

# Računalno jezikoslovlje

Damir Čavar

Sveučilište u Zadru, Odjel za lingvistiku u.o.



# Plan izlaganja

- Jezične tehnologije
  - Govor (*speech*)
  - Jezik (*language*)
- Uloga jezikoslovlja



# Govor

- Prepoznavanje govora (*speech recognition*)
- Sinteza govora (*speech synthesis*)



# Prepoznavanje govora

- Tehnološko stanje:
  - Standard: Prepoznavaći govora u svim sustavima: *Windows XP / Vista, Mac OS X, Linux*
  - Profesionalni sustavi za telefonske servere, npr. Nuance (kupili *SpeechWorks*), *Microsoft*, itd.
  - Otvoreni i slobodni alati za modeliranje i generiranje prepoznavaća (npr. HTK)



# Prepoznavanje govora

- Tehnologija:
  - Skriveni Markovljevi modeli
  - Neuronske mreže
- Sustavi:
  - Operativni sustavi
  - npr. IBM ViaVoice, Dragon NaturallySpeaking



# U Hrvatskoj

- npr. FER:
  - Laboratorij za sustave i signale
    - doc. dr.sc. Davor Petrinović



# Govor kao sučelje

- Skoro svatko čuje i zna govoriti
  - ukidanje tehnološke barijere
- Interfejs u situacijama bez mogućnosti manualne instrukcije ili manipulacije strojevima



# Primjena

- Naredbe
  - Strojevi
  - Elektronski uređaji
  - Automobili i zrakoplovi
  - Telefonske aplikacije i informacijski sistemi
- Diktiranje i prevođenje
- Studije i usvajanje stranih jezika
- Sustavi za invalide



# Sinteza govora

- Sinteze:
  - Difon modeli, baze sa frazama, skriveni Markovljevi modeli
- Primjeri:
  - OS X, AT&T Natural Voices
  - Virtualni govornici ili izlagači:
    - Charamel.de



# Jezik

- ima središnje mjesto: komunikacija je jezik
  - za gospodarstvo
  - za politiku i nacionalnu sigurnost
  - znanstveni rad
  - svakodnevni život



# Jezik

- Tekst kao:
  - komunikacijsko sredstvo
  - arhivirana informacija i znanje
- Problemi:
  - kvantiteta
  - kvaliteta (informacija vs. znanje)



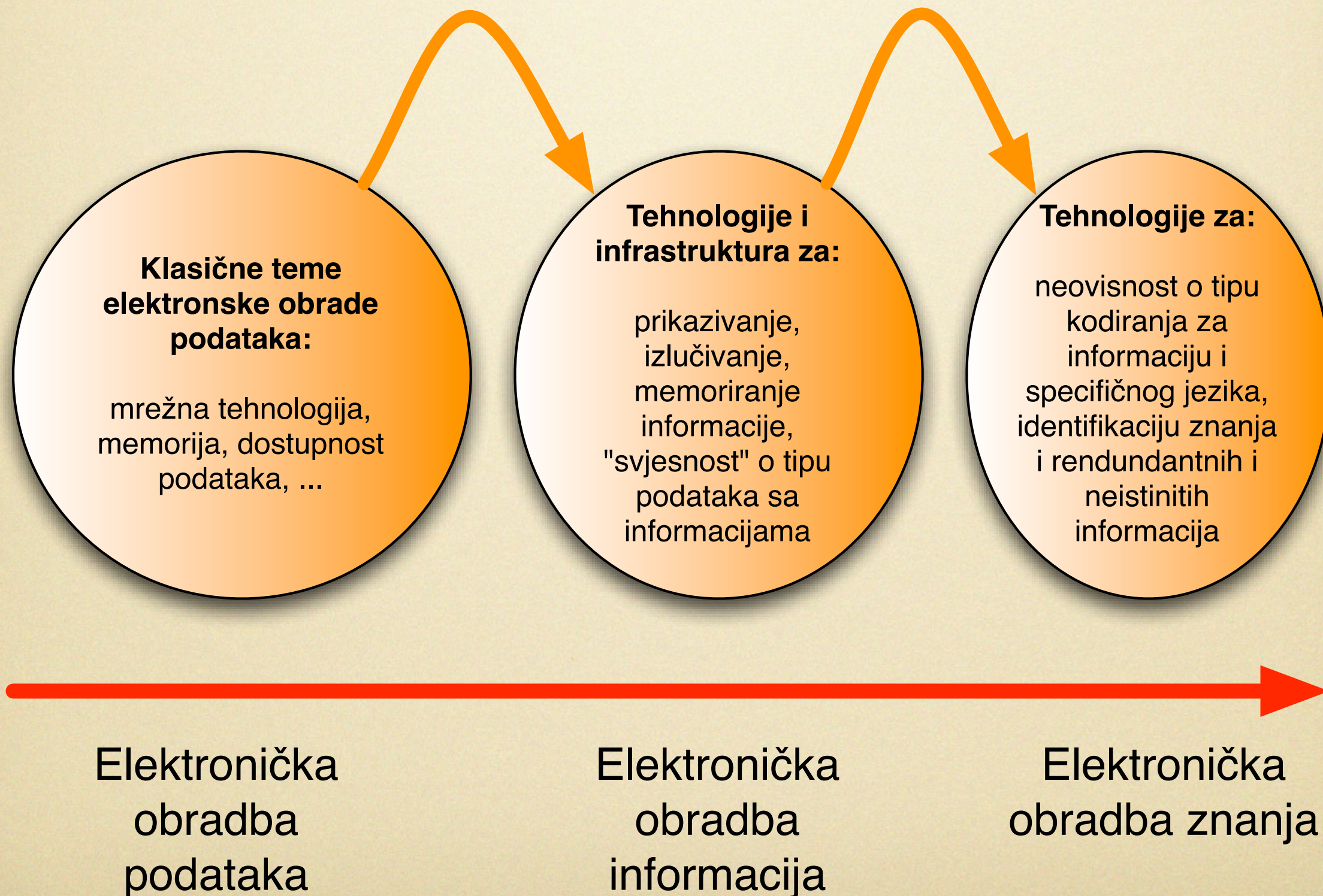
# Jezik i komunikacija

- Problemi:
  - Višejezičnost
  - Specifične domene jezika i teksta
  - Količina i kvaliteta
  - Dinamika jezika
  - itd.



**poticaj:** kvantiteta,  
kvaliteta, učinkovito  
pronalaženje,  
transparentnost

**poticaj:** zatrpanost  
informacijama, redundancije,  
proturječnost, pouzdanost,  
sažimanje





# Jezik i komunikacija

- Rješenje
  - Računalno jezikoslovlje:
    - rječnici, tezaursi, snimke, ...
    - modeli jezika na svim razinama
    - alati za analizu jezika (a i vizualnih i akustičkih informacija)
    - za SemanticWeb itd.



# Računalno jezikoslovlje

- Zadatci:
  - Stvaranje baze jezičnih podataka
    - rječnike, korpus, gramatike, itd.
  - Stvaranje osnovne tehnologije
    - prepoznavanje jezika, obilježavanje riječi, sintaktička analiza rečenica, semantičko obilježje



# Osnove

- Standard:
  - Prepoznavanje jezika, pravopis (*Spell-checker*), obilježje i morfologija (*tagger*), gramatika (*grammar checker*)
- Nema u standardnim sustavima:
  - Prepoznavanje imena (*named entity recognition*), semantička analiza, klasifikacija teksta, itd.



# Hrvatski

- Ima:
  - rječnike, korpus, donekle alate za strojno prevođenje
- Nema:
  - formalne gramatike za jezičnu tehnologiju, osnovne otvorene i dostupne tehnologije



# Hrvatski

- Što radimo (u suradnji s Institutom za hrvatski jezik i jezikoslovlje):
  - korpus (jedno- i višejezični)
  - rječnike (jedno- i višejezične)
  - alate
- gdje: Hrvatska jezična riznica



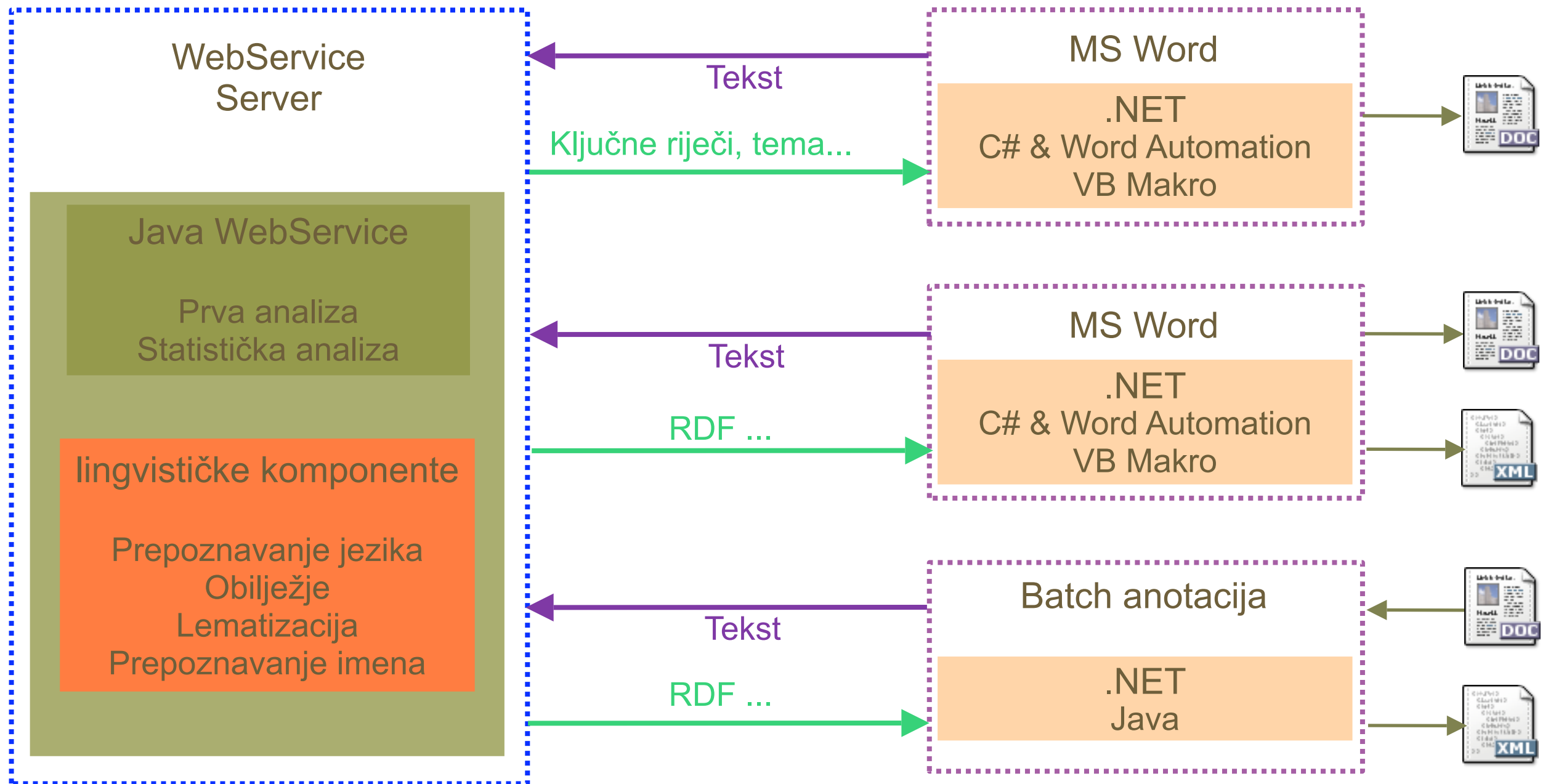
# Primjeri

- Prevođenje dokumenata
- Pretraživanje i ekstrakcija podataka
- Meta-obilježje
- Klasifikacija dokumenata



# Pilot

SOAP





# K-Net

**Dresdner Bank Knowledge Net - Netscape**

File Edit View Go Window Help

Back Forward Reload Home Search Netscape Print Security Stop

## Knowledge Net der Dresdner Bank

```
graph LR; DresdnerBank[Dresdner Bank] --- Beraterbank[Beraterbank]; DresdnerBank --- Allianz[Allianz]; DresdnerBank --- DeutscheBank[Deutsche Bank]; Beraterbank --- Kompetenz[Kompetenz]; Beraterbank --- Selbstverstandnis[Selbstverständnis]; Allianz --- AGAktie[AG-Aktie]; Allianz --- Uebernahmeangebot[Übernahmeangebot]; DeutscheBank --- Dreyers[Dreyer's]; DeutscheBank --- MCIWorldCom[MCI WorldCom]; Allianz --- Gewinn[Gewinn]; Allianz --- USDollar[US-Dollar];
```

Maximale Grade pro Stufe  
auf Stufe 1   
auf Stufe 2

Darstellung der Beziehungen  
☐ numerisch  
☒ visuell

Suchbegriff:    
 Suche mit:

Version vom 18.04.2001 - 22:00 Uhr

[Hilfe](#)



# Istraživanja

- Kvalitativna i kvantitativna analiza jezika:
  - Usvajanje jezika
  - Evolucija jezika
  - Dijalektološke studije
  - Baze za psihološka i kognitivna istraživanja



# Istraživanja

- Interdisciplinarno povezano:
  - Jezik
  - Računarstvo
  - Psihologija i kognitivne znanosti