

Information Day on “Audio-Visual Search Technologies”

Jean-Pierre Chanod

Xerox Research Centre Europe

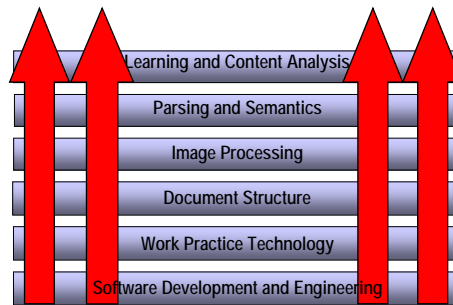
December 19th, 2005



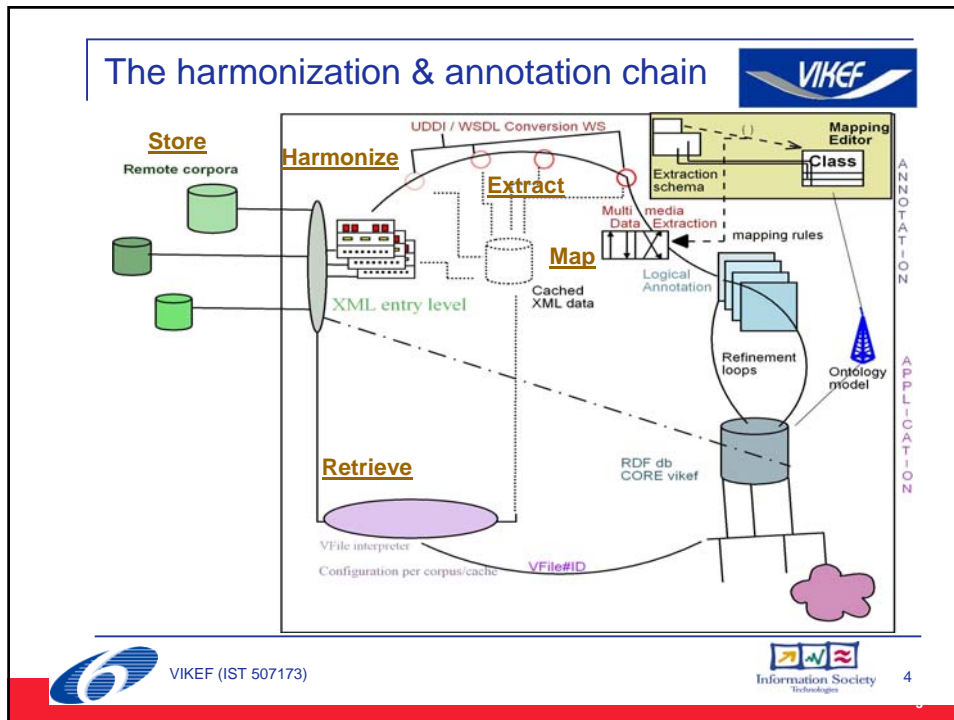
1

Introduction

- Xerox Research Centre Europe
- Content & Structure Program Manager:
“Use novel content technologies to increase the automation of doc. intensive business processes”
- Document Structure area manager
 - ◆ Document conversion to XML
- Parsing & Semantics area manager (until 2003)
 - ◆ FST, Syntax, multilingual language resources
- VIKEF, InfoMagic



2



Directions for advanced search

Some drivers

- Compliance and risk mitigation (litigation)
- Need to manage and unify access to information across very large repositories
- Need to automate information work to reduce costs
- Multilinguality

Consumer search vs. Search in the enterprise workplace

- Infrastructure
 - ◆ Provide access to unstructured information across multiple repositories, high volumes, long term archiving
- Search practices
 - ◆ The intent before the content (who and why before what)
 - ◆ Intuitive search interfaces for naive users (SMEs)
 - ◆ Security, controlled access
 - ◆ Customization, learning: personal views of the search space
- Content and metadata
 - ◆ Target rare, high value information into ultra large heterogeneous collections (e. g. patents related to cloning, compliance, "the smoking gun")
 - ◆ Concept diffusion vs. keywords and patterns, discourse analysis
 - ◆ Trust, confidentiality, temporal dimension, qualified source
 - ◆ Ontology: closing the gap (multimedia ontology building and linking)
 - ◆ Structural search (document understanding, multimedia convergence)
 - ◆ Information fusion (complex transformation chains, validation)