

# www.PoliDocs.nl

In de Handelingen browsen, zoeken en vinden.

Maarten Marx

Universiteit van Amsterdam

December 2008

## Wie zijn we?

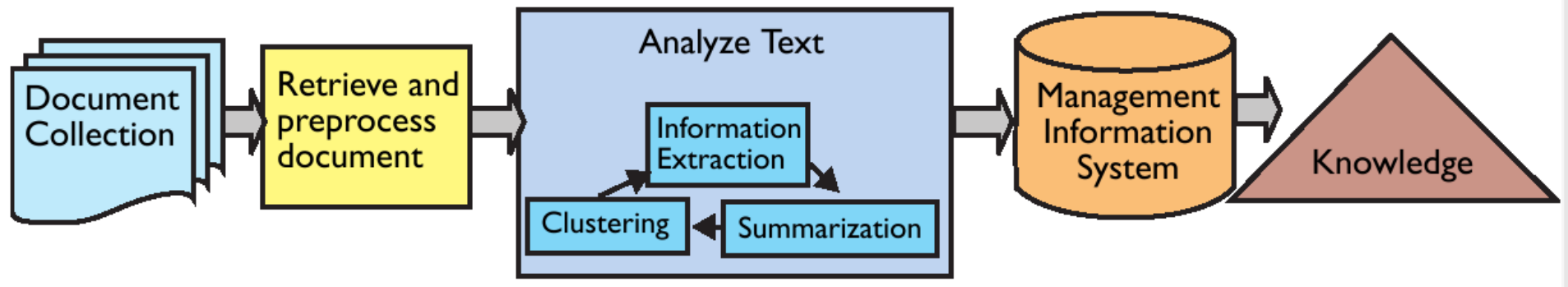


Prof Dr Maarten de Rijke.  
Leerstoel Intelligente Informatie  
Ontsluiting, Informatica, UvA.



Dr Maarten Marx  
UD, Informatica, UvA.

# Wat doen we?



# Intelligente Informatie Ontsluiting

# Sleutelwoorden

- Informatie vinden
- Semantische annotatie
- Web 2.0 en Semantic Web
- MultiMedia
- XML

Handelingen zien als Web 2.0 documenten

## Heel veel data

- Handelingen der Staten Generaal vanaf 1974 nu al digitaal beschikbaar.
- ★ meer dan 50.000 Kamervragen
  - ★ meer dan 1 miljoen uitspraken in debatten
  - ★ meer dan 200 miljoen woorden.
- Groeit dagelijks met tientallen documenten.

⇒ spannende uitdagingen

# Uitdaging 1

- **Verschillende schrijfwijzen** terugbrengen naar dezelfde persoon

Aadsted-Madsen

Aasted Madsen

Aasted-Madsen

Aasted Madsen-Stiphout

Aasted Madsen - van Stiphout

Aasted-Madsen van Stiphout

Aasted-Madsen-van Stiphout

Aasted Madsen-Van Stiphout



## Uitdaging 2

- Verbeteren van OCR-fouten in SGD data

FvdA FVdA PcdA pdA  
PdA PF P>dA Pvc)A  
PvciA PvciA PvclA  
pvdA Pvd.A PvdA  
PvDA PviiA PvriA  
PvrJA PvtiA PWD  
PydA

## Wetenschappelijk Doel

Nieuwste zoek- en web-technieken  
toepassen op de parlementaire data.

Niet zomaar, maar ...



# Onze uitdaging

- schaalbaar
- met beperkte middelen
- volledig
- betrouwbaar
- bruikbaar voor gewone man én professional

Leren door doen.

# Voorbeeld

<http://www.polidocs.nl>

- knettergek
- Geroffel op bankjes
- Joint Strike Fighter

## Toepassing: Berichten ontzenuwen door kennis

**Volkskrant** schreef op 30 Oktober: **Bron**

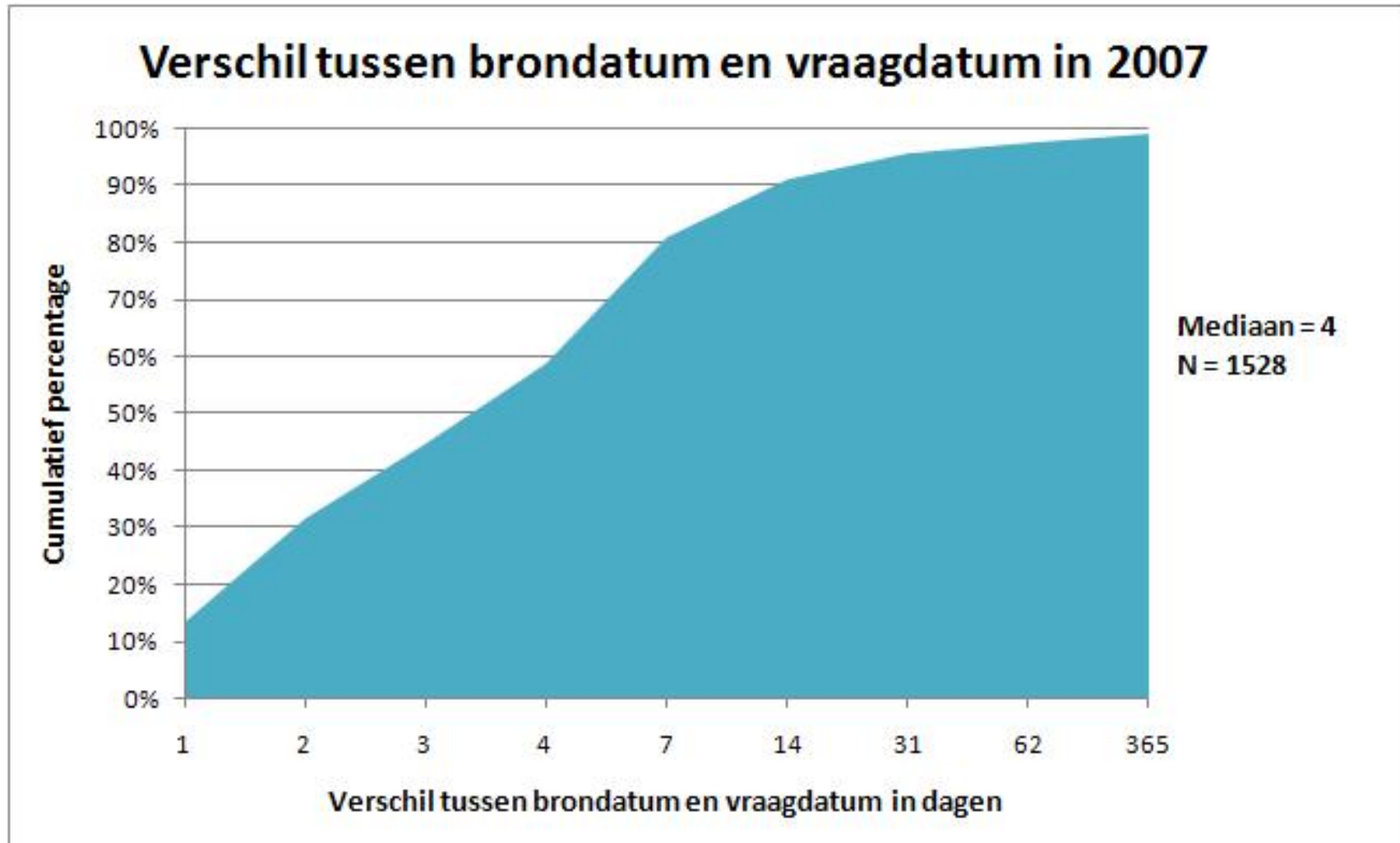
Is de Tweede Kamer misleid?

Mogelijk. De Kamer heeft sinds 2002 duizenden vragen gesteld over dit dossier.

De werkelijkheid is toch echt anders.

**Hijgerigheid Gevoel:** Kamervragen worden in meerderheid binnen een dag na een nieuwsfeit gesteld.

# Actualiteit van de Kamervragen



## Niks met de hand

- Bijna 200 jaar aan data,
- die dagelijks groeit.

Dat kán niet met de hand.

- Alles wat ik heb laten zien is automatisch gedaan ⇒
  - kosten blijven beheersbaar,
  - mensen kunnen zich storten op analyse en duiding.

## Conclusies

- Semantische verrijking van Handelingen
- is haalbaar en schaalbaar,
- biedt ongekende mogelijkheden,
- is wetenschappelijk vooruitstrevend,
- integreert diverse bronnen,
- en
- verbindt parlement met publiek.