

Das Digitale Wörterbuch der deutschen Sprache – Bestandssicherung und Aktualisierung

Alexander, Geyken, Axel Herold,
Lothar Lemnitzer

Berlin-Brandenburgische Akademie der
Wissenschaften

4. Arbeitstreffen des Forschungsnetzwerks
„Internetlexikographie“

Trier, 22. November 2012

Einleitung

- Archivierung oder Versionsverwaltung?
- Ebenen des Projektes aus Datensicht
- Verfahren der Bestandssicherung

Archivierung oder Versionsverwaltung?

Archivierung:

- Langfristige Sicherung von Daten
- Technisch niederschwellige Speicherung
- Geringere Verfügbarkeit
- Wenige Nutzer / Zugriffe

Versionsverwaltung:

- Permanente Bereitstellung von Daten im jeweils aktuellen Zustand
- Transparente Versions-Metadaten
- Hohe Verfügbarkeit
- Viele Nutzer oder Zugriffe

Ebenen des DWDS

Ebene 1: das Digitale lexikalische System –
Webseite

Ebene 2:

- (Bestands-)Wörterbücher
- Korpora
- Statistiken
- Werkzeuge

Ebene 3: Redaktionssystem für die Neuerstellung
von Daten

Digitales lexikalisches System

retro.dwds.de (altes Paradigma, viele Nutzer, nicht
(mehr) gewartet oder weiterentwickelt)

www.dwds.de (neues Paradigma, alte Versionen)

www.dwds.de (neues Paradigma, aktuelle Version,
viele Nutzer, wird gewartet)

beta.dwds.de (neues Paradigma, wenige ausgewählte
Benutzer, wird weiterentwickelt)

Wörterbuch der deutschen Gegenwartssprache (eWDG)

- Verfügbar über Webseite
- archiviert (Daten und Metadaten: XML/TEI, Dokumentation in einem Wiki)
- laufende Fehlerkorrektur, Versionierung kumulativ

Etymologisches Wörterbuch des Deutschen (EtymWB)

- Verfügbar über Webseite
- archiviert (Daten und Metadaten: XML/TEI, Dokumentation in einem Wiki)
- Laufende Überarbeitung durch den Hauptautor, Wolfgang Pfeifer und
- Anreicherung der strukturellen Annotation (kumulative Versionierung)

Deutsches Wörterbuch von Jacob Grimm und Wilhelm Grimm, Erstbearbeitung

- Verfügbar über Webseite
- Gemeinsames SVN von Trier und Berlin
- strukturelle Bearbeitung, um die Tiefenstrukturierung und Auszeichnung einzelner Datenelemente zu verbessern

DWDS-Wörterbuch

- verfügbar über Webseite
- archiviert (Daten und Metadaten: XML/TEI, Dokumentation in einem Wiki)
- laufende Arbeiten aller Art – Ergänzungen, Löschungen, Korrekturen – Versionierung kumulativ nach Arbeitsschritten

Lexikalische Ressourcen

Openthesaurus („Fremd“anbieter):

- Verfügbar über Webseite
- Pflege / Archivierung beim Anbieter
- Regelmäßige Aktualisierung aus den Quellen

DLS und Ressourcen sind unabhängig voneinander

- in die aktuelle Version des DLS werden neue Versionen der Wörterbücher „eingespielt“, wenn dies sinnvoll erscheint;
- „einspielen“ = Übernahme der Daten aus den XML-Dokumenten in eine rel. Datenbank
- In jedem Panel wird die Versionsnummer der eingespielten Version angezeigt

- Kernkorpus 20 /21, Zeitungs- und Spezialkorpora
- Verfügbar über Webseite
 - Primärdaten und Metadaten sind archiviert
 - Linguistische Annotation wird dynamisch erzeugt
 - Stabil, Versionierung in sehr großen Zeitabständen

Wortprofil, Wortverlauf und „gute Beispiele“

- Verfügbar über Webseite
- Dynamisch generiert auf stabiler Korpusbasis (vorgenerierte Daten als Datenbanken)
- Stabil, Versionierung in sehr großen Zeitabständen

Überarbeitung / Neubearbeitung

Redaktioneller Arbeitsplatz, basierend auf oXygen, wobei die Möglichkeiten der jeweils neuesten Version dieser Software genutzt werden

Daten, die sich in Bearbeitung befinden, werden über ein Versionierungssystem verwaltet (SVN über Oxygen-Schnittstelle)

XML-Schema in einer Werkstattsprache, die nicht mit der XML/TEI-Version des DWDS-Wörterbuches identisch ist (Konversion ist ein Bestandteil der Integration beider (Teil)Ressourcen)

Archivierung statischer Daten, unter der Maßgabe, die Daten und Metadaten möglichst unabhängig von technischen Bedingungen und Gegebenheiten zu halten. Ausführliche Dokumentation

Versionsverwaltung dynamischer Daten. Die Versionierung ist für die Nutzer transparent. Dokumentation der Versionen

Ziel: bei einem heterogenen Informationssystem mit vielen Komponenten den Überblick zu behalten und über Änderungen Rechenschaft geben zu können