



Anatomy of Commercial CLIR Applications

*CLEF Workshop 2002
Rome, Italy
September 19, 2002*

**David A. Evans¹, Gregory Grefenstette¹,
Joop van Gent², Yan Qu¹**

¹Clairvoyance Corporation & ²Irion Technologies



Many Thanks!

- **Carol Peters (CLEF)**
- **Susan Feldman & Steve McClure (IDC)**
- **Páraic Sheridan (MNIS-TextWise Labs)**
- **Peter Schäuble (Eurospider)**
- **Debbie Moran & Lynnae Evans
(Clairvoyance)**



**The World is (Finally)
Clamoring for CLIR...**

NOT!!



Reflecting on Commercial CLIR

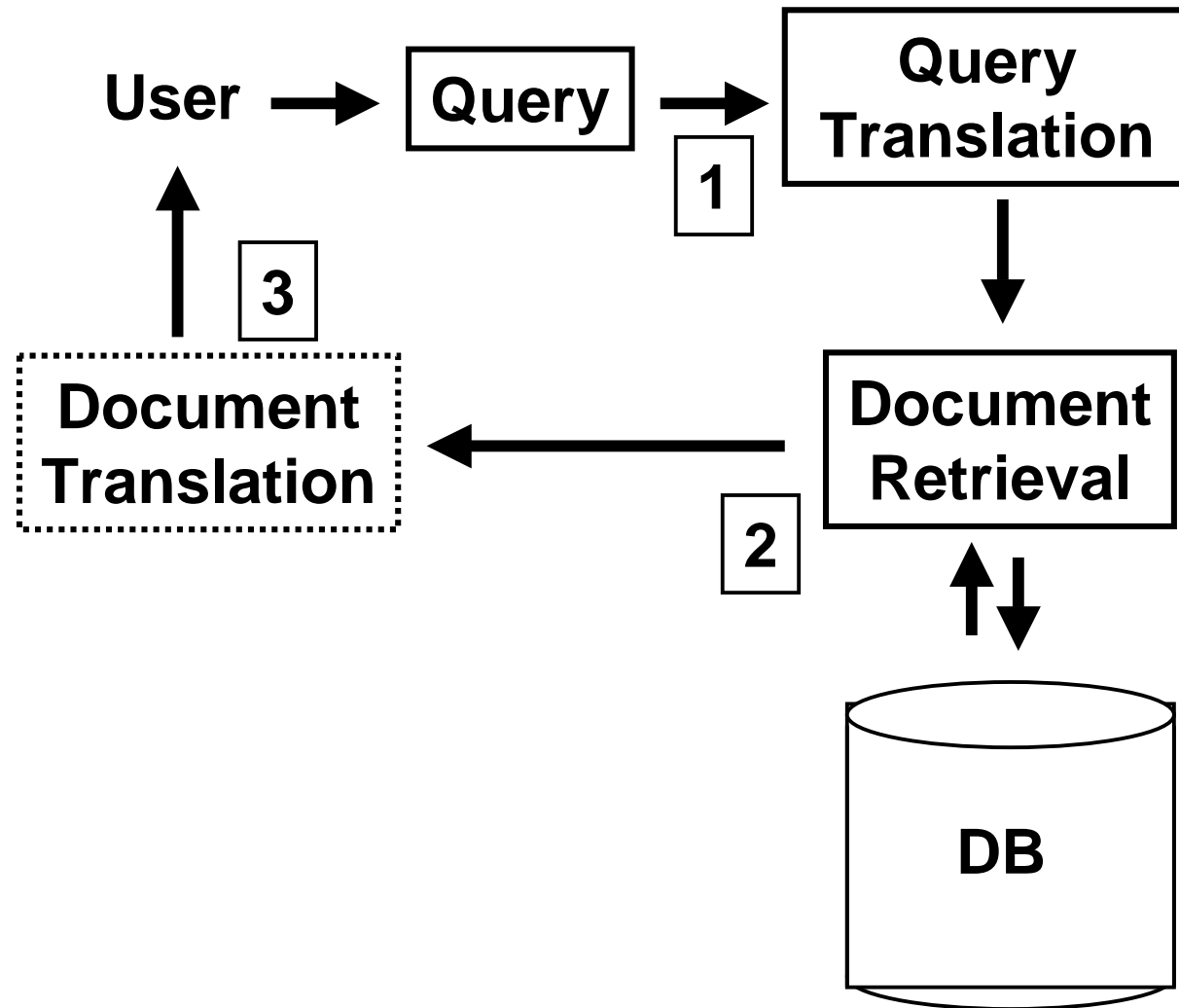
- **What is CLIR?**
- **What is the state of the market?**
- **What's in a commercial application?**
- **Specific Cases**
 - Cindor
 - AnswerWorks
 - Lirix (Grefenstette)
 - TwentyOne (van Gent)
- **Future directions**
 - Pidgin (van Gent)
 - InSiteProxy (Qu)
- **Concluding thoughts**
- **Discussion?**



What is CLIR?



CLIR Functional Architecture



A complete CLIR system will do 1+2+3.

A minimal CLIR system will do 1+2.

Other combinations of functions do not yield CLIR systems.

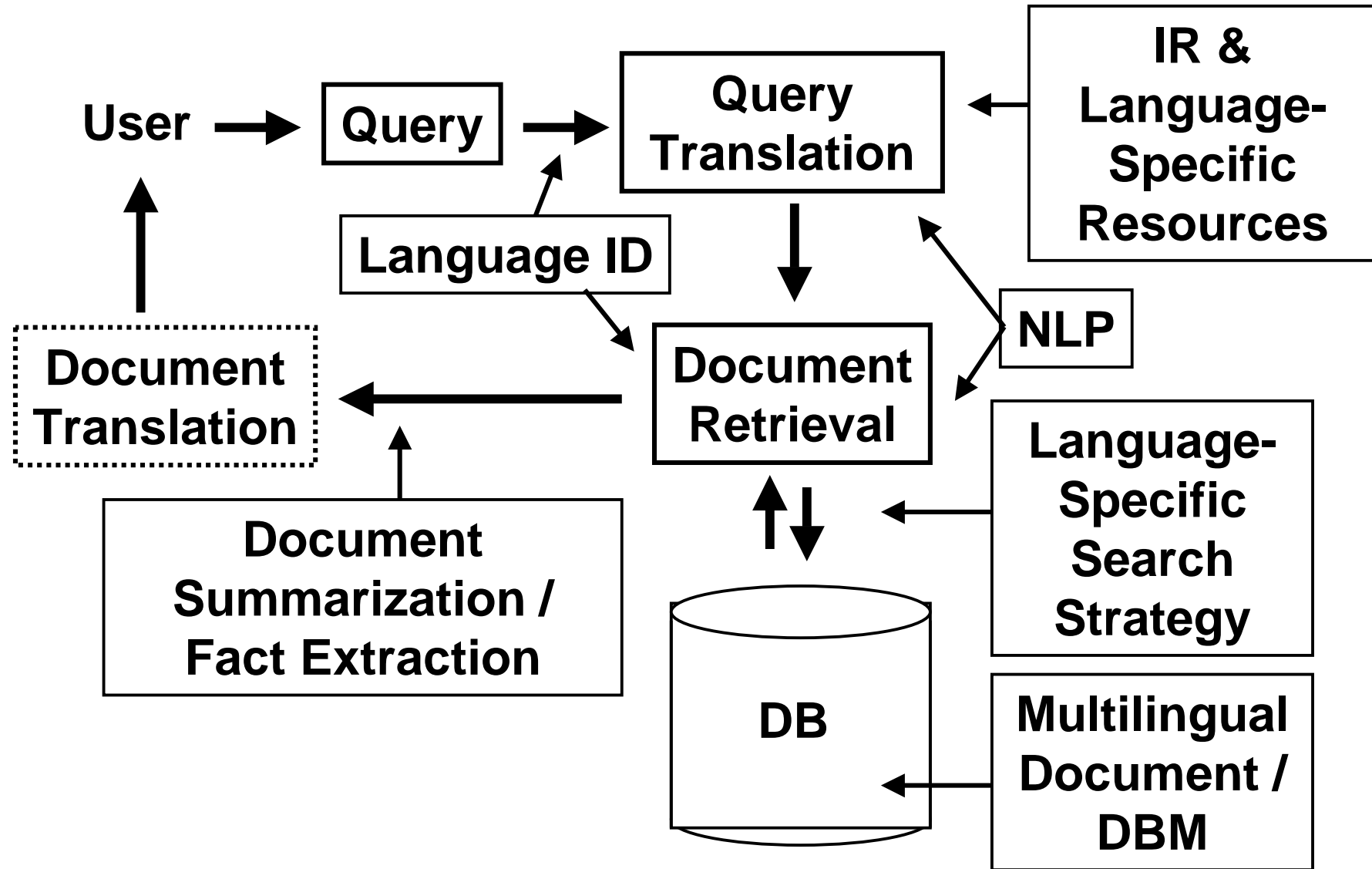


CLIR \neq MT + IR...

- **Problems for Machine Translation (MT)**
 - Queries are minimal texts
 - Alternative interpretations of a query are sometimes better than one
 - Best methods for MT not always available for all language pairs
- **Problems for Information Retrieval (IR)**
 - Most “efficient” IR may not be applicable to all languages
 - Need for language-specific coordination of indexing and retrieval



CLIR Functional Architecture





What is the state of the CLIR Market?



Market Trends

Multi-Lingual Environment

- **English becoming less dominant on WWW**

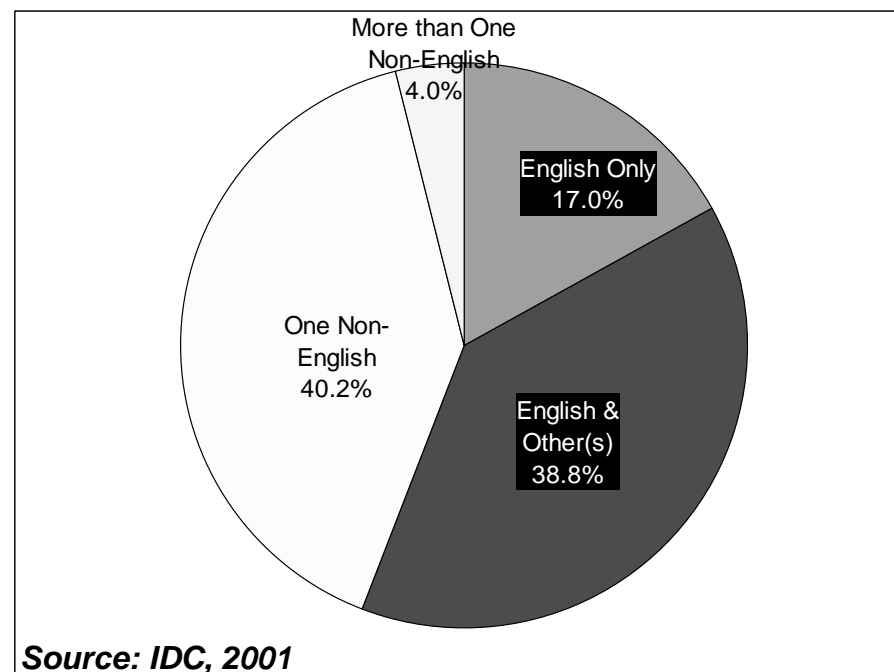
- Many “words” of non-English languages now on web pages

	Oct 1996	Ratio to English	Aug 1999	Ratio to English	Mar 2001	Ratio to English
English	6,082,090,000	1.000	28,222,100,000	1.000	76,598,718,000	1.000
German	228,938,428	0.038	1,994,229,409	0.071	7,035,850,000	0.092
French	223,316,023	0.037	1,529,795,169	0.054	3,836,874,000	0.050
Spanish	104,319,158	0.017	1,125,646,460	0.040	2,658,631,000	0.035
Italian	123,555,682	0.020	817,270,444	0.029	1,845,026,000	0.024
Portuguese	106,167,245	0.017	589,391,943	0.021	1,333,664,000	0.017
Finnish	20,647,404	0.003	107,260,274	0.004	326,379,000	0.004

Source: Grefenstette, TIA 2001

- **Commercial / Business Web Sites increasingly in languages other than English**

- Almost 45% of business web sites worldwide are in languages other than English
- Only about 17% of all business web sites are exclusively in English



Source: IDC, 2001



Market Trends

General Trends

- **Continuing growth of the Web**
 - Increasing number of sites
 - Increasing number of users
- **Continuing orientation of business to “self service”**
- **Internationalization of trade, consumer activities**
- **Improvements in technology (including HLT)**
- **Government emphasis on multi-language efforts**



Market Trends

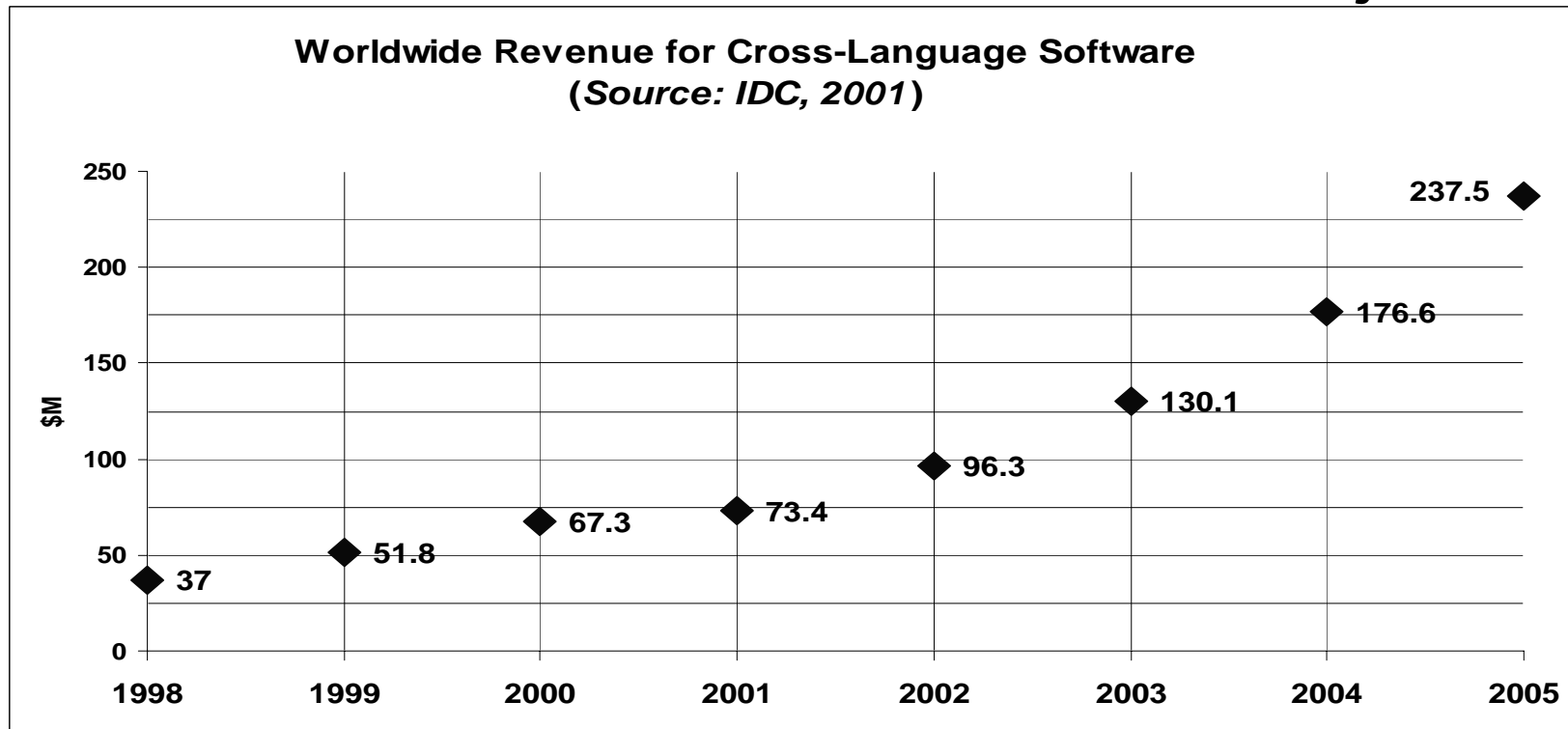
Demand for CLIR

- **Distinct sectors and applications**
 - Government (e.g., intelligence agencies; legal/legislative requirements)
 - Technical (e.g., research organizations— pharmaceutical, chemical, engineering groups; patent attorneys)
 - General business (e.g., competitive intelligence; business communications)
 - Services (e.g., customer support)
- **Public-sector (government) spending is up**
 - U.S.: TIDES, Communicator, ROAR, others
 - E.U.: Euromap, Elsnet, etc.
- **Revenue for “cross-language software” (including MT) growing 30% annually (*IDC*)**



Market Trends

Revenue Projections



- **CLIR-specific (non-MT) revenue currently < 10% total**
- **Worldwide revenue for CLIR products 2002–2003 likely under \$15M**



Market Trends

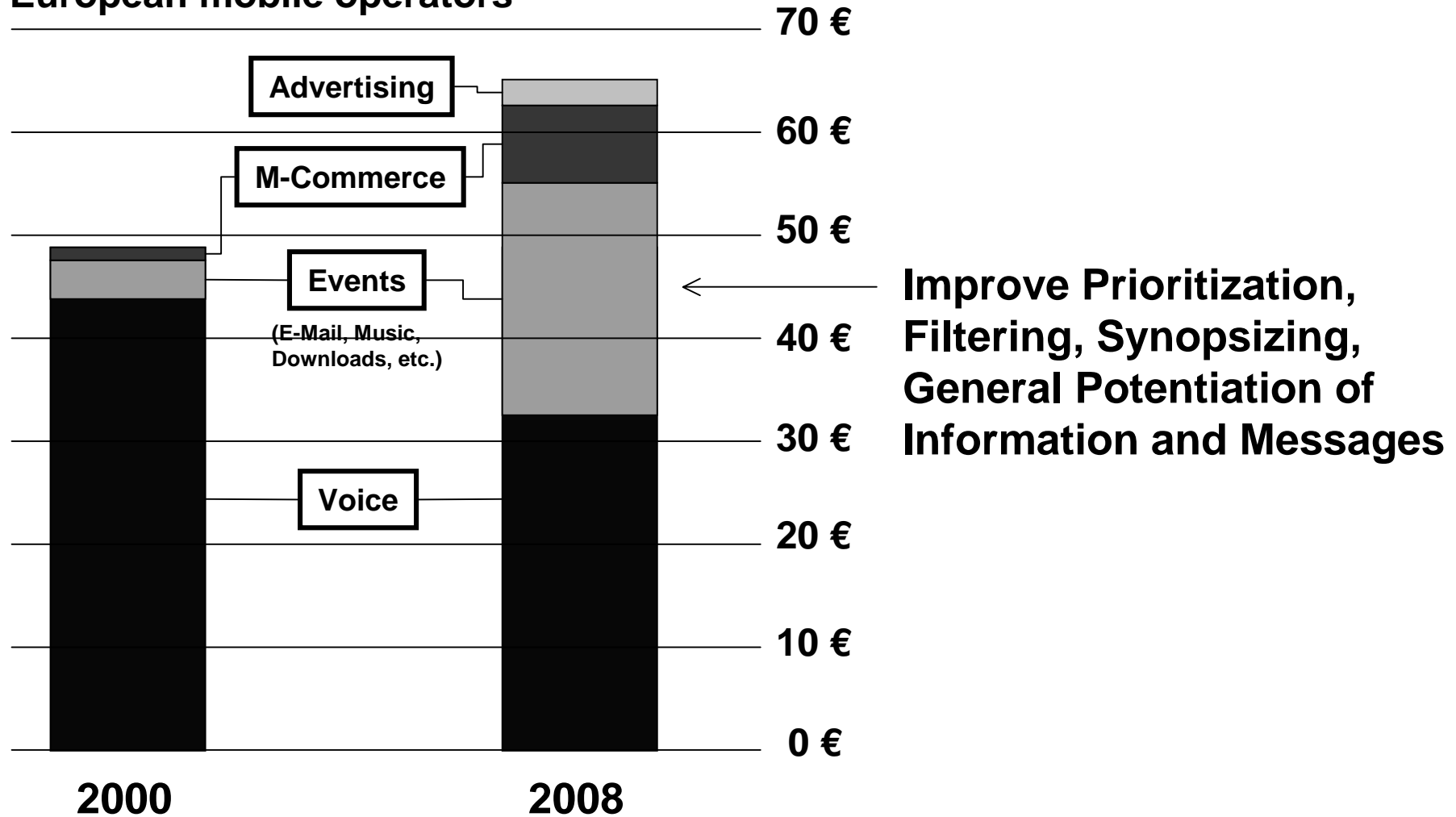
CLIR Application Requirements

- **Asian Languages**
 - Japanese, Chinese, ..., Arabic
- **Effective translation of retrieved information**
- **Focus on task-specific applications**
 - (FA)QA
 - Patent interpretation
 - Customer support
- **Transparency**
 - Minimal user interaction
 - Speed
 - Fluency
- **Speech \Rightarrow (Text \Rightarrow Text \Rightarrow) Speech**



Momentum in Wireless IM

Forecast average revenues per user per month,
European mobile operators



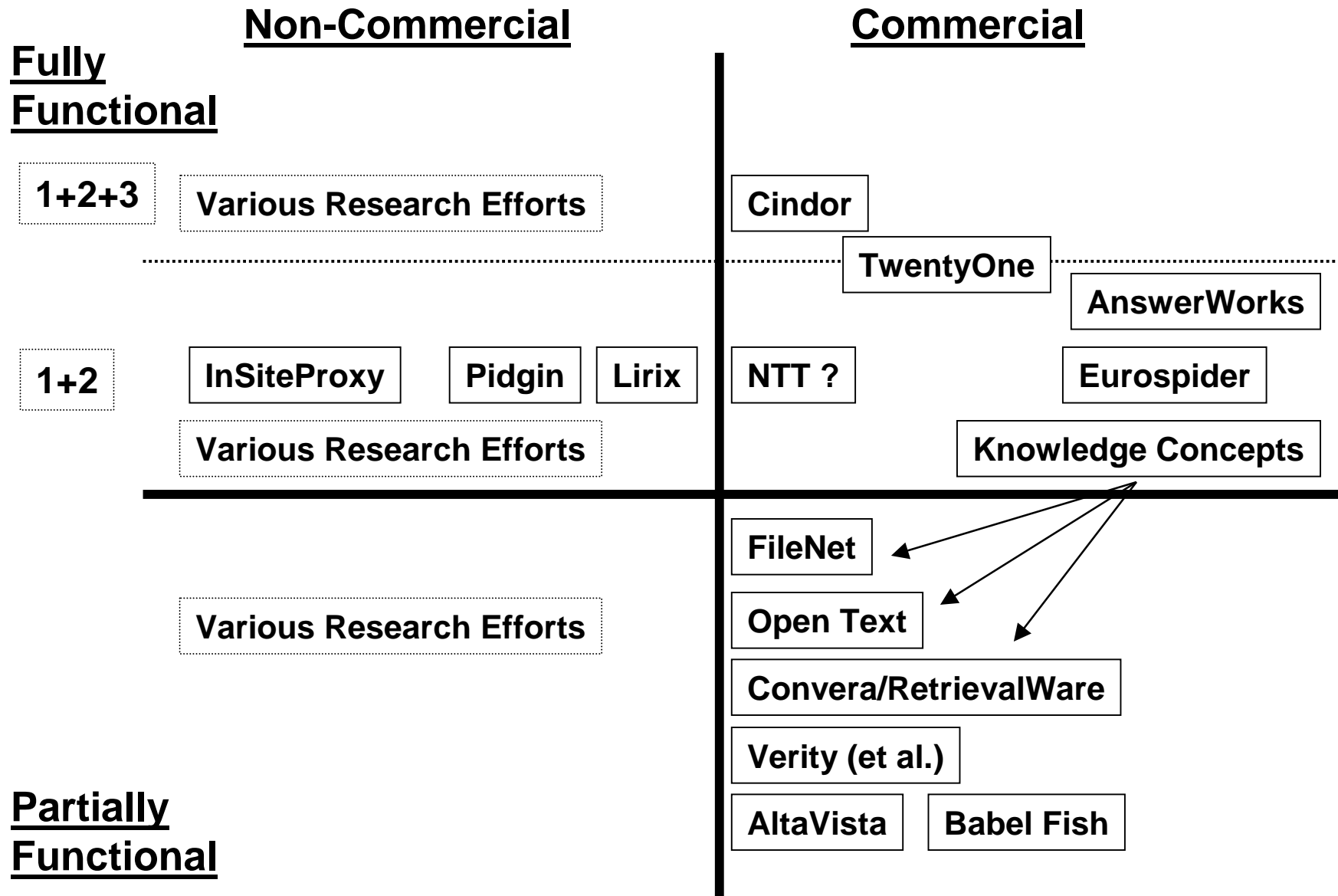
Source: Wireless Internet Report, Morgan Stanley Dean Witter, via *The Economist*, October 14, 2000



A brief survey of CLIR systems...



Partial Survey of CLIR Systems





Cindor (MNIS-TextWise Labs)



General Characterization

- **Full CLIR (1+2+3), targeting document retrieval**
- **Core technology components**
 - Language analysis—proprietary; using InXight LinguistX for tokenization, stemming, POS tagging in several (foreign) languages
 - Conceptual Interlingua—proprietary; language-neutral lexical representation, using modified version of WordNet; includes genre/domain typing and supports word-sense disambiguation, proper-noun ID, phrase detection
 - Search Management—proprietary; includes query analysis; may be replaced by generic SE
- **Document translation via “Gist-in-Time” (Alis)**



- **Six languages**
 - English, French, German, Italian, Japanese, Spanish
 - Chinese under development
- **Query interpretation *not* based on MT system**



Commercial Observations

- **Major effort to market system and toolkit in early 2001**
- **Extended trial at Unilever (NL) Food Science research group to search Japanese patents**
 - Gisting of retrieved documents inadequate
 - Complete (quality) MT for Japanese not possible
- **Experience with Hong-Kong-based financial services company**
 - Chinese→English MT not adequate
- **Market “not there yet” in 2001**
 - Infrastructure issues still in early stages
 - MT functionality critical, but not sufficient
- **Suspended commercial push mid-2001**



Cindor Illustration

CINDOR: Ask a question: - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

Back Forward Reload Stop <http://demo.cindorsearch.com/cindor/Cindor?FUNC=QRY> Search Print

Home Netscape Search Shop Bookmarks Net2Phone

CiNDOR HELP

Username: parak

- ▶ New question
- ▶ Modify question
- ▶ Return to headlines
- ▶ Exit Cindor

Ask a question:

Step: 1	Collections: (Use ctrl or shift to select multiple corpora) Agence France Presse (5/94 - 1/96)--es Associated Press 2/88 - 12/90--en ELNORTE_TREC--es KYODO News Service/Dow Jones Telerate (11/94 to 6/95)--ja NACSIS English--en NACSIS Japanese--ja Nihon Kezai Shimbun News data (12/93 to 11/94)--ja Schweizerischen Depeschenagentur (1/88 - 12/90)--fr	Language codes: en = English ja = Japanese
Step: 2	Ask a question in: English	
Step: 3	Documents retrieved per language: 25	
Step: 4	Automatically translate all results: <input type="checkbox"/>	
Step: 5	Enter a question in the form of a sentence, capitalizing all proper nouns Status of negotiations on world trade agreements.	
Step: 6	Perform search	

Document Done [1.051 secs]



Cindor Illustration

CINDOR: Search Results - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

Back Forward Reload Stop Search Print

Home Netscape Search Shop Bookmarks Net2Phone

CINDOR HELP

Username: paralc

▶ New question

▶ Modify question

▶ Exit Cindor

Your query: Status of negotiations on world trade agreements.

Translate headlines to: English

View results in: Japanese

Relevance

Rank	Length	Result
1	2337	WTO 1月発足を正式決定 設立実施会議 <i>Date: 1994年12月8日 Source: Kyodo News Service/Dow Jones Teleterm</i>
2	5806	ニュースサマリー・ 1 (27日) <i>Date: 1995年6月26日 Source: Kyodo News Service/Dow Jones Teleterm</i>
3	5465	日米は世界の繁栄に共同責任——米・ 進歩的政策研究所ロバート・ J・ シャピロ氏。 <i>Date: 1993年12月26日 Source: Nihon Kezai Shinbun</i>
4	13616	94年版通商白書 (要旨) <i>Date: 1994年5月17日 Source: Nihon Kezai Shinbun</i>
5	6429	日米明日への視点 (1) ガット事務局長サザーランド氏—寛容な貿易、世界へ責務。 <i>Date: 1994年10月25日 Source: Nihon Kezai Shinbun</i>
6	2270	期限迫る金融市場開放交渉 WTOにもう一つの試練 <i>Date: 1995年6月20日 Source: Kyodo News Service/Dow Jones Teleterm</i>
		自由貿易の重さかみしめ再出発を (社説)

Document Done [5.628 secs]



Cindor Illustration

The screenshot shows a Netscape browser window with the Cindor search interface. The search query is 'Status of negotiations on world trade agreements.' The results are sorted by relevance and displayed in a table with columns for Rank, Length, and Result. The results include news summaries and articles from 1993 and 1994, primarily sourced from the Kyodo News Service and Nihon Kezai Shimbun.

CINDOR HELP

Username: paralc

- ▶ New question
- ▶ Modify question
- ▶ Exit Cindor

Your query: Status of negotiations on world trade agreements.
Translations provided by: Alis Gist-in-Time™
View results in: Japanese

Relevance

Rank	Length	Result
1	2337	WT O1 month start formal decisive establishment execution meeting <i>Date: December 8th of 1994 Source: Kyodo News Service/Dow Jones TeleRate</i>
2	5806	News summary * 1 (27 days) <i>Date: June 26th of 1995 Source: Kyodo News Service/Dow Jones TeleRate</i>
3	5465	As for Japan and America joint responsibility to prosperity of the world -- the United States & progressive policy laboratory Robert J Shapiro <i>Date: December 26th of 1993 Source: Nihon Kezai Shimbun</i>
4	13616	94 year edition commerce white paper (gist) <i>Date: May 17th of 1994 Source: Nihon Kezai Shimbun</i>
5	6429	Point of view (1) GATT Executive Office length fU1[land person to Japanese-American tomorrow - magnanimous trade, to world duty <i>Date: October 25th of 1994 Source: Nihon Kezai Shimbun</i>
6	2270	In the financial market open negotiation WTO which time limit is approached another ordeal <i>Date: June 20th of 1995 Source: Kyodo News Service/Dow Jones TeleRate</i>
7	3705	Weight of free trade you bite and close restarting (editorial) <i>Date: December 16th of 1993 Source: Nihon Kezai Shimbun</i>
		WTO financial market negotiation the United States is dissatisfied in measure of the

Document Done [1.713 secs]



Cindor Illustration

The screenshot shows a Netscape browser window titled "CINDOR: Full Document - Netscape 6". The address bar contains the URL: <http://demo.cindorsearch.com/cindor/Cindor?FUNC=DCLO&CID=9412080145&DL=ja&SL=ja&CORPUS=K>. The browser interface includes a menu bar (File, Edit, View, Search, Go, Bookmarks, Tasks, Help), a toolbar with navigation buttons (Back, Forward, Reload, Stop), and a search bar. The main content area displays the Cindor search results for the query "Status of negotiations on world trade agreements." The results include a document titled "WTO 1月発足を正式決定 設立実施会議" (WTO 1st Jan launch formally decided, establishment implementation meeting). The document's citation date is 1994年12月8日 (December 8, 1994) and the source is Kyodo News Service/Dow Jones Teletype. The document text discusses the establishment of the WTO, mentioning the Uruguay Round and the GATT agreement.

Previous Next document

Translate this document to: using

Your query:

WTO 1月発足を正式決定 設立実施会議

Citation Date: 1994年12月8日
Source: Kyodo News Service/Dow Jones Teletype

【ジュネーブ8日共同】ウルグアイ・ラウンド合意署名各国代表は8日、ジュネーブで世界貿易機関(WTO)設立協定実施会議を開き、来年1月1日のWTO発足を全会一致で正式決定した。1948年の発足以来、世界貿易の発展を助け、ルールの監視をしてきた関税貿易一般協定(GATT)に代わり、紛争処理などでより強力な機能を備えたWTOがその役割を担う。

WTOは、86年から7年以上のラウンド交渉で設立が決まり、1月発足を目指して各国が国内批准手続きを進めてきた。第2次大戦後に設立寸前まで行きながら流れた国際貿易機関(ITO)の構想が、冷戦後にやっと実現にこぎつけたことになる。

ITOの挫折によって暫定的な機関として存在してきたGATTと違い、WTOは正式な国際機関としての体裁を整える。全加盟国が反対しない限り、一定期間を区切って次々と手続きや対抗措置の発動を進める強力な紛争処理機能を持つのが特徴。これにより、WTOの手続きを経ない一方的制裁措置の発動に歯止めをかける。

Document Done [2.654 sec]



Cindor Illustration

CINDOR HELP

Username: paraic

- ▶ New question
- ▶ Modify question
- ▶ Return to headlines
- ▶ Exit Cindor

Previous Original Document Next document

Translations provided by: Alis Gist-in-Time™

Your query: Status of negotiations on world trade agreements.

WT O1 month start formal decisive establishment execution meeting

Citation Date: December 8 of 1994 day
Source: Kpodanewsservice/lowjonestelerat B

< Geneva 8 day cooperation > Uruguayan round agreement signature each national representation on the 8th, opened international trade system (WTO) establishment agreement execution meeting in Geneva, decided the WTO start January 1st of next year formally in all the meeting agreements. Since starting of 1948 years, development of international trade is helped, the WTO which has more powerful function in such as dispute treatment in place of the tariff trade general agreement (GATT) which watches rule, bears the role.

As for the WTO, from 86 establishment was decided with round negotiation of 7 years or more, you aimed toward January start and each country advanced domestic ratification procedure. While going to before establishment sum/size after the Second World war, conception of the world trade system (ITO) which flowed, means after cold war to reach at last to actualization.

The WTO arranges appearance as the formal international institution different from the GATT which exists as the provisional system depending upon the collapse of the ITO. If all the participation countries do not oppose, dividing fixed period, one after another having the powerful dispute treatment function which advances the motion of procedure and opposition measure feature. Because of this, it does not pass the procedure of the WTO, the brake

Document Done [5.329 secs]



AnswerWorks (WexTech \Leftarrow Knexys)



AnswerWorks

General Characterization

- **CLIR (1+2(+3?)), targeting customer support**
- **Core HLT components (via Knexys)**
 - Language analysis—proprietary; language-specific tokenization, normalization/disambiguation, phrase identification
 - ConceptNet—proprietary; multi-lingual lexicon (mapping to English?); weak and strong synonyms supported
 - Indexing and retrieval—proprietary; index of data based on “linguistic image” (= terms processed under language analysis); matching queries to documents (answers) based on linguistic-image similarity
- **Eight languages**
 - English, French, German, Italian, Dutch, Spanish, Portugese, Japanese; (others under development)



Lirix (Xerox)



Linguistic Information Retrieval Technology

- **Mono-lingual and cross-lingual search engine**
- **LIRIX uses advanced linguistic techniques:**

- Query expansion:

e.g., “election” ⇒ “elect, elector, elected, etc.”

- Multi-word dictionary lookup:

e.g., “ignition key” ⇒ “clé de contact”

- Relation detection:

e.g., Query = “presidential election”

“...to elect its first President.”

OK (verb/obj)

“The President has been elected...”

OK (subj/verb)

“The elected government of President X”

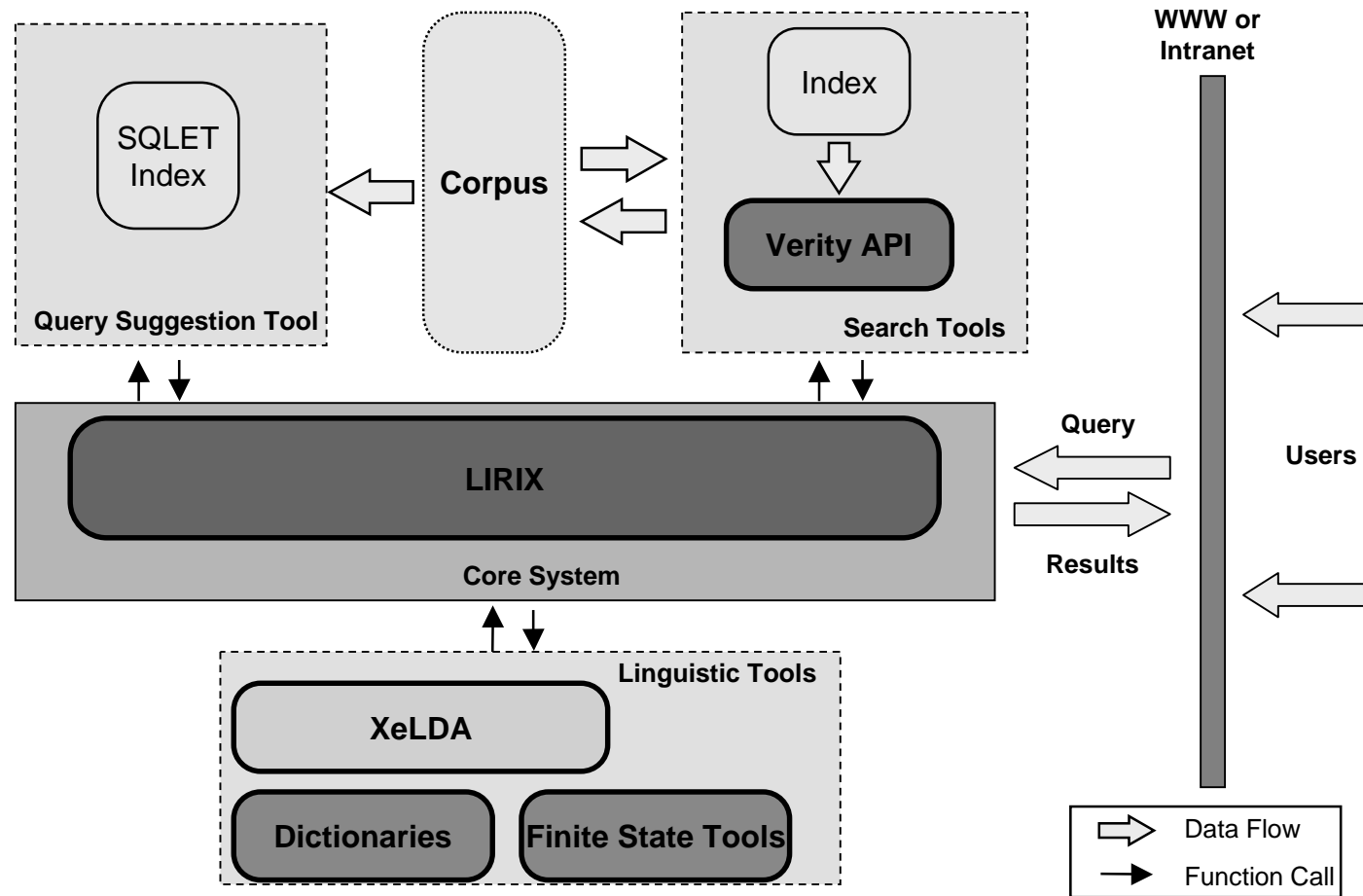
No relation

- **XRCE’s web site search engine:**

<http://www.xrce.xerox.com/search.html>



Lirix Architecture





Lirix Illustration

Recherche XRCE - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Address <http://www.xrce.xerox.com/search.fr.html> Go

Liens [Accueil Wanadoo] [Assistance] [Bazar bleu] [Mon compte] [Pages Perso] [Tout Minitel]

xrce Recherche

XRCE a mis en place un **moteur de recherche multilingue** appelé **LIRIX** pour vous aider a trouver les informations qui vous intéressent sur ce site. LIRIX permet d'améliorer considérablement la pertinence des résultats par rapport au moteur de recherche de Verity, sur lequel il se base. Il a également la capacité unique d'effectuer des recherches multilingues: tapez quelques mots en Français, et vous obtiendrez des documents à la fois en Français et en Anglais!

Une page de présentation de LIRIX est disponible (en anglais). N'hésitez pas à consulter [l'aide en ligne](#) et à [nous contacter](#) pour plus d'informations.

English **Français** Español Italiano Deutsch

Tapez une requête courte en Français [Recherche Avancée](#)

gestion de connaissances

[Aide](#) **Rechercher**

Utilisation : Tapez un ou plusieurs mots en Français, éventuellement séparés par une virgule, puis pressez le bouton "Rechercher".

Internet



Lirix Illustration

The screenshot shows a Microsoft Internet Explorer browser window displaying search results for 'xrce'. The browser's address bar contains the URL: `anking=n&postfilter=distance&queryexpand=y&coll=1%2C2&query=gestion+de+connaissances&Rechercher.x=54&Rechercher.y=11`. The search results page features the 'xrce' logo and the title 'Résultats de la Recherche'. A sidebar on the left lists navigation links under 'Xerox Links' and 'Liens XRCE'. The main content area displays three search results, each with a title and an 'Extraits' (snippets) section.

xrce Résultats de la Recherche

Anglais : Résultats

Xerox Links

- [Xerox Home](#)
- [Xerox Research & Technology](#)
- [Related Business Groups](#)

Liens XRCE

- [Accueil](#)
- [Showroom](#)
- [R&D](#)
- [Publications](#)
- [Membres](#)
- [Chercher](#)
- [Contact](#)
- [Commentaires](#)
- [Emploi](#)
- [Plan du Site](#)

- [%IPS-Adobe-3.0](#)
Extraits: ...interface, members of the Pump can ask questions like: ?What has such-and-such well-known expert in **Knowledge Management** submitted to that community recently?? How have other people reacted to that article I recommende...
...lassified into the CSCW community and is returned a list where many items are cross-classified into **Knowledge Management** and WWW programming, s/he may well be motivated to explore what else these communities and its memb...
...f Agents?97 (Marina del Rey CA, February 1997), ACM Press. Glance, N., Arregui, D. and Dardenne M. **Knowledge Pump: Supporting the flow and use of knowledge**, in Information Technology for **Knowledge Management**. Eds. U. Borghoff and R. Pareschi, New York: Springer-Verlag, 1998. Glance N.S. and Huberman B.A. ...
- [Sheet: 8,5? x 11?](#)
Extraits: ...of a single, isolated user?s interaction with an information base. The recent surge of interest in **Knowledge Management**, however, has contributed to enlarging the view on processes involving information discovery, shari...
- [Campiello - New user interface approaches for community networks](#)



TwentyOne (Irion Technologies)



Irion CLIR End-User Applications

- Adjust:**
cross-lingual filtering and classification
- TwentyOne:**
cross-lingual information retrieval
- Pidgin:**
cross-lingual dialog & chat



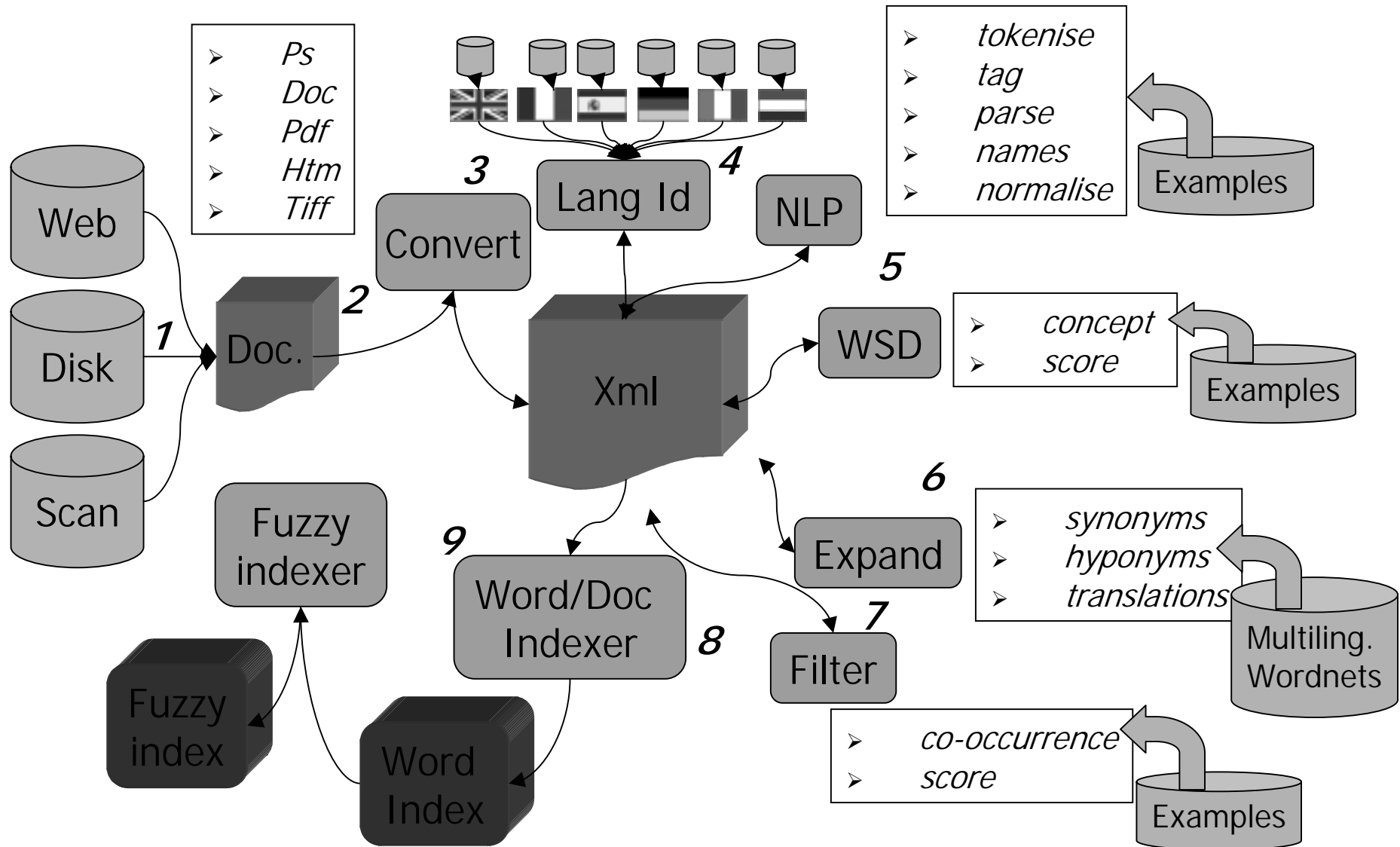
TwentyOne

General Characterization

- **Cross-lingual information retrieval system for 6 languages**
- **Automatic language detection**
- **Linguistic analysis and index enhancement**
- **Document retrieval and phrase retrieval**
- **Fuzzy search**

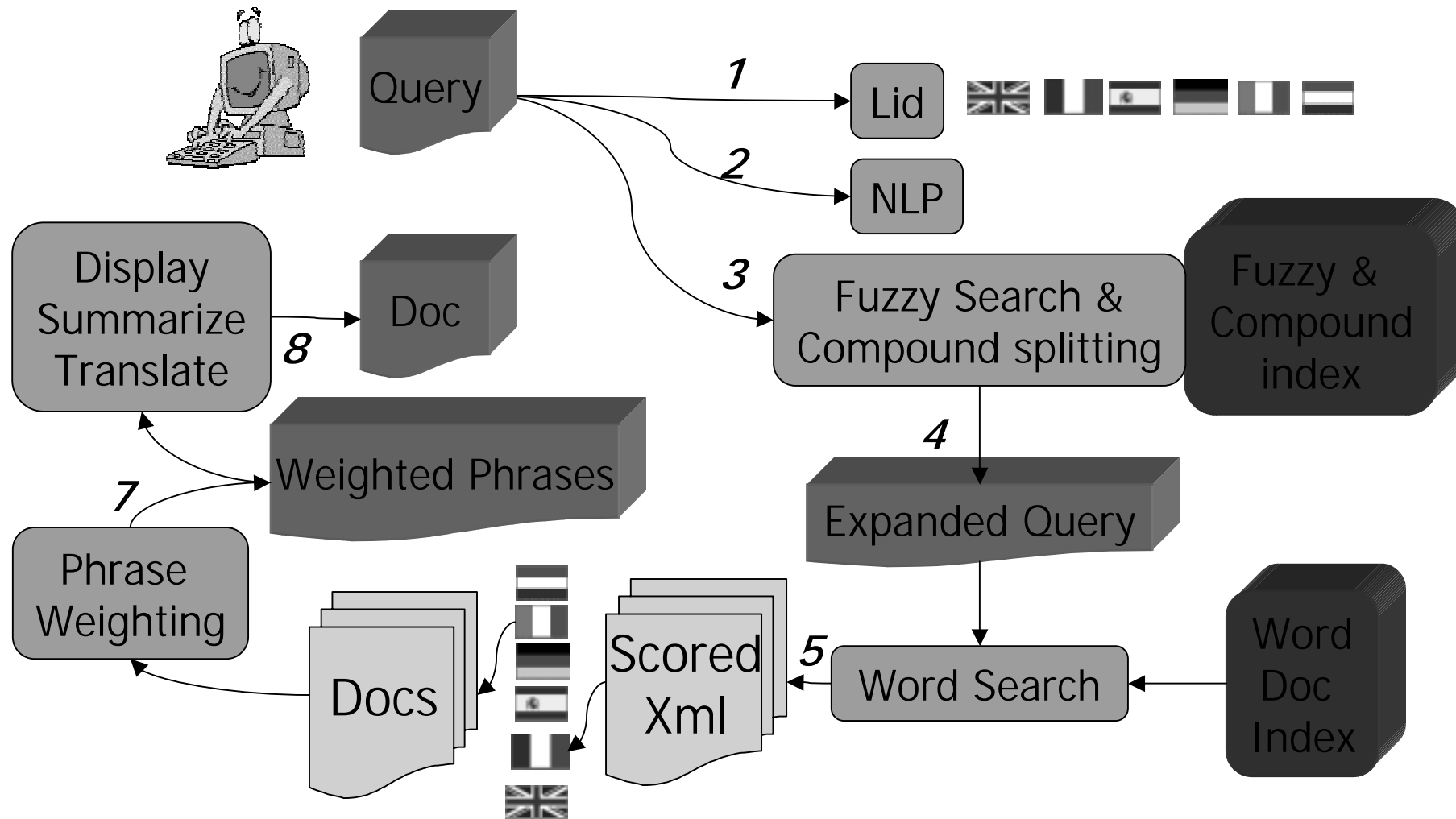


Capturing and Indexing





Cross-Lingual Search





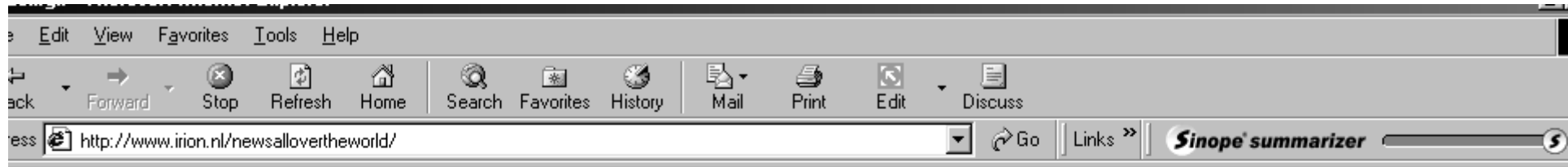
Document and Phrase Scores

- **Showing evidence in context, *Google* style with linguistic phrase marking**
- **Best matching document does not necessarily contain the best matching phrase.**
- ***Exact semantic relation may not be expressed in the document:***
 - *toxic medication*
 - *medicines for toxication*



Phrase Matching

- **Fuzzy matching**
- **Origin: un-translated, synonyms, translations**
- **Focus**
 - Number of query words in phrase
 - Number of phrase words in query
- ***Structure: Head or modifier***
- ***Concept score (WSD)***
- ***Co-occurrence score***



**Search in newspapers
all over the world
across the borders of language!**

with

News all over the World





Search in newspapers all over the world in your own language



gend

Our system detected the query language as being either Dutch or English.

Id - No translations or synonyms

There were no results with your settings. Another search process with a lower fuzzy matching level was started.

[1](#) [2](#) [3](#) [4](#) [5](#) [Next](#)

Id *Italic* - translations or synonyms

1. Figaro. fr États Unis rendent hommage aux victimes



Normal - Context rounding phrase

nouvelles menaces d'attaques terroristes A la veille de la **commémoration** , de nombreux visiteurs continuent de venir se recueillir sur ... elles rappelaient par leur forme celles émises avant le **11 septembre** 2001 . Le vice-président Dick Cheney a été ... organisées aujourd'hui . New York : Jean-Louis Turlin [**11 septembre** 2002] « United We Stand. » Telle est , depuis le 11 ... septembre 2002] « United We Stand. » Telle est , depuis le **11 septembre** 2001 , la devise d'une Amérique qui ... Zero , le président Bush , qui aura fatalement manqué la **commémoration** du matin , déposera une gerbe à la mémoire ... qui , considéré comme un des « héros » du **11 septembre** (sa cote de popularité dépasse celle de Bush
<http://www.lefigaro.fr/international/20020911.FIG0039.html>

- Quality of the document

2. September Gedenkfeiern den Globus Politik SPIEGEL ONLINE



11 . **September** : Gedenkfeiern rund um den Globus- Politik- SPIEGEL ONLINE Aktuell Wahl ... 11 .
 September : **Gedenkfeiern** rund um den Globus- Politik- SPIEGEL ONLINE Aktuell Wah ... Videos - - - - - Der Milzbrandterror - - - - - Zur Startseite des Dossiers 11 . **SEPTEMBER** Gedenkfeiern rund um den Globus Die Welt trägt ... Der Milzbrandterror - - - - - Zur Startseite des Dossiers 11 . **SEPTEMBER** **Gedenkfeiern** rund um den Globus Die Welt trägt ... der australischen Goldkiste , um den Opfern des 11 . **September** zu gedenken . Sydney/Wellington- In den Hauptstdten ... dem Botschaftsgelände , um an die Opfer des 11 . **Septembers** zu erinnern ." Dieser Tag hat sich fr
<http://www.spiegel.de/politik/ausland/0,1518,213425,00.html>





Search in newspapers all over the world in your own language



gend

Id - No translations or synonyms

Id *Italic* - translations or synonyms

Normal - Context rounding phrase

- Quality of the document

Our system detected the query language as being either Dutch or English.

[1](#) [2](#) [3](#) [4](#) [5](#) [Next](#)

1. NRC Handelsblad Buitenland Eén jaar na 11 september



aan *herdenkingen van de aanslagen op New York en Washington*. Deel één van een serie over de erfenis van **11 september**. naar

<http://www.nrc.nl/buitenland/11september.html>

2. The Moscow Times Photo of The Week



The Moscow Times Photo of The Week **09 September** - - 14 September , Vladimir Filonow/ MT Some of Moscow's youngest fashion

<http://www.themoscowtimes.com/photos/photo-of-week/24.html>

3. TDN-ee Commentaries and Analysis Section



TDN-ee , Commentaries and Analysis Section **10 September 2002** , Copyright © Turkish Daily News Editorial Mehmet Ali Birand Cuneyt Ulsever

<http://www.turkishdailynews.com/FrTDN/latest/article.htm>

4. NRC Handelsblad Buitenland Soedan



Soedan Soedan ' Soms weten we niet wie we zijn' **10 september 2002** Als het volk mag spreken , valt Soedan vrijwel

<http://www.nrc.nl/buitenland/soedan.html>

5. NRC Handelsblad Discussie Bede Troonrede





Search in newspapers all over the world in your own language



gend

Id - No translations or synonyms

Id *Italic* - translations or synonyms

Normal - Context rounding phrase

- Quality of the document

Our system detected the query language as being either Dutch or English.

[1](#) [2](#) [3](#) [4](#) [5](#) [Next](#)

1. NRC Handelsblad Buitenland Eén jaar na 11 september



aan *herdenking* York en Washington. Deel één van een serie over de erfenis van 11 september . naar

<http://www.nrc.nl/buitenland/11september.html>

2. The Moscow Times Photo of The Week



The Moscow Times Photo of The Week **09 September** - - 14 September , Vladimir Filonov/ MT Some of Moscow's youngest fashion

<http://www.themoscowtimes.com/photos/photo-of-week/24.html>

3. TDN-ee Commentaries and Analysis Section



TDN-ee , Commentaries and Analysis Section **10 September 2002** , Copyright © Turkish Daily News Editorial Mehmet Ali Birand Cuneyt Ulsever

<http://www.turkishdailynews.com/FrTDN/latest/article.htm>

4. NRC Handelsblad Buitenland Soedan



Soedan Soedan ' Soms weten we niet wie we zijn' **10 september 2002** Als het volk mag spreken , valt Soedan vrijwel

<http://www.nrc.nl/buitenland/soedan.html>

5. NRC Handelsblad Discussie Bede Troonrede





Search in newspapers all over the world in your own language



gend

Id - No translations or synonyms

Id Italic - translations or synonyms

Normal - Context rounding phrase

- Quality of the document



Our system detected the query language as being English.

[1](#) [2](#) [3](#) [4](#) [5](#) [Next](#)

1. SPIEGEL ONLINE



SERVICE IMPRESSUM Mittwoch , 11 . September2002 **DER ERSTE JAHRESTAG** Amerika zwischen Trauer und Angst Am Jahrestag der ... **DER ERSTE JAHRESTAG** Amerika zwischen Trauer anniversary; annual celebration; @ \$ % ^ * A % \$ @ **Terrorangriffe** vom 11 . September herrscht in den USA ... auf die US annual feast; birthday **POLITIK Union reitet heftige Attacke gegen Schröder** Union und FDP reagieren immer gereizter auf ... auml;rkt Schröder den Rücken **NEUE OFFENSIVE** Israel schickt Panzer in den Gazastreifen Die israelische Armee ... Panzer in den Gazastreifen **Die israelische Armee @ \$ % ^ * A % \$ @ hat eine neue Offensive in den Palstinensergebieten gestartet** . Mindestens 60 Panzer , gepanzerte <http://www.spiegel.de/>

2. Der erste Jahrestag Amerika zwischen Trauer Angst Politik SPIEGEL ONLINE



Der erste Jahrestag : Amerika zwischen Trauer und Angst- Politik- SPIEGEL ONLINE Aktuell ... Videos - - - - - Der Milzbrandterror - - - - - Zur Startseite des Dossiers **DER ERSTE JAHRESTAG** Amerika zwischen Trauer und Angst Am Jahrestag der ... sident Cheney suchte Zuflucht an einem sicheren Ort. REUTERS **Jahrestag der Terroranschläge** : Plakat am Ground Zero Washington- Die Furcht vor ... am Ground Zero Washington- **Die Furcht vor @ \$ % ^ * A % \$ @ neuen Terrorangriffen ist berall in den Vereinigten Staaten sprbar** . Prsident ... Autobomben , Selbstmordanschlägen und **anderen Attentaten im Zusammenhang mit dem Jahrestag die Rede gewesen** . Neun Auslandsvertretungen der USA wurden <http://www.spiegel.de/politik/debatte/0,1518,213378,00.html>

3. elpais



del 11-S(PDF) **ANIVERSARIO DEL 11-S** : Todos los gráficos animados de los atentados ... S : Todos los gráficos



Search in newspapers all over the world in your own language



Nederland grondwatervervuiling Alle woorden OK ?

Gezocht naar **grondwatervervuiling** 1 Resultaten

Verklaring

Het systeem detecteert als vraagtaal Nederlands.

Gezocht - Geen
 talingen of
 toniemen

Met de huidige instellingen is geen antwoord gevonden. Een nieuw zoekproces met een ruimere instelling is gestart.

t-en
ruingedrukt -
 rtaligen of
 toniemen

1. NRC Handelsblad Columns Anil Ramdas neme een wasvrouw



rmaal - Kontekst
 d de treffer

redeneren . Neem nou de bijeenkomst in Johannesburg. Het gaat over **water** en armoede en duurzaamheid en
 globalisering : allemaal even saai ... plek gebracht waar een aantal vrouwtjes gehurkt zat met **een kruik water** en
 een lat . Een middelgrote dhobi had twintig ... maakten het goed nat en begonnen met de lat het **vuil** eruit te slaan .
 Na een uur meppen werden de ... er geen struiken of takken waren , dan zomaar op **de grond** . Dat de net
 gewassen kleren meteen weer vuil werden ... op de grond . Dat de net gewassen kleren meteen weer **vuil** werde @\$
 %^*%\$@ , daar stond niemand bij stil . ' s Avonds bracht ... op menselijke schaal . Een ouderwetse bak nog , die je
 met **water** vult , waarin je zeep gooit , dan op een knopje ... pittiger is dan het nat maken van **de vloer met een**
vuile lap . De wasvrouwen kwamen in rang en loon
<http://www.nrc.nl/ramdas/artikel/1030944335165.html>

- Kwaliteit van document

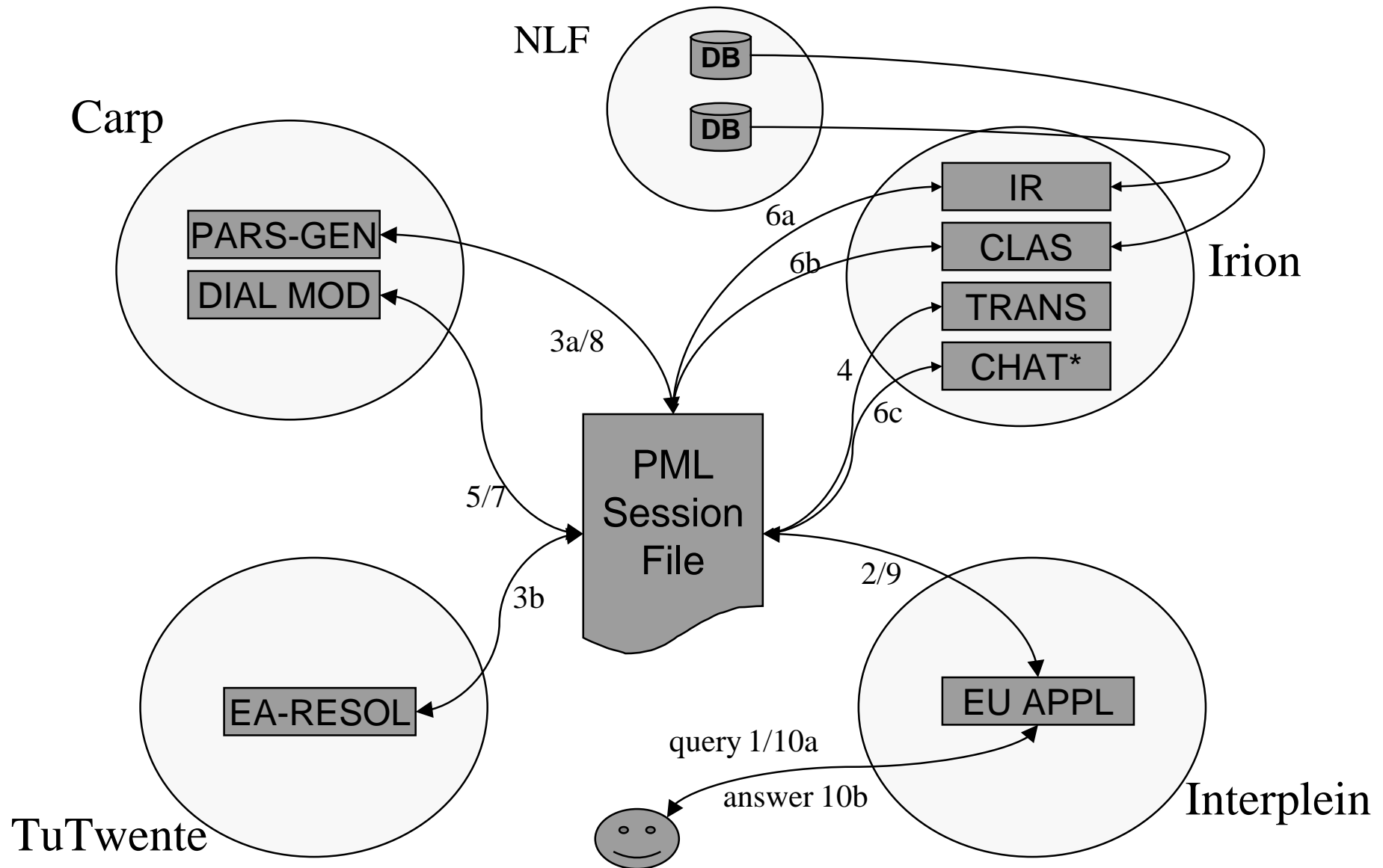




Pidgin (Iriion Technologies)



Pidgin Server Model





Market Areas

- **Within one country:**
 - Communication / providing information through the web between governmental organizations and minority groups and expats
 - Company intranets in multinational organizations
- **Abroad:**
 - Communication / providing information through the web in so-called “Euregions”
 - Information sharing between residences of international NGOs
 - European Commission and Union
 - Communication between companies and their customers abroad
 - Company extranets in multinational organizations

EUROPE'S LINGUISTIC DIVERSITY

ASSET OR BURDEN?



The mission of Irion : Equal access to
the information society for everybody



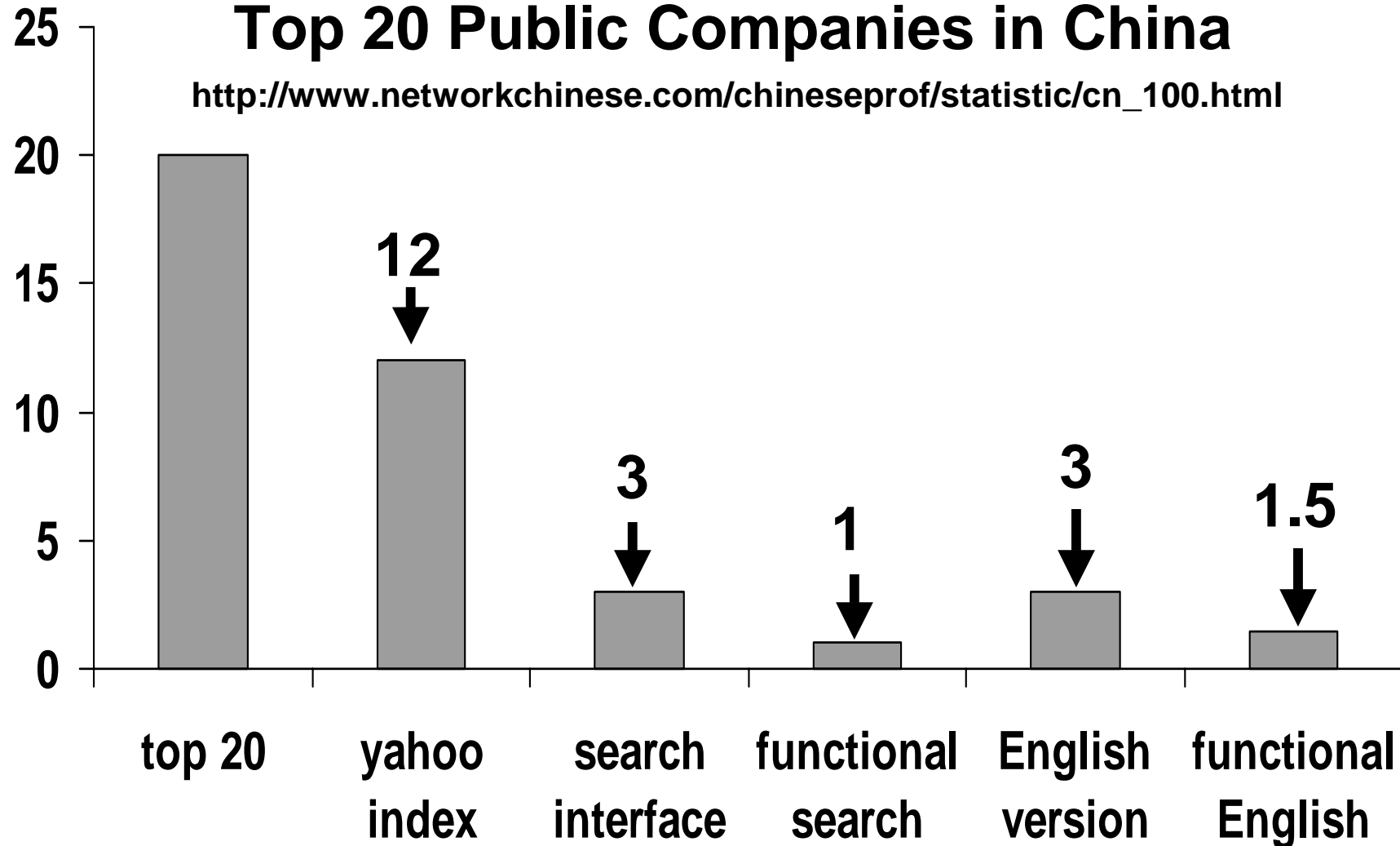
InSiteProxy™ (Clairvoyance)



The Need

Web Site Analysis of Top 20 Public Companies in China

http://www.networkchinese.com/chineseprof/statistic/cn_100.html





The Solution—InSiteProxy™

- **Information access to (foreign) Web sites**
 - **Not indexed by Web portals (e.g., Yahoo!China)**
 - **Missing their own search interface**
 - **Having poor search functionality**



Starting with an English Query

Obtain query terms in source language: *information, technology*

Enter your query:

Specify source language:

Specify target language:

Enter search site:

Specify search options Query cached site Query the site directly

Obtain translations from bilingual lexicons:
见闻|消息|信息|讯, 技术

Select best translations:
信息, 技术



Starting with an English Query

搜索在线 InSiteProxy

Enter your query:

Specify source language:

Specify target language:

Enter search site:

Specify search options Query cached site Query the site directly

**Fetch all subpages
at this URL**

**Then,
For each subpage, create
a CLARIT document;
Index database of documents**




Retrieval Results

InSiteProxy Demo - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

← → ↶ ✕ Search

Home My Netscape Search Shop Bookmarks Net2Phone

 搜索在线 InSiteProxy

Query string: information technology
Query site: <http://www.centek.com.cn/>

Query Results:

- 68.247 <http://www.centek.com.cn/park/9xiangcuoshi.htm> centek (Translate)
- 41.733 <http://www.centek.com.cn/news/20020806n7.htm> centek (Translate)
- 41.733 http://www.centek.com.cn/product/prod_cendate_idczengzhi.htm centek (Translate)
- 41.733 <http://www.centek.com.cn/baodao/20010318.htm> centek (Translate)
- 41.733 <http://www.centek.com.cn/park/200208211.htm> centek (Translate)
- 41.733 <http://www.centek.com.cn/baodao/我国信息化任重道远.htm> centek (Translate)
- 31.300 http://www.centek.com.cn/product/prod_centek_aoxingtianwang.htm centek (Translate)
- 21.404 <http://www.centek.com.cn/park/chuangyeziben.htm> centek (Translate)
- 21.404 <http://www.centek.com.cn/news/20020806n4.htm> centek (Translate)
- 21.404 <http://www.centek.com.cn/news/20020820d1.htm> centek (Translate)
- 21.404 <http://www.centek.com.cn/news/20020806n5.htm> centek (Translate)
- 21.404 <http://www.centek.com.cn/news/20020806n6.htm> centek (Translate)
- 21.404 <http://www.centek.com.cn/news/20020820d6.htm> centek (Translate)
- 21.404 http://www.centek.com.cn/product/prod_centek_yinxintong.htm centek (Translate)

- Retrieve from database
- Wrap results in HTML form



Click to See a Result Page

centek - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

http://www.centek.com.cn/product/prod_cendate_idczengzhi.htm Search

Home My Netscape Search Shop Bookmarks Net2Phone

CENTEK 北京中关村科技发展（控股）股份有限公司

首页 | 关于CENTEK | 投资者关系 | 最新消息 | 旗下企业 | 业务介绍 | 网上交流 | 求才若渴 | 网站导航 | 联系我们 |

数据中心增值业务
集团邮件系统服务
电子邮件是Internet用户使用最为广泛的应用服务，它的方便，快速，
高效，成为...

证券信息平台业务
北京中关村数据科技有限公司提供证券营业部解决方案。经对证券行业
联网现状进行分析...

教育信息平台业务
当今社会，随着信息技术的发展和计算机应用的迅速推广，互联网络也已经渗透到社会生活的各个层面，同
样地...

移动信息平台服务
北京中关村数据科技有限公司的数据中心拥有移动信息增值平台，在短信息服务方面可为企业进行...

中天村数据
点击进入

首页 | 关于CENTEK | 投资者关系 | 最新消息 | 旗下企业 | 业务介绍 | 网上交流 | 求才若渴 | 网站导航 | 联系我们 |
北京中关村科技发展（控股）股份有限公司版权所有



Translated Result Page

AltaVista Translations - Netscape 6

File Edit View Search Go Bookmarks Tasks Help

http://babelfish.altavista.com/tr?doit=done&urltext=http://www.centek.com.c Search

Home My Netscape Search Shop Bookmarks Net2Phone

altavista THE SEARCH COMPANY

ROLL OUT YOUR 5-PAGE WEB SITE FOR PENNIES A DAY!

www. .com Go!


VeriSign

Home > Tools > Babel Fish Translation > Translated Web Page

[View page in its original Language](#)

Translation powered by **SYSTRAN**

BEIJING CENTERGATE TECHNOLOGIES (HOLDING) CO., LTD.

 Home About Investor Most recent Under flag Service On-line Asks only then if is Website Contacts page | CENTEK | relations | news | enterprise | introduction | exchange | thirsty | navigation | with us |

 Clicks enters

Data center increment service

Group mail system service
The electronic mail is the Internet user uses the most widespread application service, its convenient, fast, highly effective, becomes...

Negotiable securities information platform service
The Beijing 中关村 data science and technology limited company provides the negotiable securities business department solution. After carries on the analysis to the negotiable securities profession networking present situation ...

Education information platform service
Now the society along with the information technology development and the computer application rapid promotion the





Some Concluding Thoughts...



Summary

- **The Market is “not there yet”**
 - But the Underlying Drivers are in Place
- **Quality of MT is a Gating Factor**
 - But Services and Consumer eCommerce may be Ripe Targets
- **Commercial CLIR Systems are “Complex”**
 - But Hybrid Systems may be Viable
- **Funding for Research & Development is “Healthy”**
 - But use it Wisely! ...
 - Develop Asian Language (and Arabic) Support
 - Don’t Focus on Document Retrieval alone



The End