CHORUS

CA on Multimedia Search Engines

ROADMAP & RECOMMANDATIONS

Nozha Boujemaa

CHORUS Scientific coordinator

INRIA

CHORUS Objectives and organisation

- Bridge the gap between *researchers view* (academia and industry) and the *new services* and applications *prospective* for every day life needs

- Identify and derive *critical issues* involving *cross-disciplinary* aspects: recommendations for technological approaches together with socio-economic and legal issues,

- CHORUS community: EU projects, national initiatives and key players in the domain of multimedia search engines

- Different structures and events: working groups, Think-Tank, D2 "A-V Search" cluster, information exchange platform and workshops

Search Definition

- Search is about *making best use of available meta-data* to provide the user with *useful information* in spite of the fact that his request is possibly poorly formulated and typically *unanticipated.*

- Keeping the *"user in the loop",* maximize its efficiency

Metadata?

- Metadata: all information besides the raw content that make the content searchable:
    - manual annotations,
    - automatically generated low-level signatures,
    - automatically generated semantic labels,
    - device generated media context
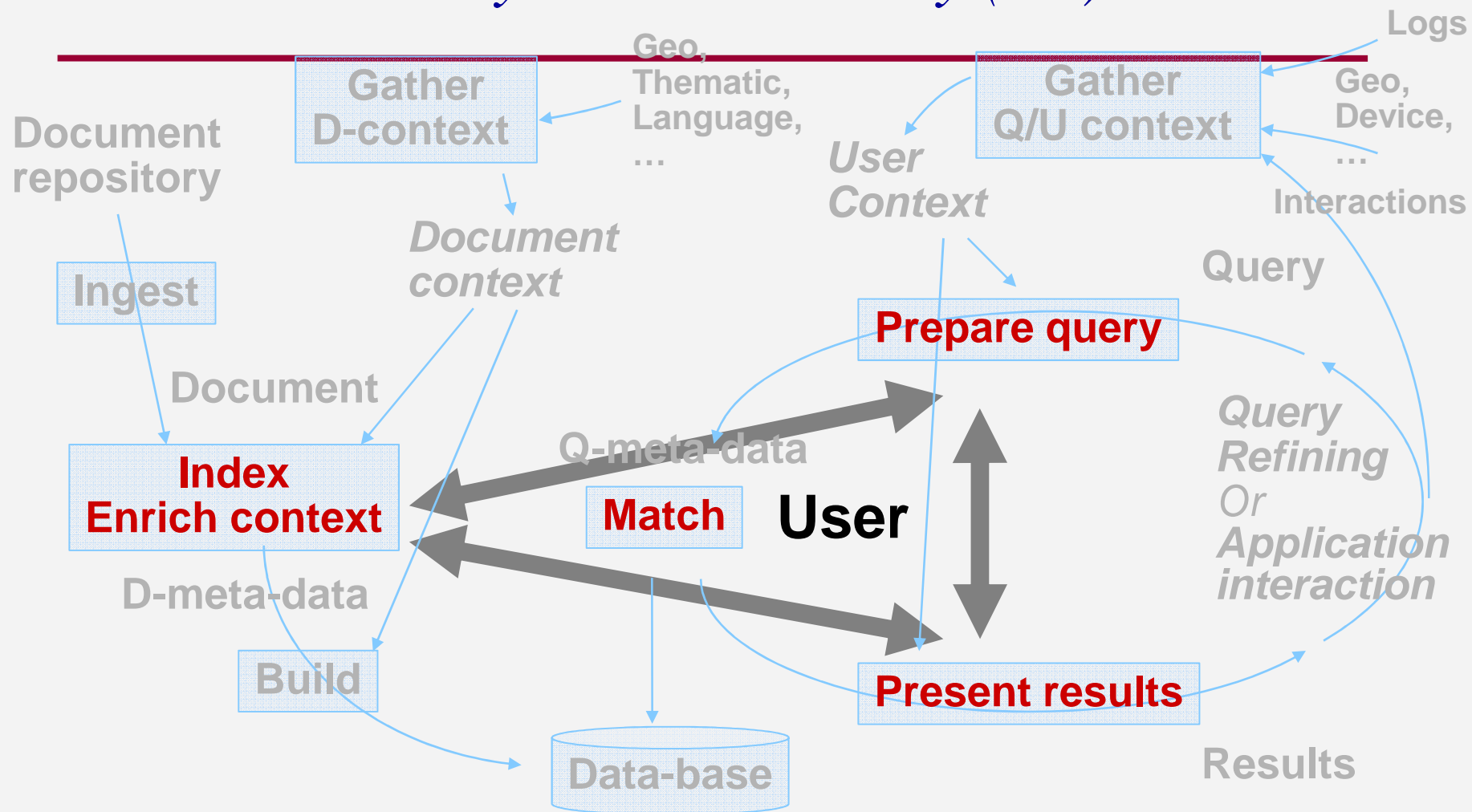   - …

Functional description of MSE

- CHORUS starts with elaborating a **functional breakdown** of a generic search engine regardless of the application domain or business sector.

- It presents the benefit of shared projects' description and vocabulary across industry and academia.

Chorus - Search Engines: Functional Description
Shared vocabulary across Industry (TT) & Academia

Logs

Gather
D-context

Geo,
Thematic,
Language,
…

Gather
Q/U context

Geo,
Device,
…

Document
repository

*User
Context*

Interactions

*Document
context*

Query

Ingest

Prepare query

Document

*Query
Refining
Or
Application
interaction*

Q-meta-data

Index
Enrich context

Match

User

D-meta-data

Build

Present results

Data-base

Results

Search evolution

*Search will be a pervasive and ever
present function fully integrated into all
applications and oftentimes invisible to the
user - users will not realize they are posing
queries, e.g.!*

Six prototypical areas of business

1. *Web search (billions, millions)*
2. *Personalised TV (video volumes)*
3. *Enterprise Content search (variety, business)*
4. *Library search (value, old)*
5. *Personal Content Search (variety, local)*
6. *Monitoring, Detection & Alert (flux vs base)*

- differences in characteristics attributes: content management, content ownership, access rights and the revenue model

- Vision, Trends and Challenges presented in the "Vision -Think-Tank" outcome session.

Cross-disciplinary challenges and Recommendations

Gap analysis studies conducted by CHORUS allows identifying directions that deserve more European effort.

CHORUS recommendations toward **more efficient search engines** that will help to makes the implicit knowledge reachable and in fair and attractive ways to the user:

- Push for more efficient indexing techniques:

  - multimedia content enrichment and automatic meta-data creation.

  - Socially-enriched automated indexing will empower the robustness of the indexing techniques.

Cross-disciplinary challenges and Recommendations

- Develop new media search paradigms based on content/context/event, to go beyond current retrieval systems

  => definition of new query models outperforming traditional keyword-based or query-by-example-based) event structures are expected to be the main driver for media contextualization.

- Model efficiently the implicit and explicit feedback to empower the personalization and recommendation abilities of a search engine (including collaborative tags filtering, user preference detection ...)

Cross-disciplinary challenges and Recommendations

- Develop more informative user interfaces (too little overlap between networked media technology providers and UI designers today)

  => toward smart visualization of media delivery and enhanced user quality of experience (QoE)

- Breaking complexity and afford scalability: besides the amount of input data and generated features, complexity need to be managed for other growing quantities such as number of users, number of information sources, number of data attributes / features dimension

Cross-disciplinary challenges and Recommendations

- Develop interoperable meta-data standards: open ended content, association with object and preservation through its life

    => preservation of metadata across transfer, migration issues, postproduction and restoration

    => Beyond Internet of Media

- Make available and develop open annotated multimedia corpora which is a key enabler for MSE scientific and commercial success,

Cross-disciplinary challenges and Recommendations

- Address privacy concerns to afford guarantees to the users: minimum regulation is necessary for consumer protection, privacy protection or unfair competition

    => Current EU regulation does not cover adequately or are not applicable to search engines

- Address security, integrity and trust issues related to search and networked storage (international cooperation needed)

- Support Pan-European privacy certification of IT products or IT-based services compliance with European data protection regulation

Organizational Recommendations

-     CHORUS allows positioning various EU efforts among the technological landscape dimensions

- The mapping reveals that Europe is **rich with very sharp expertises** in many separated topics in the field of search engines.

- **Observation**: innovative commercial services are feasible today using exiting pieces of research results for some niche markets - mainly for business search market (EO, Enterprise, …)

Organizational Recommendations

- Regarding **consumer search market**: Europe is lacking today an integrated program that gathers all needed expertises for building competitive real life search engines

- **Empower aggregation and orchestration of such expertises into an operational end-to-end search organisational structure** (idea such creating a "Center for New Generation European Search Engine" needs to be investigated).

- **Foster user-centric design** (market-pull) requirements for EU funded projects against **the technology-driven design** (techno-push)

How should the Future Internet be in 2020? Search perspective

Media will be plentiful all over the Future Internet (FI), often in distributed form

=> making them searchable and accessible (metadata generation and structuring) is the major challenge for search in the FI.

- Users will be accustomed to seeing what they need delivered just-in-time,

- They will deal with data and services in the cloud (storage and applications …).

How should the Future Internet be in 2020? Search perspective

- Search is at the application level: the network should be neutral (agnostic) with respect to applications

- Only one exception: Media delivery stage; eg: search in the context of P2P architecture => need for logical structure of multimedia document rather than neutral packets - for optimization reasons
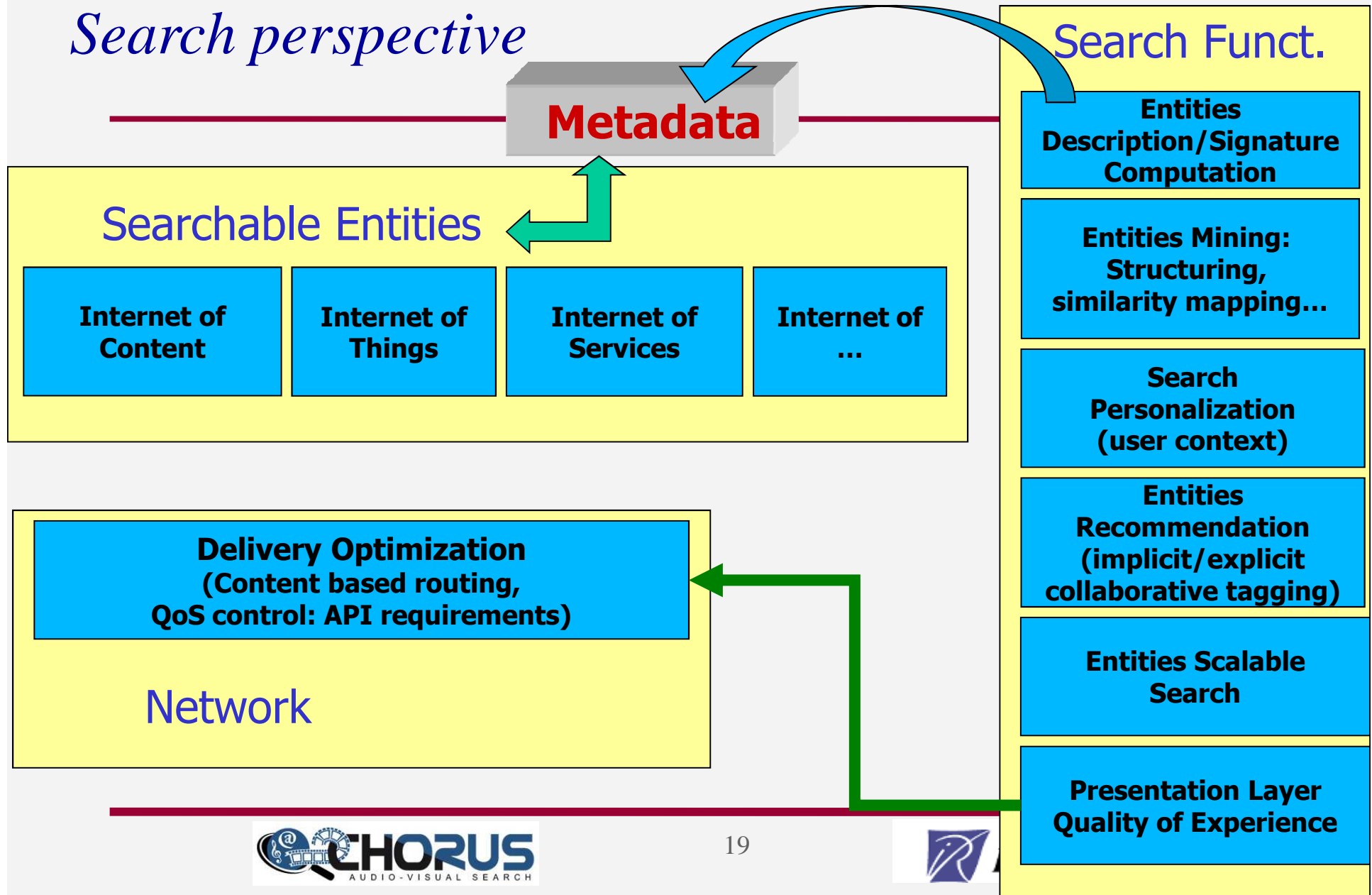
How should the Future Internet be in 2020? Search perspective

- Search being omnipresent – most applications (including communication applications) will involve search in some form – it will not have significant differentiating impact on FI

  ⇒ either unique or multiple Internets: Internet of Things, Internet of Services, Internet of Media Content

  ⇒ Search operates on METADATA of SEARCHABLE ENTITIES

How should the Future Internet be in 2020?
Search perspective

Search Funct.

Metadata

Searchable Entities

Internet of Content	Internet of Things	Internet of Services	Internet of ...

Entities Description/Signature Computation

Entities Mining: Structuring, similarity mapping...

Search Personalization (user context)

Entities Recommendation (implicit/explicit collaborative tagging)

**Delivery Optimization
(Content based routing,
QoS control: API requirements)**

Network

Entities Scalable Search

Presentation Layer Quality of Experience

How should the Future Internet be in 2020?
Search perspective

- Growing impact of social networking for deriving new knowledge

- It may represent an alternative to information retrieval through trusted recommendation mechanism

  (=> See Panel discussion)

Interoperability appear as major concern that need to be carefully addressed in the coming years

=> beyond the Internet of Media reaching all networked entities including "Services" and "Things"

Thanks for your attention

http://www.ist-chorus.org/

Topics to be addressed (1)

1. **Power of the crowd vs search engines?**

   ➢ Recommendation, ratings mechanisms and folksonomy vs automatic indexing and searching techniques

2. **User-generated content vs. editorial professional content?**

   ➢ YouTube, daily motion, Wikipedia vs: TV broadcaster, Encarta, ...

3. **How to foster competition in the search engines market?**

   ➢ In the domain of web search ? Niche applications ?

Topics to be addressed (2)

4. **Interoperability standards** vs core domain standards?

   ➢ Loosely-coupled heterogeneous systems vs tightly-coupled fully standardized systems?

   ➢ Too much standards kill standards?

   ➢ Allow more creativity and flexibility along with holistic interoperability?

   ➢ Are we into the era where de facto standards precede de jure standards?

5. What are the most suitable topics for international cooperation?

   ➢ interoperability, benchmarking, regulation ...?

Thanks for your attention

http://www.ist-chorus.org/