CoSyne

Multilingual Content Synchronization with Wikis

A. Bronner
University of Amsterdam
Netherlands Institute for Sound & Vision

Wikimania 2011, Haifa Israel
Introduction

- CoSyne: Multilingual Content Synchronization with Wikis

- Research & Development Project, Co-Funded by the European Union under the 7th Framework Programme (FP7- ICT)

- March 2010 – February 2013
Consortium

- **4 Research Institutions**
  - University of Amsterdam (UvA)
  - Fondazione Bruno Kessler (FBK)
  - Dublin City University (DCU)
  - Heidelberg Institute for Theoretical Studies (HITS)

- **3 End User Partners**
  - Deutsche Welle
  - Netherlands Institute for Sound & Vision
  - Wikimedia Netherlands
Objective

The CoSyne project will develop a system for synchronizing wiki pages in multiple languages and for maintaining consistency while pages are edited in the different languages.
Motivation

- Why do we want to synchronize wiki pages?
  - Keep content consistent across languages
  - Learn about content in other languages
  - Expand to new languages

A. Bronner
University of Amsterdam / Netherlands Institute for Sound & Vision
Wikimania 2011, Haifa Israel
Soccer Example

- 28 May 2011, 19:45
- Wembley stadium, London
- Barcelona vs. Manchester United
- Barcelona wins (3:1)
- Goalkeeper Edwin van der Sar, at the age of 40, plays his last game for Manchester United and retires from football
Soccer Example

- 28 May 2011, 20:35 (already at the half-time ???)


- “Van der Sar officially retired from football following a 3-1 defeat ...”

---

Revision as of 18:56, 28 May 2011 (edit)
Ntadams (talk | contribs)
← Previous edit

and [[Fabio Pereira da Silva|Fabio]]. He was included in the [[PFA Team of the Year]], making his third appearance there, others in 2007 and 2008.

Revision as of 20:35, 28 May 2011 (edit) (undo)
ShadowTerror10 (talk | contribs)
Next edit →

On 22 May 2011, he played his final game at Old Trafford in a 4–2 victory over [[Blackpool F.C.|Blackpool]] resulting in Blackpool's relegation from the Premier League. He captained United as a farewell treat. Previous to the last game of the Premier League season he had announced that after the [[2011 UEFA Champions League Final]] against [[FC Barcelona|Barcelona]] he would retire.

Van der Sar officially retired from football following a 3-1 defeat to Barcelona in the 2011 UEFA Champions League Final.
Soccer Example

- 28 May 2011, 22:47

- Italian user edit: http://it.wikipedia.org/wiki/Edwin_van_der_Sar

- “He will play played his last game … ”

Versione delle 22:39, 28 mag 2011 (modifica)
79.42.63.220 (Discussione)
← Differenza precedente

Versione delle 22:47, 28 mag 2011 (modifica) (annulla)
151.24.72.126 (Discussione)
Differenza successiva →

----- Stagione 2010-2011 -----
Il [[23 dicembre]] [[2010]] annuncia che a fine stagione si ritirerà. Il [[22 maggio]] seguente ha raggiunto le 605 partite di campionato disputate in vent’anni: 226 con l’Ajax, 66 con la Juventus, 127 con il Fulham e 186 con il Manchester.<ref>{{Cita news|autore = Massimo Perrone|titolo = Huntelaa a quota 201|pubblicazione = La Gazzetta dello Sport|Extra Time|giorno = 24|mese = 04|anno = 2011|pagina = 6}}<ref> Di sjuta La sua ultima partita il [[28 maggio]] nella finalissima di [[Champions League]] allo [[Wembley Stadium (2007)|Stadio Wembley]] contro il [[F.C. Barcelona]]

----- Nazionale -----
Soccer Example

- 28 May 2011, 23:10

- German user edit: http://de.wikipedia.org/wiki/Edwin_van_der_Sar

- “On May 28 2011 he was defeated in his last professional game … ”
Soccer Example

- 29 May 2011, 13:33
- Dutch user edit: http://nl.wikipedia.org/wiki/Edwin_van_der_Sar
- “Van der Sar played his last game as goalkeeper …”
Keep Content Consistent (made-up example)
Content from Another Language
(made-up example)

Amsterdam
From Wikipedia, the free encyclopedia

This article is about the Dutch capital. For other uses, see Amsterdam (disambiguation).

Amsterdam (/æmˈstərˌdam/; Dutch: [aːmˌstɑr.dəm] (listen)) is the largest city and the capital of the Netherlands. The current position of Amsterdam as capital city of the Kingdom of the Netherlands is governed by the constitution of 24 August 1815 and its successors. The city is the province of North Holland in the west of the country. It comprises the northern part of the Randstad, one of the larger urban agglomerations of Western Europe.

Náhuatl
Dorerin Naoero
Nederlands
Nedersaksisch
Nepali

→ Synchronize

Amsterdam's mayor, Eberhard van der Laan, and the Maastricht city council. Others have argued that the plan may indeed violate not only the constitution of the Netherlands but also that of the European Union. It is thus expected that the measure will fail before it is more closely considered. As of July 2011, it is safe to say that no "wietpas" will be implemented until 2015 - at the earliest.

A. Bronner
University of Amsterdam / Netherlands Institute for Sound & Vision
Wikimania 2011, Haifa Israel
Overlapping Details (Same Info)

Amsterdam

Amsterdam is the largest city and the capital of the Netherlands.

De naam van de stad is afgeleid van de ligging bij een in de 13e eeuw aangelegde dam in de Amstel. De plaats kreeg stadsrechten rond het jaar 1300 en groeide later uit tot één van de grootste handelssteden ter wereld in de Gouden Eeuw. Bevolkingsgroei leidde vanaf het eind van de 16e eeuw tot grote stadsuitbreidingen, waaronder de grachtengordel, die op de Werelderfgoedlijst is opgenomen en tot de belangrijkste bezienswaardigheden van de stad behoort. Andere attracties in de stad onder meer de musea, zoals het Rijksmuseum, het Stedelijk Museum en het Van Gogh Museum; het prachtige Slot Zeeland en de Amsterdam Oosterpark; en andere bezienswaardigheden zoals het Museumplein, de Rembrandtplein en het Prinsengracht.

Amsterdam

From Wikipedia, the free encyclopedia

This article is about the Dutch capital. For other uses, see Amsterdam (disambiguation).

Amsterdam (ˈɑmstərˌdɑm; Dutch (ˈɑmstərˌdɑm) (listen)) is the largest city and the capital of the Netherlands. The current position of Amsterdam as capital of the Kingdom of the Netherlands is governed by the constitution of 24 August 1815 and its successors. Amsterdam has a population of 780,152 within city limits, an urban population of 1,209,419 and a metropolitan population of 2,158,592. The city is in the province of North Holland in the west of the country. It comprises the northern part of the Randstad, one of the larger conurbations in Europe, with a population of approximately 7 million.

Its name is derived from Amsteldamme, indicative of the city's origin: a dam in the river Amstel. Settled as a small fishing village in the late 12th century, Amsterdam became one of the most important ports in the world during the Dutch Golden Age, a result of its innovative developments in trade. During that time, the city was the leading center for finance and diamonds. In the 19th and 20th centuries, the city expanded, and many new neighbourhoods and suburbs were formed. The 17th-century canals of Amsterdam (in Dutch: 'Grachtengordel'), located in the heart of Amsterdam, were added to the UNESCO World Heritage List in July 2010.
Amsterdam is the largest city and the capital of the Netherlands. The city is in the province of North Holland on the rivers Amstel and IJ.
Amsterdam is the largest city and the capital of the Netherlands. The city is in the province of North Holland on the rivers Amstel and IJ. The city is also called 'Mokum' (מוקם) in the local dialect.
Amsterdam is the largest city and the capital of the Netherlands.

The city is in the province of North Holland on the rivers Amstel and IJ.

The city is also called 'Mokum' (мест) in the local dialect.

Amsterdam's coffee shops will not be closing their doors to tourists any time soon.
Amsterdam is the largest city and the capital of the Netherlands.

The city is in the province of North Holland on the rivers Amstel and IJ.

The city is also called 'Mokum' (منح) in the local dialect.

Amsterdam's coffee shops will not be closing their doors to tourists any time soon.

Amsterdam has a population of 780,152 within city limits. On April 1st 2011 the municipality counted 783,363 residents.
Take Action? (Not Necessarily...)

Amsterdam, also called 'Mokum' in the local dialect, is the largest city and the capital of the Netherlands.

The city is in the province of North Holland on the rivers Amstel and IJ.

Amsterdam's coffee shops will not be closing their doors to tourists any time soon.

Amsterdam has a population of 780,152 within city limits.

On April 1st, 2011 the municipality counted 783,363 residents.

Its name is derived from 'Amstelredamme', indicative of the city's origin: a dam in the river Amstel. 

Important ports in the world during the 16th century, Amsterdam was the leading center for finance and trade. In the 19th and 20th centuries, jobs and suburbs were formed.

A. Bronner
University of Amsterdam / Netherlands Institute for Sound & Vision
Wikimania 2011, Haifa Israel
Challenges

- **How** do we synchronize wiki pages?
- Research domains:
  - User Edits Classification
  - Structural Analysis
  - Content Entailment
  - Machine Translation
Synchronization Process

A. Bronner
University of Amsterdam / Netherlands Institute for Sound & Vision
Wikimania 2011, Haifa Israel
User Edits Classification

- Decides whether the meaning has changed:
  - Factual change → Synchronize
  - Fluency change → Improve (self learning)
Structural Analysis

- Decides which segments to compare
- Instructs how to reconstruct the wikitext
- Decides where to insert the modifications
Cross-lingual Content Entailment

- Determines whether text segments are equivalent
  - Equivalent → No change
  - More specific → Replace
  - Unrelated → Add
  - Contradiction → Ask the user

A. Bronner
University of Amsterdam / Netherlands Institute for Sound & Vision
Wikimania 2011, Haifa Israel
Machine Translation

- Translates segments that need to be added or replaced
- Collects “fluency” corrections and learns improved models
Status of Development

- 3 prototypes planned
  - 1st prototype released
  - Deployed on end user's (private) wiki

- Target Languages:
  - English, German, Italian, Dutch (core)
  - Bulgarian, Turkish

- Machine Translation Quality
  - Evaluation work
  - Ongoing improvement
  - Open source & open data
Use Cases 1 & 2

- The wiki of the Netherlands Institute for Sound & Vision
- Deutsche Welle's Kalenderblatt (Today in History)
  - Demo
1. Januar

Themen
- 1.1.1863: USA schaffen Sklaverei ab
- 1.1.1802: Berliner Schauspielhaus eröffnet
- 1.1.1801: Entdeckung des Ceres

Gedenktage
- 1.1.2002: Neue Währung für Europa
- 1.1.1993: Europäischer Binnenmarkt
- 1.1.1967: "Kommune 1"
- 1.1.1959: Fidel Castro in Havanna
- 1.1.1957: Saarland 10. Bundesland der Bundesrepublik
- 1.1.1901: Australien
- 1.1.1900: "BGB" in Kraft
- 1.1.1896: Grundstein des Jugendstills
- 1.1.1801: Union mit Irland

Geburtstage
- 1.1.1939: Michèle Mercier
- 1.1.1920: Anna Langfus (†12.5.1966)
- 1.1.1919: Jerome David Salinger (†27.1.2010)

In anderen Sprachen
The user (DW) edits German content for Kalenderblatt

There is no English page

New entry: Bartolomé Esteban Murillo, a Spanish painter
Born 1.1.1618, Died 3.4.1682
Welcome to CoSyne

Content Synchronization with Wikis

This is a prototype version. It is still under development.
You will be redirected to the modified page in a few seconds.
Executing analysis .. in_process

The user triggers CoSyne
CoSyne creates a new English page and adds language links between the German & English versions.
CoSyne provides German translation to English and reconstruction of Wikitext formats (hyperlinks, headings, bold/italic, images, templates, ...)

1.1.1618: Bartolomé Esteban Murillo (†3.4.1682)

This page has been edited by CoSyne. It contains machine generated content. It requires copy editing for grammar, style, cohesion, tone and/or spelling. You can assist by editing it.

bartolomé Esteban Murillo was a Spanish painter of the baroque.
bartolomé Esteban Murillo was born in Seville;

Category: CoSyne generated content
CoSyne adds a warning message about machine generated content and asks users to review & correct the content
The German version is now linked to an English version.
The user provides additional information to the German version
adding a new sentence & changing an existing sentence.
Welcome to CoSyne

Content Synchronization with Wikis

This is a prototype version. It is still under development.

You will be redirected to the modified page in a few seconds.

Executing analysis .. in_process

The user triggers CoSyne
CoSyne creates a new version of the English page
CoSyne does not change the 1st sentence, it replaces the 2nd by a translation of the modified German & it adds a translation of the new German sentence
Use Case 3

- Wikimedia Netherlands
  - Options:
    - MediaWiki manuals
    - Copyright lobby texts (?)
    - Specific domains (technical, soccer, …)
  - Languages:
    - EN, DE, IT, NL
  - CoSyne users mailing list
Want to Get Involved?

Visit our website:

cosyne.eu