

**Datenbank basiertes Wörterbuch
Neugriechisch-Deutsch mit Web-Interface
(Arbeitstitel)**

Betreuer: Dr. Sosna

Bearbeiter: Christian Helmchen

05.11.2005

1. Anforderungen
2. Entwurf und Modellierung
3. Suchfunktionalität
4. Ausblick

1. Anforderungen

1. Aufgabenstellung
2. Allgemeine Anforderungen
3. Nutzerinterface
4. Unicode-Unterstützung
5. Analyse des Ist-Zustandes

- Erstellung eines Wörterbuches
Neugriechisch ↔ Deutsch
- Erstellung eines Tools zur Dateneingabe
- Integration administrativer Funktionen in die Web-Oberfläche
- möglichst verlustfreie Übernahme bestehender Daten aus früheren Versionen/Prototypen
- Analyse der Unicode-Unterstützung und Eingabemöglichkeiten (griechische Tastatur, grafisches Hilfsmittel o.Ä.)

- Plattform- und Browserunabhängigkeit
- Verwendung freier Software (Apache, MySQL, PHP und Java)
- hohe Suchgeschwindigkeit
- „4-Augen-Prinzip“ bei der Dateneingabe
- Fehlermanagement bei der Datenübernahme
- Fokus auf deutschsprachigen Nutzer
 - ➔ Eingabe auch ohne griechische Tastatur möglich
 - ➔ detaillierte Angaben zu griechischen Einträgen

- Benutzer
gewöhnlicher Browser
Java-Applet zur Eingabe auch auf Griechisch
- Tutor/Autor
Java-Applikation zur Dateneingabe/-bearbeitung
- Administrator
Webinterface für verschiedene Aufgaben, z.B.
Anpassungen der grammatikalischen Regeln
- Nachbearbeitung der übernommenen Daten entweder
über das Webinterface oder ins Eingabetool integriert

Java	volle Unterstützung
MySQL	volle Unterstützung ab MySQL 4.1.x
PHP	Unterstützung nur teilweise (Auslesen aus DB und Senden an Browser)
Browser	Unterstützung durch alle modernen Browser
Betriebs- system	volle Unterstützung in neueren Linuxdistrib. + Unterstützung in neueren Windowsversionen

- griechische Daten liegen in TeX-Notation vor
- leider unzureichende Dokumentation
- manuelle Analyse der Datenstruktur und der konkreten Daten, teilweise durch „Ausprobieren“
- Ergebnis
 - Konvertierung in Unicode nötig
 - Daten zum Teil fehlerhaft und inkonsistent
 - Vielzahl an Parametern bei der Konvertierung zu beachten
 - manuelle Nachbearbeitung unvermeidlich (Logging)

2. Entwurf und Modellierung

1. Aufbau eines Wörterbuches
2. Klassendiagramm
3. Datenkonvertierung

- Eintrag eines gedruckten Wörterbuches

Stichwort in
„Nennform“

Grammatik

Bedeutung

ελληνικ|ός, ή, ό [elinik'os] griechisch

Phonetik

Orthographie

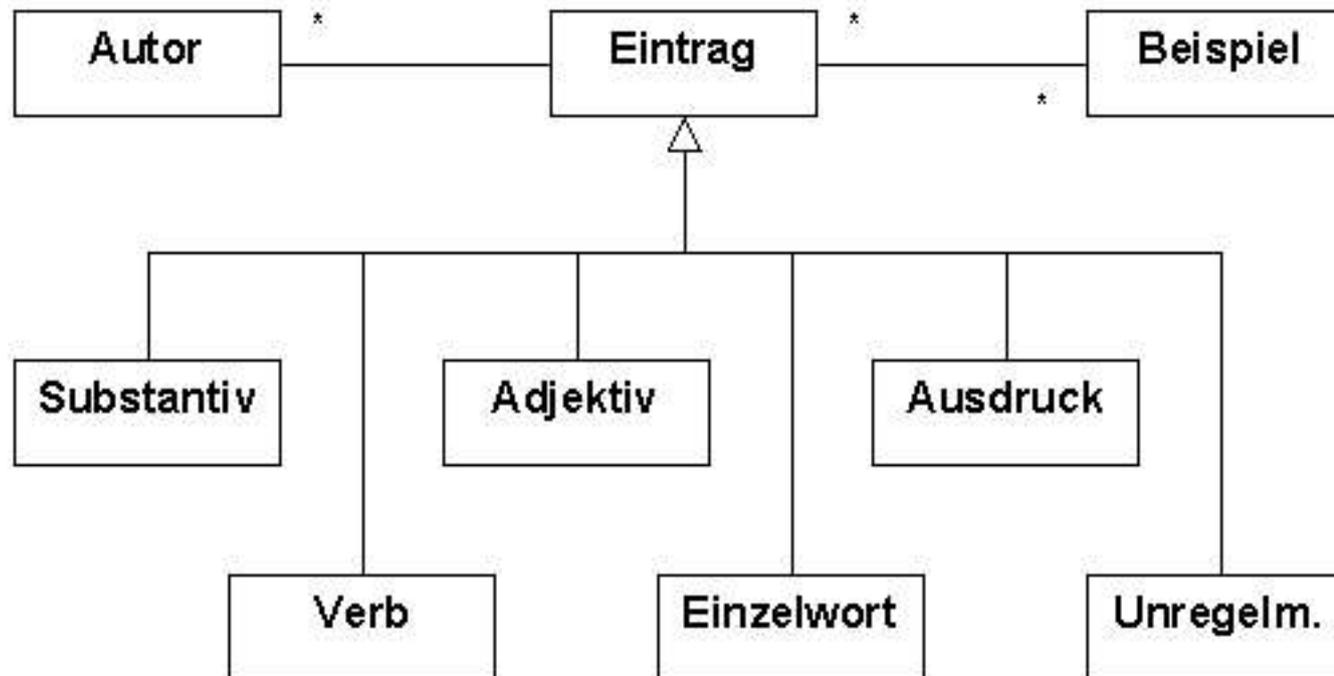
Diasystematische
Angaben

Varianten

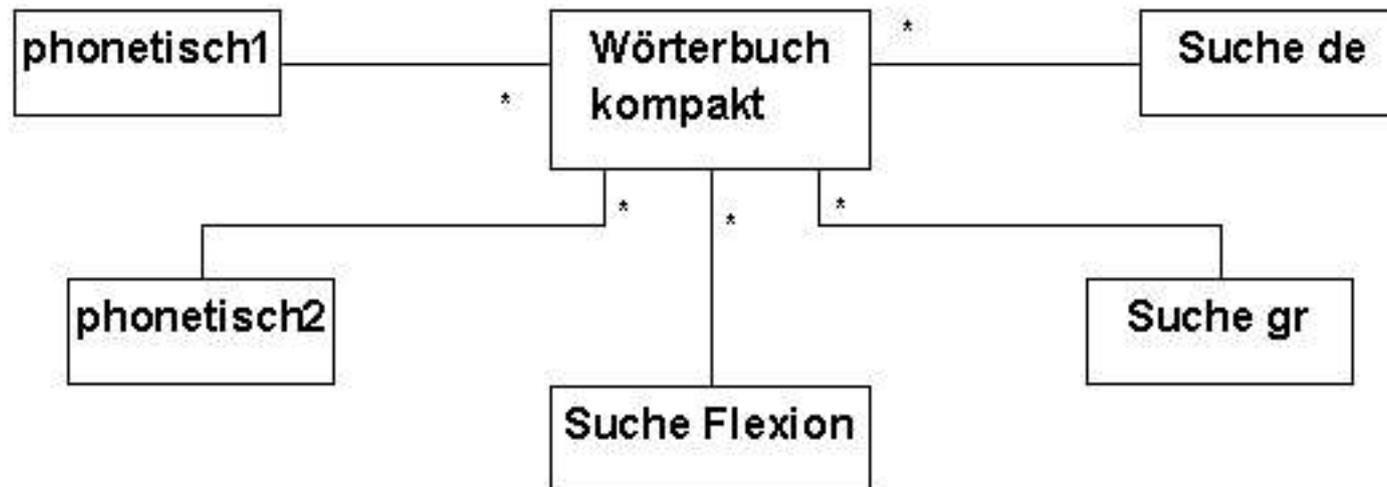
Belege (Zitate)
ggf. Verweise

- Eintrag des elektronischen Wörterbuches
 - ➔ Dynamik und Interaktionsmöglichkeit
 - (a) Kompakte Darstellung
Generierung eines Data Warehouse/Mart sowie Lookup Tables aus den vorhandenen Daten
 - (b) Detaillierte Darstellung
greift auf die Originaldaten zurück
 - (c) Verweise
...auf Beispiele und weitere Einträge
 - (d) geeignete Suchmechanismen
...für Redewendungen, Wortfamilien und Synonyme

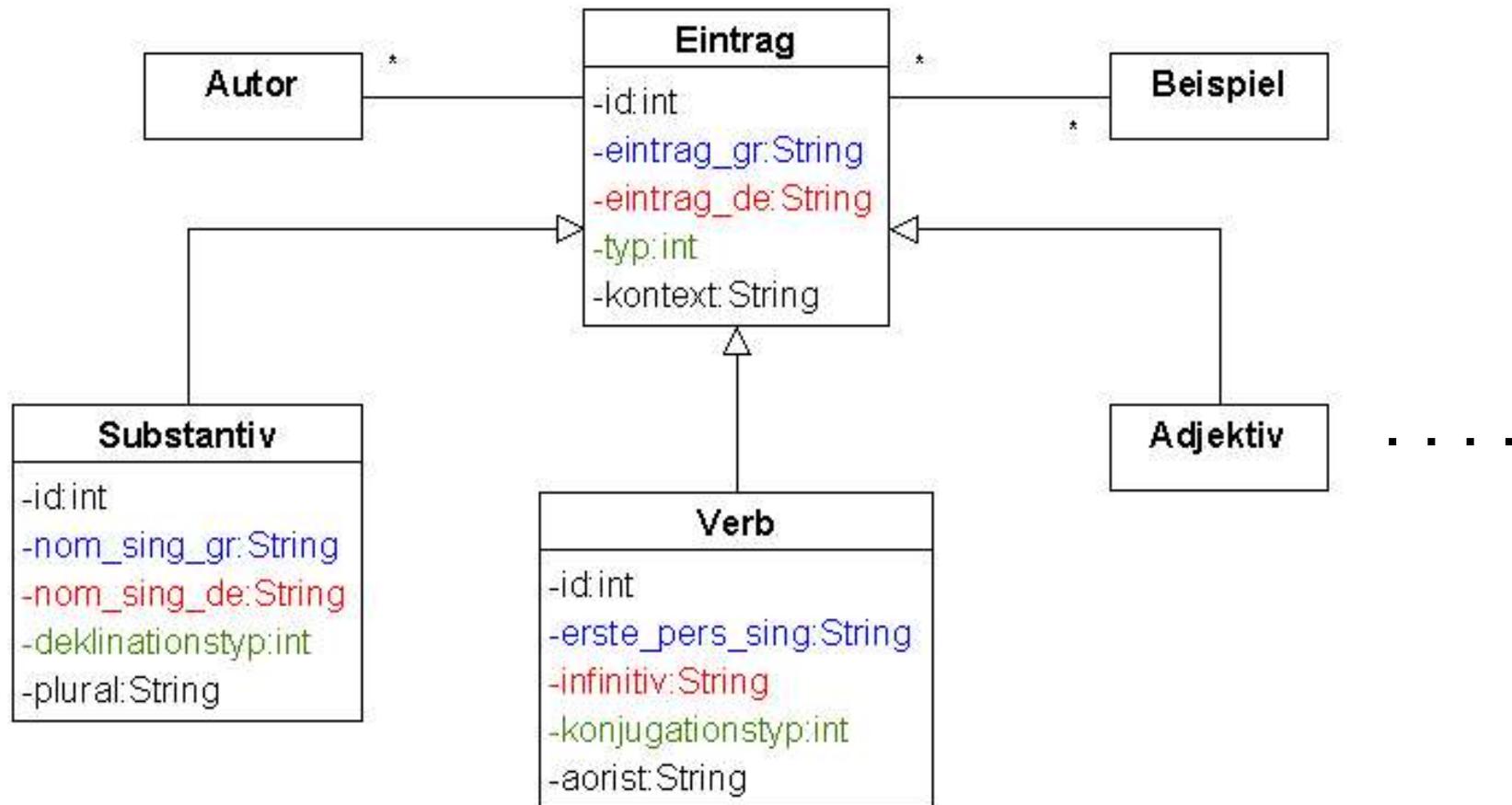
- Daten des Wörterbuches



- Suchfunktion:
Data Warehouse/Mart und Lookup Tables



- Generalisierung bei den Wortarten



- ein Eintrag des Data Warehouse/Mart

dictionary	
🔑	entry_id: INT(10) (FK)
🔹	greek: VARCHAR(255)
🔹	importantNotes: VARCHAR(255)
🔹	german: VARCHAR(255)
🔹	context: VARCHAR(255)
🔹	greek_type: ENUM('noun', 'verb', 'adjective', ...)

eindeutige ID

griechisch

Kontext

deutsch

Wichtige Hinweise

Eintragstyp

- Substantiv: Genus
- Adjektiv: Deklination
- Einzelwort: Wortart
- ...

- Übertragen der Daten in das neue Schema
- neue Felder können automatisch gefüllt werden (durch Parsen vorhandener gramm. Daten)
- technische Angaben werden neu generiert
- Probleme werden mitgeloggt

Konvertierungsoptionen

allgemein

Autor

Substantive

Numerus

Verben

Diathese (aktiv - passiv)

Adjektive

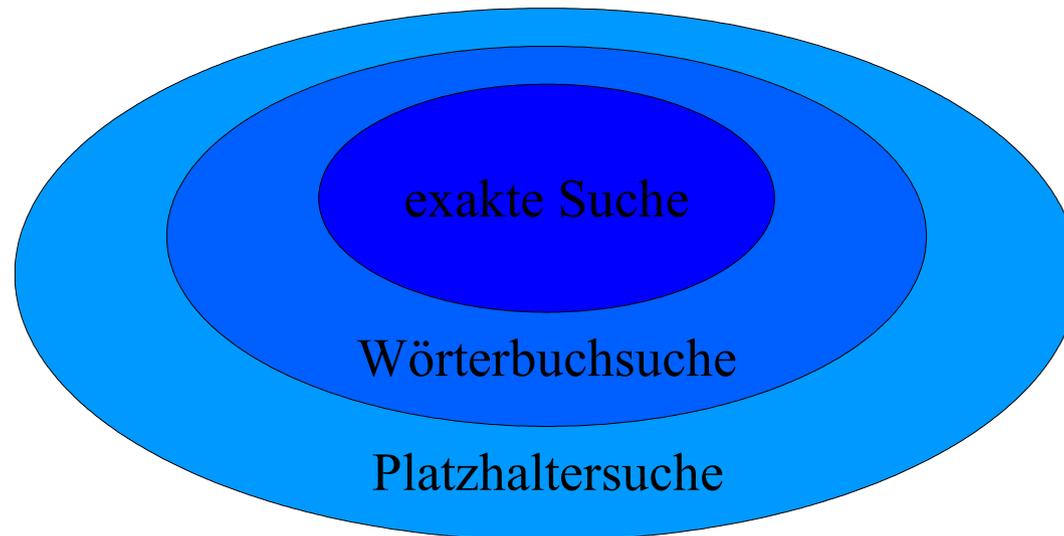
Genus Numerus

3. Suchfunktionalität

1. Suchvarianten
2. Phonetische Suche
3. Trefferrelevanz

- Exakte Wortsuche
 - ➔ exakte Treffer
 - ➔ auch „Stopworte“
- Wörterbuchsuche
 - ➔ alle Einträge mit demselben Beginn
 - ➔ Betonung wird ignoriert
- Platzhaltersuche
 - ➔ alle Einträge, die den Suchbegriff enthalten
 - ➔ Betonung wird ebenfalls ignoriert

- Ergebnismengen



- Wortfamilien/zusammengesetzte Worte
 - ➔ müssen vom Autor manuell eingegeben werden

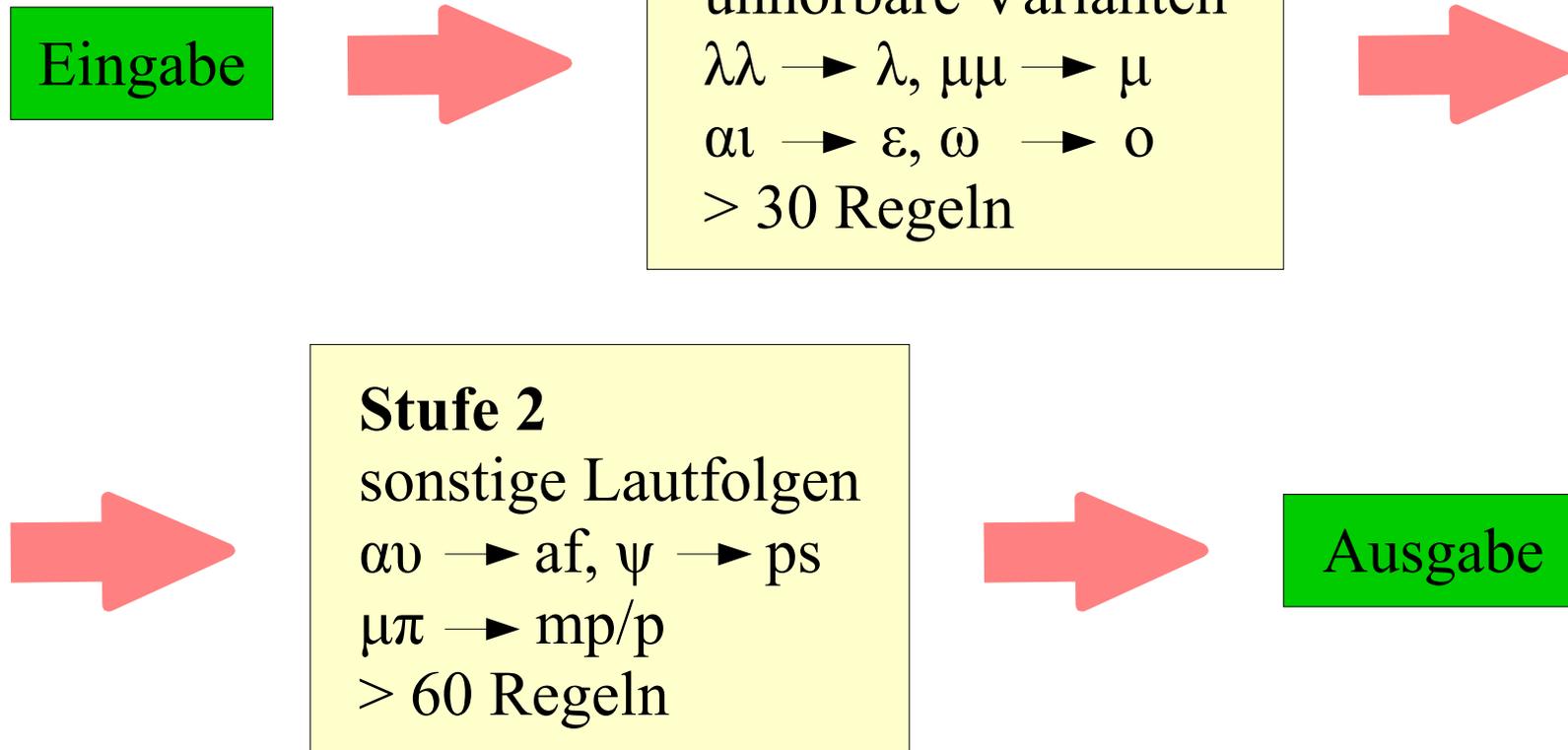
- Suche ohne Flexionsendungen
 - ➔ die Flexionsendungen aller relevanten Worte werden bei der Dateneingabe automatisch abgeschnitten und die resultierenden Wortteile gespeichert
 - ➔ der Autor kann jedoch Änderungen vornehmen
 - ➔ der Suchbegriff wird ebenso behandelt und mit dem Ergebnis (meist mehrere Wortteile) wird die Suche durchgeführt

καλός	→	ός, ή, ό	→	καλ	Dateneingabe
καλή	→	ή	→	καλ	Suche

- **Phonetische Suche**
 - ➔ Wort anhand der Aussprache erkennen
 - ➔ 2 Stufen
 - ➔ sowohl die Worte bei der Dateneingabe als auch der Suchbegriff werden nach einer Reihe von Regeln automatisch umgeformt
- **Phonetische Suche kombiniert mit der Suche ohne Flexionsendungen**
 - ➔ Kann erforderlich sein, da die Aussprache eines Wortes von der Flexionsendung abhängen kann

- Soundex-Algorithmus
 - ➔ keine Vokale, Längenbegrenzung
 - ➔ funktioniert gut im Englischen
 - ➔ leider nicht fürs Griechische adaptierbar, da kein Zusammenwirken mehrerer Buchstaben abgebildet werden kann (vgl. dazu im Dt.: ich - ach, ai - ei ...)
- einige Besonderheiten im Griechischen
 - ➔ Phonem **i**: ι, η, υ, ει, οι
 - ➔ μπ, ντ, γκ (λά**μ**πα → lam**μ**pa/lap**μ**a)
 - ➔ Lautverschiebungen (πτ → φτ)

- mehrstufiges Verfahren



- Exakte Treffer haben höchste Relevanz
- Ranking bei Wortfamilien
 - ➔ „Wörter**buch**“, „**Buch**“, „Wort“
 - ➔ „**Buch**handel“, „Handel“, „**Buch**“
- Wortlänge bei Platzhaltersuche
- alphabetische Reihenfolge bei Wörterbuchsuche

- Phonetische Suche (insbesondere Stufe 2) in Kombination mit Suche ohne Flexionsendungen verursacht noch Probleme
- Möglichkeit für Nutzer Bemerkungen/Fragen zu bestimmten Einträgen zu hinterlassen
- weitere Projekte denkbar
 - ➔ Wortschatzlisten
 - ➔ Vokabeltrainer
 - ➔ andere Sprachen ?

ENDE