QUAERO

Quaero … Ergo Sum
Presentation of an industrial research program on Multimedia Search and Navigation

Presentation for Chorus Event on National Initiatives.

Pieter van der Linden & Henri Gouraud

pieter.vanderlinden@thomson.net
henri.gouraud@exalead.com

October 10th 2007

Why Quaero ?

Trends for media market evolution

QUAERO

- Information will continue to evolve towards more and more multimedia content. Volumes will explode.
- Content will be accessed from a variety of devices through different channels and models.
  - Triple Play, PC Mobile.
  - Integration of mobile devices (phone/camera/MP3, PDA/GPS/Wifi/3G, …)
    - Content and service integration by Portal/Search
- Traditional distinction between consumers and producers will progressively vanish
  - Increasing portion of content will be produced by consumers or independent professionals
    - Blogs, podcasts, corporate sites, ….
  - In the professional area, similar to what happened in the Consumer markets, search and selection tooling will become the standard tool for accessing information.
- Distribution will evolve from static (TV, Radio, …) programs broadcasting toward personalized content aggregation.
  - Ex: Chime TV, news service aggregators
- Internet is expected to become a privileged information space.
  - Video : Youtube, Joost, ….

Why Quaero

Media consumption is migrating to Internet

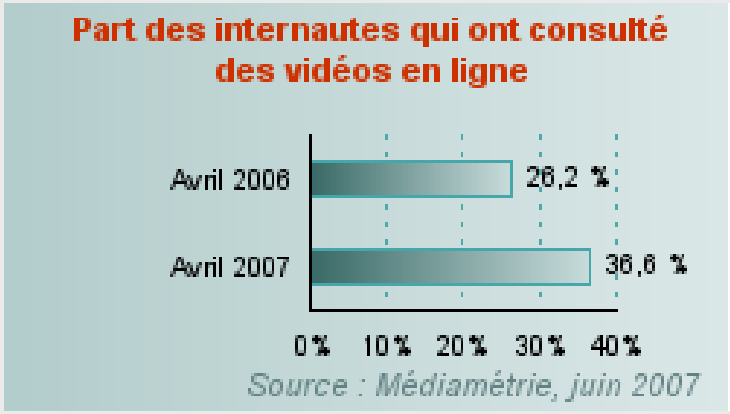
QUAERO

Media consumption in US – Hours per person.

Filmed entertainment	2002	2006	%change
Cable & Satellite TV	828	989	19.4
Broadcast TV	744	684	-8.1
Consumer internet	138	177	28.3
Home video	57	63	10.5
Box Office	14	12	-14.3
In-flight & Mobile	4	12	200
Total	1,785	1,937	8.5

Source : Motion Picture Association 2006, Veronis Suhler Stevenson.

Internet video penetration in France – Share of users having viewed online videos



Source : Mediametrie 2007,

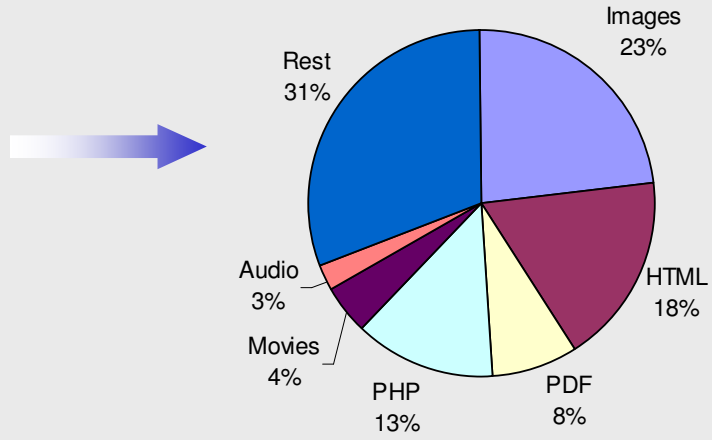
Why Quaero

But there still a long road to go.

QUAERO

How large is the internet ?

The size of the Internet 2003 (Tera Bytes).		2007
Surface Web	167	494
Deep Web	91,850	271,487
Email (originals)	440,606	852,786
Instant messaging	274	1,202
TOTAL	532,897	1,125,968



Source : How much information, 2003

And how much multimedia data is there around ?

World annual production of original broadcast media			
		Total Annuel (TB)	
Media Type	Unique Items per Year	Lower	Upper
Radio	70 million hours of original programming	3,488	3,488
Television	31 million hours of original programming	39,841	68,955
	Total:	**43,329**	**72,443**

Source: Raw data from CIA World Factbook, 2002.
Table, How much information 2003.

Program overview

QUAERO

Goals

- Develop demonstrators of applications corresponding to identified use cases in the domain of access and manipulation of multimedia and multilingual content
  - Search, navigate, distribute, produce
- Develop the corresponding enabling technologies for multilingual and multimodal content processing

Organization

- Gathers the whole development chain in a single, large, structured, collaborative program: applied research, basic research, corpus development and evaluation infrastructure
- Structured into application-specific projects, a core technology cluster, a corpus development cluster and an evaluation group, with transversal per technology committees

Consortium

Balanced between

- private companies (large, medium and small): Thomson, France Telecom, Jouve, Exalead, Bertin Technologies, LTU Technologies, Vecsys, Synapse Development;
- public research labs: LIMSI-CNRS, RWTH-Aachen, Karlsruhe University, INRIA, LIG-UJF, IRCAM, ENST-GET, IRIT, INIST-CNRS, MIG-INRA, LIPN;
- public institutions: INA, BNF, LNE, DGA.

Contacts with other European potential participants

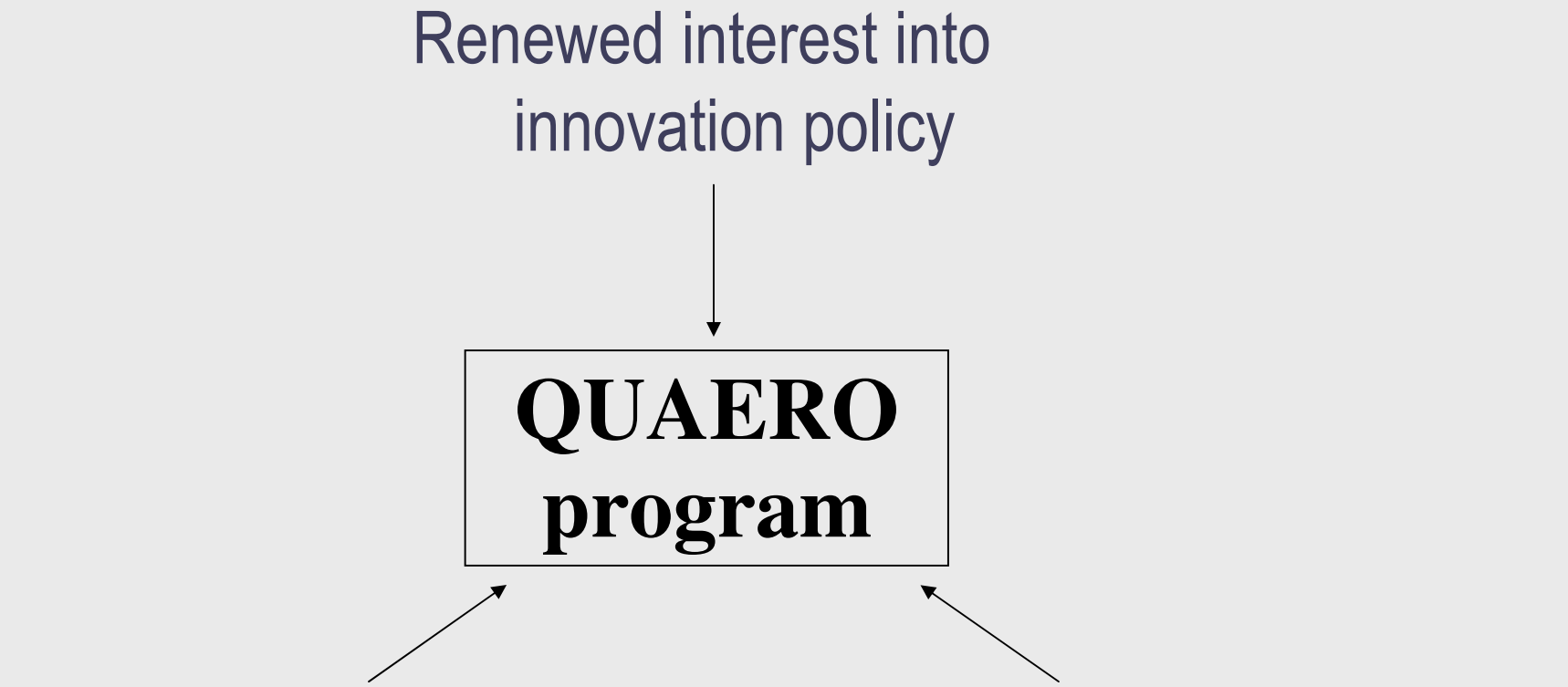
Funding

- Close to €100M for > 5 year and > 20 partners
- Has been granted by French "Agence de l'Innovation Industrielle"
- State aid to be authorized by DG Competition of European Commission

Agence de l'innovation industrielle

Context – *To be finalised*

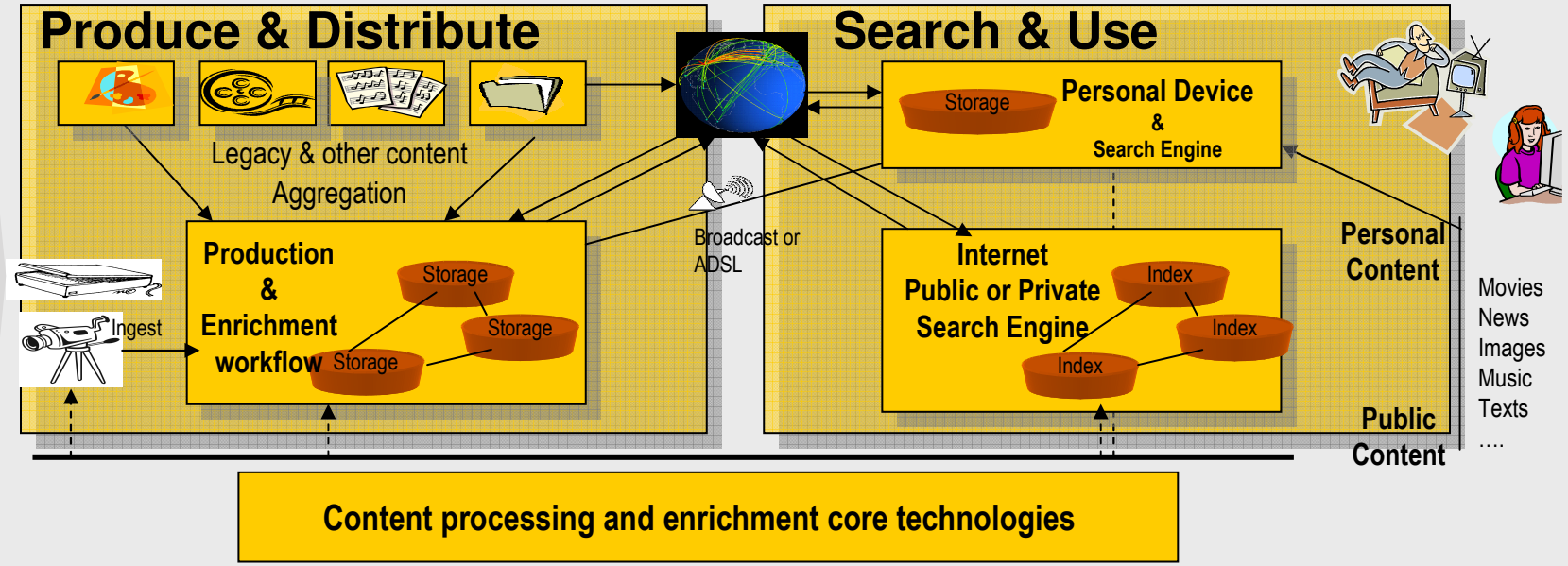
QUAERO



Quaero Scope

QUAERO

What



Structure

- Currently five application projects addressing complementary business cases.
  - Business target ranging from B2B technology to B2C service provision
  - The projects are lead by industrial "champions"
- And two technology project
  - Core Technology Cluster : Application driven advanced technological research.
  - Corpus project : Large annotated multimedia corpus used by other projects.

How

- Leverage existing know-how, product and market position of participants
- Foster cooperation and shared developments between activity areas of participants
- Focus on multimedia content
- *And take advantage of Europe's specificities (rich cultural background, multiple languages) ….*

Quaero Use Cases

addressing emerging but identified markets

QUAERO

B2C Search

1. **Consumer Multimedia Search Engine**
   - Search through Podcast and broadcast Radio and TV using speech to text transcription
   - Search images and video on name, context and metadata annotations
2. **Multimedia Search Services to enrich European portals**
   - Enhance the user experience by more convenient in interfaces.
3. **Personalized Video on interactive consumer networked devices Anytime and Anywhere.**
   - Combine IPTV networks, set top box, PC and mobile for customizing and specializing TV and Video programs.

B2B Content creation

4. **Recondition the Audivisual Cultural Heritage**
   - Document once, publish multiple - improved annotation and encoding by a combination of automatic and manual means.
   - Structure audiovisual content and digital books for access from Public and Professional Marketplace portals.
5. **Professional Digital Media Asset Management for Broadcasting Industry**
   - Ingestion, (post) production, aggregation, storage, search and repurposing of video material

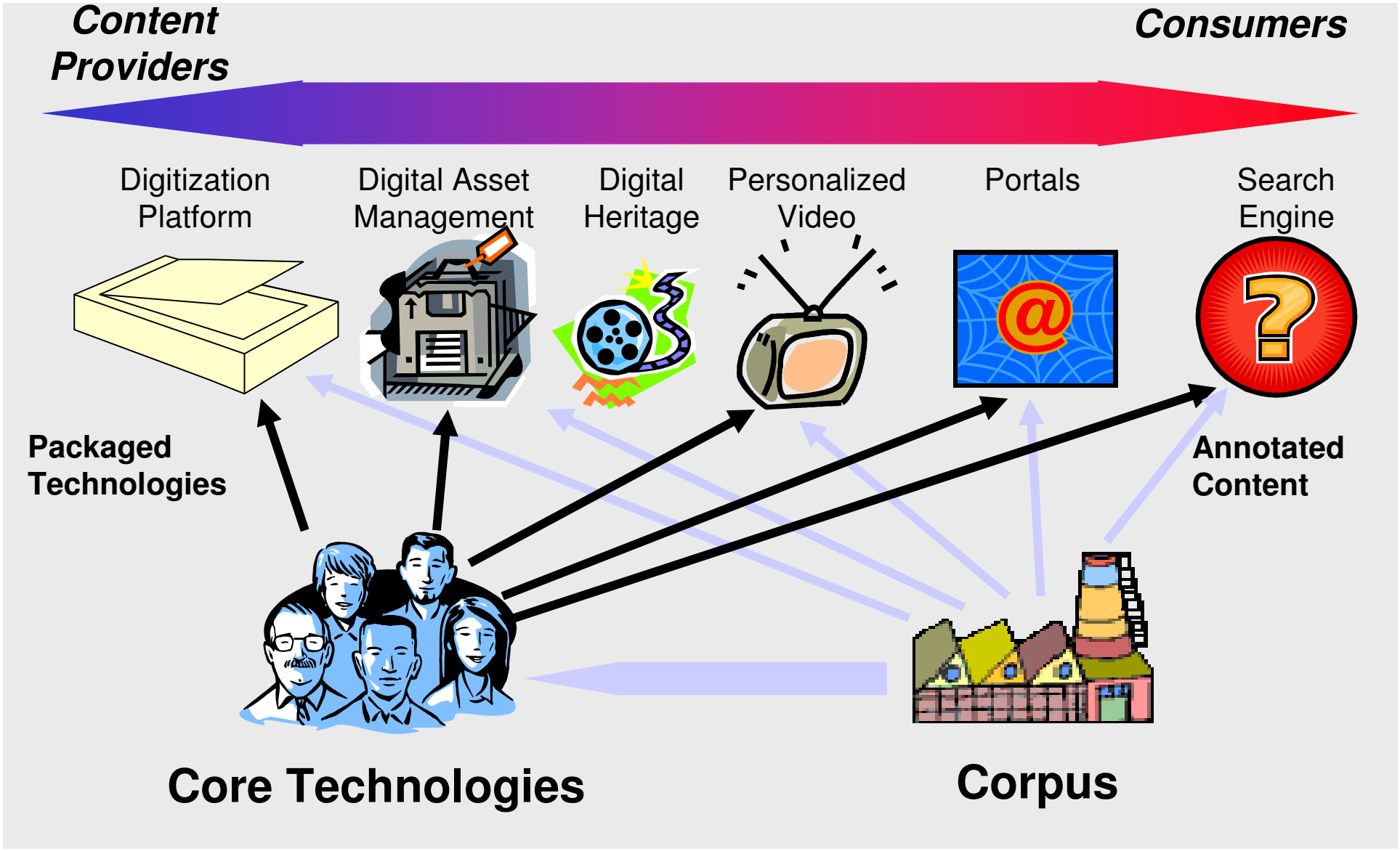
B2B Search

6. **Platform for Text and Image Annotation.**
   - Digitalization of large amounts of paper information (digital libraries, patent offices)
   - Translation, classification and structuring of information into knowledge.

Sharing technology

Resources and know-how.

QUAERO



Main Technological Areas

QUAERO

Infrastructure

- Huge scale document and index storage distribution.
- Communication and storage abstraction layer for interconnecting systems and devices.
- Workflow procedures for managing and aggregating content and metadata.

Metadata ingestion and structuring

- Pivotal formats and standards for storing metadata

Automatic annotation

- Speech to text transcription
- Video segmentation.
- Image recognition.
- Data fusion
- OCR
- Speaker recognition
- Language recognition
- Summarization
- Translation
- Categorization
- Named entity recognition
- Thesaurus construction

User Interface

- Natural language queries
- Cross lingual queries
- Context management and profiling.
- Summarization
- Translation
- Question Answering

Search & Extraction

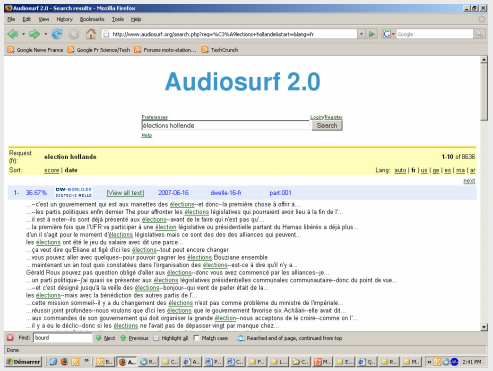
- Indexing of billions of documents, ranking, storage
- Use of web document statistics for approximate, phonetic and multilingual search.
- Question answering

Security

- Video and Audio Fingerprinting

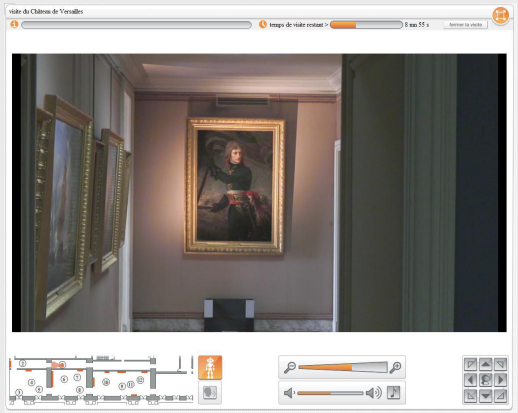
Technologies : what are we talking about?
Some examples

QUAERO



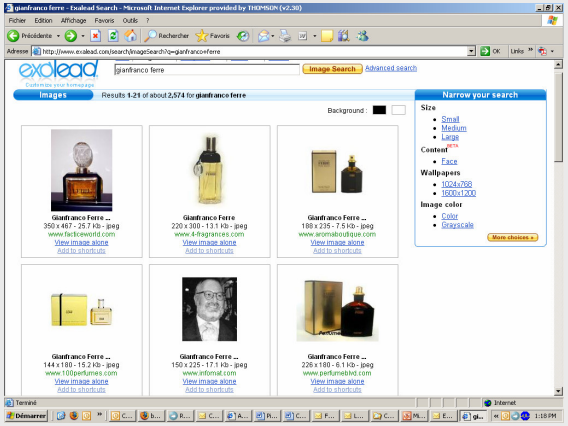
Automated speech annotation & translation

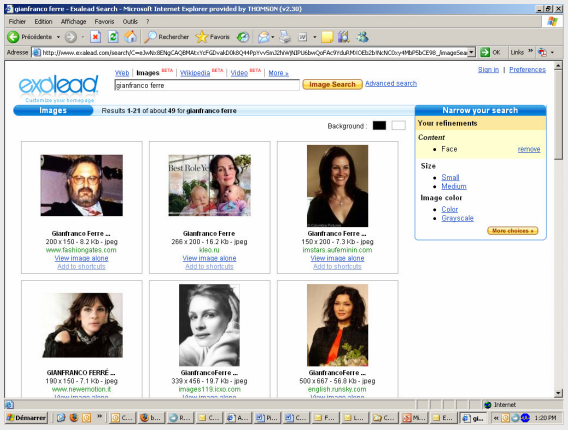
A fully operation prototype by Limsi & Vecsys



A robot assisted remote visit to the Versailles Castle

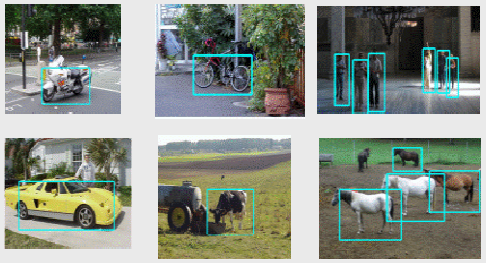
A new service for fiber subscribers by Orange





Face detection in images

An Beta service by Exalead & LTU



Object detection & Tracking

A research result by INRIA (Vista project)

Video repurposing

A research prototype by Thomson R&D

Project Guidelines and Development Approach.

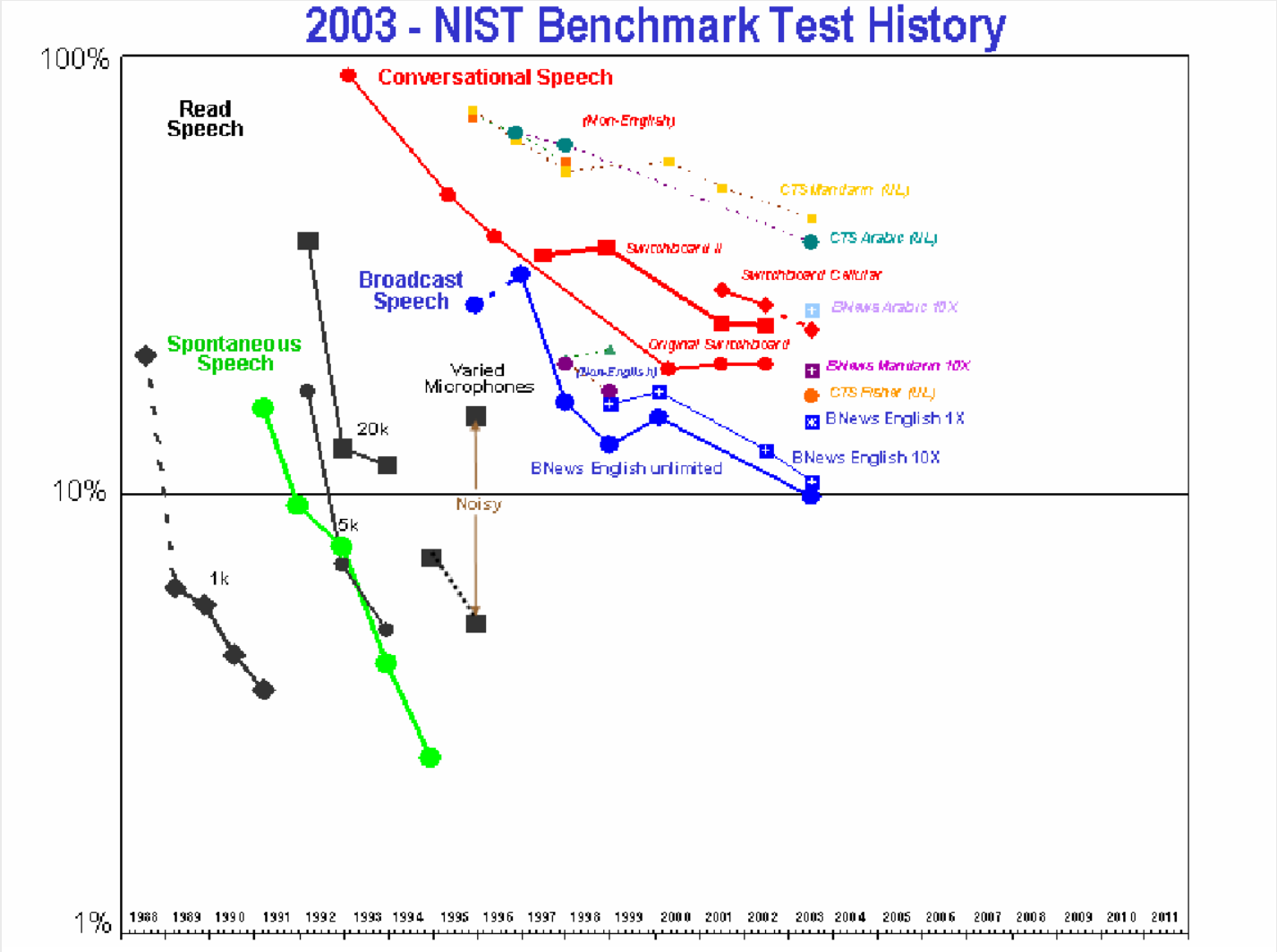
QUAERO

- **Derive technical and scientific objectives directly from the application / use case level requirements**
  - **Application driven technology roadmap**
  - **Yearly alignment of technological objectives between partners**
- **Foster co-opetition on strategic technological subjects**
  - **Multiple partners may develop concurrently on the same subjects using their own approaches with regulars workshops to openly exchange findings and evaluation results**
- **Conduct periodical performance evaluation to assess progress**
  - **Build on most advanced advanced evaluation procedures developed by national and international evaluation bodies and programs (e.g. NIST/TREC, CLEF, FRVT, Technolangue, Techno-Vision, …)**
- **Use evaluation to facilitate technology transfer**
  - **Discussion on tangible objectives and results between technology developers and industrials**

Strong believers in progress by evaluation

Example: NIST evaluations for speech recognition

QUAERO



Various potential collaborations

QUAERO

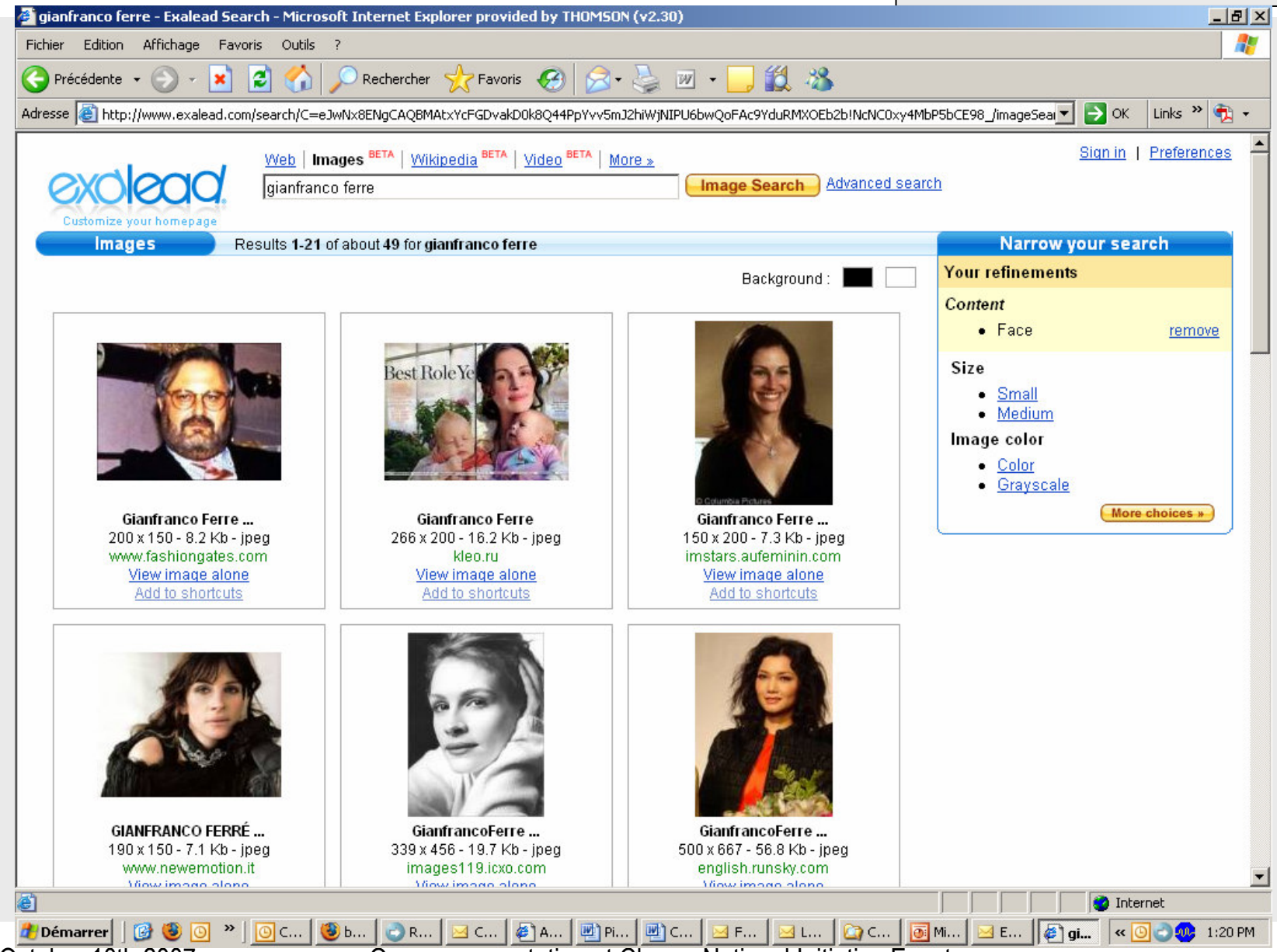
- R&D conducted in Quaero can be evaluated through existing campaigns
- Evaluations campaigns organized by Quaero can be opened to external participants bringing in their know-how
- Presentation of results obtained in Quaero to larger forums

Bye !

Hope to see you soon

***with some demos*.**

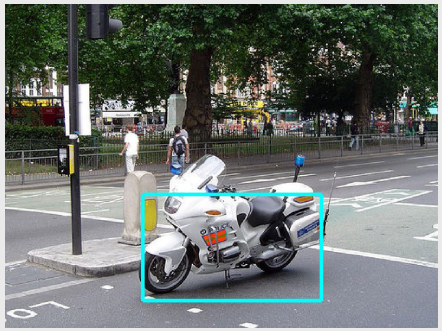
QUAERO


gianfranco ferre - Exalead Search - Microsoft Internet Explorer provided by THOMSON (v2.30)
Fichier Edition Affichage Favoris Outils ?
Précédente Rechercher Favoris
Adresse http://www.exalead.com/search/C=eJwNx8ENgCAQBMAtxYcFGDvakD0k8Q44PpYvv5mJ2hiWjNIPU6bwQoFAc9YduRMXOEb2b!NcNC0xy4MbP5bCE98_/imageSea
OK Links
exalead
Customize your homepage
Web | Images BETA | Wikipedia BETA | Video BETA | More »
gianfranco ferre
Image Search
Advanced search
Sign in | Preferences
Images
Results 1-21 of about 49 for gianfranco ferre
Background :
Narrow your search
Your refinements
Content
Face remove
Size
Small
Medium
Image color
Color
Grayscale
More choices »
Gianfranco Ferre ...
200 x 150 - 8.2 Kb - jpeg
www.fashiongates.com
View image alone
Add to shortcuts
Gianfranco Ferre
266 x 200 - 16.2 Kb - jpeg
kleo.ru
View image alone
Add to shortcuts
Gianfranco Ferre ...
150 x 200 - 7.3 Kb - jpeg
imstars.aufeminin.com
View image alone
Add to shortcuts
GIANFRANCO FERRÉ ...
190 x 150 - 7.1 Kb - jpeg
www.newemotion.it
GianfrancoFerre ...
339 x 456 - 19.7 Kb - jpeg
images119.icxo.com
GianfrancoFerre ...
500 x 667 - 56.8 Kb - jpeg
english.runsky.com
Internet
Démarrer
1:20 PM

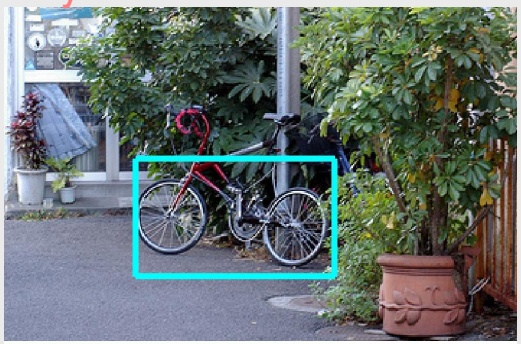
Audiosurf 2.0 - Search results - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://www.audiosurf.org/search.php?req=wimbledon+vainqueur&action=Search
Google
Google News France
Google Fr Science/Tech
Forums moto-station...
TechCrunch
Audiosurf 2.0
Preferences
Login/Register
wimbledon vainqueur
Search
Help
Request (fr): wimbledon vainqueur
1-10 of 4747
Sort: score | date
Lang: auto | fr | us | ge | es | ma | ar
next
1- 13.54% DW-WORLD.DE DEUTSCHE WELLE [View all text] 2007-07-09 dwelle-17-fr part:001
...le coureur le plus combatif de cette épreuve qui au vainqueur final laser vert et une prime de quarante quatre cent...
...euros aux côtés du FC Nuremberg pour la compétition le vainqueur de la coupe fédérale et les promus de Karlsruhe au...
...America--la qualification pour les demi-finales--le Brésil vainqueur du Chili si ça on essaie de régler aux dépens...
2- 12.55% france 3.fr [View all text] 2007-07-09 france3-national-soir3 part:001
[source]
...pour la cinquième année consécutive le tournoi de tennis de Wimbledon l'un des événements sportifs du dimanche dix et rouler outre-Manche...
...sport et d'exploits ce dimanche Roger Federer a gagné Wimbledon pour la cinquième année consécutive Kimi Raikkonen a remporté le...
3- 47.89% France inter [View all text] 2007-07-08 franceinter-8h009h00 part:001
...reçu message on avait rêvé l'inédit la finale de Wimbledon ne sera un grand classique bien Nadal Federer--les...
du soleil une foule considérable et un beau vainqueur ce prologue londonien aura donné des airs de fête aussi...
...vendredi historique--le tennis français a déchanté hier à Wimbledon
...imposé sa puissance et son expérience pour remporter son quatrième Wimbledon le score est sans appel le six quatre six un...
...elle--ce qui restera inscrit au palmarès événements du vainqueur--pas non mais finaliste et jordanienne et Chérèque réadmis...
...à Roland Garros c'est comme l'an passé à Wimbledon--Roger Federer sera opposé à Rafael Nadal vainqueur du serbe Novak que Joe communique sur abondent
Done
Démarrer
Boîte ...
Com...
Rech...
"A la ...
Histor...
Audi...
RE: U...
Micro...
2007...
5:23 PM

QUAERO

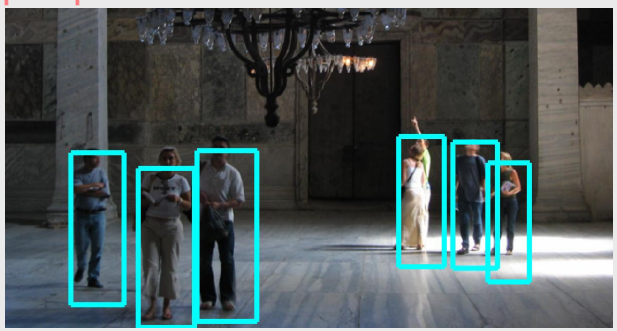
motorbikes



bicycles



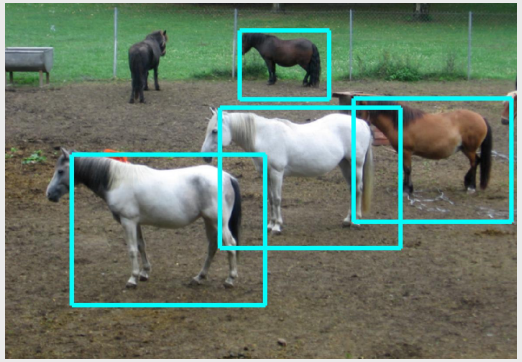
people



cars

cows

horses



Evaluation:

Method	Motorbikes	Bicycles	People	Cars
INRIA-Laptev	0.896	0.370	0.250	0.663
TU-Darmstadt	0.886	–	–	0.489
Edinburgh	0.453	0.119	0.002	0.000
INRIA-Dalal	0.490	–	0.013	0.613

Average precision for object detection on "test1" VOC 2005 image set

Comparison to the best results in PASCAL VOC'05 Challenge