Presentation outline

- The Quaero program
  - Context, scope and approach
  - Consortium, budget
  - History and timeline
- The Quaero evaluation infrastructure
  - Organization
  - Current activities
- Coordination with other campaigns
  - Types of coordination
  - The Quaero/imageCLEF cooperation
- Conclusions and perspectives
Context and overview

Need for advanced search applications

Evaluation-driven technology progress

Renewed industrial innovation policy

Political awareness of the strategic importance of information dominance

• Program on automatic multilingual and multimedia document processing

• Application-aware, corpus-based, evaluation-driven

• Large, addressing identified emerging markets

• Specific history and media coverage
Scope

- A collaborative research and development program
  - Focused on automatic extraction, analysis, classification and use of multimedia, multilingual content
  - To develop access to content

- Five application projects
  - Multimedia search on the internet
  - Enrichment of access services to audiovisual content on portals
  - Personalised video selection and broadcasting
  - Professional audiovisual asset management
  - Digitisation and content enrichment for libraries, audiovisual heritage and scientific publishing

- A shared research structure
  - A broad research scope
  - Systematic evaluation of scientific and technical progress
  - Extensive resources for annotating large collections of multimedia data
Five projects with application targets
From content providers to consumers
Sharing resources and know-how

Who steers

1- Digitization and content enrichment
Jouve
Software and services for editors, patent offices and libraries

2- Digital media asset management
Thomson / INA
Software for broadcasters, media companies, audiovisual archives

3- Personalized video
Thomson
Software for telecom operators, retailers and enterprise video

4- Search engines
Exalead
Multimedia search engine

5- PC, Mobile portals
France Télécom
New generation of access services to audiovisual content

Shared research structure
Coordinated by CNRS and RWTH

Technologies for analyzing audio, speech, music, image, video content.
Technologies for natural language analysis and translation
Content protection technologies

Sept 16th, 2008
Two research projects

- **CTC project** (coordinated by CNRS)
  - Improve the state of the art in terms of structuring and indexing multimedia, multilingual documents
  - Develop generic models and technologies for program application projects

- **Corpus project** (coordinated by RWTH Aachen)
  - Collection and annotation of data required for developing and benchmarking (evaluating) technologies
    - “Ground truth” for statistical and machine learning based algorithms.
Covered technologies

Translation

Written language processing

Search & navigation

Coding Protection

Image Processing

Speech Transcription

Audio & Music

Video Processing

Multimodal structuring

Structuring and mgmt of multimedia and multilingual documents

Corpus for Evaluation & Demonstration
Examples of targeted innovations

- Extensions of information search to a wide range of media
  - e.g. radio broadcasting, podcasts, images, video and music

- Search for multimedia documents in all kinds of terminals
  - e.g. television sets, cell phones

- Automatic text generation from audiovisual documents
  - e.g. transcription of the sound track of a film

- Automatic selection of relevant video extracts
  - e.g. selection of goals scored during a soccer match

- Cross lingual search of information
  - e.g. search a sound track in Polish using French key words
An integrated program

Advanced research
Generic technologies
Statistical models
Automatic learning

Integration

JOUVE (digitisation)
Thomson / INA media management
Thomson (video on demand)
Exalead (multimedia search)
France Telecom (portals)

Corpus
Audio, images, music, text, video, speech, multimedia, multilingualism

Application projects
Development approach

- Derive technical and scientific objectives directly from the application requirements
  - Application-driven technology roadmap
  - Yearly alignment of the technological objectives between partners
- Foster coopetition on strategic technological subjects
  - Multiple partners may develop concurrently on the same subjects using their own approaches with regular workshops to openly exchange findings and evaluation results
- Conduct periodical performance evaluation to assess progress
  - Build on most advanced evaluation procedures developed by national and international bodies and programs
- Use evaluation to facilitate technology transfer
  - Discussion on tangible objectives and results between technology developers and integrators
Consortium

- Coverage of the whole value chain
  - Technologies benchmarked in international assessments (Technolangue, Techno-Vision, NIST/TREC, CLEF…)
  - Experience in technology transfer
  - Access to large markets

- Balanced composition
  - Large companies and SMEs
  - Public research laboratories
  - Public institutions
Consortium members

- **Private companies**
  - Bertin, Exalead, France Télécom, Jouve, LTU Technologies, Synapse Développement, Thomson, Vecsys

- **Public research laboratories**
  - CNRS-LIMSI, CNRS-IMMI, CNRS-INIST, INRIA, IRCAM, IRIT, Institut Telecom, LIPN, MIG-INRA, Université Joseph Fourier, University of Karlsruhe, RWTH university, Aachen

- **Public institutions**
  - BnF, DGA, Ina, LNE

Coordinated by Thomson
Budget

- 99 M€ funding from oseo
- 200 M€ total budget
- over 5 years

Budget by type of activity
**History**

- **Conception phase**
  - Proposal elaboration

- **Evaluation phase**
  - Proposal by independent experts

- **Preparation phase**
  - Program setup and organization
  - Acceptance by AII

- **Operational phase**
  - R&D activities

- **First announcements**

- **Establishment of the Agence de l’innovation industrielle (AII)**

- **EC approval for Quaero aid**

- **All is integrated in OSEO**
Timeline

2005 2006 2007 2008 2009 2010 2011 2012 2013

Baseline evaluation

Yearly evaluations

Final evaluation

Consortium proposal

All acceptance

EC approval

Operational phase

R&D activities

Sept 16th, 2008 Evaluation in Quaero
Evaluation infrastructure

- Over 35 tasks (or tracks), each of them including one or more sub-tasks
- About 20 persons involved (8 FTE) for the program duration, distributed over 6 partners (DGA, LNE, IRIT, UJF/LIG, INRIA/LEAR, LIPN)
- Acting together with corpus creators, technology developers and integrators in per-technology committees
- Organizing yearly evaluation campaigns
Example of active tasks

- Q&A (Fr, En)
  - Using a fixed web corpus
  - Questions correspond to real user requests
- Speech to text (Fr, En, Ge)
  - On broadcast and podcast
- Translation
  - On web pages
- Beat detection in music
  - Evaluated at MIREX
- Face detection
  - Cf. www.exalead.com
- …
Current activities

- **Baseline technology assessment**
  - Testing the existing technologies at program start
    - On Quaero specific data when already available
    - On existing similar data otherwise

- **Preparation of upcoming evaluation campaigns**
  - Corpus production is ramping up
  - Evaluation plans are expected early 2009
Coordination with other evaluation campaigns

- Need for coordination
  - If tasks is identical or similar, results can be compared only if common evaluation

- Types of interaction
  - R&D conducted in Quaero can be evaluated through existing campaigns
  - Evaluation campaigns organized by Quaero can be opened to external participants bringing in their know-how
  - Quaero and another program can organize common evaluation campaigns
The Quaero/imageCLEF cooperation

- Quaero members already involved in imageCLEF
  - RWTH in the “Visual Concept Detection” track
  - UJF/LIG in the “Photo Retrieval” track
- Quaero program sponsors the imageCLEF workshop
  - Fostering the CLEF evaluation efforts
  - Investigating possibilities of common campaigns
Conclusions

- Quaero is a collaborative program on multilingual and multimedia processing
- Evaluation is a strategic approach for Quaero
  - Use evaluation to select core technologies, foster innovation and measure progress
  - Built-in evaluation infrastructure
  - Cooperate with other evaluation campaigns
Perspectives

- 5-year work program just starting
- Quaero members participating in various evaluation campaigns
- Various Quaero-organized evaluation campaigns welcoming new participants
- Joint evaluation campaigns
Thank you for your attention!

More information on

http://www.quaero.org