





Complementarity of information found in media reports across different countries and languages



Ralf Steinberger

& the JRC's *OPTIMA* team – Open Source Text Information Mining and Analysis

Technical details and publications: Applications:

http://langtech.jrc.ec.europa.eu/ http://emm.newbrief.eu/overview.html



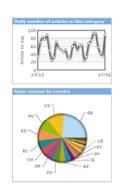
Agenda

Itilingual Web Workshop, Pisa, Italy, 4 April 2011

- 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
- · How to get access to this complementary information
 - Multilingual category definitions and alerts
 - Linking of related news across languages
 - · Multilingual information gathering on named entities
 - · Multilingual event scenario template filling
- Ongoing work & Summary







Latest News About - Climate Change

rticles published more than 10 minutes ago El Gobierno de Aragón pone en marcha el Manual de Buenas Prácticas Hadional hadional de marcha el Manual de Buenas Prácticas

que Monday, January 17, 2011 419100 PM CET | Info ©
Los Departamentos de Otras Públicas, Urbanismo y Transportes y de Medio Ambiente del
Gobierno de Aragón han puesto en marcha el Manual de Buenas Prácticas Medioambientales,
cuyo objetivo es conseguir mejoras en la integración de los aspectos medioambientales en las
obras relacionadas con las careteras de bitulandad de la Comunidad autónomia....

Articles published more than 20 minutes ago

Ban calls for global revolution to ensure clean energy for all unlorg Monday, January 17, 2011 4:11:00 PM CET | Info ©

With 1.6 billion people in developing countries still lacking access to electricity, Secretary-General Ban Ki-moon today called for a global revolution to address the world's energy challenges....

app-en Monday, January 17, 2011 4:05:00 PM CET | Info ○
ISLAMABAD. Jan 17 (APP): The Competition Commission of Pakistan (CCP)

of the exclusive rights grainted by the Pakistan Electronic Media Regulatory Authority (PEMMAto "Goo Super" to telecast all Cricket World Cup 2011 matches on cable and satellite under as exclusive agreement with ESPN....

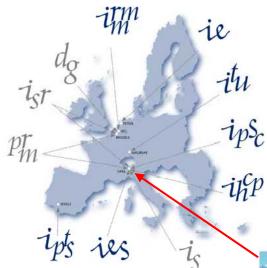
Gas drilling done 'by the book'

A firm which is drilling for shale gas in Lancashire seeks to reassure residents and environmentalists that it is safe.





Joint Research Centre - Who we are



European Commission (scientific-technical arm of public administration)

- Non-commercial
- Multi-disciplinary / multilingual
- Relatively small team working on Language Technology and media monitoring

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)

KARLSRUHE (DE)

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)

The Institute for Prospective Technological Studies (IPTS)





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:
 - Ca. 20 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

- ~ 2500 Sources (world-wide, with focus on Europe)
 - ~ 2300 news sources (web portals)
 - ~ 200 specialist medical sites
 - ~ 20 commercial newswires
 - Specialist pay-for sources (LexisMed)
 - 24/7, updated every 10 minutes
- ~ 100,000 articles / day in ~ 50 languages
- Converts dirty html with adverts, menus, html tags, 'related stories', etc. into clean and standardised UTF-8 encoded RSS format.
- Articles are fed into the various EMM applications:





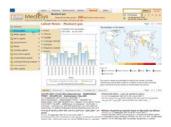


Steinberger Ralf, Bruno Pouliquen & Erik van der Goot (2009). An introduction to the Europe Media Monitor Family of Applications. In: Fredric Gey, Noriko Kando & Jussi Karlgren (eds.): Information Access in a Multilingual World - Proceedings of the SIGIR 2009 Workshop (SIGIR-CLIR'2009), pp. 1-8. Boston,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

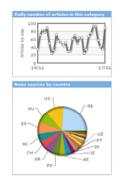


Agenda

- 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
- How to get access to this complementary information
 - Multilingual category definitions and alerts
 - Linking of related news across languages
 - Multilingual information gathering on named entities
 - Multilingual event scenario template filling
- Ongoing work & Summary







Latest News About - Climate Change

Articles published more than 10 minutes as





Multilinguality: coverage of medical news in various languages

Multilingual Web Workshop, Pisa, Italy, 4 April 2011

Locations mentioned in MedISys medical articles across languages – complementary coverage



Italian - German





English - French





Spanish - Portuguese





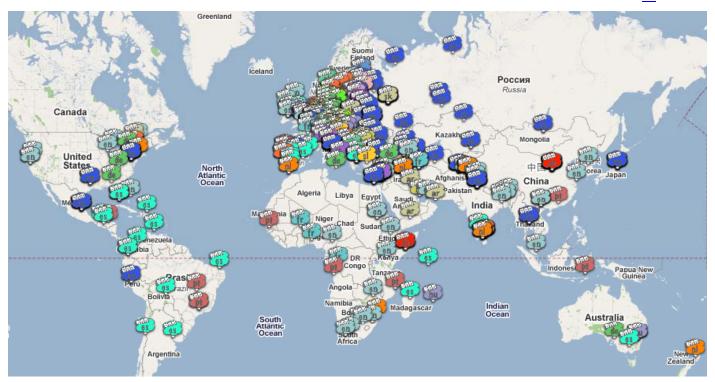
NewsBrief Live Cluster Map

Multilingual Web Workshop, Pisa, Italy, 4 April 2011



Display of latest geo-located news clusters

live



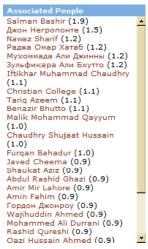


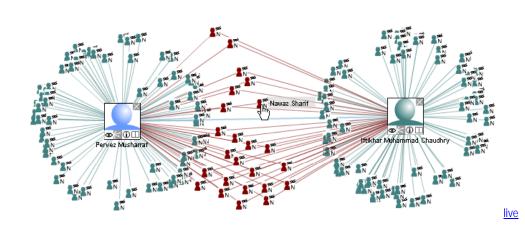
Multilinguality: More information about relations between people

Multilingual Web Workshop, Pisa, Italy, 4 April 2011



Co-occurrence relation between people produced on the basis of many languages is **less biased**.

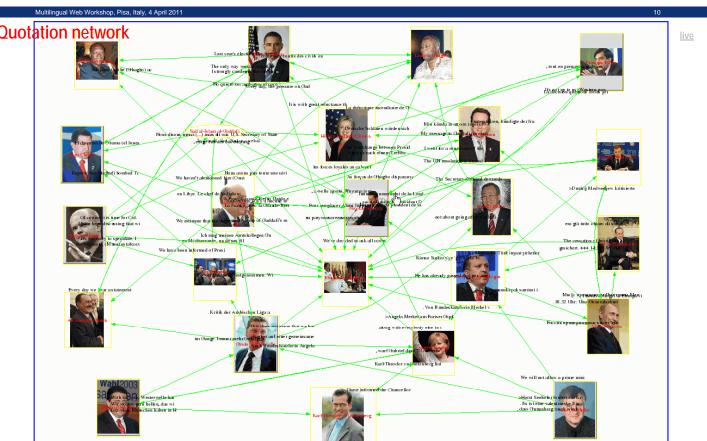




Pouliquen Bruno, Ralf Steinberger, Jenya Belyaeva (2007). Multilingual multidocument continuously updated social networks. Proceedings of the Workshop Multisource Multilingual Information Extraction and Summarization (MMIES'2007) held at RANLP'2007, pp. 25-32. Borovets, Bulgaria, 26 September 2007.(PDF) Hristo Tanev (2007). **Unsupervised Learning of Social Networks from a Multiple-Source News Corpus.** Proceedings of the Workshop *Multi-source Multilingual Information Extraction and Summarization* (MMIES'2007) held at RANLP'2007, pp. 33-40. Borovets, Bulgaria, 26 September 2007. (PDF)

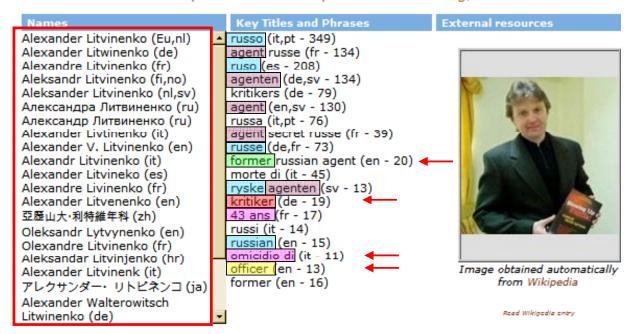


Multilinguality: less-biased centrality in social networks



Alexander Litvinenko

Information about this person was last updated on Dienstag, 20. März 2007.



Steinberger Ralf & Bruno Pouliquen (2007). Cross-lingual Named Entity Recognition. In: Satoshi Sekine & Elisabete Ranchhod (eds.), Journal Linguisticae Investigationes, Special Issue on Named Entity Recognition and Categorisation, LI 30:1, pp. 135-162. John Benjamins Publishing Company. ISSN 0378-4169.

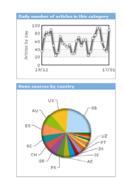


Agenda

- 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
- How to get access to this complementary information
 - Multilingual category definitions and alerts
 - Linking of related news across languages
 - Multilingual information gathering on named entities
 - Multilingual event scenario template filling
- Ongoing work & Summary





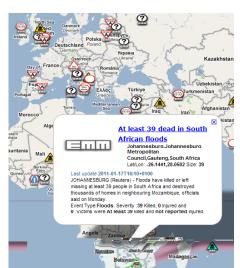


Latest News About - Climate Change

Articles published more than 10 minutes ago

offitis freatomasis con as cerements of the man 20 minutes ago
Ban calls for global revolution to ensure clean energy for all
unrong Mandra, harvary 17, 2011 41100 PM CET [Info 0
With 1.8 billion people in developing cour

Gas drilling done 'by the book'
bbr Manday, January 17, 2011 4-02:00 DM CET | Info O
A firm which is drilling for shale gas in Lancashire seeks to reassure reside
environmentalists that it is safe....





EMM – NewsBrief & MedISys (up to 50 languages)

- Public sites: http://emm.newsbrief.eu/ & http://emm.newsbr
- Categorises news into over 1000 categories, using:
 - Boolean search word combinations
 - vicinity operators
 - optional weights
 - regular expressions
- Clusters and tracks news live (multi-monolingually)
- Sends out email notifications for each category
- Detects breaking news
- Lookup of known entities
- **Quotation recognition**

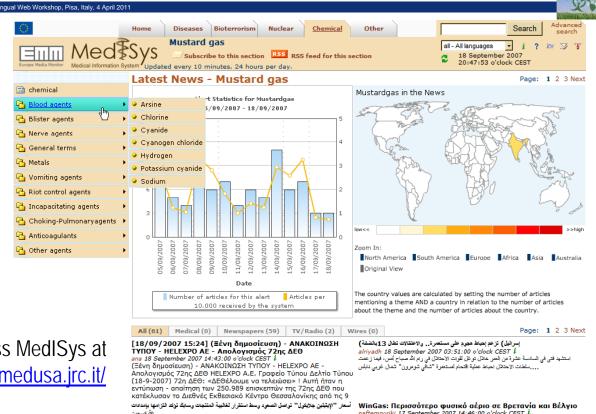


Medisvs





MedISys - Filtering and classification in up to 50 languages



afriyadh 18 September 2007 03:51:00 °lock CEST أ وراشر أرقام البتروكيماريات الله يقيل مركة أسطر البتروكيماريات السلة من العواد التي يتم إنتاجها في طلبح العربي، عند مستوى 252.9نطقة منهيا فترة تسعة أسليع من الارتفاع المتثلي، منطوعا بالمستويات

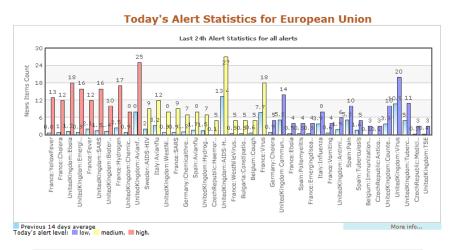
Access MedISys at http://medusa.jrc.it/ إسرائيل) ترجم إدينظ هيوم على مستصرق. والاعتقالات تطال 13 ياششة) #airiyadh 18 September 2007 03:51:00 o'clock CES7 أ السادسة عظرة من المسر طائع على قالت الاستعالات إلى براها شمياً إلى من أيضا والمساللطات الاحتلال احباط عطية اقتمام المستصرة "لدائي ترمرون" شمال غربي نابلس

WinGas: Περισσότερο φυσικό αέριο σε Βρετανία και Βέλγιο naftemporiki 17 September 2007 14:46:00 ο'clock CEST έ εταιρεία WinGas. BASF Συγκεκριμένα, η εταιρεία θα αυξήσει φέτος τις προμήθειες στους Βέλγους πελάτες της στα 12,3 δια. κιλοβατώρες αερίου. Για το 2008 έχει ήδη υπογράψει συμφωνίες, οι οποίες

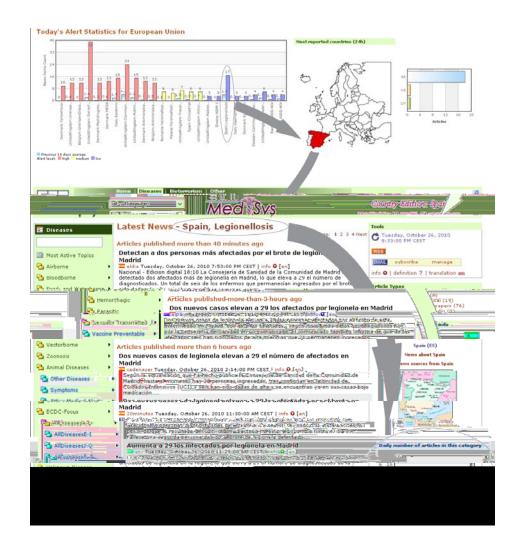
Multilingual Web Workshop, Pisa, Italy, 4 April 2011

15

- Documents from all languages get classified according to the same countries and categories.
- An increase of the number of media reports on any country-category combination is detected,
- independently of the reporting language.
- Graphs and alerts may show events not yet reported in your own language.

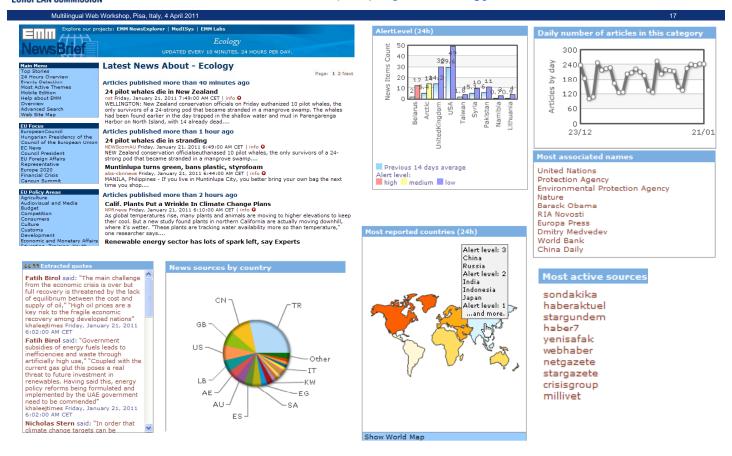


Ralf Steinberger, Flavio Fuart, Erik van der Goot, Clive Best, Peter von Etter & Roman Yangarber (2008). **Text Mining from the Web for Medical Intelligence**. In: Fogelman-Soulié Françoise, Domenico Perrotta, Jakub Piskorski & Ralf Steinberger (eds.): Mining Massive Data Sets for Security. pp. 295-310. **IOS Press**, Amsterdam, The Netherlands





EMM-NewsBrief – Example page: <u>Ecology</u>





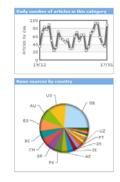
Agenda

Multilingual Web Workshop, Pisa, Italy, 4 April 201

- 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
- · How to get access to this complementary information
 - · Multilingual category definitions and alerts
 - Linking of related news across languages
 - · Multilingual information gathering on named entities
 - Multilingual event scenario template filling
- Ongoing work & Summary







Latest News About - Climate Change

Page: 1 2 3 4 Ne

Articles published more than 10 minutes ago El Gobierno de Aragón pone en marcha el Ma

El Gobierno de Aragón pone en marcha el Manual de Buenas Prácticas Medioambientales en obras de carreteras due Modday, January 17, 2011 4119100 PM CET | Info 9

que Monday, January 17, 2011 419100 DM CET | inho or commented of Monday, January 17, 2011 419100 DM CET | inho or commented of Otras Publicas, Urbanarmo y Transportes y de Medio Ambienta del Sobietario de Aregolf han puesto en marcha el Hannardo de Buenar Prácticas Mediodomientales, sobietario de Aregolf han puesto en marcha el Hannardo de Duenar Prácticas de Mondado en las objetas reglacionadas, son Jas carreteras de Bublandad de la Compunidad Authoriama.

Articles published more than 20 minutes ago

Ban calls for global revolution to ensure clean energy for all

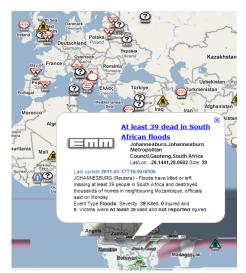
with 1.5 billion people in developing countries still lacking access to electricity, Secretary-General Ban Ki-moon today called for a global revolution to address the world's energy challenges....

Environmental degradation needs immediate attention: Afrid

ISLANABAD, Jan 17 (APP): The Competition Commission of Paisistan (CCF) has taken notice of the exclusive nights granted by the Paisstan Electronic Media Regulatory Authority (PEMRA) to "Geo Super" to telecast all Cricket World Cup 2011 matches on cable and satellite under an exclusive agreement with ESFN....

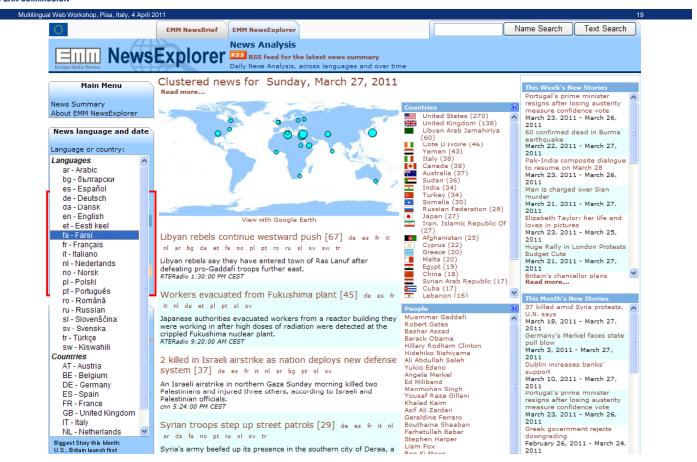
Articles published more than 30 minutes ago

bbs Monday, January 17, 2011 4:02:00 PM CET | Info O A firm which is drilling for shale gas in Lancashire seeks to reassur





NewsExplorer - Multilingual daily news overview





epl 23:17:00 CET

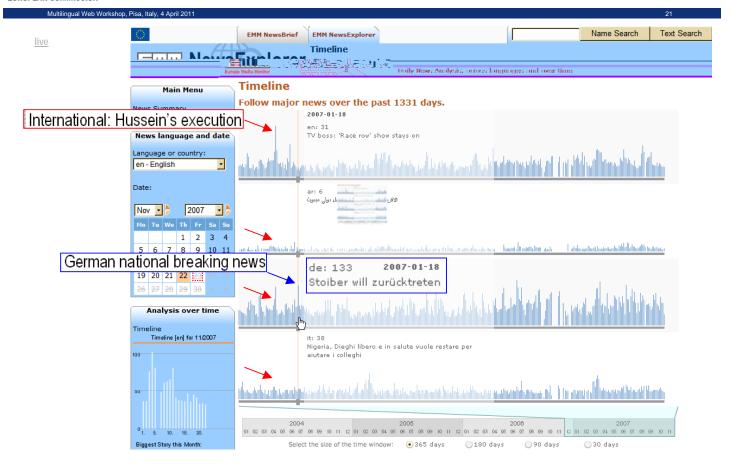
NewsExplorer - Cross-lingual cluster linking

NewsExplorer

Multilingual Web Workshop, Pisa, Italy, 4 April 2011 Castro quits as president, state-run paper reports [72] de nl fa no pl pt ro ru sl tr de en fr it nl گزارش تلویزیون فرانسه از کناره گیری فیدل کاستری on Tuesday, according to a letter published by iranpressnews 13:36:00 o'clock CET state-run newspaper Granma. Kuba: Fidel Castro gibt das Zepter ab en es fr it nl ar bg da et fa no pl pt ro ru sl cnn 9:23:00 AM CET Der legendäre kuhanische Staatschef verzichtet laut Opline Ausgabe d Parteizeitung a **Fidel Castro zrezygnował!** de en es fr it nl kleinezeitung 10. Fidel Castro går av de en es fr it nl Przywódca kubański Fidel Castro po 49 latach rządów zrezygnował we wtorek z funkcji Partiav Fidel Castro renuncia a la Presidencia del Consejo de Estado de przewodniczącego Rady Państwa Kuby. Fidel Castro renunciou à presidência de Cuba de en es fr it nl et no pl pt ro sl sv tr Anúncio no órgão oficial do Partido Comunista cubano Fidel Castro anunciou hoje que se retira da vg A Cuba, Fidel Castro renonce au pouvoir de en es it nl ar bg da et Fidel Castro se retrage de la presedintia Cubei de en es fr it nl ru sl sv tr Fidel Castro a anuntat, marti, ca renunta la presedintia Cubei, in editia electronica a cotidianului Cuba, Fidel Castro rinuncia alla presidenza de en es fr nl ar bg Fidel Castro se je odpovedal položaju kubanskega predsednika de en es fr it nl ru sl sv tr Kubanski voditelj Fidel Castro je danes sporočil, da se odpoveduje položaju predsednika države. Kot je L'/ Cubaanse president Fidel Castro afgetreden de en es fr it ar Castro še zapisal v sporočilu, objavljenem v spletni izdaji uradnega glasila Granma, se ne poteguje in Värmby skakade hand med Fidel Castro. de en es fr it nl ro ru sl sv tr Ja. Jag har inte tvättat högernäven sedan dess, 1983. Jag var på en stor sammandragning på Kuba de en fr it كاسترو يستقيل ويوش يدعو للتحول الديمقراطي في كوي med uppmaningen att USA skulle häva blockaden mot landet. smp kl 19:51 CET : أُولِّي مؤتمر صحفي في رواندا، إلى مساعدة كوبا على البدء بعملية "انتقال ديمقراطية"، وذلك إشر قرار الزعيم -- (CNN) مافاتا، كوبا فيدل كاسترو، بالتنحي عن منصبه كرئيس للبلاد. وقال بوش: "إن على المجتمع الدولي أن يحل مع الشعب الكربي للبدء بإقامة مؤسسات Фидель Кастро отказался от поста председателя Госсовета Кубы de en fr it cnnarabic CET 01:21:00 a Kactdo се оттегли от президентския пост de en es fr it nl ГАВАНА, 19 февраля. /ИТАР-ТАСС/. Фидель Кастро отказался от поста главы государства и правительства - председателя Государственного совета Кубы. Об этом он сообщил в обращении к Фидел Кастро обяви, че се отказва от президентския пост, съобщава АФП Fogh vil ikke savne Castro de en es fr it nl Bir dönemin sonu de en es fr it nl REPLIIK: Castro-aja lõpu algus de en es fr it nl
Üks 20. sajandi menukam vabadus-võitleja ja tuntum diktaator Fidel Cameliyattan beri iktidar koltuğundan uzak kaldı. Ülke yönetimine, ağabeyi Fidel Castro'ya vekalet eden kui Nõukogude "gerondid", kelle jaoks tähendas lahkumine võimult ka Raul Castro bakıyordu.
Võim kestab vähemalt esiotsa edasi, sest esimeseks asendajaks peetaluurivetim 10:15:00 CET "Polit REPLIIK: Castro-aja lõpu algus de



NewsExplorer - Time line: biggest clusters per day





NewsExplorer – Aggregation of clusters into longer 'stories'

live





Name variants found in 16 hours of multilingual news analysis

(25.3.2011)

2190. Muammar Gaddafi

2011-03-24T23:59+0200

Muammar Gaddafi / Muammar al-Gaddafi / Mouammar / معمر القذافي Kadhafi / Moammar Gadhafi / Муамара 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 / Moamerja Gadafija / 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 / فَذَافَي 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 /

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

2011

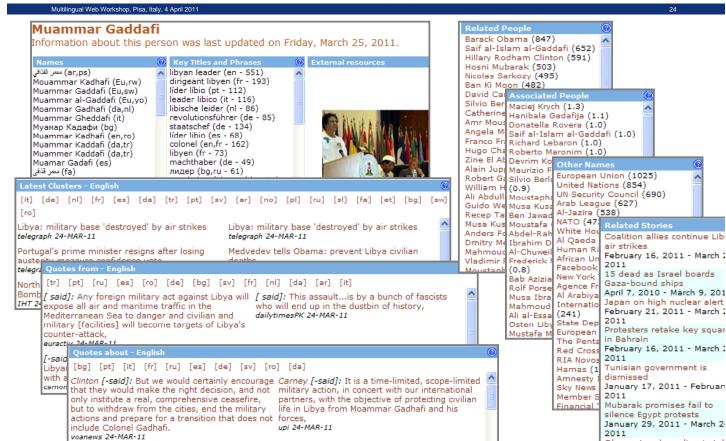
Obama toughens line in tal

live



NewsExplorer –Information about people

collected from multiple languages and over time





NewsExplorer – Relation exploration



Key tilles and Phrases son (en,sv - 125) fils (fr - 63) son (en, 46) dirigeant libyen (fr - 13) zoon (nl - 26) libyen (fr - 13) filho (pt - 22) hijo (es - 19) sin (sl - 12) président libyen (fr - 6) libyan leader (en - 7) moammar gadhafi#'s so 4)

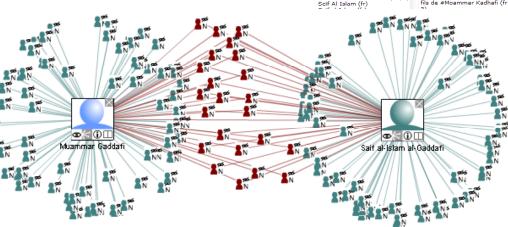
Information about this person was last updated on Sat



Example:

Muammar Gaddafi &

son Saif al-Islam al-Gaddafi





<u>live</u>



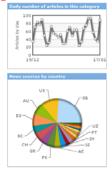
Agenda

Multilingual Web Workshop, Pisa, Italy, 4 April 2011

- 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
- How to get access to this complementary information
 - Multilingual category definitions and alerts
 - · Linking of related news across languages
 - · Multilingual information gathering on named entities
 - · Multilingual event scenario template filling
- Ongoing work & Summary







Latest News About - Climate Change

Page: 1 2 3 4 Ne

Articles published more than 10 minutes ago El Gobierno de Aragón pone en marcha el M

El Gobierno de Aragón pone en marcha el Manual de Buenas Prácticas Medioambientales en obras de carreteras

que Monday, January 17, 2011 4119100 PM ERT Jinfo y Transportes y de Medio Ambiente del Los Departamento de Corea Publicas, Urbanasmo y Transportes y de Medio Ambiente del Codo de Company de Compa

Articles published more than 20 minutes ago

Ban calls for global revolution to ensure clean energy for all

un-org Monday, January 17, 2011 4111:00 PM CET | Into © With 1.6 billion people in developing countries still lacking access to electricity, Secretary General Ban Ki-moon today called for a global revolution to address the world's energy

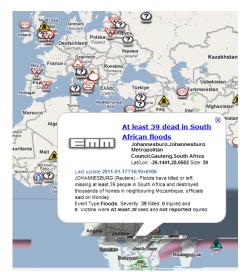
Environmental degradation needs immediate attention: Africa

appren Monday, January 17, 2011 4:05:00 PM CET | info ©
ISLAMARBAD, Jan 17 (APP) The Competition Commission of Pakistan (CCP) has taken notice
of the exclusive rights granted by the Pakistan Electronic Media Regulatory Authority (PEMRA)
to "Goo Super" to telecast all Circlect World Cup 2011 matches on cobile and satellite under an

Articles published more than 30 minutes ago

Gas drilling done 'by the book'

A firm which is drilling for shale gas in Lancashire seeks to reassure residents and





EMM-NEXUS Event Extraction System

EMM-Labs

Multilingual Web Workshop, Pisa, Italy, 4 April 2011



Access NEXUS at: http://emm-labs.jrc.it/ or

http://emm.newsbrief.eu/geo?type=event&format=html&language=all

Atkinson Martin, Jakub Piskorski, Bruno Pouliquen, Ralf Steinberger, Hristo Tanev & Vanni Zavarella (2008). **Online-monitoring of security-related events.** In Proceedings of the 22nd International Conference on Computational Linguistics (CoLing'2008). Manchester, UK, 18-22 August 2008. (PDF)



EMM-NEXUS - Event Extraction System

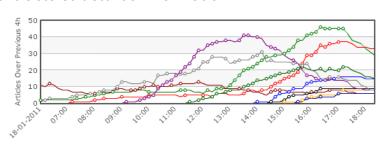
Multilingual Web Workshop, Pisa, Italy, 4 April 201

NEXUS:

Multilingual Information Extraction system for the extraction of **structured event descriptions** from online news referring to conflicts, crimes and disasters.

TYPE
PLACE
TIME
DEAD COUNT
DEAD DESCRIPTION
WOUNDED COUNT/DESC
DISPLACED COUNT/DESC
HOMELESS COUNT/DESC
ARRESTED COUNT/DESC
PERPETRATOR
WEAPONS

- Currently 7 Languages:
 English, French, Portuguese, Arabic, Spanish, Italian, Russian (and Chinese).
- Near real-time: every 10 minutes, EMM clusters the latest articles about the same event and NEXUS extracts structured information.
- Objective:
 Global crisis monitoring
 (Live situation or long-term trend).





Event Extraction Output (English, French and Portuguese)

Multilingual Web Workshop, Pisa, Italy, 4 April 2011

۷:

Baghdad car bombs kill at least 127

Event Type: Terrorist Attack

Severity: 127 killed 448 injured

Weapons: car bomb
Place: Baghdad

Johannesburg: cinq suspects arrêtés pour le meurtre du curé français

Event Type: Arrest

Severity: 1 killed 0 injured

Victims: prêtre français/ Louis Blondel killed

Place: Johannesburg

Police search for killer bus driver

Event Type: Man-Made Disaster

Severity: 1 killed 6 injured Victims: passenger killed

Place: London

Timor-Leste: Indonésios estão a fazer

"cortina de fumo" sobre morte dos "5 de Balibó" - viúva (C/ÁUDIO)

Severity: 5 killed, 0 injured Victims: jornalistas killed

Place: Timor-Leste.



Aggregating information extracted from various articles

Multilingual Web Workshop, Pisa, Italy, 4 April 2011

Car bomber strikes north Pakistan

ech-chorouk-en Tuesday, November 10, 2009 2:23:00 PM CET

A car bomb has exploded in Pakistani's northwestern town of Charsadda killing at least 10 people....

Bomb explodes in northwestern Pakistani town

yediotaharonot Tuesday, November 10, 2009 1:58:00 PM CET

A bomb exploded in the northwestern Pakistani town of Charsadda on Tuesday causing an unknown number of casualties, police said. "It was a bomb blast....

10 killed in Pakistan bomb

RTERadio Tuesday, November 10, 2009 1:57:00 PM CET

A bomb has exploded in the north-western Pakistani town of Charsadda, killing 10 people....

TYPE Bombing

PLACE Charsadda, Pakistan

TIME Tuesday, November 10, 2009

DEAD COUNT 10

DEAD DESCRIPTION people

WOUNDED COUNT/DESC DISPLACED COUNT/DESC HOMELESS COUNT/DESC ARRESTED COUNT/DESC

PERPETRATOR

WEAPONS Bomb



Event extraction – Text Version



Wednesday, January 26, 2011 6:30:00 AM CET | info Q



Event extraction - Display on a map



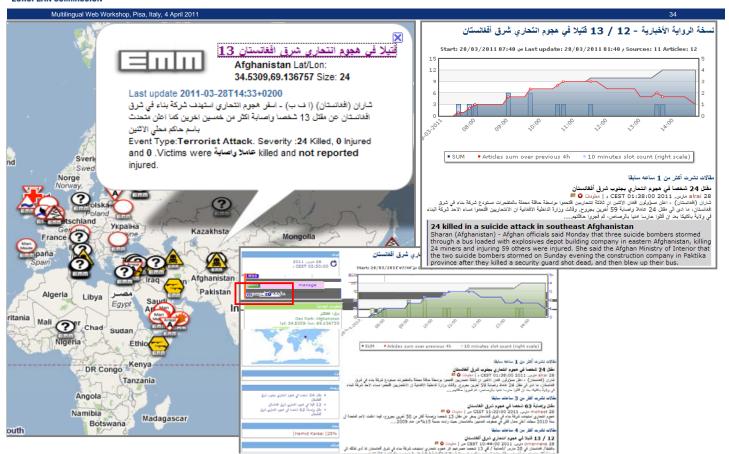


Event extraction - Display on a map - click on one event





Event extraction – View news cluster and translation





Event types currently recognised



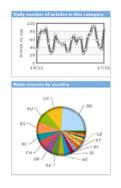


Agenda

- 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
- How to get access to this complementary information
 - Multilingual category definitions and alerts
 - Linking of related news across languages
 - Multilingual information gathering on named entities
 - Multilingual event scenario template filling
- **Ongoing work & Summary**







Latest News About - Climate Change

Articles published more than 10 minutes as

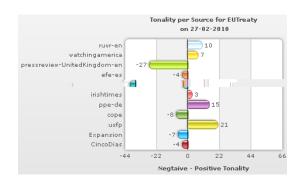


Multilingual Web Workshop, Pisa, Italy, 4 April 2011

17

- E.g. Detect opinions on
 - European Constitution; EU press releases;
 - Entities (persons, organisations, EU programmes and initiatives);
- Detect and display opinion differences across sources and across countries;
- Follow trends over time.







Ongoing: Monitoring social media

Facebook:

Keyword searches on publicly available posts e.g. search for *Chikungunya* on openbook.org

extract publicly available friend networks.

Twitter:

Keyword searches on publicly available tweets e.g. search for *Chikungunya* on twitter.com

Blogs







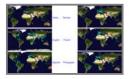


Summary – News complementarity



Multilingual Web Workshop, Pisa, Italy, 4 April 201

News content (and internet content in general) is complementary across languages.





- EMM gathers and processes multilingual news, etc.
 - Multilingual category definitions and alerts → alert and produce statistics











Linking of related news across languages

Libyan rebels continue westward push [67] de es fr i

Multilingual information gathering on named entities





 Multilingual event scenario template filling



