

Hungarian Media Monitoring and Text Analysis with the Europe Media Monitor (EMM)

Ralf Steinberger
European Commission – Joint Research Centre (JRC)
http://ipsc.jrc.ec.europa.eu/?id=179
EMM: http://emm.newsbrief.eu/overview.html

Hungarian Academy of Sciences, Budapest, 18 January 2013



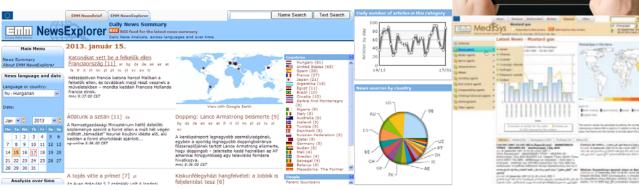
Agenda



- JRC: Who we are what we do our customers.
- Europe Media Monitor (EMM) family of applications
 - Publicly accessible at http://emm.newsbrief.eu/overview.html
- Motivation for multilingual text processing
- Adapting EMM to Hungarian Challenges
- Hungarian EuroVoc thesaurus indexing

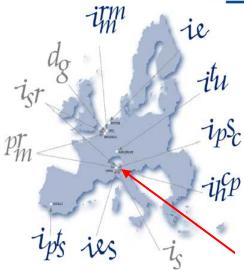






JRC - Who we are





- **European Commission** (scientific-technical arm of public administration)
- Non-commercial
- Multi-disciplinary / multilingual

BRUSSELS (BE)

The Directorate General (DG)

The Institutional and Scientific Relations Directorate (ISR)

The Programme and Resource Management Directorate (PRM)

The Institute for Reference Materials and Measurements (IRMM)

The Institute for Transuranium Elements (ITU)

ISPRA (IT) Download the Ispra site Brochure (English - Italian) The Institute for the Protection and Security of the Citizen (IPSC) The Institute for Environment and Sustainability (IES) The Institute for Health and Consumer Protection (IHCP) The Ispra site Directorate (IS)

PETTEN (NL)

The Institute for Energy (IE)

SEVILLE (E)

The Institute for Prospective Technological Studies (IPTS)



EMM users





EMM media monitoring users - wide coverage, world-wide

European Commission (most DGs) and other EU Institutions **EU Agencies**:



- e.g. Public Health (ECDC), Food Safety (EFSA), Chemicals Bureau (ECHA), etc.
- EU Member State organisations: e.g.
 - Public Health,
 - law enforcement authorities,
 - parliaments,
 - crisis management/humanitarian

International and extra-European organisations: e.g.

- various UN organisations
- Centres for Disease Prevention and Control in the US, Canada, China,

The public:

- ~30,000 anonymous internet users of publicly accessible EMM systems.
- Combined between 1 and 2 Million hits per day









Europe Media Monitor (EMM) news gathering - A few facts

- ~ 175,000 articles / day in over 70 languages (>40 active)
- ~ 4000 Sources (world-wide, with focus on Europe)
 - news sources (web portals)
 - · specialist medical sites
 - ~ 20 commercial newswires
 - 24/7, updated every 10 minutes



Articles are fed into the various EMM applications:



Steinberger Ralf, Bruno Pouliquen & Erik van der Goot (2009). An introduction to the Europe Media Monitor Family of Applications. In: Fredric Gey, Noriko Kando & Jussi Karlgren (eds.): Information Access in a Multilingual World - Proceedings of the SIGIR 2009 Workshop (SIGIR-CLIR'2009), pp. 1-8. Boston, USA. 23 July 2009.

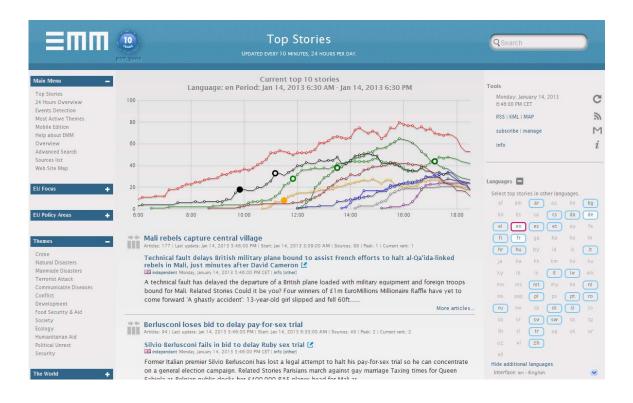


EMM functionality



Monitoring live news

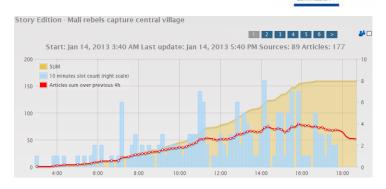
Clustering of related news; showing biggest current news clusters;



EMM functionality



Cluster information





- · For selected news cluster, show:
 - Show short-term timeline,
 - mentioned countries and persons

Mali [100%]
France [89%]

Names

Jean-Yves Le Drian [56%]

François Hollande [53%]

Laurent Fabius [24%]

- Related news categories
- •

Categories

Conflict [77%]

Terrorist Attack [68%]

UN Bodies [50%]

EMM functionality



News categorisation

- Categorise the news into over 1,000 news categories
 - Some generic, e.g. 'Themes'
 - Many customer-specific classes.
- Same categories for all languages,
 - allowing joint statistics for all languages
 - Allows alerts for important events even if reported in other languages

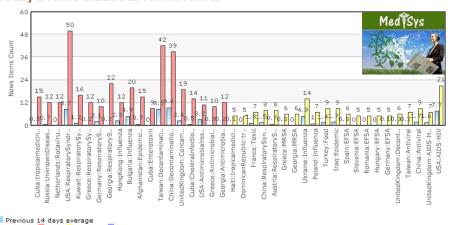
Top Stories 24 Hours Overview Events Detection Most Active Themes Mobile Edition Help about EMM Overview Advanced Search Sources list Web Site Map EU Focus EU Policy Areas Themes Crime

Manmade Disasters
Terrorist Attack
Communicable Diseases
Conflict
Development
Food Security & Aid

Ecology Humanitarian Aid

Political Unrest

Today's Alert Statistics for All Alerts



The World

Offices & Agencies

EMM functionality



Statistics on categories

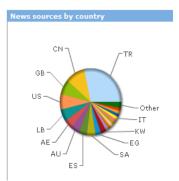


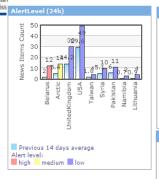
Renewable energy sector has lots of spark left, say Experts

Fath Birol said: "The main challenge from the economic crisis is over but full recovery is threatened by the lack of equilibrium between the cost and supply of oil," "High oil prices are a key risk to the fragile economic recovery among developed nations" khaleejtimes Friday, January 21, 2011 6:02:00 AM CET Fath Birol said: "Government subsidies of energy fuels leads to inefficiencies and waste through artificially high use," "Coupled with the current gas glut this poses a real policy reforms being formulated and implemented by the UAE government need to be commended."

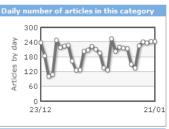
6:02.00 AM CET

Nicholas Stern said: "In order that climate change targets can be









United Nations Protection Agency Environmental Protection Agency Nature Barack Obama RIA Novosti Europa Press Dmitry Medvedev World Bank

sondakika haberaktuel stargundem haber7 yenisafak webhaber netgazete stargazete crisisgroup millivet

China Daily







Daily overview



News Summary About EMM NewsExplore News language and date Language or country: en - English Languages ar - Arabic bg - български

es - Español de - Deutsch da - Dansk

en - English et - Eesti keel fa - Farsi

fr - Français it - Italiano

nl - Nederlands no - Norsk pl - Polski pt - Português

ro - Română ru - Russian sI - Slovenščina sv - Svenska

tr - Türkçe sw - Kiswahili Countries

AT - Austria BE - Belgium DE - Germany ES - Spain

FR - France GB - United Kingdom IT - Italy

View with Google Earth

No UK army boots in Mali, says PM [24] ar bg da de es fa fr it nl no pl pt ru sl sv sw tr hu

British troops will not go onto the ground in Mali, the prime minister says, despite the UK offering logistical support to France in the West African country.

bbc 4:51:00 PM CET

Residents in Beijing warned to stay indoors as air quality hits 'worst on record' after city is engulfed in thick smog [24] ar da de es it nl pt sl sv tr hu ru

The Chinese capital has seen pollution rise to 30 to 45 times above the recommended safety levels after being engulfed in a thick smog since Friday. dailymail 4:22:00 AM CET

World's biggest religious festival [14] da de fr it nl pt

IOA intends to build road isle in Khader town without Palestinian consent [13] ar fr pt tr

The Israeli occupation authority (IOA) intends to establish a big road isle at the crossroads of a main street in Al-Khader town west of Bethlehem city without Palestinian consent.

United States (291)
United Kingdom (154)
Pakistan (108)
India (101)
France (51)
Canada (49)
Russian Federation (48) Canada (49)
Russian Federation (48)
Armenia (41)
Japan (36)
Italy (28)
Iran, Islamic Republic Of (27)
Greece (24)
Turkey (23)
Azerbaijan (22)
Palestinian Territory,
Occupied (21)
Korea, Democratic
People's Republic Of (21)
South Africa (21)
China (20)
Israel (20)
Libyan Arab Jamahiriya
(20)

(20) Philippines (20) François Hollande David Cameron Jean-Yves Le Drian Barack Obama Bashar Assad Hosni Mubarak Hosni Mubarak
Christoph Waltz
Anne Hathaway
Jodie Foster
Lance Armstrong
Hugh Jackman
Daniel Day-Lewis
Kathryn Bigelow
Damian Lewis
Steve Webb
Joseph Bidan
Gail-Islam al-Gaddafi
Osama bin Laden
Benjamin Netanyahu
Jennifer Lawrence

Jennifer Lawrence

January 8, 2013 - January 14, If you got the flu vaccine, are you covered? January 9, 2013 - January 13, January 9, 2013 - January 1: 2013 French forces begin military intervention in Mali January 10, 2013 - January 11, 2013 Savile report to show victims were 'as young as 10' January 11, 2013 - January 12, 2013 Petertive describes Colo Detective describes Colo. Detective describes Colo. shooting suspert's odd police station behavior January 7, 2013 - January 12, 2013 - Gram donor to Barretstown revealed January 8, 2013 - January 13, 2013 US to review Boeing 787 Read more...

This Month's New Stories
French president says Mali air
strikes necessary
December 24, 2012 January
14, 2013
Connecticut school shooting:
Adam Lanza shot each child
between three and 11 times
December 14, 2012
December 25, 2012
President of Iraq suffers
stroke stroke
December 15, 2012 - January
12, 2013
Northern Ireland: Draped in
Union flags, protesters march
in Belfast
December 30, 2012 - January
13, 2013
Barcelona coach 17 ia coach Vilanova has Barcelona coach Vilanova has 2nd throat operation December 18, 2012 - January 10, 2013

Pakistan court dismisses final Musharraf challenge de es fr it nl ar bg da pl pt ru

Pakistan's new-look Supreme Court has, as expected, dismissed the last challenge to President Pervez

Emergency rule 'destroying the judiciary

Exiled ex-PM Sharif plots return to Pakistan usaToday 10:58:00 PM CET

'Bush unembarrassed by Pakistan emergency' dailytimesPK 12:28:00 AM CET

Pakistan's Commonwealth suspension in sorrow,

PM in Uganda for Commonwealth summit TorontoStar 5:40:00 PM CET

Commonwealth to drop Pakistan guardian 1:03:00 PM CET

not anger: Britain (AFP) news-yahoo 11:34:00 PM CET Pakistan: Emergency and chaos

Pakistan Court Rules for Musharraf ABCnews 2:43:00 PM CET

BangkokPost 8:26:00 PM CET

Pakistan (26) United States (13) 146 Saudi Arabia (10)

Islamabad(PK) Rawalpindi(PK) Karachi(PK) Washington(US) Ar Riyad(SA) Jiddah(SA)

Pervez Musharraf (26)

Pervez Musharraf (26)
Imran Khan (5)
George W. Bush (5)
Gordon Brown (4)
Don McKinnon (4)
Nawaz Sharif (4)
Benazir Bhutto (3)
Malik Mohammad Qayyum (3)
Iftikhar Muhammad Chaudhry
(2)

(2)
Altzaz Ahsan (2)
Ahsan Iqbal (2)
Thomas Jefferson (2)
Asma Jahanqir (2)
David Cameron (2)
Louise Arbour (1)
Qazi Hussain Ahmed (:
Stephen Harper (1)
Najam Sethi (1)
Khalid Hassan (1)
Hina Jilani (1)

Hina Jilani (1) Wajihuddin Ahmed (1) John Negroponte (1) Rashid Qureshi (1) Jemima Khan (1) Benazir Bhutto (1)

Other Names

Supreme Court (27) Human Rights Watch (3) Pakistan Muslim League (2) Pakistan Musiim League (2) High Court (2) High Commission (2) GEO Television (2) Human Rights Commission (1) Al Qaeda (1)

Stories consist of time-linked news clusters with overlapping keywords. eywords: Pakistan, Afghanistan, United states / Pervez Musharraf, Supreme Court / pakistani, Ilamabad, bhutto, military, government, army Importance: 2833 articles in 125 clusters Start date: Saturday, June 23, 2007 End date: Thursday, November 8, 2007 Bhutto to fight on despite Karachi attac

This cluster belong to the following story: Pakistan defends emergency rule

islamabad, bhutto, military, government, army Start date: Saturday, June 23, 2007 7 days before. Musharraf unveils new interim PM similarity: 0.87

Tendulkar sets up series win over Pakistan similarity: 0.34 6 days before. Musharraf swears in caretaker cabinet similarity: 0.83 5 days before.

US envoy meets Musharraf similarity: a.85 Pakistan-India tennis series: Pakistan win second leg to level series 1-1 similarity: 0.34

4 days before. US tells Musharraf to step back similarity: 0.86

EMM NewsExplorer



Name variants (found in 16h!)

2190. Muammar Gaddafi

2011-03-24T23:59+0200

Muammar Gaddafi / Muammar al-Gaddafi / Mouammar / معمر القذافي Kadhafi / Moammar Gadhafi / Myamapa Каддафи / Муаммара Каддафи / Muamar Gadafi / Muammar Gheddafi / 卡扎菲 / Muamar al Gadafi / Muamar el Gadafi / Muammara Kadafiego / Muammer Kaddafi / Muammar al-Ghadhafi / Muammar Kadhafi / Муамар Кадафи / Muammara Kaddáfí / Moamer Kadhafi / Muammar al Gaddafi / Moamer Gaddafi /

Muammar al Gaddafi / Moamer Gaddafi / Muammar Gaddafin / Moammar Kadhafi / Muammar el Gaddafi / Muammar Qaddafi / Moameria Gadafiia / Муамару Каддафи / Muammar Kaddafi / Muamaro Gaddafi / Муаммару Каддафи / Muamar Kadafi / Муамар Каддафи / Moammar Qaddafi / Moamar Gadafi / Muammar Kaddáfí / Moamer Gadafi / أمعمر / Muammar Kadafi / Muammar Khadafi / القذافي Muammar Gadhafi / Muammar al Gadafi /



Муаммар Каддафи / Muammar Gadafy / Moammer Kadhafi / Muammar el-Qaddafi / Muamar Kadhafi / معمر / Moammar Kadafi Муаммаром / قَذَافَي Каддафи / Moammar Gaddafi / Muammar Khaddafi / Muamaro Kadhafi / Muamar Khadafi / Muammar Khadaffi / Muammer Gaddafi / Muammaro Gaddafi / Moamar el Kadafi / Muhammar Gheddafi / Mouammar al-Kadhafi / Muammar Kadhaffi / Muamar El Gadafi / Moamar

Gaddafi / Muammar el-Gaddafi / Muammar al-Gadafi / Mouammar Kaddafi / Muammarem Kaddáfí / Muammar Gadafi / Muammar al-Gadhafi / Muammar al-معمر القذافي / Gaddafi Muammar al-Qadhafi / Muamarui Gaddafi / Muammar Gadaffi / Muamaro Kadafio / Muamar Gaddafi / Muammar Qadhafi / Muammaro Gaddafio / Muamarui Gaddafiui / Muammar al-Gadaffi / Muamar al Gaddafi / Муамаром Каддафи / Muammar Ghadhafi /

live

Moamerju Gadafiju / Muammarowi Kadafiemu / Moammar Khadafy / Moammer Gaddafi / Muammar El / מועמר קדאפי / Gadaffi / Mouamar Kadhafi / Mouammar El Kadhafi / Muammar Al-Gaddafi / Mummar Gaddafi / Muamar Gadaffi / Moammar Ghadafi

EMM NewsExplorer



Multilingual Info about people

Pervez Musharraf

Information about this person was last updated on Friday, November 23, 2007.

Pervez Musharraf (Eu,sv) General Pervez Musharraf (da,sv) Gen Musharraf (en) Pervez Mušaraf (sl)

Gen Pervez Musharraf (en) Pervez Musharrafs (da.sv) Pervez Musharaf (da,sv) Первез Мушарраф (ru) Perwez Musharraf (fr,sv) Pervez Moucharraf (fr)

(ar) برویز مشرف Perveza Mušarafa (sl) Pervez Muscharraf (de) Perves Muscharraf (de) pakistani president (en - 827)

president (de.sv - 3230) president gen (en - 506) président pakistanais (fr - 398) pakistaanse president (nl - 235) presidente paquistanês (pt -277)

pakistani president gen (en -181)

presidente paquistaní (es - 147) präsident (de - 617) general (en,sv - 450) præsident (da - 210) presidente (es,pt - 589) president, gen (en - 62)



Co Al Qaeda (855)
Wa Supreme Court (505)
Aft White House (312)
(20 NATO (207)
Ayı GEO Television (197)
Im Lal Masjid (187)
Abı Tautas Partija (176)

at Daily Times (153)
sociated People Salman Bashir (1.9) Джон Негропонте (1.5) Nawaz Sharif (1.2)

Раджа Омар Хатаб (1.2) Мухоммада Али Джинны (1.2) Зульфикара Али Бхутто (1.2) Iftikhar Muhammad Chaudhr

Benazir Bhutto (910)

Co Al Qaeda (855)

Nawaz Sharif (552) George W. Bush (537)

Iftikhar Muhammad Chaudhry Osama bin Laden (430) Shaukat Aziz (395) Tario Azeem (251) Ha Other Names

51)

(1.1) Ch Ta Related Stori Be Pakistan defends emergency

Ma rule (1. June 23, 2007 - November 22, Ch 2007

(1. Pakistan's president urges calm Ful May 5, 2007 - June 23, 2007 Ja\ Toops launch hostage rescue

Sh bid

who challenged his power-grab, by keer Ra 42 die in bomb judge's family under house arrest, Oa 'lucky' weekend July 15, 2007 - October 25,

Demonstrations at UK embassy

[-has repeated]: The (presidential) oath can be taken ... by the weekend or immediately

omatically

▲ dia

Ab June 23, 2007 - October 25, Arr 2007 Arr Protests in Pakietan take ain

Musharraf 't Quotes from - English

thereafter.

 $[de] \ [pt] \ [es] \ [nl] \ [fr] \ [ar] \ [sv] \ [it] \ [da] \ [pl] \ [sl] \ [ro] \ [ru] \ [no] \ [bg]$

Pakistan court dismisses final Musharraf challenge Struggle to help cyclone surivors in Bangladesh cnn 21-NOV-07

 $[es] \ [ru] \ [sv] \ [pt] \ [nl] \ [de] \ [fr] \ [no] \ [bg] \ [it]$ [-has said]: Where it fails to live up to those values, it needs to act with credibility and

consistency. I think Pakistan is a test in that duotes about - Er

[said]: The m Rice [-said]: And look, a lot of that was done by instead, (Pervez) Musharraf himself. And so for him at this Musharraf is punishing Chief Justice Chamber Musharraf who challenged his power-arch by least 10 mush at the least point to help put his country back on the road to democratic reform is important. We're looking for him to take off his uniform, expressindia 22-NOV-07

Brad Adams [said]: Rather than making

Brown [-said]: He (Musharraf) has assured in Iran

HumanRightsWatch 22-NOV-07

EMM NewsExplorer



Relation exploration

Salman Bashir (1.9) Джон Негропонте (1.5)

джон негропоне (1.0)
Nawaz Sharif (1.2)
Раджа Омар Хатаб (1.2)
Мухоммада Али Джинны (1.2)
Зульфикара Али Бхутто (1.2)
Iftikhar Muhammad Chaudhry

(1.1)

Christian College (1.1) Tariq Azeem (1.1) Benazir Bhutto (1.1) Malik Mohammad Qayyum

(1.0)Chaudhry Shujaat Hussain (1.0) Furqan Bahadur (1.0)

Javed Cheema (0.9) Shaukat Aziz (0.9) Abdul Rashid Ghazi (0.9) Amir Mir Lahore (0.9)

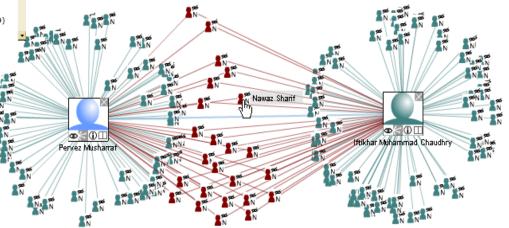
Amin Fahim (0.9) Гордон Джонроу (0.9) Wajihuddin Ahmed (0.9) Mohammed Ali Durrani (0.9) Rashid Qureshi (0.9) Oazi Hussain Ahmed (0.9)

Example: Pervez Musharraf & Iftikhar Chaudhry





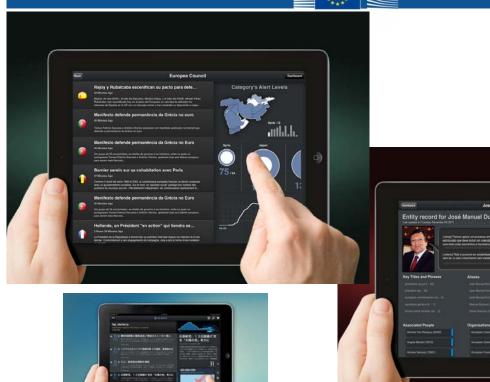




EMM mobile version



Coming soon. Free



EMM



Multilinguality

Complementary coverage of medical news in various languages



Italian - German





English - French





Spanish - Portuguese



EMM Multilinguality

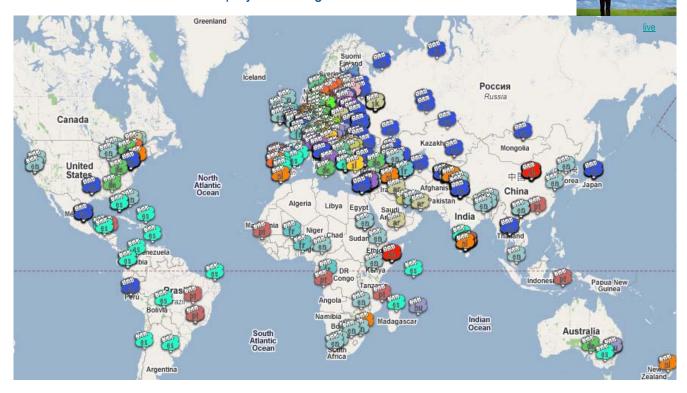


Complementarity of info.

NewsBrief

NewsBrief Live Cluster Map

Display of latest geo-located news clusters





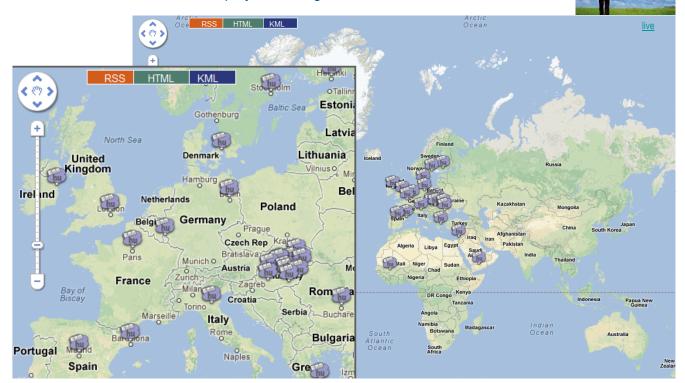


Hungarian news coverage

NewsBrief

NewsBrief Live Cluster Map

Display of latest geo-located news clusters





Adding Hungarian

- Initially, restricted coverage of Hungarian (gathering, clustering some classes in NewsBrief only)
- Very fruitful and pleasant collaboration with competent colleagues from the Research Institute for Linguistics at the Hungarian Academy of Sciences
 - Adding many Hungarian news sources







• Development of text mining tools for information extraction



101 Hungarian



news sources

168 Ora	km-168-Ora-HU
A baranya megyei online	bama
ATV	km-ATV-HU
BOON	boon
Blikk	blikk
Budapest Online	BudapestOnline
Bumm Szlovákiai magyar	bumm
Békéscsaba	bekescsaba
Duna TV	km-Duna-TV-HU
Délmagyarország	delmagyar
Délmagyarország Online	tv2-hu
EU vonal	euvonal
Echo TV	km-Echo-TV-HU
Erdély Ma	erdely
FELVIDÉK.ma	felvidek
FN	fn
Fejér Megyei Hírlap	fmh
Fejér Megyei Hírlap	km-Fejer-Megyei
Figyelo	km-Figyelo-HU
Figyelő	figyelo
GYŐR ONLINE	gyorn
HVG	km-HVG-HU
Hargita Népe	hnepe
Heti Világgazdaság	hvg
Heti Válasz	km-Heti-Valasz
Heti Válasz	hetivalasz
Heves Megyei Hirlap Online	hevesmegyeihirlap
HirTV	km-HirTV-HU
Hirszerzo	hirszerzo
Hungarian Radio News	klubradio
Hungarian news	baon
Hungarian news	dunaujvaros
Hungarian news	dunatv
Hungarian portal Debrecen	dip-hu
Háromszék	3szek
Hír TV	hirty
Hírek és szórakozás - Hir24	hir24
la dans	te dess

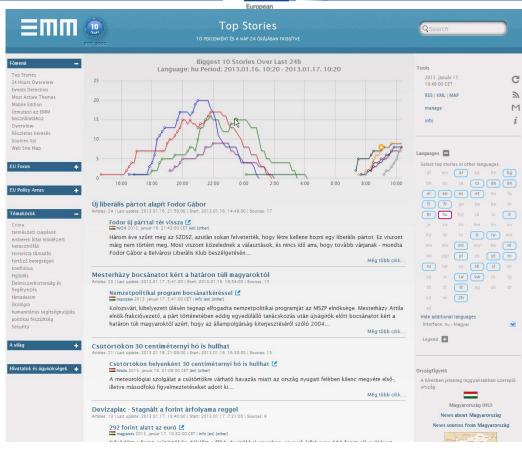
IIIORadio	KIII-IIII OKAGIO-NO
Kecskeméti Televízió	kecskemetitv
Kisalföld	km-Kisalfold-HU
KlubRádió	km-KlubRadio-HU
Kossuth Radio	km-Kossuth-Ra
Krónika.ro	kronika
Lánchíd Rádió	km-Lanchid-Rad
MNO Magyar Nemzet rovat	magyar_nemzet
Magyar ATV	magyaratv
Magyar Demokrata	demokrata
Magyar Hírlap	magyarhirlap
Magyar Hírlap	km-Magyar-Hirla
Magyar Narancs	km-Magyar-Nar
Magyar Nemzet	km-Magyar-Nem
Magyar Nemzet Online	mno
Magyar Rádió	MagyarRadio
Magyar Televizio	km-Magyar-Tele
Metropol	km-Metropol-HU
Napi Gazdaság	napi
Nyugati Jelen	nyugatijelen
NÉPSZAVA	nepszava
Népszabadság	km-Nepszabads
Népszabadság	nol
Népszava	km-Nepszava-HU
Népújság	e-nepujsag
Origo	origo
Paraméter	parameter
Portfolio.hu Online Gazdas	portfolio-hu
RTL Klub	km-RTL-Klub-HU
RTL klub	rtlklub
Rádió Q	km-Radio-Q-HU
Somogyi Hírlap	somogyihirlap
Szabadság	szabadsag
Szabolcs Online	szabolcs
Szatmári Magyar Hírlap	hirlap
Székelyhon	szekelyhon
Sólyom: az előre hozott vá	hirado
TV2 (Budapest)	km-TV2-Budape

TV2 (Hungary)	km-TV2-Hungar
The Voice of Russia	ruvr-hu
TriplexRegioNet	triplexregio
Vagy	vagy
Vajdaság MA	vajma
Vas Népe	vasnepe
Vasárnapi Hírek	km-Vasarnapi-H
Világgazdaság	km-Vilaggazdas
Világgazdaság Online	vg-online
Zalai Hírlap	zalaihirlap
hvg.hu	km-hvg.hu-HU
index.hu	km-index.hu-HU
karpatalja	karpatalja
kemma	kemma
kisalfold.hu - A címoldalon	kisalfold
msn hu	msn-hu
naplo legyen	naplo
naplo legyen nol.hu	naplo km-nol.hu-HU
nol.hu	km-nol.hu-HU
nol.hu prigo.hu	km-nol.hu-HU km-origo.hu-HU
nol.hu origo.hu oortfolio.hu	km-nol.hu-HU km-origo.hu-HU km-portfolio.hu-HU
nol.hu origo.hu oortfolio.hu (eo)	km-nol.hu-HU km-origo.hu-HU km-portfolio.hu-HU teol
nol.hu prigo.hu portfolio.hu deol transindex	km-nol.hu-HU km-origo.hu-HU km-portfolio.hu-HU teol transindex

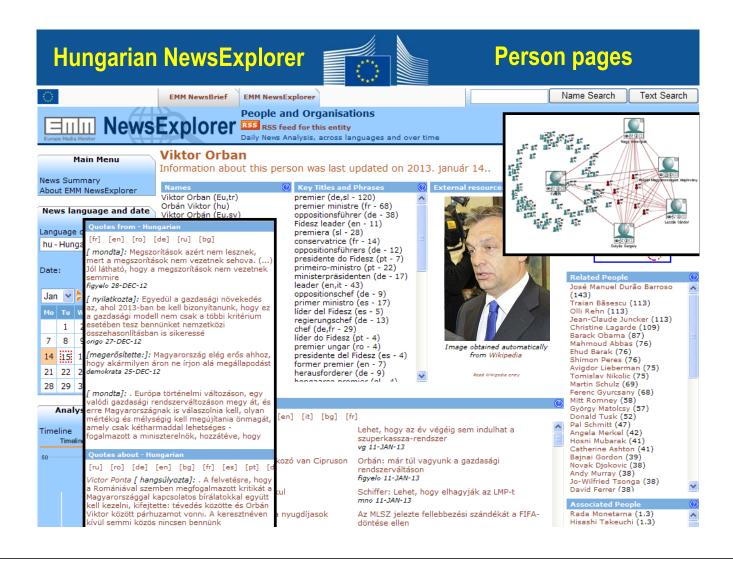
EMM-NewsBrief



Hungarian page







Adding Hungarian to EMM



Challenges

- 21st NewsExplorer language Hungarian was difficult to add.
- Challenges: mostly related to rich Hungarian morphology: (We have no access to morphological analyser software)
 - 1. Recognition patterns (many more title words need to be found)
 - EN: diplomat Uppercase1 Uppercase2
 - HU: Uppercase1 Uppercase2 diplomata/diplomatát/diplomatával/...
 - 2. Lookup of known names (e.g. Barack Obama):
 - Barack Obamával
 - Barack Obamának

Barack Obamáról

How to recognise these?

- 3. De-inflection of new names:
 - Paizs Júliával
 - Pajzs Júliának

Pajzs Júliáról

What is the base form?



Languages

ar - Arabic

bg - български es - Español

de - Deutsch

da - Dansk

en - English

et - Eesti keel fr - Français

fa - Farsi

hu - Hungarian it - Italiano

nl - Nederlands

no - Norsk

pl - Polski

pt - Português

ro - Română

ru - Russian

sl - Slovenščina

sv - Svenska

tr - Türkçe

sw - Kiswahili





A great collaboration Thank you

to Tamás Váradi, Júlia Pajzs and colleagues.



JOINT RESEARCH CENTRE

The European Commission's in-house science service

EuroVoc Indexing (JEX)



Hungarian performance

(Unrelated) work on automatic categorisation of legal documents in 22 EU languages:

• Using the controlled vocabulary from **EuroVoc** (>6,000 classes)

http://eurovoc.europa.eu/)

- Used by parliamentary libraries and more.
- JEX tool available for free download.
- 2.2 Az ilyen típusú szennyezésnek esetlegesen kitett ivóvízrendszerek esetében kötelező széles körű ellenőrzést végrehajtani a bennük található ivóvíz biztonságosságának biztosítása érdekében. Számos olyan régió is található azonban Európában, ahol a természetben előforduló radioaktív anyagok jelenléte ad okot aggodalomra.

EUroVoc Multilingual Thesaurus of the European Union

2.3 A lakosság egészségének az ivóvízben található radioaktív anyagokkal szembeni védelmére vonatkozó, uniós szintű technikai előírásokat már több

mint öt éve, az Euratom-Sz az ivóvízről szóló irányelvv Szerződés 35. és 36. cikkél bizottság bevonásával véglegesítették. A trícium water pollution vonatkozó, az emberi fogy irányelv szerinti előírások melléklet (ellenőrző rendsze vonatkozó előírások) módos

EUROVOC descriptors:

health policy food consumption radioactive materials

Steinberger Ralf, Mohamed Ebrahim & Marco Turchi (2012). **JRC EuroVoc Indexer JEX - A freely available multi-label categorisation tool.** Proceedings of the 8th international conference on Language Resources and Evaluation (**LREC'2012**), pp. 798-805, Istanbul, 21-27 May 2012.

EuroVoc Indexing



