

Web-based Question Answering

Gosse Bouma

Information Science
University of Groningen

Natuurlijke Taalverwerking II, 2006/2007

- ▶ Gebruik van het Web voor Question Answering
- ▶ Voor- en nadelen
- ▶ Gebruik van on-line databases
- ▶ Gebruik van web-search
- ▶ Demo's



Web-based Question Answering

- ▶ Vrijwel alle informatie die we zoeken staat on-line
 - ▶ Wikipedia,
 - ▶ Citeseer, Google Scholar, Wolfram MathWorld, ACL Antology, ...
 - ▶ KiesBeter.nl, Merck Manual, ICD 9, ...
 - ▶
- ▶ Question Answering moet op het web naar antwoorden kunnen zoeken
 - ▶ In combinatie met web-search
 - ▶ Door te zoeken in on-line databases



Data Intensive Question Answering

- ▶ Web bevat veel **redundante** informatie
- ▶ Kans dat je het antwoord met eenvoudige technieken kunt vinden is veel groter
 - ▶ Who killed Abraham Lincoln?
 - ▶ **John Wilkes Booth** killed Abraham Lincoln
 - ▶ **John Wilkes Booth** is perhaps America's most infamous assassin. He is best known for having fired the bullet that ended Abraham Lincoln's life.
- ▶ Kennis van synoniemen, syntactische variatie, coreferentie, etc. niet nodig, als het antwoord ook direct te vinden is.



Data Intensive Question Answering

- ▶ Web bevat VEEL redundante informatie
- ▶ Kans dat je het goede antwoord vaak vindt is veel groter
 - ▶ Who killed Abraham Lincoln
 - ▶ John Wilkes Booth killed Abraham Lincoln
 - ▶ John thinks Andrew Jackson killed Abraham Lincoln
- ▶ Het goede antwoord is meestal het meest frequente antwoord



Eenvoudig Web-based QA systeem

- ▶ What is relative humidity
- ▶ Answer Patterns:
 - ▶ X is relative humidity
 - ▶ relative is humidity X
 - ▶ relative humidity is X
- ▶ Zoek naar meest frequente woord (of N-gram) op positie van X



Eenvoudig Web-based QA systeem

- ▶ Gegeven de vraag
- ▶ Verzin een patroon voor het antwoord
- ▶ Stuur key words naar Google
- ▶ Zoek naar het patroon in de snippets van eerste N resultaten



Eenvoudig Web-based QA systeem

- ▶ Wie is de moordenaar van John Lennon
- ▶ is de moordenaar van John Lennon
- ▶ Google:
 - ▶ John Lennon's moordenaar Mark Chapman heeft onthuld
 - ▶ Moordenaar John Lennon blijft vast Mark Chapman denkt dat hij van John vrijgelaten mag worden.
 - ▶ De moordenaar van de Britse popartiest en componist John Lennon, Mark Chapman, komt niet vervroegd vrij.
 - ▶ De moordenaar van John Lennon moet nog zeker twee jaar in de gevangenis blijven. De 51-jarige David Chapman komt niet voorwaardelijke vrij
 - ▶ Over de laatste vijf jaren van ex-Beatle John Lennon, die in 1980 werd vermoord, ... Sommige geleerden beweren dat Lennons moordenaar Mark Chapman geen ...



- ▶ Fill in the blanks
 - ▶ * is de moordenaar van John Lennon
 - ▶ Antwoord in **bold** in de snippets
- ▶ Define
 - ▶ define dementie
 - ▶ [Internetdefinities voor dementie](#): Dementie (L. demens), ook dementia of dementerig, is een progressieve afname van de cognitieve functies....
(nl.wikipedia.org/wiki/Dementie)

- ▶ Veel vragen gaan over plaatsen, acteurs en films, definities,
- ▶ Antwoord op zulke vragen vind je op gespecialiseerde web-sites
 - ▶ Geonames, Internet Movie Database, Wikipedia, ...

geonames.org

- ▶ Waar ligt Bedum?
- ▶ Hoeveel inwoners heeft Bedum?
- ▶ In welk(e) gemeente/provincie/land ligt Bedum?

wikipedia

```
<geoname>
  <name>Bedum</name>
  <lat>53.3</lat>
  <lng>6.6</lng>
  <geonameId>2759356</geonameId>
  <countryCode>NL</countryCode>
  <countryName>Netherlands</countryName>
  <fcl>P</fcl>
  <fcode>PPL</fcode>
</geoname>
```

Bedum

From Wikipedia, the free encyclopedia

Your continued donations keep Wikipedia running!

Bedum (ⓘ pronunciation (help·info), Gronings: *Beem*) is a municipality and a town in the northeastern Netherlands. Populated by around 10,000 (2005) inhabitants, Bedum is one of the larger of Groningen's several satellite towns. Bedum is the site of three supermarkets, several pubs, and its infamous leaning church tower, dubbed "the leaning tower of Bedum". Footballer Arjen Robben was born in Bedum (January 23, 1984).

[edit]

The Liberation of Bedum

Bedum was liberated on Tuesday the 17th 1945 by the 8th Reconnaissance Regiment (VIII RECCE), 2nd Canadian Infantry Division.

Population centres

- Bedum
- Noordwolde
- Onderdendam
- Zuidwolde
- Ellerhuizen
- Westerdijkshorn

[edit]

External link

[edit]

Coordinates: 53°18′N, 6°36′E

Bedum

Country	Netherlands
Province	Groningen
Area (2006)	
 - Municipality	44.96 km² (17.4 sq mi)
 - Land	44.58 km² (17.2 sq mi)
 - Water	0.38 km² (0.1 sq mi)
Population (1 January 2007)	
 - Municipality	10,617
 - Density	238/km² (616.4/sq mi)
Source: CBS, Statline	
Time zone	CET (UTC+1)
 - Summer (DST)	CEST (UTC+2)