

# Alexandria Demonstrations

The Alexandria Team

# **HistDiv – History by Diversity: Helping Historians Search News Archives**

Jaspreet Singh

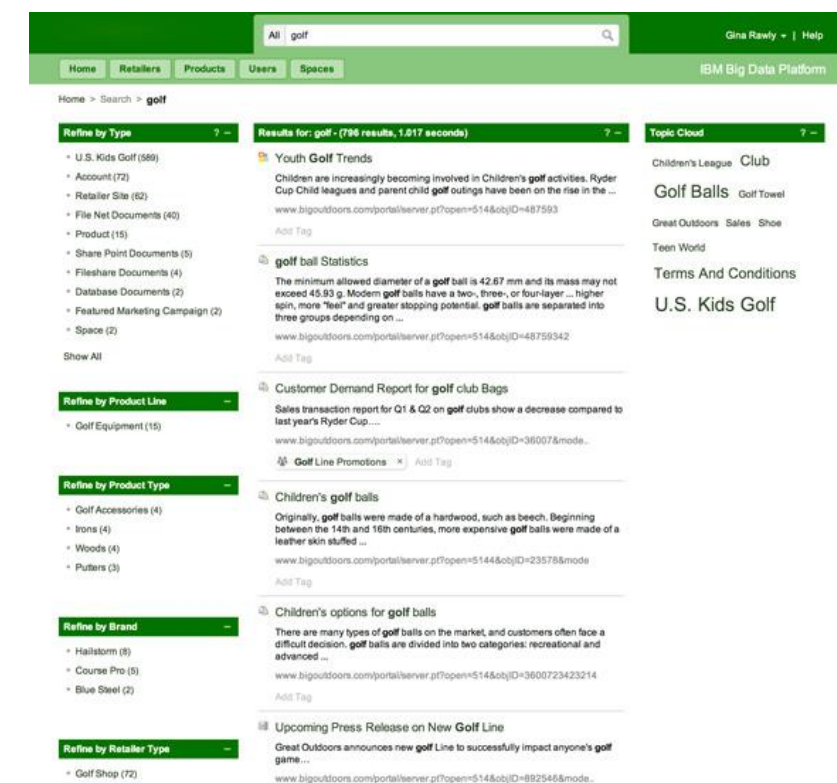
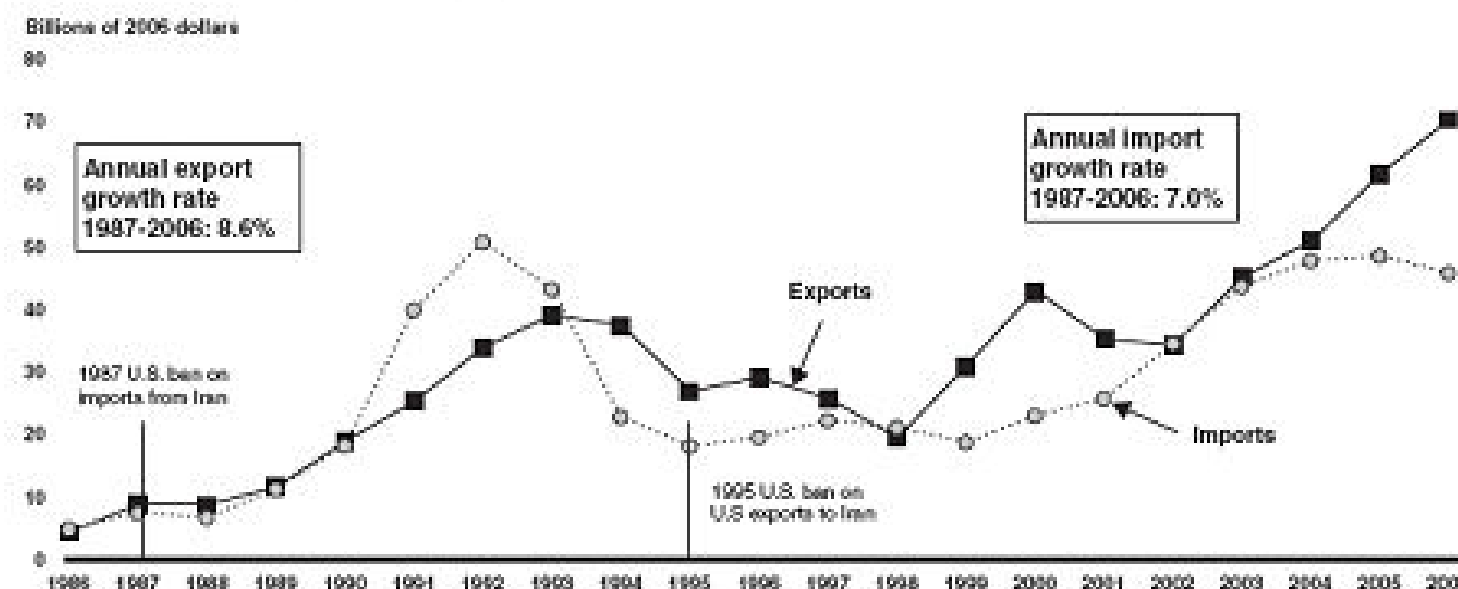
# Exploring News Archives

Complex  
Information need

Pan & Zoom

Focus & Context

Figure 2: Iran's Total Exports and Imports, 1986-2006



Why not support both?

# **The iCrawl Wizard**

## **Supporting Interactive Focused Crawl Specification**

**Gerhard Gossen, Elena Demidova, Thomas Risse**

# Focused Web Crawling for Web Science

Creation of large Web datasets through Focused Crawling:

- Automatic collection through Web Crawler
- Collector specifies **topic of crawl** through keywords and entities
- Only documents related to topic are collected

iCrawl: Help scientist create Web datasets **easily**

- Integrated crawling of Social Media
- Better monitoring & analysis of running crawls
- **Simplified creation of (focused) crawls:**  
**Crawl Specification Wizard**

Gossen et al., “iCrawl: Improving the freshness of web collections by integrating social web and focused web crawling,” JCDL '15  
Gossen et al., “The iCrawl Wizard – Supporting Interactive Focused Crawl Specification”, ECIR '15



Demo: <http://icrawl.l3s.uni-hannover.de:8090/>

Add crawl - Mozilla Firefox

Add crawl x +

localhost:8080/campaign/1/add

Suchen

iCrawl Wizard

refugees Search

Search sources: ☒ Web ☒ Twitter

**Web**

**UNHCR - The UN Refugee Agency**  
<http://www.unhcr.org/cgi-bin/texis/vtx/home>

Keywords: [camping](#), [children](#), [people](#), [Women](#), [home](#), [Liberia](#), [countries](#), [Gambia](#), [focus](#), [Kenya](#)

Entities: [UNHCR](#), [Europe](#), [Greece](#), [Germany](#), [Yugoslav Republic of Macedonia](#), [UN Refugee Agency](#), [South Africa](#), [China](#), [Asia](#), [Liberia](#)

More support needed for Somali refugees going back home from Kenya; UNHCR backs increased protection for people fleeing disasters and climate change;

**Topic - ABC News (Australian Broadcasting Corporation)**  
<http://www.abc.net.au/news/topic/refugees>

Keywords: [reports](#), [centre](#), [rights](#), [ABC](#), [Lateline](#), [EU](#), [parenting](#), [border](#), [service](#), [charities](#)

Entities: [Australia](#), [Europe](#), [australia](#), [Slovenia](#), [Nauru](#), [germany](#), [slovenia](#), [Germany](#), [Sydney](#), [Turkey](#)

Refugees. Macedonian police fire tear gas, stun grenades after migrants rush border. Posted August 22, 2015 06:22:27 | Updated August 22, 2015 07:41:03

**Refugees of the Syrian Civil War -**

**Twitter**

**Temporary reception of refugees in University Sport Centre - News - News & Events**  
<http://news.leiden.edu/news-2015/temporary-reception-of-refugees-in-...>

Keywords: [event](#), [exams](#), [Reception](#), [October](#), [week](#), [University](#), [Leiden](#), [request](#), [students](#), [More information](#)

Entities: [USC](#), [Leiden University](#), [P O](#), [Hooglandse Kerk](#), [Carel Stolker](#), [Europe](#), [Zoeterwoude](#), [Einsteinweg](#), [Municipality of Leiden](#), [University Sport Centre](#)

Leiden University will be hosting 120 refugees in the Sport Centre. Would you like to help as a volunteer? <https://t.co/zwTY5WKV7H>

**European migrant crisis: Refugees torch Brezice camp in Slovenia in protest at conditions | World | News | Daily Express**  
<http://www.express.co.uk/news/world/613809/European-migrant-crisi...>

Keywords: [Daily](#), [conditions](#), [jet](#), [World](#), [Putin](#), [Drone](#), [men](#), [contact](#), [s](#), [newspaper](#)

Entities: [ISIS](#), [Slovenia](#), [Syria](#), [UK](#), [US](#), [Europe](#), [Serbia](#), [Croatia](#), [Iraq](#), [Calais](#)

**Seeds (URLs)**

+ Add

<http://www.unhcr.org/page...>

<http://www.1000bruecken....>

<http://info.arte.tv/de/refug...>

<http://www.nytimes.com/2...>

**Keywords**

+ Add

Damascus

Immigrants

human trafficking

migrant crisis

**Entities**

UNHCR

Syria

Turkey

Lebanon

Jordan

Start time:    Repetition:  ONCE  End time:    ☐ Download resources

# Conclusions and Future Work

## Implementation of Wizard for Focused Crawling

- Easy to use
- Extensible
- Puts crawling into the hands of researchers

## Current Work

- Re-crawling of Web archives to create topical sub-collections (submitted)
- Additional sources
  - Social Media
  - Directories (e.g. German newspapers)

# Temporal Analysis of Social Networks in Archives

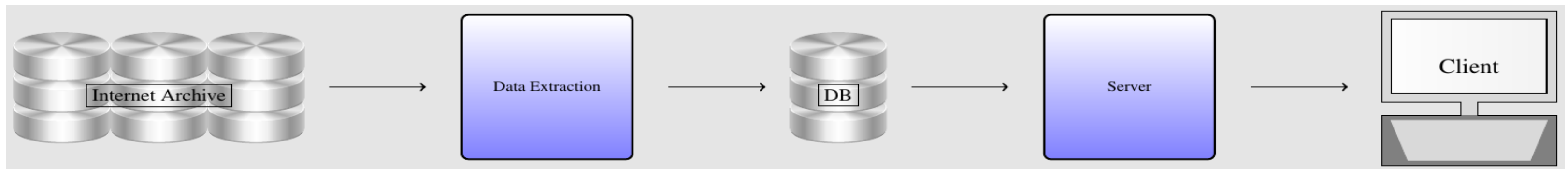
Miroslav Shaltev



# Temporal Analysis of Social Networks in Archives

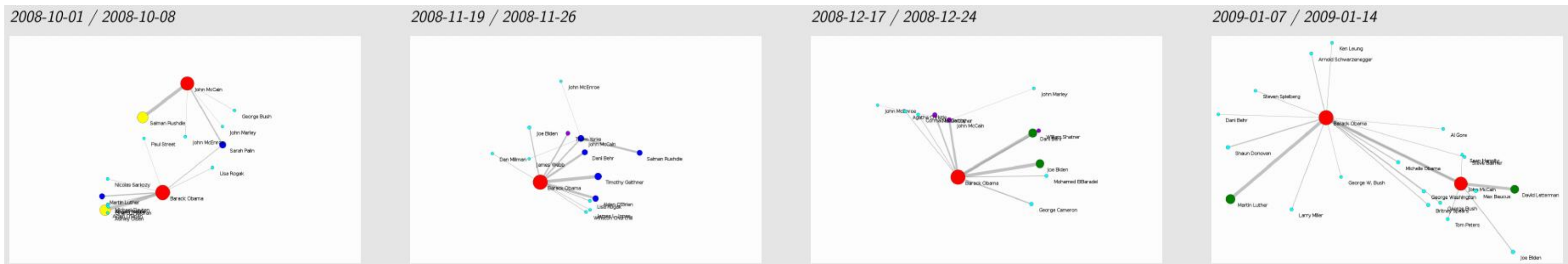
## Approach

- Background: computationally intensive cluster process.
- Foreground: server / client solution.



## Example

Temporal evolution of a graph constructed with seeds Barack Obama, John McCain from a small portion of the British archive data in the period October 2008 – October 2009.



# **Analysing Language-specific Differences in Multilingual Wikipedia**

Simon Gottschalk



WIKIPEDIA

# Ironworkers Memorial Second Narrows Crossing



Collapsed spans, August 1958



On June 17, 1958, as a crane stretched from the north side of the new bridge to join the two chords of the unfinished arch, several spans collapsed. Seventy-nine workers plunged 30 metres (100 ft) into the water. Eighteen were killed either instantly or shortly thereafter, possibly drowned by their heavy tool belts. A diver searching for bodies drowned later, bringing the total fatalities for the collapse to 19. In a subsequent Royal Commission inquiry, the bridge collapse was attributed to miscalculation by bridge engineers. A temporary arm, holding the fifth anchor span, was deemed too light to bear the weight.<sup>[2]</sup>

- ① Appearance
- ② Collapse
- ③ Official Opening

①

Die Länge beträgt 1292 Meter, die Spannweite des Mittelteils 335 Meter. Die Feldweiten der Brücke betragen 85,86 Meter – 2 x 85,91 Meter – 86,08 Meter – 142,24 Meter – 335,00 Meter – 142,09 Meter.

Darüber führt der [Trans-Canada Highway](#).

Unmittelbar östlich davon befindet sich die [Second Narrows Bridge](#), die heute eine reine Eisenbahnbrücke ist.

Die Bauarbeiten begannen im November 1957.

②

Am 17. Juni 1958 brachen mehrere Brückenbögen und 79 Arbeiter stürzten 30 Meter tief ins Wasser. Achtzehn von ihnen kamen ums Leben, da sie wegen ihrer schweren Werkzeuggürtel in die Tiefe gerissen wurden. Später ertrank auch ein Taucher, der nach den Leichen suchte.

Eine Untersuchungskommission kam zum Schluss, dass menschliches Versagen die Unfallursache gewesen war. Ein Ingenieur, der ebenfalls zu den Todesopfern gehörte, hatte eine Berechnung falsch durchgeführt und das Gewicht des Baumaterials unterschätzt. Als ein Baukran herumschwang, um ein Element einzusetzen, konnte die Brücke das zusätzliche Gewicht nicht mehr halten und stürzte ein.

③

Nach dem Unfall wurde die Brücke fertiggestellt und 25. August 1960 eröffnet.



# Tempas – Temporal Archive Search Based on Delicious and Twitter

Helge Holzmann



## Tempas

Tags / keywords  
+ URLs + times  
extracted from  
Delicious posts  
and Tweets



@user

2010-12-16

archive internet reference library video music  
media history web search <http://www.archive.org/index.php>



- Search a Web archive based on external (temporal) data sources (e.g., social networks)
- Custom index structure allows for temporal query suggestions and captures temporal relevance
- Links to Internet Archive's Wayback Machine
  - No need to index the entire archive
  - Naturally filters less important content
  - Search how a website is described in contrast to how it 'describes' itself

# **Suggesting News Articles for Enriching and Expanding Wikipedia Entity Pages**

Besnik Fetahu

# Automated News Suggestion to Wikipedia Entity Pages

## Task#2

feature extraction

one classifier per  
entity type

article section  
placement

entities

THE WALL STREET JOURNAL. INDIA

4:19 pm IST  
Oct 12, 2013 POLITICS

### As It Happened: Cyclone Reaches Orissa

Some half a million people were evacuated from the southeastern Indian coast as Cyclone Phailin, a tropical storm from the Bay of Bengal, bore down on India. The states of Orissa and Andhra Pradesh, both of which have large coastal populations, were on high alert ahead of the storm's expected arrival.

news article

[state]:geography

[city]:climate

article entity  
placement

✓ Odisha

✗ Bay of Bengal ✗ Phailin

## Task#1

### Odisha

Odisha (formerly known as Orissa <sup>[10]</sup> /oˈrɪsə/ [oʊˈrɪsə] the west and north-west and Pradesh to the i by the Mauryan Emperor Ashoka in 261 BCE) resultir speakers.<sup>[12]</sup> 1 April is therefore celebrated as Utkala of the state for over eight centuries until 13 April 1941 Odisha is the 9th largest state by area in India, and ti

Contents	[hide]
1 Etymology	
2 History	
3 Geography	
4 Districts	
5 Government and politics	
6 Economy	
6.1 Macro-economic trends	
6.2 Industrial growth	
6.3 Infrastructure development	
7 Transportation	
7.1 Air	
7.2 Seaports	
7.3 Railways	
8 Demographics	
9 Culture	
10 Notable people	
11 Literature	
12 Dance	
13 Music	
14 Other links	

wikipedia  
entity page



# ArchiveWeb: Working with Web Archive Collections

Trevor Fernando Zeon



# ArchiveWeb: working with web archive collections

Collaborative search and sharing platform


Building workable collections of resources

Sharing of collections between a group of people

User annotations for resources in the form of tags and comments

Flexibility to build your own personal archive of a resource

## Archive-It Interface



University of  
Texas Libraries  
THE UNIVERSITY OF TEXAS AT AUSTIN

### Human Rights Documentation Initiative

Collected by: [University of Texas at Austin Libraries, Human Rights Documentation Initiative](#)

Archived since: Apr, 2009

Description: The University of Texas Libraries' Human Rights Documentation Initiative Collection features fragile websites containing human rights documentation and related content from human rights organizations and advocates across the globe. Browse collection by topic, geographic area, language or do a keyword search at [http://www.lib.utexas.edu/hrdi/hr\\_archive](http://www.lib.utexas.edu/hrdi/hr_archive).

Subject: [Society & Culture](#), [Politics & Elections](#), [Blogs & Social Media](#)

### Narrow Your Results

Group

Sort By: **Count** | (A-Z)

Association africaine de Défense des droits de l'Homme (ASADHO) (2)

Cambodian Center for Human Rights (CCHR) (2)

Change for Equality (2)

Chinese Human Rights Defenders (2)

East Turkistan Information Center (2)

More ▾

Subject

Sort By: **Count** | (A-Z)

advocacy, activism, and responses to persecution--activists (54)

civil and political rights--women's rights (48)

armed conflict and persecution--torture (34)

advocacy, activism, and responses to persecution--non-governmental organizations (33)

laws, justice, and judicial proceedings--prisons and prisoners (29)

More ▾

Creator

Sort By: **Count** | (A-Z)

Change for Equality (3)

Association africaine de Défense des droits de l'Homme :: ASADHO (2)

Cambodian Center for Human Rights (CCHR) (2)

Chinese Human Rights Defenders (CHRD) (2)

Sites for this collection are listed below. Narrow your results at left, or enter a search query below to find a site, specific URL or to search the text of archived webpages.

Sites

Search Page Text

◀ Prev Page

Page 2 of 3 (266 Total Results)

Next Page ▶

Sort By: **Title (A-Z)** | Title (Z-A) | URL (A-Z) | URL (Z-A)

Title: La Strada Moldova

URL: <http://lastrada.md/>

Description: In Moldovan and some English. Focuses on women's rights, domestic violence, and human trafficking. Government continues to allow and in some cases condone discrimination based on ethnicity, gender, and religion; in some regions NGOs are routinely threatened and harassed. Captured 32 times between Apr 29, 2009 and Apr 9, 2015

Subject: [civil and political rights--women's rights](#), [civil and political rights--discrimination](#), [gender](#), [armed conflict and persecution--rape and sexual abuse](#), [economic, social, and cultural rights--economic and labor rights--human trafficking](#), [armed conflict and persecution--domestic violence](#)

Creator: [La Strada Moldova](#)

Publisher: [La Strada Moldova](#)

Source: <http://lastrada.md>

Language: [English](#); [Moldovan](#)

Coverage: [Europe--Moldova](#)

Identifier: [http://wayback.archive-it.org/1475/\\*http://www.lastrada.md](http://wayback.archive-it.org/1475/*http://www.lastrada.md)

Collector: [UT Libraries](#)

Title: Ligue des Droits de la Personne dans la Région des Grands Lacs / League for the Defense of Human Rights of the Great Lakes

URL: <http://ldgl.org/>

Description: In French and some English. Human rights umbrella group, with members from

Resource



# ArchiveWeb Interface

## Collections

## Resource Preview

ArchiveWeb

My Resources

Group overview

Profile

Group overview

[CaS] Cultures

[HRDI] Human Rights

Overview

Resources

Members

Presentations

Links

[CaS] Museums

Resources of Human Rights

Overview

Resources

Members

Presentations

Links

Upload Resource

La Strada Moldova

Text added by ArchiveDemo

Ligue des Droits de la P...

Text added by ArchiveDemo

Cambodia League for th...

Text added by ArchiveDemo

International Committe...

Text added by ArchiveDemo

Memorial: International ...

Text added by ArchiveDemo

Helsinki Committee for ...

Text added by ArchiveDemo

Moscow Helsinki Group

Text added by ArchiveDemo

Movimento Nacional de ...

Text added by ArchiveDemo

Movimento Nacional de ...

Text added by ArchiveDemo

Movimento dos Trabalh...

Text added by ArchiveDemo

NGO Coordination Com...

Text added by ArchiveDemo

Office of United Nations...

Text added by ArchiveDemo

Den Norske Helsingfors...

Text added by ArchiveDemo

NISGUA: Network in Soli...

Text added by ArchiveDemo

None on Record: Stories...

Text added by ArchiveDemo

Omo "Iliniden" - Pirin: M...

Text added by ArchiveDemo

< Previous Page

1

2

3

4

5

Next Page >

La Strada Moldova

Source: Internet

Group: Human Rights

Description

In Moldovan and some English. Focuses on women's rights, domestic violence, and human trafficking. Government continues to allow and in some cases condone discrimination based on ethnicity, gender, and religion; in some regions NGOs are routinely threatened and harassed.

Current Version:

Archive Now

View Archived Versions:

List view

Timeline View

Tags

Civil and Political Rights

Cultural Rights

Armed Conflicts

Add Tag

Comments

ArchiveDemo

Collected by University of Texas Libraries

Edit Delete

Comment

User Annotations

## User Annotations

## User Annotations

# WikipEvent: How we get events from the crowd?

Binh Giang Tran



## Our Target and Solution

- **Tracking the evolution of event from Wikipedia**
  - *How Michael Phelps becomes the holder of the record of most medals ever?*
  - *How was “the death of Osama Bin Laden”*
- **Finding similar events by event similarity measurement**
  - *Events are collected from Wikipedia revisions and WCEP*
  - *Represent an event (text, entities, references) as a high-dimensional vector*
  - *Vector similarity measurement: Cosine / Jaccard / ...*
- **Improvement**
  - *Would like to use hashing technique for a fast and effective implementation (Mitzenmacher et al. 2014)*

# Information overload problem

- A thousand of daily reported events
- Tons of daily news articles
- A lot of information sources:
  - *Mainstreams*
  - *Social Media*
  - *Blogs*
  - ...



According to **Chartbeat**, over **92,000** articles are posted to the web every 24 hours.

Source: [A Day in the Life of Web Content](#)

# A Bing-inspired Ranking for Entity-centric Web Archive Search

Tu Nguyen



[Angela Merkel](#) Translate this page

[www.angela-merkel.de](http://www.angela-merkel.de)

Die persönliche Internetseite der Vorsitzenden der CDU Deutschlands, **Angela Merkel**.

Politik

Die persönliche Internetseite der  
Vorsitzenden der CDU ...

See results only from [angela-merkel.de](http://angela-merkel.de)

Images of **angela merkel**

[bing.com/images](http://bing.com/images)



See more images of **angela merkel**

[Angela Merkel - SPIEGEL ONLINE](#) Translate this page

[www.spiegel.de](http://www.spiegel.de) › Politik › Deutschland

Deutschlands erste Kanzlerin. Früher war sie "Kohls Mädchen", heute ist sie die mächtigste Frau Deutschlands: **Angela Merkel** hat sich an der Männerriege ...

[Angela Merkel - Bundeskanzlerin - News ...](#) Translate this page

[www.welt.de](http://www.welt.de) › Politik › Deutschland › Themen

**Angela Merkel** im Themenspecial. "Die Welt" bietet Ihnen aktuelle News und Hintergründe über die CDU-Politikerin und Bundeskanzlerin **Angela Merkel**. **Angela** ...

News about **Angela Merkel**

[bing.com/news](http://bing.com/news)



**Angela Merkel** ist verschwunden! Mit dieser Vermisstenanzeige wird nach ihr gefahndet

[The Huffington Post](#) · 4 hours ago

Es gibt weiterhin keine Neuigkeiten im Falle der vermissten **Angela Merkel**. Die Satiriker von Extra3 haben daher begonnen bei ...

HORIZONT-Auflagencheck "Eiskönigin" **Angela Merkel** pusht Kioskverkauf des "Stern"

[Horizont](#) · 7 hours ago

Politik **Angela Merkel** schweigt, Schäuble wirbt

[Südkurier](#) · 3 hours ago

[Angela Merkel – Wikipedia](#) Translate this page

[https://de.wikipedia.org/wiki/Angela\\_Merkel](https://de.wikipedia.org/wiki/Angela_Merkel)

- Search Engines provide a good quality ranking for general web search.
- Web archive search and general web search have similar information needs for some types of queries.
- There is no state-of-the-art ranking model for web archive in literature.
- Lack of user feedbacks, what are the real user intents?

