

Turn-key platform Newz
Big Content & Semantics

Turn-key platform Newz Big Content & Semantics

Introduction

Michel de Ru

- Solution architect @ Dayon
- 16 years experience in publishing
- Among others Wolters-Kluwer, Sdu (ELS) and Dutch Railways
- Specialized in Content related Big Data challenges
- Specialized in added value through Semantic Technology



Dayon, part of the HintTech Group

- We design, build and maintain content driven online and mobile applications
- We help customers develop their Content Strategy
- We realize it using Content Technology
- Partners include MarkLogic, Ontotext, Alfresco, Hippo CMS, Solr and OpenText
- Big Data projects for Dutch Public Library, Kluwer, Newz

Turn-key platform Newz

Big Content & Semantics

Contents

1. Short intro to Newz
2. Machine readable news articles / Linked Open Data
3. How we put it together
4. Use-cases

michel.de.ru@dayon.nl

+31 6 38 507 567

Turn-key platform Newz

Big Content & Semantics

See video on newz.nl

Turn-key platform Newz

Big Content & Semantics

The Project

Within 3 months

- First production functionality

After another 6 month

- Semantic enrichment

October 2013

- Newz B.V. started it's organization



Turn-key platform Newz Big Content & Semantics

Rutte in American top 3 best dressed world leaders

29/03/13, 09:37 - bron: Vanity Fair



© ANP.

Volgens het Amerikaanse Vanity Fair kleedt premier Rutte zich uitstekend. Hij scoort een derde plek op de lijst best geklede wereldleiders. Hij verslaat onder andere president Obama, die op plek vijf staat. Onze premier doet de samenstellers zelfs denken aan James Bond acteur Pierce Brosnan.



About: Barack Obama

An Entity of Type : [office holder](#), from
Named Graph : <http://dbpedia.org>, within
Data Space : dbpedia.org



Turn-key platform Newz Big Content & Semantics



Instellingen...









Uitloggen

Data Journalistiek Applicatie









Zoeken

Actuele concepten

Personen

-  Mike Havenaar
-  Tom Boonen
-  Dick Advocaat
-  Frank de Boer
-  Theo Janssen
-  Wilfried Bony
-  Fabian Cancellara
-  Fred Rutten









Gebeurtenissen

-  Ronde van Vlaanderen
-  Parijs-Roubaix
-  Eredivisie
-  Serie A
-  GDC
-  Bundesliga
-  Amstel Gold Race
-  Jupiler League

Organisaties

-  PSV
-  Ajax
-  Feyenoord
-  FC Twente
-  Roda JC
-  PEC Zwolle
-  Willem II
-  FC Utrecht

Locaties

-  Nederland
-  Amsterdam
-  Arnhem
-  China
-  Longa
-  Zwolle
-  DEN BOSCH
-  Europa

Turn-key platform Newz

Big Content & Semantics



barack obama

Zoeken

Instellingen...

Uitloggen

Gerelateerde concepten

Personen

▲ 8

- Presidency of Barack Obama
- Illinois Senate career of Barack Obama
- United States Senate career of Barack Obama
- Barack Obama 15px
- Barack Hussein Obama, Sr.
- Early life and career of Barack Obama
- Jake Barack
- Moises Barack

Gebeurtenissen

- Thomas More Law Center v. Obama
- Battle of Keltis barracks

Organisaties

▼ 8

- Barack Obama Academy
- Barack Obama Charter
- Barack Obama Academy of International Studies 6-12

192 resultaten

<< < 1-10 > >>

Obama op lijst in Kenia
Reformatorisch Dagblad 1 maart 2013

Obama uit kritiek op beleid Israël
Reformatorisch Dagblad 21 maart 2013

Moeder gaat voor Obama
Wegener verzamel 22 maart 2013

Moeder gaat voor Obama
Wegener verzamel 22 maart 2013

Moeder gaat voor Obama
Wegener verzamel 22 maart 2013

Obama wil vertrouwen winnen in Israël
Reformatorisch Dagblad 15 maart 2013

Moeder gaat voor Obama
Wegener verzamel 22 maart 2013

Moeder gaat voor Obama
Wegener verzamel 22 maart 2013

Moeder gaat voor Obama

Verfijnen

Datumfilter

Start:

Eind:

Zoeken

Titel

▼ 5

- Wegener verzamel (122)
- Reformatorisch Dagblad (36)
- Het Financieele Dagblad (19)
- De Telegraaf (14)

Trefwoord

▼ 25

- obama (38)
- israel (22)
- joost sijtsma (21)
- verenigde staten (18)

Uitgever

▼ 5

- Wegener NV, Koninklijke (122)

Turn-key platform Newz Big Content & Semantics

[Zoeken](#)[Instellingen...](#)[Uitloggen](#)[bookmark](#) [e-mail](#)

Zoekresultaat



Barack Obama

1961-08-04

Verenigde Staten

Barack Hussein Obama II is de 44e en huidige president

van de Verenigde Staten. Hij is de eerste Afro-Amerikaan in deze functie. Tussen 3 januari 2005 en 16 november 2008 was Obama lid van de Senaat als vertegenwoordiger van de staat Illinois en voordien was hij staats senator in de wetgevende vergadering van zijn thuisstaat.

Beroepen/functies

Politicus
Auteur
Advocaat (beroep)
Schrijver
Hoogleraar
President

Barack Obama op het web

http://nl.wikipedia.org/wiki/Barack_Obama
<http://rdf.freebase.com/ns/m.02mjmr>

Personen

- Ann Dunham
- Barack Hussein Obama, Sr.
- Michelle Obama

▼ 10

Locaties

- Honolulu
- Hawaï
- Verenigde Staten

▼ 8

Organisaties

- Witte Huis (Washington D.C.)
- Columbia-universiteit
- Occidental College
- Harvard Law School
- Democratische Partij (Verenigde Staten)
- Barack Obama Academy

▲ 6

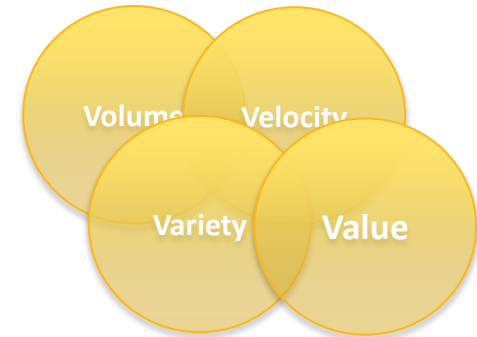
How we put it together

Contentstrategie & technologie

dayon[®]

Big Content & Semantics

Dutch news = Big Data



Volume

- 15.000 news articles a day

Velocity

- Delivery spike during 2 hours a day (just before the morning starts)
- Usage is continuously (through API, Search and Subscription interfaces)

Variety

- News articles without metadata and no structure whatsoever
- Linked Open Data

Value

- Facilitate new News business solutions for integrators, app suppliers, etc.
- Deliver a standardized (NITF NewsML) and enriched format

Turn-key platform Newz Big Content & Semantics

Key aspects

Big Data Content Store

Volume

Velocity

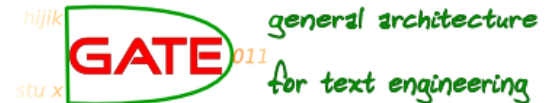
- Enterprise NoSQL
- Structured/unstructured
- ACID compliant (Atomicity, Consistency, Isolation, Durability)



Semantic Technologies

Variety

- Concept extraction
- Linked (Open) Data
- Graph databases / Inferencing



Content Lifecycle Management

- Part of Application Lifecycle Management



Turn-key platform Newz

Big Content & Semantics

Volume, Velocity



Interface with News publishers

- Content Processing Framework
- Added a Java layer for full ETL and trailing capabilities

Storage of News articles

- In cooperation with IPTC a Dutch version of NewsML-G2 has been defined
- Interface with Semantic Extraction framework
- Full search capabilities

Enterprise grade

- We also calculated a MongoDB/Lucene solution
- ML won on: TCO, Success rate of business implementations, Enterprise resilience

Big Content & Semantics

Variety



Semantic Extraction

- Existing news vocabularies and taxonomies + Linked Open Data
- World class Semantic Extraction (NLP, Golden Standard, Rules, etc.)
- Conversion to an ontology (similar to semantic web)
- Triples stored in OWLIM Enterprise

Enrichment of news articles

- Organizations
- Persons
- Locations
- Events
- Keywords
- Mentions

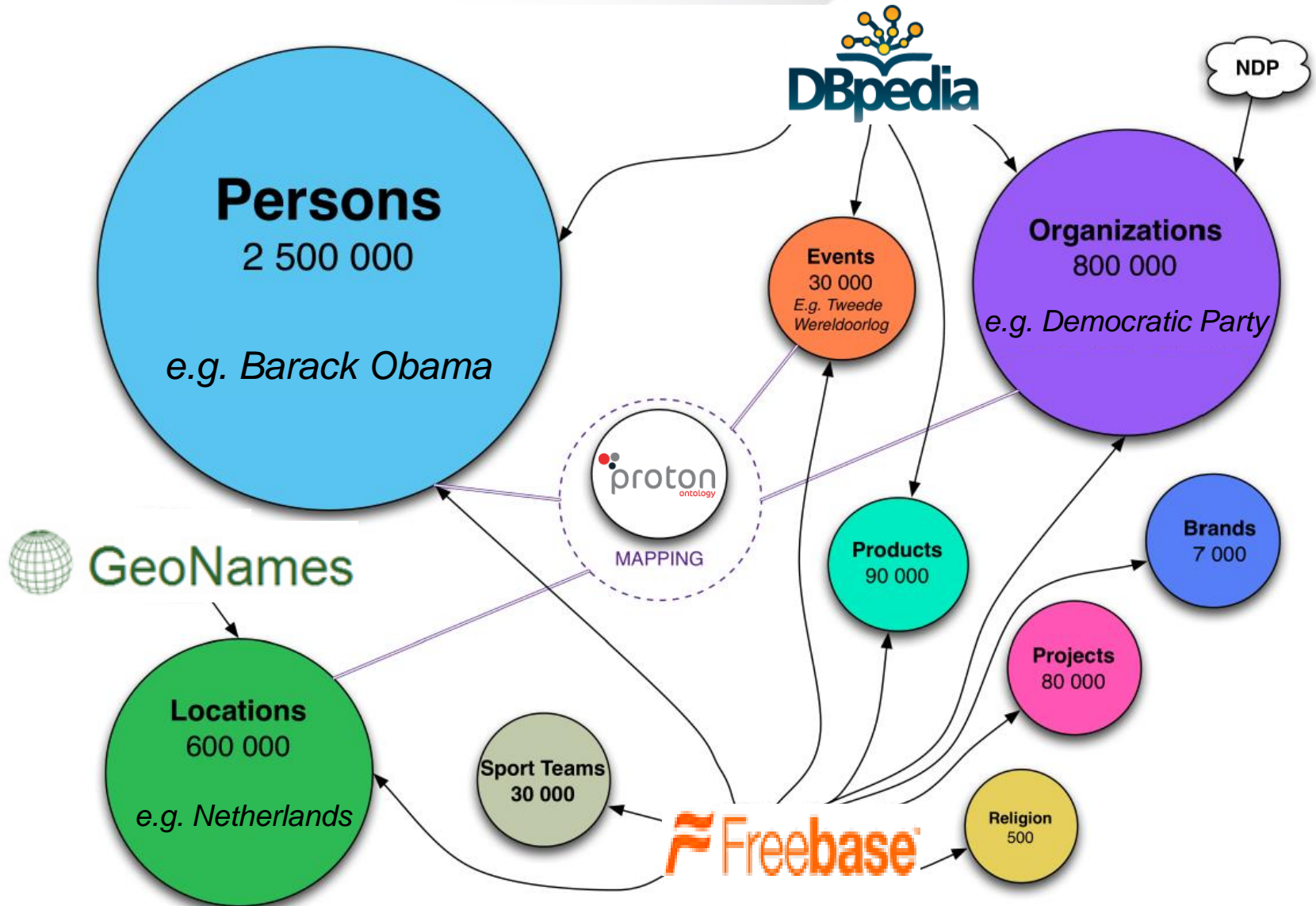


*From a lot of data...
... To even more data!*



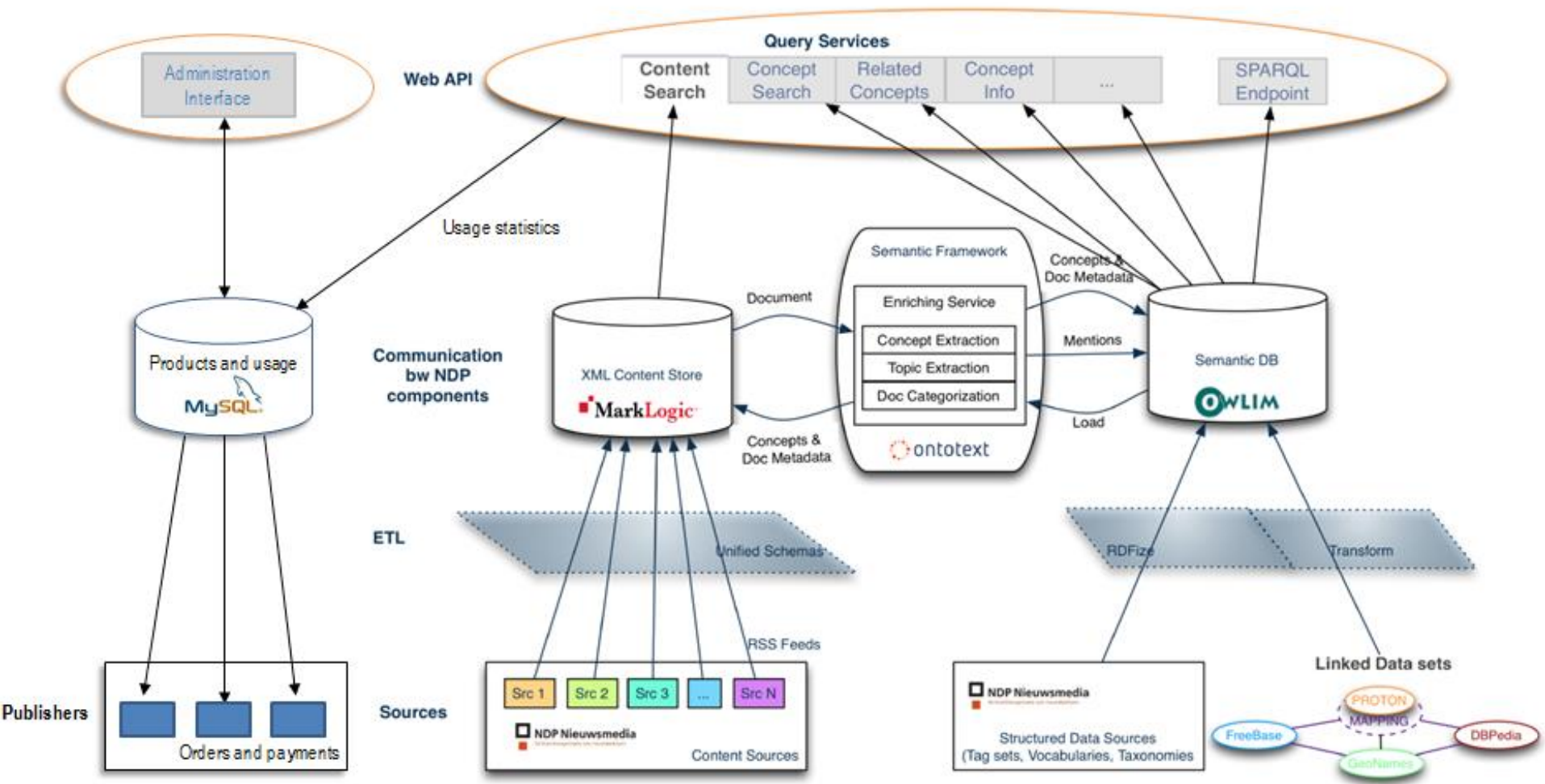
Turn-key platform Newz

Big Content & Semantics



Turn-key platform Newz Big Content & Semantics

Architecture overview



Use cases

Turn-key platform Newz Big Content & Semantics

Voorbeeld: Automatische geo taxonomie

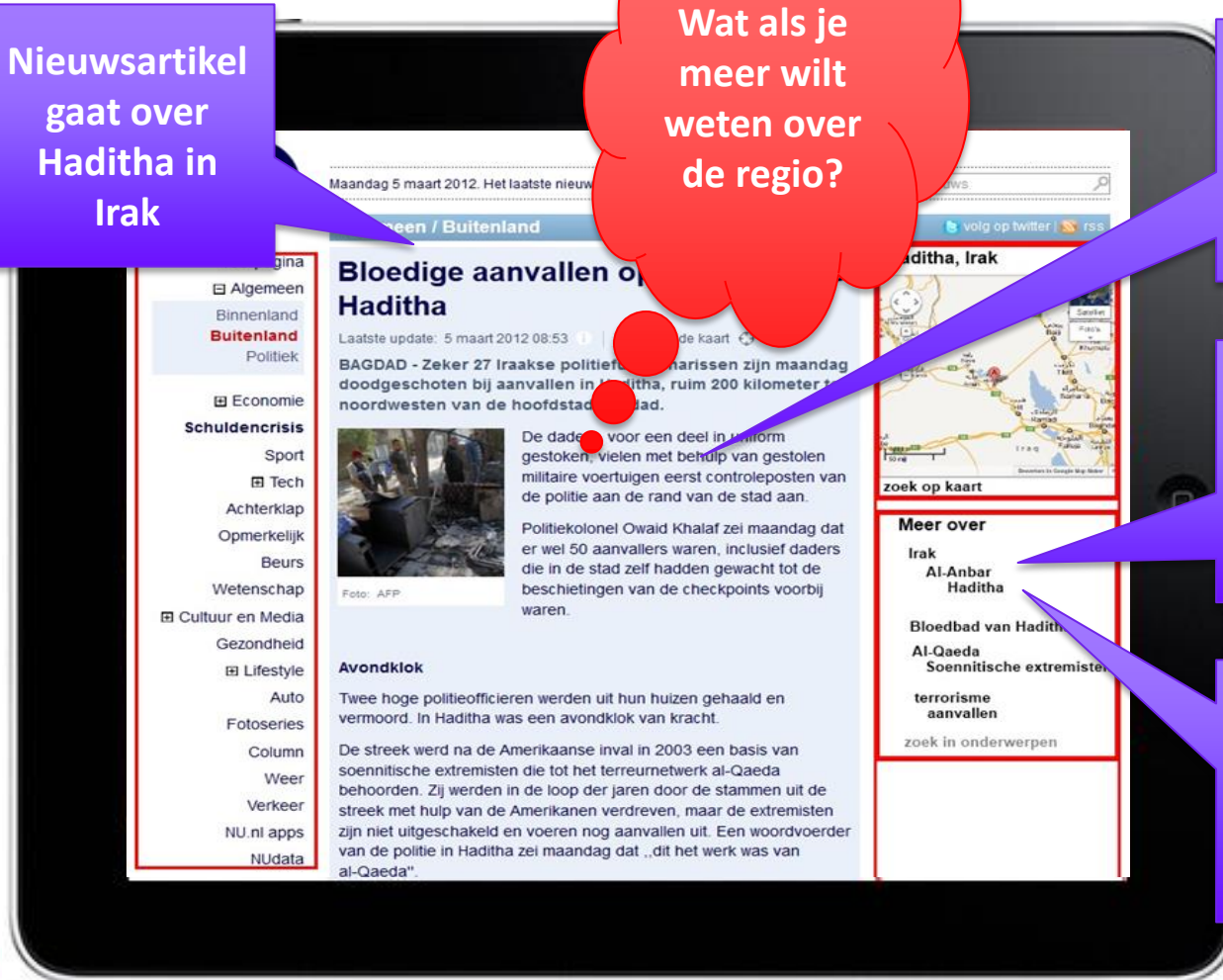
Nieuwsartikel
gaat over
Haditha in
Irak

Wat als je
meer wilt
weten over
de regio?

1. Artikel is
semantisch verrijkt
met de plaatsnaam

2. Op basis van
Linked Open Data
wordt een taxonomie
getoond

3. Daarmee kan alle
content die over de
regio gaat gevonden
worden



Turn-key platform Newz Big Content & Semantics

Nieuws gekoppeld aan boeken

Books for Understanding
Scholarship Beyond the Headlines

Book Lists | About | Contact | AAUP

Email Alerts | RSS | Contact

Google™ Custom Search [Search]

Welcome

Books for Understanding is a free, easy-to-use resource to help you find books on current events. New bibliographies are compiled when a major news story breaks or public debate heats up. For more information, click [here](#).

Latest Book List

Boston

In honor of our Annual Meeting host city, home to a number of university presses, the "Boston" bibliography highlights the latest and best university press scholarship on the scholarly city.

Go to: • [Book List](#) • [Press Release](#)

Chechnya

Though the motive behind the Boston Marathon bombings is still unclear, public interest has focused on the suspects' Chechen heritage. These works of scholarship shed light on the history and conflicts of this Caucasian republic and its region.

Go to: • [Book List](#) • [Press Release](#)

Guns in America

A tragic string of mass shootings inspired our latest bibliography. Scholarship on gun policy, violence, the Second

50th Anniversary of the March on Washington

On August 28, 1963, Martin Luther King, Jr. told hundreds of thousands gathered on the Washington Mall of his dream for racial and economic justice. Fifty years on the occasion is as much one for reflection and continued action as it is commemoration.

Go to: • [Civil Rights Era Book List](#) • [Economic Justice Book List](#)

Democratic Crisis in Egypt

Hundreds of deaths have been reported as Egypt's security forces brutally clear protests by supporters of deposed president Mohamed Morsi.

• Go to [Egypt Book List](#)

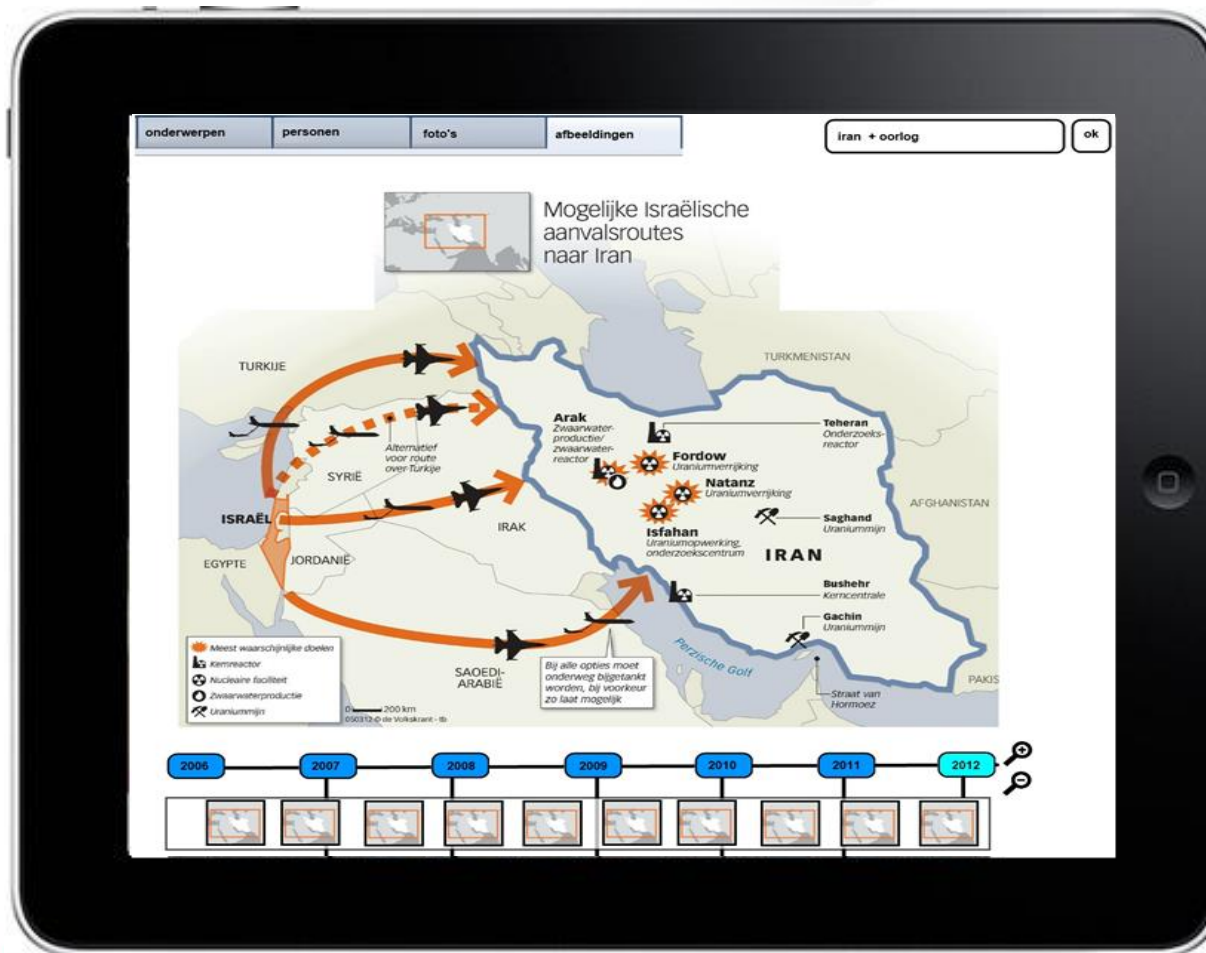
Important SCOTUS Rulings

The Supreme Court issued several important rulings this summer: striking down the Defense of Marriage Act; limiting the Voting Rights Act; and reconsidering university affirmative action policies.

Go to: • [Supreme Court Book List](#) • [Same-Sex Marriage Book List](#) • [Voting & Elections Book List](#) • [Race and Education Book List](#)

Turn-key platform Newz Big Content & Semantics

Voorbeeld: tijd reizen door infographics



Turn-key platform Newz Big Content & Semantics

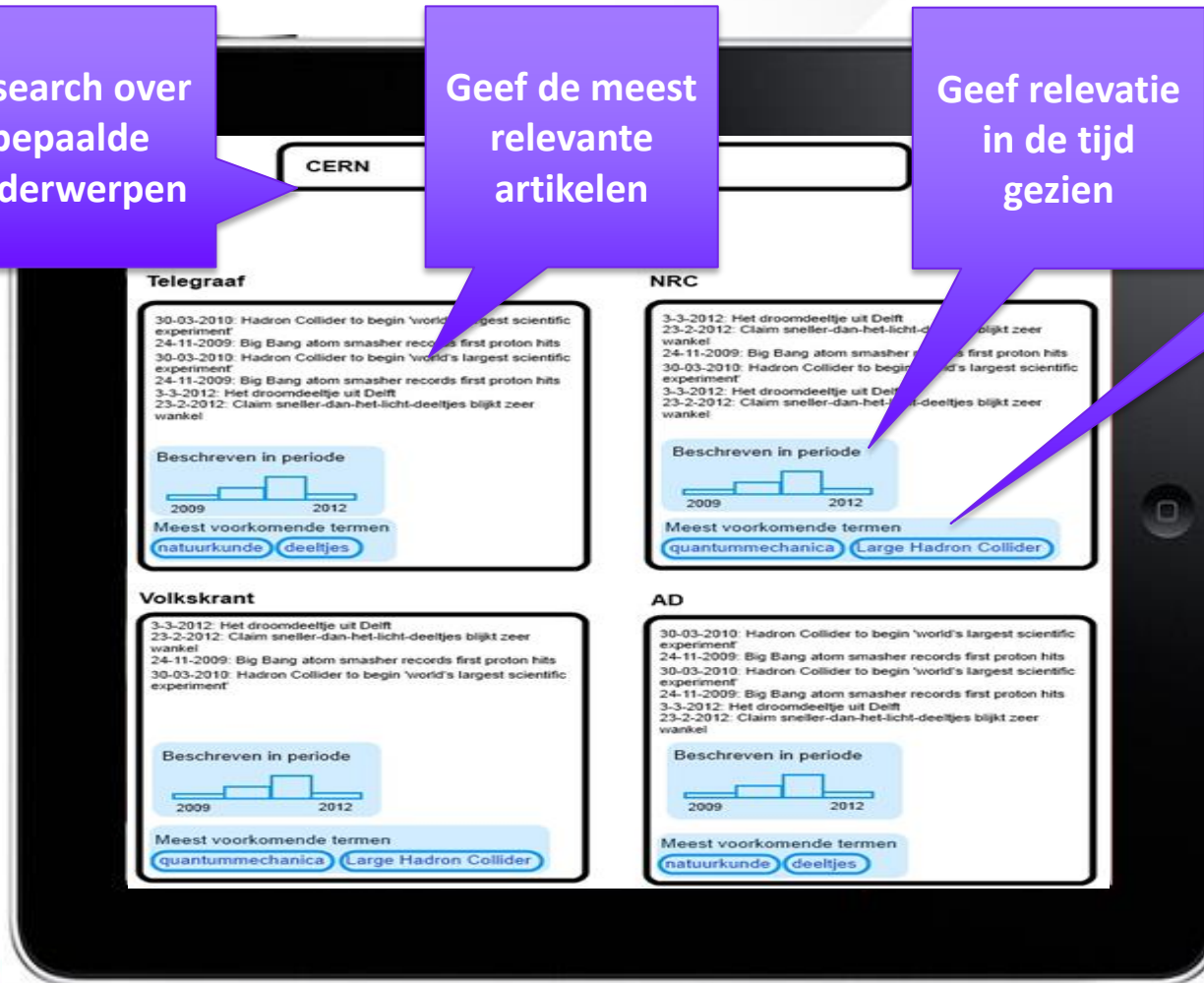
Voorbeeld: Research

Research over
bepaalde
onderwerpen

Geef de meest
relevante
artikelen

Geef relevantie
in de tijd
gezien

Geef de
mogelijkheid
tot een
verdiepende
zoektocht



Turn-key platform Newz Big Content & Semantics

Voorbeeld: Mashups

Research over
bepaalde
onderwerpen

Verrijk

Verrijk
resultaat met
Linked Open
Data

Verrijk
resultaat met
eigen
taxonomie /
ontologie

Definitie

Quantum mechanics (QM) principles describing the energy and matter that p... and subatomic scales. QM gets its... the notion of a quantum, and that... is the Planck constant...
meer...

Gerelateerde onderwerpen

Folding
Relativiteitstheorie
Dirac-formalisme
Cern
meer...

Personen

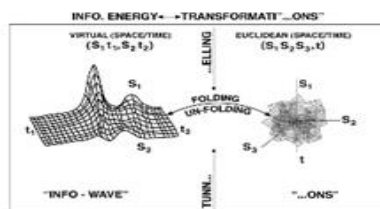
Allen, Herbert Stanley
Fritz Wolfgang London
Lev Vaidman
Pascual Jordan
Francis Crick
meer...



Universiteiten

TU Delft
Universiteit Utrecht
VU Amsterdam
California UT
meer...

Afbeeldingen



$f(x)$ Any function of position, such as x , or potential $V(x)$

p_x x component of momentum (y and z same form)

E Hamiltonian (time independent)

E Hamiltonian (time dependent)

KE Kinetic energy

L_z z component of angular momentum

$f(x)$

$\frac{h}{i} \frac{\partial}{\partial x}$

$\frac{p_x^2}{2m} + V(x)$

$i\hbar \frac{\partial}{\partial t}$

$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2}$

$-i\hbar \frac{\partial}{\partial \phi}$

