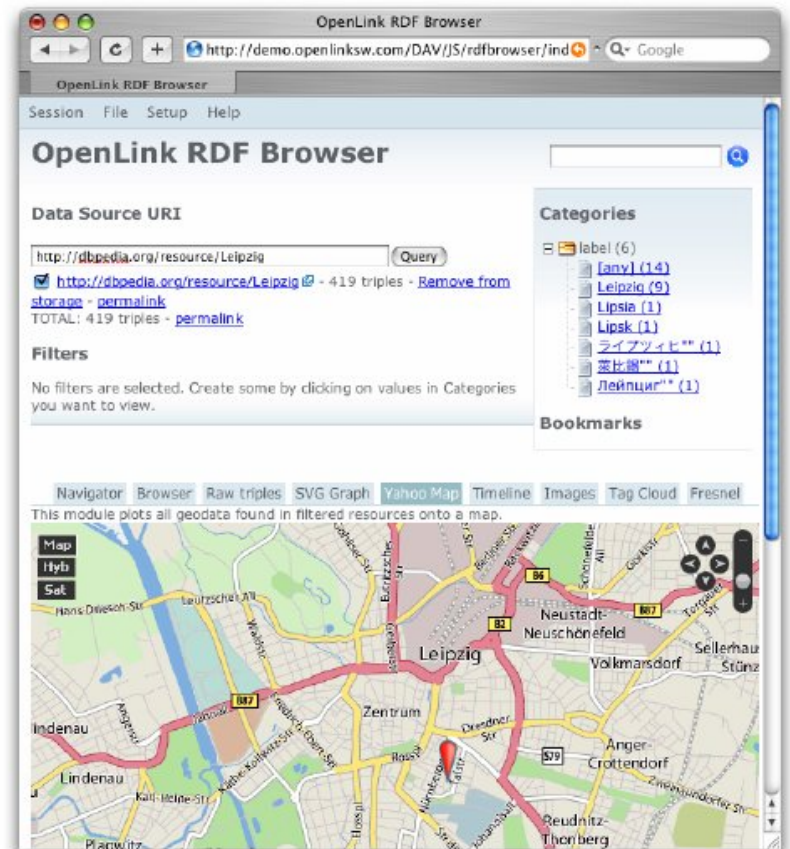
Leipzig at DBpedia.org
<http://dbpedia.org/resource/Leipzig>



Leipzig (; Sorbian/Lusatian: Lipsk from the Sorbian word for Tilia) is, with a population of over 506,000, the largest city in the federal state of Saxony, Germany. It is situated at the confluence of the Rivers Pleiße, White Elster and Parthe. *

| Property | Value |
|-------------------------|--|
| <code>p:abstract</code> | <ul style="list-style-type: none"> , Lipici o Lepicio (lit. rom. esp.) es una ciudad alemana, la mayoLeipzig on Saksassa Saksin osavaltiossa sijaitseva kaupunki. MLeipzig (; Sorbian/Lusatian: Lipsk from the Sorbian word for TiliaLeipzig ou Lipsia é uma cidade independente (kreisfreie Stadt) dLeipzig är den största staden i den tyska delstaten Sachsen ochLipsia o (in tedesco Leipzig, in sorabo-lusaziano Lipsk) é una cit |

Linked Data



OpenLink RDF Browser

Session File Setup Help

OpenLink RDF Browser

Data Source URI

<http://dbpedia.org/resource/Leipzig> @ - 419 triples - [Remove from storage](#) - [permalink](#)

TOTAL: 419 triples - [permalink](#)

Filters

No filters are selected. Create some by clicking on values in Categories you want to view.

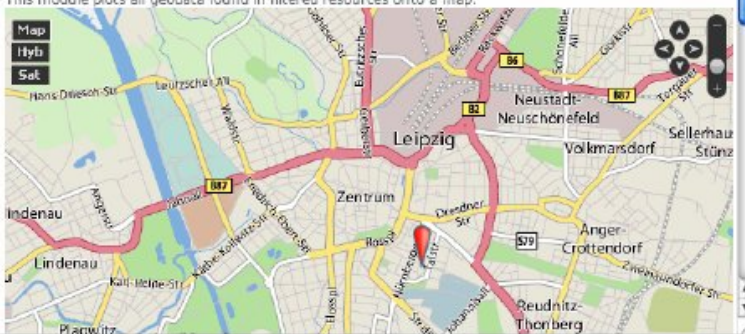
Categories

- label (6)
 - [any] (14)
 - Leipzig (9)
 - Lipsia (1)
 - Lipsk (1)
 - ライフサイト** (1)
 - 莱比锡** (1)
 - Лейпциг** (1)

Bookmarks

Navigator Browser Raw triples SVG Graph **Yahoo Map** Timeline Images Tag Cloud Fresnel

This module plots all geodata found in filtered resources onto a map.



Titel: [Bonobos - die zärtlichen Menschenaffen / Frans de Waal. Photos: Frans Lanting. Aus dem Amerikan. von Monika Niehaus-Osterloh](#)
Sonst. Personen: [Waal, Frans B. M. de *1948-*](#) ; [Lanting, Frans](#) ; [Niehaus-Osterloh, Monika \[Übers.\]](#)
Erschienen: [Basel \[u.a\]](#) : [Birkhäuser](#), 1998
Umfang: XV, 210 S. : zahlr. Ill., graph. Darst., Kt. ; 26 cm
Einheitssachtitel: [Bonobo - the forgotten ape <dt.>](#)
Anmerkung: Literaturverz. S. 197 - 202
ISBN: **3-7643-5826-2** (Pp.) : ca. DM 68.00, sfr 58.00, ca. S 497.00

Schlagwörter: [*Bonobo / Sozialverhalten / Bildband](#)
Sachgebiete: [42.66 ; Ethologie <Biologie>](#)
 SfB: BIO 954

1 von 1

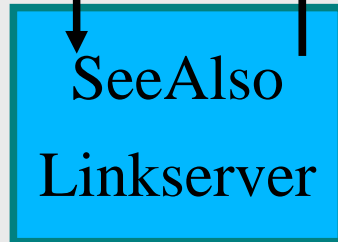
W In Wikipedia erwähnt unter: [Bonobo, Frans de Waal](#)

gehe zu

1 von 1

ISBN

Links



SeeAlso

ISBN+Artikel

aus

Datenbankdump



Bewertung

- Wikipedia (und andere Wikis) liefern eine Fülle gemeinsam von Nutzern generierter Metadaten.
- Inhalte sind eher semi-strukturiert, d.h. Nachbearbeitung ist nötig.
- Dafür sind die Inhalte unter einer freien Lizenz verfügbar.
- Metadaten in Wikipedia gehen von Inhalten statt von Formaten und Konzepten aus (Bottom-Up).