

PoliDocs.nl

Parlementaire Data Open Toegankelijk

Maarten Marx, Tim Gielissen, Robert Jan de Groot
Universiteit van Amsterdam
maartenmarx@uva.nl

Aanleiding In 2005 besloot de Tweede Kamer dat al hun data in een *open* formaat voor iedereen beschikbaar moest komen. XML werd genoemd als ideaal. In November 2005 publiceerde de politieke weblog Sargasso.nl uit onvrede met de parlementaire informatievoorziening via Parlando een lijst met 15 minimum eisen waaraan zo'n voorziening zou moeten voldoen.¹

Begin 2008 was de situatie nog net als in 2005. Ondertussen was ook het StatenGeneraalDigitaal project gestart waarin alle Handelingen der Staten Generaal vanaf 1814 worden gedigitaliseerd en publiekelijk toegankelijk gemaakt.

Dat het anders dan in Nederland kan laten bijvoorbeeld het Vlaams Parlement, de EU, en het particuliere initiatief www.theyworkforyou.com in de UK zien.

Doel van project De Nederlandse parlementaire publicaties sinds 1814 op een transparante manier ontsluiten en daarbij voldoen aan de eisen van Sargasso.nl.

Voorwaarden Het is een wetenschappelijk project met een toegepast karakter. We wilden niet alleen laten zien dat het kon, maar ook dat zo'n project uitvoerbaar is door kleine onafhankelijke organisaties zonder veel geld, expertise en machinerie. Dit mondde uit in de volgende eisen:

- **Open formaten** alleen open source software en open standaarden gebruiken: XML, XSLT en LAMP;
- **Schaalbaarheid** alle dataverwerking zonder tussenkomst van mensen; scripts moeten robuust zijn naar de toekomst en naar het verleden;
- **Transparant** alle stappen in het Extract-Transfer-Load proces zijn controleerbaar en reproduceerbaar;
- **Gebruikersvriendelijk** het eindproduct moet makkelijk toegankelijk zijn voor zowel de gewone burger met eenvoudige zoekvragen als de professionele gebruiker die complexe analyses wil uitvoeren.

Uitvoering Een team van 2 Informatiekunde Bachelor afstudeerders en 1 universitair docent hebben in 4 maanden een prototype gebouwd. De volgende stappen zijn uitgevoerd:

- 1) Opstellen van DTD's voor Kamervragen en verslagen van zittingen van de Tweede Kamer;
- 2) Ophalen van alle data uit Parlando (1995–heden) en StatenGeneraalDigitaal (1984–1995);
- 3) Transformatie van PDF's met bijbehorende metadata naar XML volgens de opgestelde DTD's d.m.v. `pdftohtml`, `sed` en XSLT 2.0;
- 4) Normalisatie van namen van personen en partijen en van datums;
- 5) Invoeren in MySQL database;
- 6) Schrijven van CSS en XSLT 1.0 stylesheets voor het tonen van de XML data;
- 7) Bouwen frontend www.polidocs.nl.

¹<http://sargasso.nl/archief/2005/11/17/actie-open-democratie-10-de-brief/>

Evaluatie De site voldoet aan 12 van de 15 Sargasso eisen. De site www.polidocs.nl is positief ontvangen door verschillende professionele gebruikers: o.a. de opsteller van de Sargasso eisen, het team achter StatenGeneraalDigitaal, de Dienst InformatieVoorziening van de Tweede Kamer, het team achter KiesKompas. De site biedt voor de Kamerverslagen en Kamervragen meer functionaliteit dan vergelijkbare sites in België, de UK en de EU. Wat betreft de gestelde voorwaarden

- **Open formaten** vrijwel geheel aan voldaan (alleen reguliere expressies zijn met sed afgehandeld);
- **Schaalbaarheid** scripts ontwikkeld voor de digitaal beschikbare Parlando data konden met kleine aanpassingen ook toegepast worden op de ingescande StatenGeneraal data.
- **Transparant** vrijwel aan voldaan; refactoring nodig en dan haalbaar;
- **Gebruikersvriendelijk** Dit is het beste te beoordelen door het zelf te proberen. Een aardige test is om deze vraag te beantwoorden: "Hoe wordt het woord *knettergek* in de loop der jaren in het parlement gebruikt?". Probeer het eerst door <http://parlando.sdu.nl/cgi/login/anonymous> en <http://www.statengeneraaldigitaal.nl/> te gebruiken, en daarna dezelfde vraag aan <http://www.polidocs.nl> te stellen.

Toekomst We zijn in overleg met de Dienst InformatieVoorziening van de Tweede Kamer om de site permanent in de lucht te houden. Onze wens is om dan *alles alléén* met XML-technologie te doen: XSLT en de gehele backend met de Nederlandse open source MonetDB/XQuery en PFTijah XML database en zoekmachine.

PoliDocs.nl
 De zoekmachine voor politieke documenten

:: Startpagina ::
:: Zoeken ::
:: Geavanceerd zoeken ::
:: Personen ::
:: Partijen ::

» Eenvoudig Zoeken - Resultaten

7 x MOT 325 x KVR 1185 x HAN

Zoek verder: Zoek
 Sorteer op: relevantie | datum (nieuwste eerst) | datum (oudste eerst)

5 meest voorkomende personen:	5 meest voorkomende partijen:	5 meest voorkomende jaren:
- Van der Vlies: 27 hits	- PvdA: 219 hits	- 2006: 98 hits
- Minister Lubbers: 17 hits	- CDA: 195 hits	- 2007: 86 hits
- Van Bommel: 17 hits	- VVD: 160 hits	- 1997: 81 hits
- Kant: 16 hits	- D66: 106 hits	- 1998: 81 hits
- De Hoop Scheffer: 15 hits	- SP: 97 hits	- 1999: 81 hits

[+/- Moties](#)

Meer moties

1
MOT



Schutte

Uitbreiding van het provinciale belastinggebied.

Uitbreiding van het provinciale belastinggebied. Wijziging regeling Provinciefonds. belastingen planning provinciale financiën wegen wetgeving

GPV

1983-06-14 Geen bron beschikbaar

2



Roemenië en Bulgarije.

