



Multilingual and multimedia information for Business Intelligence

Christian Fluhr
Director of research
CEA/LIST
Christian.fluhr@new-phenix.com



Multilingual and multimedia information for Business Intelligence



Business intelligence :

Observation of **open sources** by private or public organizations of their ever changing world environment to help deciders in their managing activity



Multilingual and multimedia information for Business Intelligence



Observe what:

- Competitors activity
- possible partners
- change in legislations
- evolution of technologies and patents
- evolution of markets in various countries
- activity of disinformation from competitors ...

Where:

- web sites
- news wires, news papers
- scientific papers and congresses
- radio, television
- groups of discussions, blogs
- reports on contacts by people going to congresses or visiting companies



Multilingual and multimedia information for Business Intelligence



World wide economy needs a worldwide observation

The assertion that all valuable information can be found in English is false

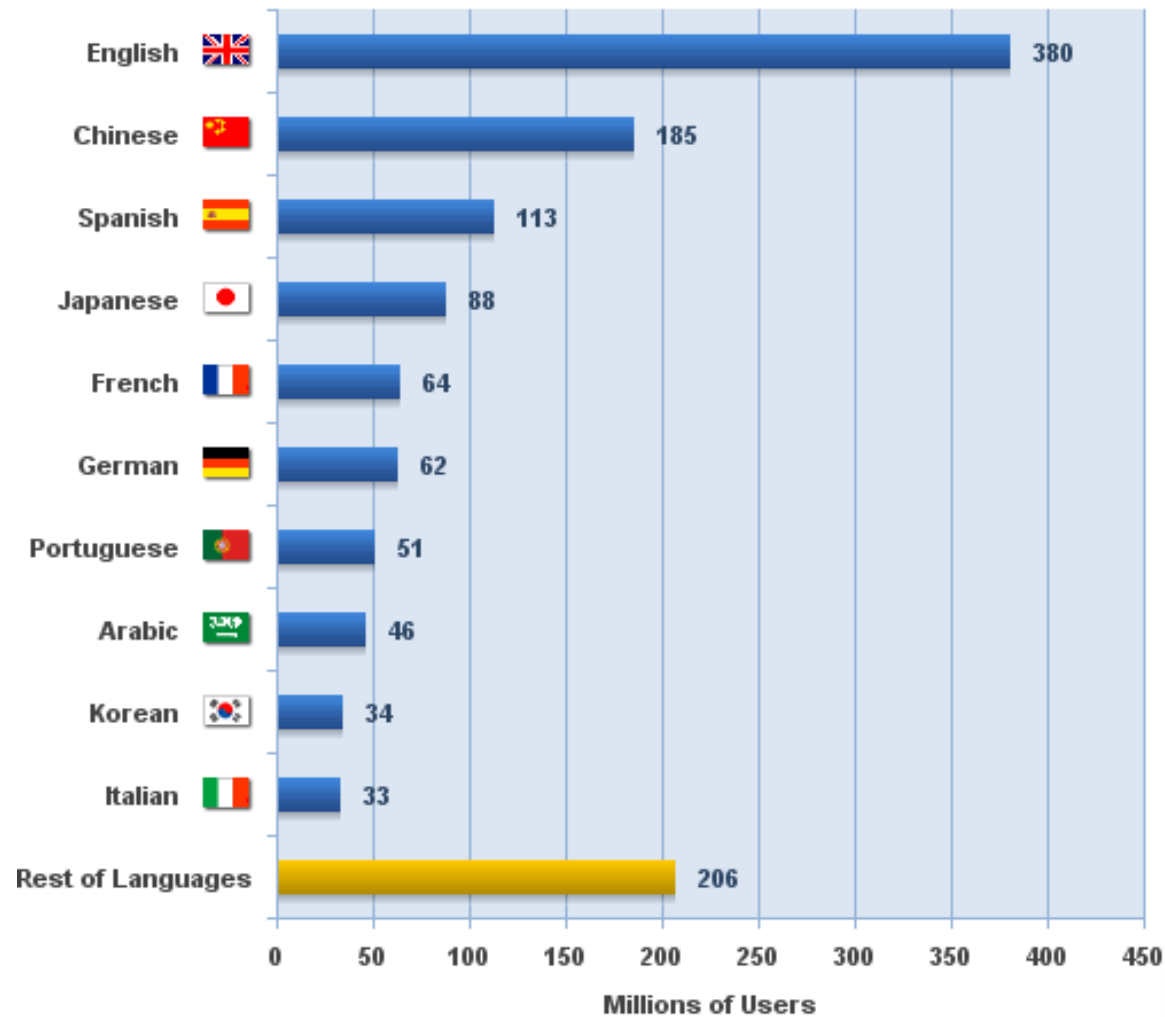
- The proportion of English in the web is statically decreasing
- In non English speaking countries, English is used to give what they want you to believe. The detailed information remains in native languages.
- Newspapers, radio, television are mainly in native language



Multilingual and multimedia information for Business Intelligence



Top 10 Internet Languages - November 2007



Source: www.internetworldstats.com
Copyright © 2008, Miniwatts Marketing Group



Multilingual and multimedia information for Business Intelligence



Europe has a long experience in **machine translation** and **crosslanguage interrogation**. Multilinguality is a key problem in Europe due to the number and diversity of languages

Ex :

➤ for MT, CETA (Grenoble) from 1961 until now
(IMAG/CLIPS)

➤ 1st Crosslingual information retrieval European project
EMIR (European multilingual information retrieval) 1990-
1994.



Multilingual and multimedia information for Business Intelligence



Second half of the 90ies : In the USA strategic importance of the multilingual information appears :

DARPA 12th Working group considers that taking the diversity of languages is strategic for the US :

« This barrier puts the United States at a distinct STRATEGIC DISADVANTAGE in the international competition for Information Dominance, because technical and military personnel in other nations have far better skills in English than we have in their languages.

They can, do, and will effectively utilizing our English-language information, while we lag enormously in effectively utilizing information in their languages. »



Multilingual and multimedia information for Business Intelligence



What Would you Pay to Know about India's Nuclear Test 3 Weeks

Prior?

Source : TIDES planning conference

Charhdi Kala 50

May 21-May 27/1998

ਭਾਰਤ ਦੇ ਨਿਊਕਲੀਅਰ ਟੈਸਟਾਂ ਦਾ ਮਕਸਦ ਘੱਟ ਗਿਣਤੀਆਂ ਨੂੰ ਧਮਕਾਉਣਾ

ਕੋਵਿਨ ਗਠਿਓਨ—(ਵੈਨਕੂਵਰ ਸਨ) ਵੈਨਕੂਵਰ—ਇੰਟਰਨੈਸ਼ਨਲ ਸਿੱਖ ਯੂਥ ਫੰਡਰੇਸ਼ਨ ਦੇ ਵੱਲੋਂ ਭਾਰਤ ਵਲੋਂ ਕੀਤੇ ਗਏ ਲਗਾਤਾਰ ਪੰਜ ਧਮਕਿਆਂ

ਹਵਾ ਤੋਂ ਵੀ, ਨਿਊਕਲੀਅਰ ਟੈਸਟਾਂ ਨੂੰ ਡਾਕਟਰ ਸਿੰਘ ਵਲੋਂ ਚਾੜ੍ਹਦੀ ਕਲਾ ਅਧਿਐਸ਼ ਵਜੋਂ ਧਮਕਿਆਂ ਦੀ ਪਲੀਨਗਰੀ ਸਬੰਧੀ ਖਬਰਾਂ ਨੂੰ ਪ੍ਰਮੁੱਖਤਾ ਨਾਲ ਛਾਪਿਆ।



ਪ੍ਰਤੀਕਿਰਿਆ—ਭਾਰਤ ਦੇ ਨਿਊਕਲੀਅਰ ਟੈਸਟਾਂ ਦਾ ਵਿਰੋਧ ਕਰਨ ਵਾਸਤੇ 200 ਦੇ ਕਰੀਬ ਸਿੱਖਾਂ ਨੇ ਵੈਨਕੂਵਰ ਸਥਿਤ ਭਾਰਤੀ ਕੌਂਸਲੇਟ ਜਨਰਲ ਦੇ ਦਫਤਰ ਸਾਹਮਣੇ ਮਾਰਚ ਕੀਤਾ।

ਦੇ ਵਿਰੁੱਧ ਆਯੋਜਿਤ ਰੱਲੀ ਵਿੱਚ ਇੱਕ ਬੁਲਾਰੇ ਨੇ ਕਿਹਾ ਕਿ ਇਹ ਧਮਕੀ ਪਾਰਲਿਕ ਘੱਟ ਗਿਣਤੀਆਂ ਨੂੰ ਦਬਾਉਣ ਲਈ ਇੱਕ ਸੰਕੇਤ ਹੈ।

ਖਾਲਿਸਤਾਨ ਅੰਦੋਲਨ ਸੰਗਤ ਵਾਸਿੰਗਟਨ ਡੀ. ਸੀ. ਦੇ ਡਾਕਟਰ ਅਮਰਜੀਤ ਸਿੰਘ ਨੇ ਇਹ ਵੀ-ਐਚ ਟੀਵੀ ਟੀਵੀ ਹੇਠ ਉੱਤਰ ਪੱਛਮੀ ਭਾਰਤ ਵਿੱਚ ਕੀਤੇ ਗਏ ਬੰਬ ਧਮਕਿਆਂ ਦਾ ਅਸਰ ਲਾਗੂ ਪਾਕਿਸਤਾਨ ਵਿੱਚ ਵੀ ਹੋ ਸਕਦਾ ਹੈ। ਉਨ੍ਹਾਂ ਕਿਹਾ ਕਿ, 'ਅਸੀਂ ਹੁਣ ਇਸ ਖਤਰੇ ਨੂੰ ਸਹੀ ਢੰਗ ਨਾਲ ਕੰਧ ਸਕਦੇ ਹਾਂ। ਭਾਰਤ ਨਿਊਕਲੀਅਰ ਬੰਬ ਧਮਕਿਆਂ ਨਾਲ ਦੁਖੀ ਹੋ ਸਕਦਾ ਹੈ ਸਗੋਂ ਦੱਸਾਂ ਨੂੰ ਧਮਕਾਉਣਾ ਚਾਹੁੰਦਾ ਹੈ। ਡਾਕਟਰ ਸਿੰਘ ਨੇ ਕੰਨੜਾ ਸਰਕਾਰ ਨੂੰ ਅਪੀਲ ਕੀਤੀ ਕਿ ਉਹ ਭਾਰਤ ਦੇ ਮਿਲਾਫ ਅੰਤਰਰਾਸ਼ਟਰੀ ਦਸਾਅ ਸਟਾਈ ਹੱਥ ਦੇ ਯਤਨਾਂ ਨੂੰ

'ਮੈਂ ਭਾਵੇਂ ਆਈ. ਐਸ. ਵਾਈ. ਐਫ. ਦਾ ਮੈਂਬਰ ਨਹੀਂ ਹਾਂ ਪਰ ਦਾਅਵੇ ਨਾਲ ਕਹਿੰਦਾ ਹਾਂ ਕਿ ਭਾਰਤ ਅਧਿਕਾਰ ਘਟੀਆ ਨਿੱਜੀ ਮੁਦਾਦ ਦੀ ਪ੍ਰਭਾਵੀ ਲਈ ਸਿੱਖਾਂ ਦੀਆਂ ਅਸਾਦ ਖਾਲਿਸਤਾਨ ਰਾਜ ਸਟਾਈਟ ਦੀਆਂ ਕੋਸ਼ਿਸ਼ਾਂ ਨੂੰ ਹਰ ਕੋਢ ਢੰਗ ਤਰੀਕੇ ਦੀ ਵਰਤੋਂ ਨਾਲ, (ਭਾਵ ਜਿਹਨੂੰ ਮਰਜ਼ੀ ਦਰਿਸਤਗਰਦ ਗਰਦਾਨ ਕੇ) ਦਾਅ ਵਾ ਧਿਰਾ ਹੋ। ਡਾਕਟਰ ਸਿੰਘ ਨੇ ਵੈਨਕੂਵਰ 325 ਹੋਏ ਸਟਰੀਟ 'ਤੇ ਸਥਿਤ ਭਾਰਤੀ ਕੌਂਸਲੇਟ ਜਨਰਲ ਦੇ ਦਫਤਰ ਮੁਹੱਰੇ ਆਯੋਜਿਤ ਇਸ ਰੱਲੀ ਨੂੰ ਸੰਬੰਧਨ ਕਰਦੇ ਹੋਏ ਡਾਕਟਰ ਅਮਰਜੀਤ ਸਿੰਘ ਅਤੇ ਹੋਰ ਬੁਲਾਰਿਆਂ ਨੇ ਭਾਰਤ ਸਰਕਾਰ ਦੀ ਸਾਰ ਬਾਰ ਅਲੋਚਨਾ ਕੀਤੀ।

ਭਾਰਤ ਵਿੱਚ ਸੀ. ਜੇ. ਪੀ. ਦਾ ਰਾਜ ਹੈ। ਭਾਰਤ ਵਿੱਚ ਇੱਕ ਕੰਨੜਾ ਪੰਜਾਬੀ ਜਾਗਤ ਹੈ। ਭਾਰਤ ਵਿੱਚ ਸ਼ਹੂਰਗੀ

ਭਾਰਤ ਦੇ ਨਿਊਕਲੀਅਰ ਟੈਸਟ ਵਿਰੁੱਧ ਧਮਕਾਏ ਮੁਜਾਹਰਾ!

ਵੈਨਕੂਵਰ—ਸਲੀਮ ਜੀਵਾ—(ਸਟਾਵ ਰਿਪੋਰਟਰ)—ਇੱਕ ਸਿੱਖ ਖਾਤਕ ਜਬਬੰਦੀ ਜਿਹੜੀ ਕਿ ਆਪਣੇ ਹਿੱਸਕ ਖਿੱਕੜ ਵਜੋਂ ਮਸ਼ਹੂਰ ਹੈ, ਨੇ ਪਿਛਲੇ ਸੋਮਵਾਰ ਹੋਏ ਸਟਰੀਟ ਡਾਕਟਰ ਸਿੰਘ ਵਲੋਂ ਭਾਰਤੀ ਕੌਂਸਲੇਟ ਜਨਰਲ ਦੇ ਦਫਤਰ ਮੁਹੱਰੇ ਇੱਕ ਧਮਕਾਏ ਮੁਜਾਹਰਾ ਕੀਤਾ।

ਕਮਰੀ ਦੇ ਪੰਜ ਮਬਰਾਂ 'ਚੋਂ ਇੱਕ ਹੈ, ਜੋ ਕਿ ਸਮੁੱਚੇ ਪਾਰ ਖਾਲਿਸਤਾਨ ਰਿਪਬਲਿਕ ਵਾਸਤੇ ਜੂਝ ਰਹੇ ਸਿੱਖ ਗਰੁੱਪ ਦੀ ਅਗਵਾਈ ਕਰ ਰਿਹਾ ਹੈ, ਨੇ ਕਿਹਾ ਕਿ, 'ਇਨ੍ਹਾਂ



ਡਾ. ਅਮਰਜੀਤ ਸਿੰਘ ਜਿਨ੍ਹਾਂ ਨੇ ਟੈਸਟਾਂ ਸਬੰਧੀ ਭਵਿੱਖਬਾਣੀ ਕੀਤੀ, ਮੁਜਾਹਰੇ ਨੂੰ ਸੰਬੰਧਨ ਕਰਦੇ ਹੋਏ

ਇਹ ਪੁੱਛ ਜਾਣ 'ਤੇ ਕੀ ਤੁਸੀਂ ਉਸ ਸਿੱਖ ਗਰੁੱਪ ਵਲੋਂ ਬੁਲਾਰੀ ਗਈ ਭਾਰਤ ਵਿਰੁੱਧ ਰੱਲੀ 'ਚ ਕਿਉਂ ਬੋਲ ਰਹੇ ਹੋ, ਜਿਸ ਨੂੰ ਇੱਕ ਸਮੱ ਕੰਨੜਾ ਸਰਕਾਰ ਨੇ ਦਰਿਸਤਵਾਦੀ ਗਰਦਾਨਿਆ ਸੀ, ਤਾਂ ਉਨ੍ਹਾਂ ਕਿਹਾ ਕਿ, "ਮੈਂ ਖਾਲਿਸਤਾਨੀ ਹਾਂ ਅਤੇ ਭਾਰਤ ਨੂੰ ਦੁਨੀਆ ਦੇ ਨਕਸ਼ੇ 'ਤੇ ਮਿਟਿਆ ਦੱਖਣਾ ਮੱਗ ਇਸਕ ਹੈ।"

ਅਗੇ ਚੱਲ ਕੇ ਡਾਕਟਰ ਅਮਰਜੀਤ ਸਿੰਘ ਨੇ ਕਿਹਾ ਕਿ, 'ਮੈਂ ਭਾਵੇਂ ਆਈ. ਐਸ. ਵਾਈ. ਐਫ. ਦਾ ਮੈਂਬਰ ਨਹੀਂ ਹਾਂ ਪਰ ਦਾਅਵੇ ਨਾਲ ਕਹਿੰਦਾ ਹਾਂ ਕਿ ਭਾਰਤ ਅਧਿਕਾਰ ਘਟੀਆ ਨਿੱਜੀ ਮੁਦਾਦ ਦੀ ਪ੍ਰਭਾਵੀ ਲਈ ਸਿੱਖਾਂ ਦੀਆਂ ਅਸਾਦ ਖਾਲਿਸਤਾਨ ਰਾਜ ਸਟਾਈਟ ਦੀਆਂ ਕੋਸ਼ਿਸ਼ਾਂ ਨੂੰ ਹਰ ਕੋਢ ਢੰਗ ਤਰੀਕੇ ਦੀ ਵਰਤੋਂ ਨਾਲ, (ਭਾਵ ਜਿਹਨੂੰ ਮਰਜ਼ੀ ਦਰਿਸਤਗਰਦ ਗਰਦਾਨ ਕੇ) ਦਾਅ ਵਾ ਧਿਰਾ ਹੋ। ਡਾਕਟਰ ਸਿੰਘ ਨੇ ਵੈਨਕੂਵਰ 325 ਹੋਏ ਸਟਰੀਟ 'ਤੇ ਸਥਿਤ ਭਾਰਤੀ ਕੌਂਸਲੇਟ ਜਨਰਲ ਦੇ ਦਫਤਰ ਮੁਹੱਰੇ ਆਯੋਜਿਤ ਇਸ ਰੱਲੀ ਨੂੰ ਸੰਬੰਧਨ ਕਰਦੇ ਹੋਏ ਡਾਕਟਰ ਅਮਰਜੀਤ ਸਿੰਘ ਅਤੇ ਹੋਰ ਬੁਲਾਰਿਆਂ ਨੇ ਭਾਰਤ ਸਰਕਾਰ ਦੀ ਸਾਰ ਬਾਰ ਅਲੋਚਨਾ ਕੀਤੀ।

ਨਿਊਕਲੀਅਰ ਟੈਸਟਾਂ ਨੇ ਪੰਜਾਬ ਵਾਸਤੇ ਖਤਰਾ ਪੈਦਾ ਕੀਤਾ ਹੈ। 1984 ਵਿੱਚ ਆਪਣੇ ਜਨਮ ਤੋਂ ਹੀ ਆਈ. ਐਸ. ਵਾਈ. ਐਫ. ਨੂੰ ਭਾਰਤ ਅਤੇ ਸੀ. ਸੀ. ਵਿੱਚ ਹਿੱਸਕ ਕਾਰਵਾਈਆਂ ਨਾਲ ਜੋੜਿਆ ਜਾ ਰਿਹਾ ਹੈ, ਇਨ੍ਹਾਂ ਵਿੱਚ ਸਾਲ 1986 ਵਿੱਚ ਪੰਜਾਬ ਦੇ ਇੱਕ ਕਬਲਿਟ ਮੰਤਰੀ ਦਾ ਵੈਨਕੂਵਰ ਵਿਖੇ ਗੋਲੀ ਕੱਢਾ। ਮੁਜਾਹਰਾਕਾਰੀਆਂ ਵਿੱਚ ਅਮਰਜੀਤ ਦੇ ਵਾਸਨੀਕ ਅਤੇ ਪੰਥਕ ਕਮਰੀ ਦੇ ਬੁਲਾਰੇ ਡਾਕਟਰ ਅਮਰਜੀਤ ਸਿੰਘ ਸ਼ਾਮਲ ਸਨ। ਉਨ੍ਹਾਂ ਨੇ ਆਪਣੇ ਇੱਕ ਲੇਖ ਵਿੱਚ ਜੇ 7 ਮਈ ਨੂੰ (ਆਈ. ਐਸ. ਵਾਈ. ਐਫ. ਦੇ ਸਫੀ ਤੋਂ

\$1.25



Charhdi Kala Punjabi Weekly Vol 12, No. 13, 21-27 May 1998, p. 50.



Tools to help in multilingual multimedia information observation



- Language identification for written and spoken language
- Multilanguage speech to text processing
- Crosslanguage information search
- Crosslanguage information filtering
- Crasslanguage information clustering
- Machine translation



Multilingual vs crosslingual systems



- **Multilingual system** : a system that can process separately various languages by changing language resources
- **Crosslingual system**: it is a multilingual system that can in addition process a set of documents in various languages like if they were in a unique language

For example: a crosslanguage search can be interrogate in one query in one language and gives as answer documents in various languages.



Named entities play a key role in business intelligence activity



Named entities recognition :

- Locations (Paris)
- Person's name (Bill Clinton)
- Company name (Microsoft)
- Product name (LINUX)
- Dates (13 September 2007, yesterday)
- Other numeric information (ten kilos, 20%, 3 injured people)

But also domain specific : symptom, disease, test, part of the body, drugs for the medical domain



Named entities play a key role in business intelligence activity



L'**AFP** prend une participation minoritaire dans **Citizenside** (photos amateurs)

PARIS, 27 nov 2007 - L'**Agence France-Presse (AFP)** a pris une participation minoritaire au capital de **Scooplive**, plateforme technologique diffusant des images fixes et animées en provenance du public, dont le nom devient **Citizenside**, ont annoncé **mardi** l'**agence de presse** et la **plateforme**.

L'**AFP** et la société **IAM**, créée entre autres par l'ancien président de **France Télévisions** **Xavier Gouyou-Beauchamps**, ont chacune pris une participation de 30% dans **la société**. Les 40% restants sont détenus par les fondateurs de la plateforme: **Matthieu Stefani**, **Julien Robert** et **Philippe Checinski**.





Named entity translation

- Several spelling and transcription for cyrillic Characters
Чернобыль (RU), Чорнобiль (UA), Tchernobyl (Fr), Chernobyl (ISO US)
- Several spelling and transcriptions between Arabic and Latin Characters
فيصل , Faical, Faicel, Faisal, Faissal, Faycal, Faycel, Faysal, Fayssal
فيروز , Fairouz, Fayrouz, Ferouze, Ferouz
Jean, جون , جان



Named entities and multilingualism



Chinese name transcription

Conversion from Chinese writing into a latin-coded pronunciation produce few ambiguities but there is not several possible coding (piyin).

福建, Fuchien (Tongyong Pinyin Taiwan), Fujian (Hanyu Pinyin Chine)

The reverse transcription from piyin to ideograms is very ambiguous (up to 100 hypothesis). Example on a Hanyu pinyin representation of wen jiabao. The right spelling is validated by the web.

wen jiabao温 家宝 7,210 pages

wen jiabao文 家宝 219 pages

wen jiabao闻 家宝 32 pages

wen jiabao温 加宝 7 pages

wen jiabao文 加宝 4 pages

wen jiabao文 佳保 3 pages

wen jiabao温 家暴 3 pages

wen jiabao文 加保 2 pages

wen jiabao闻 家抱 1 page

wen jiabao闻 加宝 1 page

wen jiabao文 家暴 1 page

wen jiabao文 佳宝 1 page

wen jiabao温 家抱 1 page

wen jiabao温 加堡 1 page

wen jiabao温 假报 1 page



Crosslingual interrogation



Why crosslingual interrogation systems :

- Only one query to find all relevant documents.
- Use of mother tongue in query
- Access to document or part of documents even if the language is not understood. Such system can be linked with a MT system.



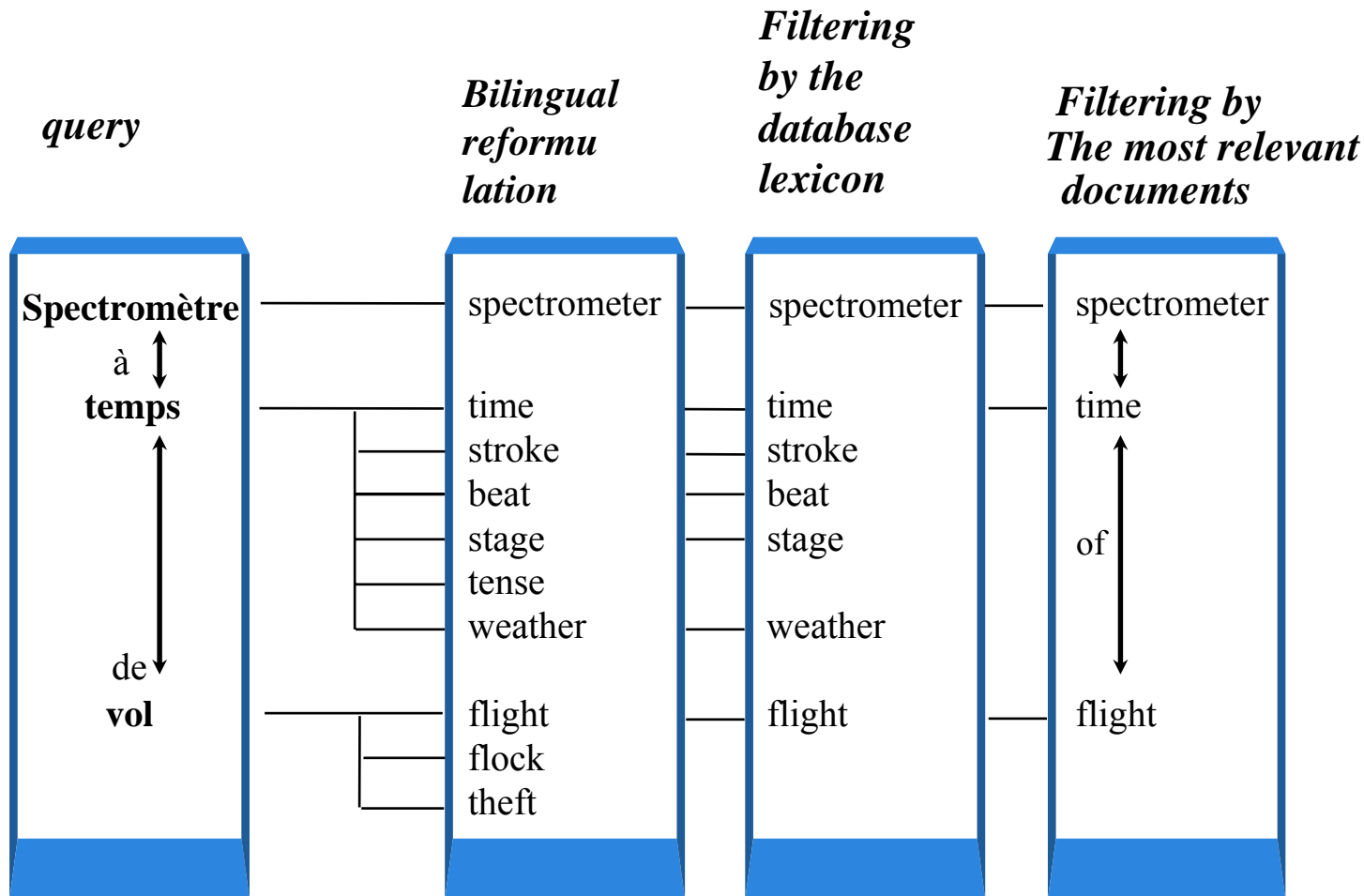


Crosslingual interrogation strategies

- use of bilingual documents or comparable documents to produce implicit equivalences between languages using statistical models (LSI approach)
- use of a Machine translation system to translate the queries or the documents
- use of bilingual dictionaries to perform bilingual reformulation
- use of a multilingual categorization



An example of the bilingual dictionary approach



Crosslingual interrogation (Alma European project)



The screenshot shows a web browser window titled "Recherche - Wanadoo". The address bar contains the URL: `http://132.167.34.85:8866/npcocoon/texte/query?config=alma&locale=fr&lang=fr&maxDocsByPage=20`. The main content area displays the logo for "NEWPHENIX RICH MEDIA" and the word "alma" in a bold font. Below this, there is a search input field containing the text "la gestion des ressources en eau". To the right of the input field is a dropdown menu set to "Français". A button labeled "lancer la recherche" is positioned below the search field. The browser's taskbar at the bottom shows several open applications, including "EPSON Scan", "Explorat...", "codageinfo...", and "2 Microsof...". The system clock indicates the time is 15:27.

Crosslingual interrogation (Alma European project)



Résultats de la recherche - Wanadoo

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris Média

Adresse <http://132.167.34.85:8866/npcocoon/texte/interface?query=la+gestion+des+ressources+en+eau&lang=fre>

NEWPHENIX
RICH MEDIA

[reformulation](#)

"la gestion des ressources en eau"

[nouvelle recherche](#)

1. [gestion_ressources_eau;](#)
(26 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))

2. [ressources_eau; gestion_ressources;](#)
(14 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))

3. [eau; gestion_ressources;](#)
(9 documents : [Arabe](#) [Anglais](#) [Espagnol](#))

4. [gestion; ressources_eau;](#)
(26 documents : [Arabe](#) [Anglais](#) [Français](#))

5. [gestion_ressources;](#)
(1 documents : [Anglais](#))

6. [ressources_eau;](#)
(8 documents : [Arabe](#))

7. [gestion; ressources; eau;](#)
(40 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))

8. [gestion; eau;](#)
(13 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))

9. [gestion; ressources;](#)
(6 documents : [Arabe](#) [Anglais](#))

10. [ressources; eau;](#)
(6 documents : [Arabe](#) [Français](#))

11. [eau;](#)
(5 documents : [Anglais](#) [Français](#) [Espagnol](#))

12. [ressources;](#)
(1 documents : [Anglais](#))

classe de pertinence n° 1 : [gestion_ressources_eau;](#) pages : 1 2

1. [القصة العالمية المعنية بالتنمية المستدامة](#)
langue : Arabe; origine : Url
... فيما يتعلق بصكوك السياسات والنواحي القانونية وغير ذلك من الصكوك من أجل تحقيق كفاءة إدارة موارد المياه العذبة؛ '4' تحيين فريق مهام معني بالمياه للاضطلاع بدور ... في "مراقبة المياه الإقليمية" التي سيوكل

2. [الماء في دول مجلس التعاون](#)
langue : Arabe; origine : Url
... تقييم المياه الجوفية وأنظمة المراقبة الكنتية والنوعية لها * تطوير وحماية وإعادة تفعيل وإدارة موارد المياه الجوفية * تقييم الموارد السطحية ومراقبتها * تطوير المياه * ... السطحية وإدارتها * تقييم فعالية

3. [منتدى إدارة الطلب على المياه](#)
langue : Arabe; origine : Url
... أبو شرار ويطيحي؛ إدارة موارد المياه في ظل الطلب التنافسي من القطاعات، دراسة حالة من الأردن؛ مجلة "وويز إنترناشيونال"، المجلد 27، رقم ...

4. [فصحة المياه في عملية السلام الفلسطيني الإسرائيلي](#)
langue : Arabe; origine : Url
... الفلسطينية في الضفة الغربية بحصة من مياه نهر الاردن . ويضيف أن الطرفين اتفقا على التعاون في إدارة موارد المياه وانظمة الصرف في الضفة الغربية ، مع الالتزام ... المتبادل للسلطات والمسؤوليات في منطقتيها كل على حدة

5. [ندوة المياه في تونس](#)
langue : Arabe; origine : Url
... على الصعيد المؤسسي : تحسين التنسيق بين مختلف الفعاليات ذات العلاقة بالمياه - تعزيز فترات إدارة موارد المياه وإشراك القطاع الخاص في أنشطة إعادة الاستخدام

6. [BIBLIOGRAPHICAL SYNTHESIS WATER SAVING STRATEGIES IN DEVELOPED COUNTRIES WITH WATER SCARCE RESOURCES](#)
langue : Anglais; origine : Url
... Policy issues in [water resources management](#). ...

7. [TECHNICAL SYNTHESIS INTERNATIONAL WATERSHED AGREEMENTS PRINCIPLES AND APPLICATIONS](#)
langue : Anglais; origine : Url
... 114 organisations or councils have been created for allocation and [management of water resources](#) [3]. ...

8. [TECHNICAL SYNTHESIS WATER RIGHTS MARKETS: PRINCIPLES AND RELEVANCE](#)
langue : Anglais; origine : Url
... In Globalization and [Water Resources Management](#): The Changing Value of Water. ...

list

Internet

Crosslingual interrogation (Alma European project)



Résultats de la recherche - Wanadoo

Fichier Edition Affichage Favoris Outils ?

Précédente Rechercher Favoris Média

Adresse <http://132.167.34.85:8866/npcocoon/texte/interface?query=la+gestion+des+ressources+en+eau&lang=fr>

NEWPHENIX
RICH MEDIA

[reformulation](#)

"la gestion des ressources en eau"
[nouvelle recherche](#)

- [gestion ressources eau;](#)
(26 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))
- [ressources eau; gestion ressources;](#)
(14 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))
- [eau; gestion ressources;](#)
(9 documents : [Arabe](#) [Anglais](#) [Espagnol](#))
- [gestion; ressources eau;](#)
(26 documents : [Arabe](#) [Anglais](#) [Français](#))
- [gestion ressources;](#)
(1 documents : [Anglais](#))
- [ressources eau;](#)
(8 documents : [Arabe](#))
- [gestion; ressources; eau;](#)
(40 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))
- [gestion; eau;](#)
(13 documents : [Arabe](#) [Anglais](#) [Français](#) [Espagnol](#))
- [gestion; ressources;](#)
(6 documents : [Arabe](#) [Anglais](#))
- [ressources; eau;](#)
(6 documents : [Arabe](#) [Français](#))
- [eau;](#)
(5 documents : [Anglais](#) [Français](#) [Espagnol](#))
- [ressources;](#)
(1 documents : [Anglais](#))

[9. conventions RELATIVES A La gestion des Bassins versants internationaux. principes et modalités](#)
langue : Français; origine : Office international de l'eau
... 14 Etats du sud-est de l'Afrique, est en train de développer une intégration régionale pour la **gestion des ressources en eau** et a signé, en 1995, le Protocol on Shared Watercourse System. ...

[10. LA GESTION DE L'ENVIRONNEMENT DANS LES INDUSTRIES FRANCAISES : ETAT DES LIEUX DES REGLES ET PRATIQUES](#)
langue : Français; origine : Office international de l'eau
... : prélèvements et rejets La loi sur l'eau (loi n°92-3 du 3 janvier 1992) promeut la **gestion équilibrée de la ressource en eau**, à usage tant industriel que domestique. ...

[11. LA GESTION DES BOUES D'EPURATION DANS L'UNION EUROPEENNE](#)
langue : Français; origine : Office International de l'eau
... 1986 : Loi fondamentale relative à l'Environnement. - 2/07/87 : Loi relative à la **gestion des ressources en eau** (création de 14 bassins hydrographiques). - 07/91 : Décision Ministérielle (80568/4225) qui ...

[12. LES MESURES PRISES POUR LUTTER CONTRE LA SUREXPLOITATION DES NAPPES EN FRANCE ET DANS DIFFERENTS PAYS EUROPEENS](#)
langue : Français; origine : Office international de l'eau
... est souvent considéré comme étant le plus efficace et correspond à une vision moderne de la **gestion des ressources en eau** (gestion intégrée, durable et concertée au niveau d'un bassin versant), ainsi qu'au principe ...

[13. INVENTAIRE, STATUTS ET FONCTIONNEMENT DES STRUCTURES INTERNATIONALES DE GESTION DES GRANDS LACS INTERNATIONAUX ET DE LEUR BASSIN DANS LE MONDE](#)
langue : Français; origine : Office international de l'eau
... compétent pour les questions de voisinage entre Genève et la France voisine, ce qui comprend la **gestion des ressources en eau**. ...

[14. LES MARCHES DE DROITS D'EAU : PRINCIPES ET PERTINENCE](#)
langue : Français; origine : Office international de l'eau
... pour les abandonner quelques années plus tard lorsque le droit sera à nouveau vendu ? Une **gestion bien planifiée des ressources en eau** qui identifierait les besoins futurs ne permettrait-elle pas de limiter les aménagements ...

[15. CONSEQUENCES DE LA NAVIGABILITE SUR LA GESTION TECHNIQUE ET INSTITUTIONELLE DES COURS D'EAU](#)
langue : Français; origine : Office international de l'eau
... attribue aux voies navigables deux enjeux publics majeurs : les transports et la **gestion de la ressource en eau**. ...

[16. LES ORGANISMES DE GESTION DE L'EAU PAR BASSIN HYDROGRAPHIQUE DANS LES PAYS DE LA MEDITERRANEE](#)
langue : Français; origine : Office international de l'eau

Internet

Crosslingual filtering



Filtering systems are useful tool for Business intelligence because of the observation of the environment along the time.

Main difference between information filtering and information retrieval

Information Filtering	Information Retrieval
Long and elaborated queries (profiles)	short queries
Profiles are used for along the time	one short query
Search for a domain activity	precise query

But the crosslanguage relevance evaluation technology is quite the same.



Crosslanguage clustering



Because the crosslanguage interrogation technology is able to compute a semantic proximity between 2 documents that are in different languages, crosslanguage clustering can be done using classical clustering algorithms

- computation of proximity matrix between documents but only between the closest ones.
- computation of crosslingual clusters with a standard clustering algorithm

The crosslingual clustering can show event that are not visible in each language (weak signals)



Machine translation systems



Machine translation is a dream since the first years of computers. The state of the art gives imperfect solutions nevertheless this is very useful to have a good idea of the subject treated by a document.

Mainly used MT systems are based on linguistic processing that analyze the source language, perform a transfer between the two languages and produce a translation by a generation in the target language

In this category is SYSTRAN, or Reverso can tested on the web or Sakhr for Arabic-English translation.





translation by Reverso (Softissimo) of an AFP news:

L'Orchestre Philharmonique de New York, sous la direction de son chef Lorin Maazel, a donné mardi à Pyongyang un concert historique visant à contribuer à une amélioration des relations entre les Etats-Unis et le régime communiste de la Corée du Nord.

The Philharmonic Orchestra of New York, supervised by his(her) leader Lorin Maazel, gave on Tuesday in Pyongyang a historic concert to contribute to an improvement of the relations between the United States and the communist regime of North Korea.



Evolution of the MT systems



- **Translation memories:** they use already translated documents to improve human translation activity especially in case of different version of a same document.
- **Statistical MT :** they also use already translated text. Very large bilingual corpora are needed. They learn from these bitexts correlations between n grams (n words) in a source language with n grams in the target language. Pioneers in this approach was ETWH (University of Aachen). Now this technology is used by Google for English-Arabic translation
- **Translation based on a crosslanguage interrogation of the full target language.** No bilingual corpora needed. Possibility to translate from a rare language to a large one.



Example of a statistical translation by Google:

Original text :



Turkish troops pressed a major ground offensive against Kurdish rebels in neighbouring Iraq for a third day Saturday as Ankara said dozens of rebels and at least five soldiers had been killed.

Translation into Arabic:

القوات التركية من ضغوط كبيرة في الميدان الهجومية ضد المتمردين الاكراد في العراق المجاور لثالث يوم السبت انقرة كما قال عشرات من المتمردين وخمسة جنود على الاقل قد لقوا مصرعهم.

Reverse translation:

Turkish troops from great pressure on the ground offensive against rebel Kurds in neighbouring Iraq for the third day Saturday Ankara also said dozens of rebels and at least five soldiers were killed.





Speech to text systems especially to process radio or TV news are effective enough to be used to quickly locate the part of news bringing information on an event or a domain of interest.

Of course, crosslingual systems that are used for written texts can be used to process audio converted into text with a time code for each word.

Remark : text to speech processing in Arabic giving spelling with vowels, provide a much less ambiguous data for the information retrieval.



Example of multimedia crosslanguage interrogation



Example multimedia crosslingual search:

Video are taken from the web (signal is better by satellite broadcasting) in French; English and Arabic.

Text to speech by LIMSI/VECSYS,

Crosslingual information retrieval by NewPhenix

