

What you say is who you are.  
How open government data facilitates profiling  
politicians.

Arjan Nusselder and Maarten Marx  
University of Amsterdam

April 23, 2010

# Overview

- Parliamentary proceedings:
  - How they are now (human readable, not for machines)
  - How we'd like them to be (open, XML, mashupable)
- Usecase: people search
  - Politicians as documents
  - Evaluation / Committees
  - European elections example

# Parliamentary Proceedings

- In most countries published as PDF
- Proceedings contain lots of implicit structure hidden inside the raw data
- Example: Rich implicit data model

# Parliamentary Proceedings

## Rouvoet

hard bezig en ik verwijs dan ook naar het bestuursakkoord met de gemeenten om te bezien of wij een nieuwe balans kunnen vinden in hetgeen lokaal kan gebeuren – preventief, licht en dicht bij huis – en hetgeen echt op provinciaal niveau moet gebeuren, bijvoorbeeld op het gebied van ambulante zorg. Dat geldt ook voor opvoedingsondersteuning. Er gebeurt al heel veel aan opvoedingsondersteuning in het lokale domein. Daar is geen indicatie voor nodig.

De beraadslaging wordt gesloten.

De vergadering wordt enkele ogenblikken geschorst.

## Voorzitter: Verbeet

De **voorzitter**: Ik deel u mede dat het debat dat tijdens de regeling van werkzaamheden is aangekondigd door mevrouw Kant, in ieder geval niet vandaag zal plaatsvinden.

---

Aan de orde is het **debat over het kabinetsstandpunt ten aanzien van het rapport "Dynamiek in islamitisch activisme" van de Wetenschappelijke Raad voor het Regeringsbeleid (WRR) (30800-VI, nr. 115).**



De heer **Wilders** (PVV): Mevrouw de voorzitter. Om te beginnen mijn oprechte dank aan u persoonlijk omdat u op mijn verjaardag vandaag een debat over de islam heeft gepland. Een mooier cadeau had ik mij niet kunnen wensen!

Ongeveer 1400 jaar geleden is ons de oorlog verklaard door een ideologie van haat en geweld die toen ontstond en werd verkondigd door een barbaar die zichzelf profet Mohammed noemde. Ik heb het over de islam.

Laat ik beginnen met het fundament van de islam: de

De heer **Dijsselbloem** (PvdA): Mag ik de heer Wilders vragen waarom hij teksten die 1400 jaar oud zijn als volstreckte waarheid en absoluut en voor eeuwig beschouwt? Sluit hij zich daarmee aan bij Osama bin Laden en waarom doet hij dat?

De heer **Wilders** (PVV): Ik doe dat niet, maar heel veel mensen doen dit wel en niet alleen de Osama's bin Laden van deze wereld. Ik weet niet waar de heer Dijsselbloem de afgelopen tien of twintig jaar is geweest, maar als hij om zich heen kijkt in Nederland en buiten Nederland dan zal hij zien dat of het nu gaat over Madrid, Londen, de Filipijnen, Bali of de moord in Amsterdam, er vandaag de dag nog steeds moslims zijn die de teksten van de Koran letterlijk nemen en dat ook moeten doen en daardoor de meest verschrikkelijke dingen doen.

De heer **Dijsselbloem** (PvdA): De heer Wilders hoeft mij niet te vertellen dat er extremisten en idioten zijn die inderdaad met die teksten in de hand tot vreselijke extremiteiten overgaan. De vraag is waarom hij hen gelijk geeft. De vraag is waarom hij hen gelijk geeft in hun interpretatie van die teksten, in de wijze waarop zij menen met die teksten te moeten omgaan. De vraag is waarom hij hen iedere keer weer gelijk geeft.

De heer **Wilders** (PVV): Omdat de Koran een extremistisch boek is en omdat het een extremistisch boek is, moet het worden verboden. De heer Dijsselbloem heeft gelijk: dit betekent natuurlijk niet dat alle moslims extremistisch zijn. Ik maak ook een onderscheid tussen het geloof en de mensen. Ook de Partij voor de Vrijheid heeft geen probleem met moslims die zich houden aan de wet en aan onze Nederlandse rechtsstaat, maar helaas is dit niet het hele verhaal, want de islam, de Koran, de hadith deugen niet en zijn intrinsiek gewelddadig.

De heer **Dijsselbloem** (PvdA): Ik vind het een geweldige

# Hierarchical structure of Proceedings

## Meeting (1 Day)

- Topic
  - Stage direction
  - Scene (Block)
    - Stage direction
    - Speech
      - .....Paragraph

# Asking queries becomes very easy

Focused retrieval:

`http://mashup1.science.uva.nl:8080/sandbox/sandbox.xml`



## PoliticalMashup

Examples:

```
(: about knowledge and from Holland :)  
collection('/db/euparliament/data/EN')//speech[contains(., 'knowledge')][contains(./stage-direction, 'NL')]
```

Send

Clear

Check

Found 351 in 0 seconds.

Showing items 1 to 20

```
1 <speech docno="en.20021010.4.4-094" speaker="Meijer (GUE/NGL )," MPid="4508" VideoUrl=""  
  <p docno="en.20021010.4.4-094.1">  
    <stage-direction>Meijer (GUE/NGL ),</stage-direction>  
    <stage-direction>In writing.</stage-direction>  
    <stage-direction/>  
  -  
  <stage-direction>(NL)</stage-direction>
```

This non-legislative report is brimming with fine things and statements of the obvious. Of course, more women must be able to find work in knowledge-intensive sectors. It is good that a 'Statute for European Employees' should be drafted, a European system of labour legislation and social security for mobile employees. This fits in with the needs of workers who work without collective labour agreements. However, the gist of the proposal is about adapting people to suit the work, rather than adapting the work to suit the workers. It is good that the Commission is function more effectively by better tailoring education to work. It intends to make it easier for people to move thousands of kilometres to regions with a different language and culture. If people really want to move, it is self-evidently a good thing that their qualifications should be valid even far away from home and that they should be able to

## Usecase - Politicians as Documents

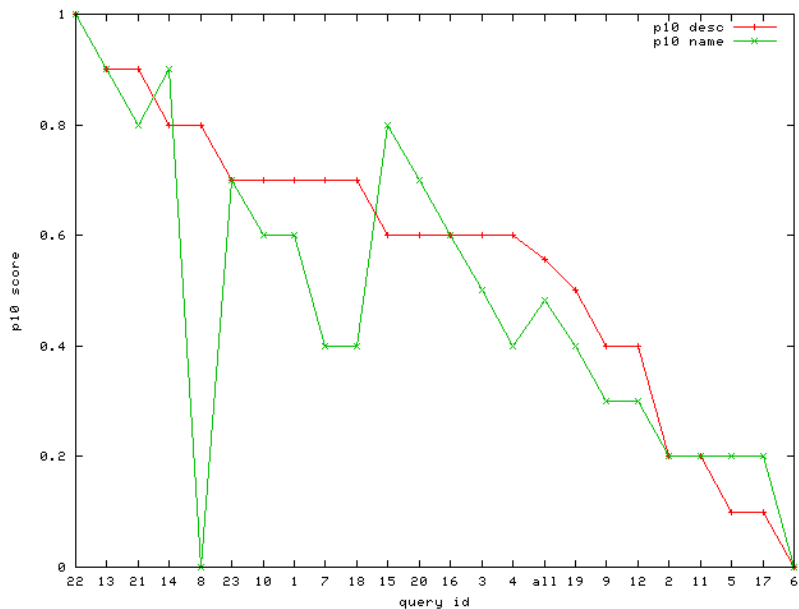
- Model a politician as one document, i.e. a collection of all text attributed to him/her.
- Reduce politician search to document retrieval.
- Basic retrieval with Kullback-Leibler divergence.
- Easy to implement with the Lemur toolkit.

# Evaluation

- **Task:** find the politicians belonging to committees
  - Justice
  - Defense
  - Agriculture, etc
- **Two runs:**
  - given the committee title (2-5 words)
  - full description of the committee (100-200 words).



# Comittee evaluation



## European elections example

Website created for the European elections. Goal was to assist users in finding background information on European politicians and search Dutch party programs.

<http://www.ikkieswijzer.nl/>