Slide 1

User needs in video

Bruno Bachimont
Institut National de l'Audiovisuel

Slide 2

Content

- Presentation of a "user" : INA
  - What are the main uses of documents?
- What is an audio-visual document?
  - Why is it a strange object?
- What is indexing?
  - Why indexing audiovisual document is difficult?
- How to manage the indexing process?
- What are the main issues?

Slide 3

Who are we?
Presentation of INA

Slide 4

Key facts

The 1974 law concerning freedom of communication, creates seven audiovisual societies from ORTF :

INA is an industrial and commercial public institution, founded on January, the 6th, in 1975

Slide 5

Main activities

- Archiving activities
  - CATALOGUING
  - PRESERVATION
  - DISTRIBUTION

- Innovation in the digital and audiovisual fields:
  - PROFESSIONAL TRAINING
  - RESEARCH AND DEVELOPMENT
  - AUDIOVISUAL AND MULTIMEDIA PRODUCTION AND POST-PRODUCTION

Slide 6

Archiving

- Public archiving:
  - preservation and exploitation of all audiovisual archives from the French public television channels and radio stations.

- Legal deposit
  - Preservation of national records to make the programmes and related documents of all French radio and television outlets readily accessible for research purposes.

Slide 7

Archives

The television archives, today:

From the professional archiving
~ 575 000 hours

From the legal deposit (since 1995)
+
12 cable & satellite TV
~ 430 000 hours

Slide 8

Archives

The television archives: heritage.

- **RTF et RTF Regions (1949-1964)**
- **ORTF (1964-1975)**
- **TF1 (1975-1982)**
- **Antenne 2 (1975-1992)**
- **FR3 national et regional (1975-1992)**

+ 2500 hours of cinematographic programming (1914-1969)
Including the « Actualités Françaises » collection (1940-1969)

Slide 9

Archives

The radio archives, today:

From the professional archiving
~ 535 000 hours

From the legal deposit (since 1995)
+
12 Non public radios since 2002
~ 500 000 hours

Slide 10

The radio archives : heritage:

- **Private and national programming from before the war (1933-1939)**
- **WWII (1939-1945) :**
  Radio Paris, Radio Vichy, Radio Alger, Radio Brazaville, BBC programmes made for France
- **RDF-RTF (1945-1963)**
- **Regional programming (1945-1974)**
- **ORTF (1964-1974)**
- **Sorafom (1945-1962), Ocora (1962-1969), DAEC (1969-1974)**
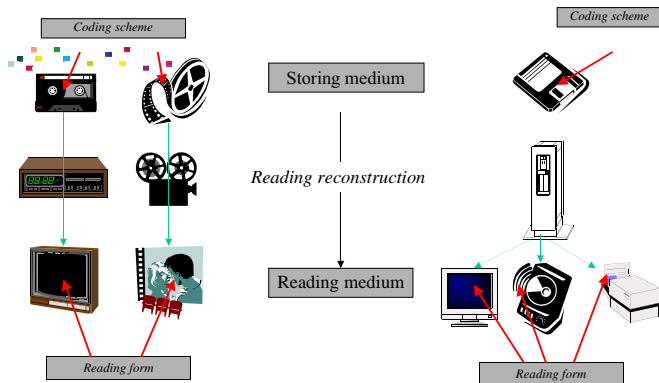
10

Slide 11

Some numbers:

- Professional Archives :
  - 1,1 million hours of radio an TV programmes;
    - 575 000 hours / Radio ;
    - 535 000 hours / TV
    - + ~ 50 000 hours / year
- Legal Deposit :
  - 930 000 hours (NB: professional archives are partially included in legal deposit).
    - 430 000 hours / TV
    - 500 000 hours / radio
    - + 500 000 hours / year
- 2,5 millions documents covering 113km of shelf space;
  - 8km / year
- 133 years for watching or listening all archives;

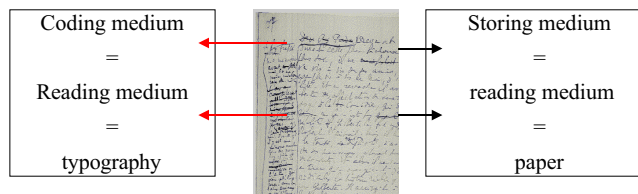11

Slide 12

What is an audiovisual document?

Why is it a strange thing?

Slide 13

What is a document ?

Coding scheme

Storing medium

Reading reconstruction

Reading medium

Reading form

Coding scheme

Reading form

13

Slide 14

A special case: paper documents

Coding medium
=
Reading medium
=
typography

Storing medium
=
reading medium
=
paper

Reading reconstruction system : the reader !

14

Slide 15

Reading form

Static / Spatial forms:
Interpretable structures are shown *simultaneously* on the reading medium ;
The order and the rhythm of the reading are determined by the reader

Dynamic / Temporal forms :
Interpretable forms are shown successively on the reading medium;
The order and the rhythm of the reading are determined by the player.

15

Slide 16

Temporal forms : reconstructing the reading

- Problem :
  - Reading is a temporal and dynamic process imposed by the document;
  - Storage is static and spatial.
- Consequences :
  - Storing form ≠ reading form ;
  - Storing form = code that programs the temporal flow
  - Reading system = a mechanical / digital player.

What is read is not what is stored !

16

Slide 17

Temporal forms : apprehending globality

- Problem:
  - Reading support = static and spatial;
  - Reading form = temporal ;
- Consequences :
  - No global nor synthetic apprehension of the document ;
  - No possibility for browsing ;

No analytic access to temporal contents

17

Slide 18

Temporal forms: indirect access

- Temporal documents :
  - Viewing 10' video takes 10' ;
  - Watching the first 9 minutes to find an information at the 10th minute.
  ➡No direct access to content.
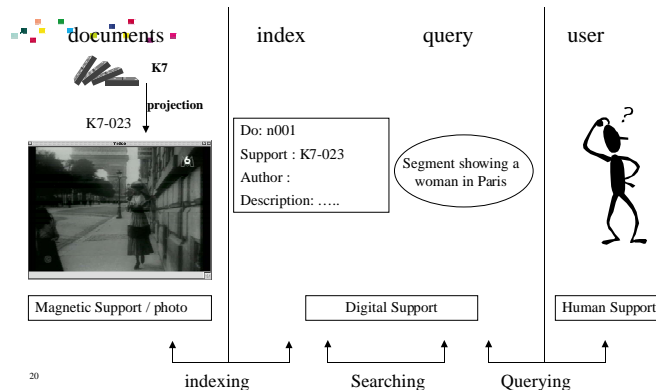- To find an information :
  - Explicitly localising information.

Needs for indexing

18

Audio-visual documents : two main applications

- Reconstruction of the reading form:
  - First application : playing the AV documents !!

- Indexing contents to retrieve them :
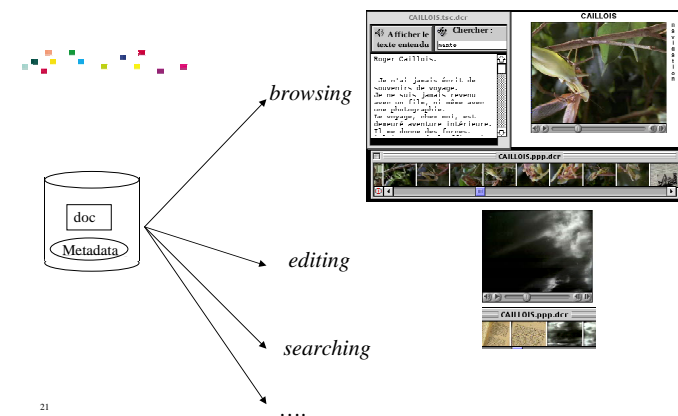 - Second application : retrieving the AV documents !!
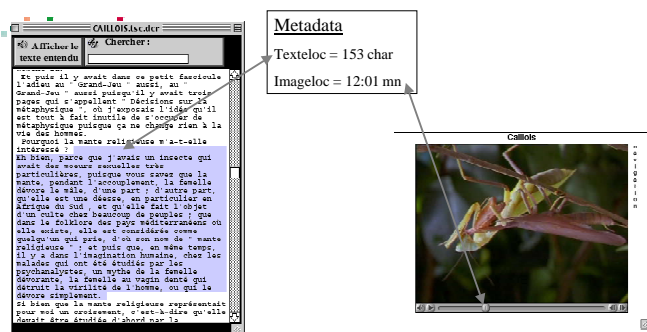
19

Analogical AV : viewing, retrieving

| documents | index | query | user |

K7

projection

K7-023

Do: n001
Support : K7-023
Author :
Description: …..

Segment showing a woman in Paris

| Magnetic Support / photo | Digital Support | | Human Support |

indexing   Searching   Querying

20

Using digital AV documents

browsing

doc
Metadata

editing

searching

….

21

Metadata for browsing

Metadata

Texteloc = 153 char

Imageloc = 12:01 mn

Caillois

22

Metadata for editing

| magazine Structure | Available Magazines | Best of | Best-of Structure |

Mag 1   Mag 2   Mag 3   Compile

people

people

people

1st people segment of each

interview

politics

politics

politics

1st politics segment of each

23

What do we need ?

Two main uses

- Retrieving AV materials for re-use by end users:
  - Productions;
  - News;
  - Pedagogical needs;
  - ….

- Proposing AV materials for research needs:
 - Historical, sociological, etc. research;

25

INA needs

- Building documented collections:
  - Editorial consistency;
  - Right management;
  - Cataloguing information
  - Information on technical quality.

- By browsing archives:
  - Detecting interesting document ;
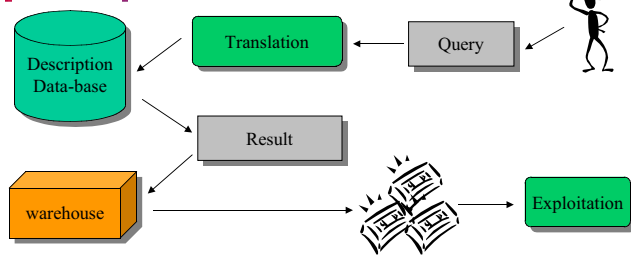 - Extracting relevant information to build automatically standard descriptions.

26

Past and present : analogical framework -1

Indexing → Description → Description Data-base

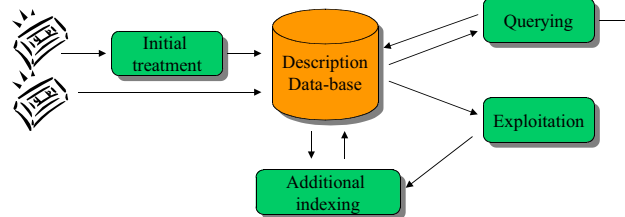Media Archiving → Warehouses
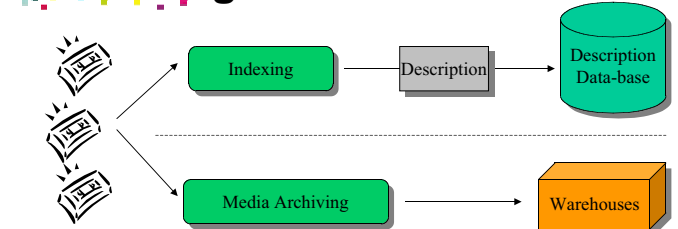
Upstream phase: feeding the bases.

27

Past and present: analogical framework - 2

Downstream phase: querying the description data-base
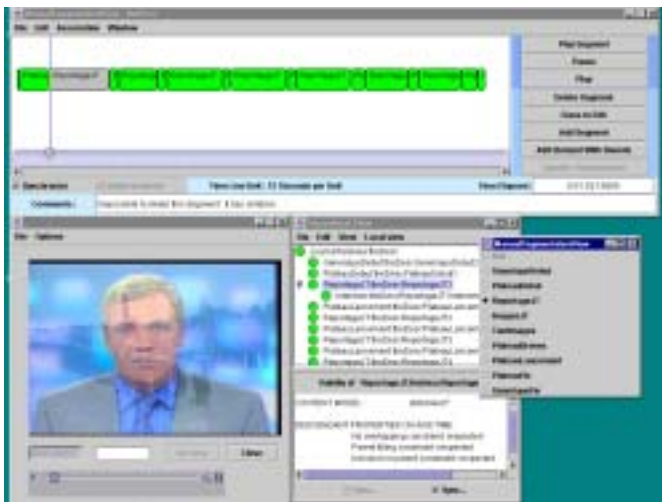
Present and future: digital framework

The digital promises

- Directly accessing to audio-visual content from meta-information;
- Dealing at the same time with audio-visual content and with metadata ;
- Applying automatic analysis on digitised videos to extract useful meta information and relevant descriptions;

  But this last promise is hard to turn into reality !

Four main uses:

Archive	Library
• Evidence or testimony Témoignage of a person activity ; • Structured according a *source principle;* • Objectives: *proof, sharing, testimony.*	• Document for consultation and reuse; • Structured according to a *use principle*; • Objectives: *adequacy, relevance* ;
Storing Avaibillity and persistency of a resource Structured according to an *accessibility and perennity principle*; Objectives: *accessing, reusing*.	*Legal Deposit* Preserving memory of what has been published, Structured according to a *medium principle*: Objectives: control, *memory, scientific research*.

What is indexing?

Why indexing audiovisual documents is so hard?

Indexing: a definition

- Indexing:
  - Reformulating content into a more exploitable form:
    - Exploiting content: retrieving it, re-using it, composing it, etc.
    - Form: natural language, controlled language, logical or conceptual formalism, ontologies, etc.
- Indexing is composed of three main steps:
  - Localizing content:
    - defining a particular part of content that deserves attention ;
  - Qualifying content:
    - Associating with the localized part a meaning: the point of view under which content is relevant or deserves attention.
  - Structuring content:
    - Structuring the relationships between qualifiers and qualified parts.

Localizing

- Objective:
  - defining what is meaningful in the document, which part may be consider as *signifier,* conveying meaning and being a sign:
    - A sign is a unit composed of a form emerging from a background;
    - A sign is a unit which is *meaningful,, i.e.* which means a *signified.*

- Problem: how to determine the signifiers?
  - Content is not built from units being a priori defined;
  - Signifiers are the result of an interpretation process.

Sign results from interpretation and does not precede it.

Example: localizing with tags

< Begin Tag 1>
  This is a text segment
< End Tag 1>
< Begin Tag 2>
  Another segment
  < Begin Tag 3 > yet another segment
  < End Tag 3 >
< End Tag 3 >

- Tags enable to localize the distinguished parts of content:
  - What should be interpreted because it is particularly relevant is bracketed by tags;
  - Tags semantics conveys a meaning that provides the bracketed content interpretation.

Qualifying content

- Objective:
  - Defining which meaning is associated with signifiers, *i.e.* the localized units:
    - In order to have meaning, one needs a sign systems where every sign is defined be its position in the system: signs are interrelated, interdependent and circularly defined;
- Problem:
  - Finding a system where units can be used to qualify content ;
    - Usually: language;
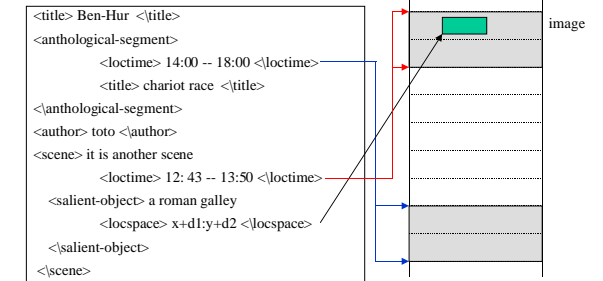 - Each unit expresses under which aspect the localized unit is relevant and meaningful.

Structuring

- Objective:
  - Gathering and structuring the qualifiers used in the qualifying phase.
    - Metadata;

- Problem:
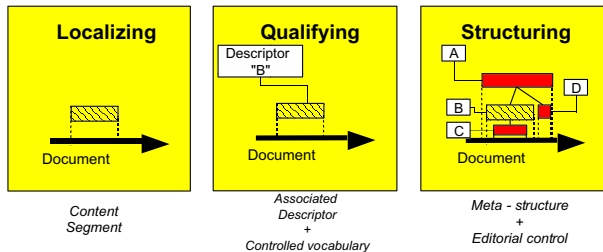 - How to structure qualifiers according to the document structure ?

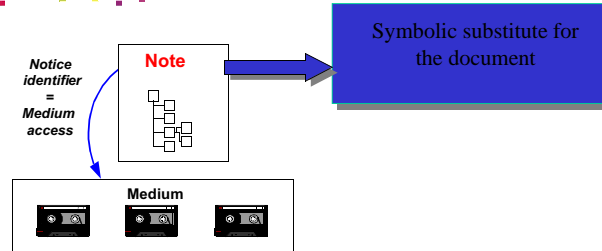# Relating document structure with description structure


```
<title> Ben-Hur  <\title>
<anthological-segment>
        <loctime> 14:00 -- 18:00 <\loctime>
        <title> chariot race  <\title>
<\anthological-segment>
<author> toto <\author>
<scene> it is another scene
        <loctime> 12: 43 -- 13:50 <\loctime>
  <salient-object> a roman galley
        <locspace> x+d1:y+d2 <\locspace>
  <\salient-object>
<\scene>
```

image

Summary

Paraphrasing content with a natural or controlled language
Three phases:

Localizing	Qualifying	Structuring
Document	Descriptor "B" / Document	A / B / C / D / Document
Content Segment	Associated Descriptor + Controlled vocabulary	Meta - structure + Editorial control

---

The analogical Context

Notice identifier = Medium access

Note
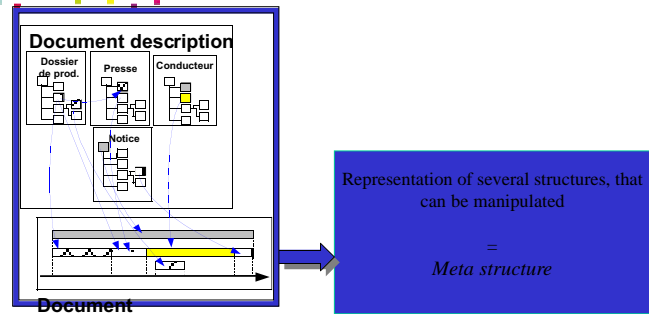
Symbolic substitute for the document

Medium

Digital born contents

- Documents, description and transmission are full digital:
  - Reading is performing through a computer or digital program;
  - Document and its description are stored on the same medium.
- Direct access to content can be operationalized:
  - Descriptor localization is a input data that can be used by the display device;
  - Those metadata can be considered as index.
- Describing content is a task that should be performed at every stage of the document lifecycle: creating, editing, broadcasting, archiving:
  - Several points of view are possible and can be used to parameterize reading.

The digital context

Document description

Dossier de prod.   Presse   Conducteur

Notice

Document

Representation of several structures, that can be manipulated

=

Meta structure

Meta structure

- The meta structure represents
  - A normalized content description taking into account several points of view
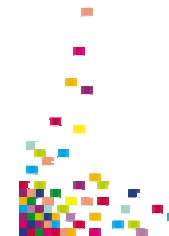 - A digital entity computationnally exploitable

It is a structure connecting test, still image, sounds and videos.

Hypermedia Structure?

From index to metadata

Making videos hypermedia objects

From index to metadata

- Index:
  - To point towards, to show where something is localised;
  - index only for information retrieval
- Metadata:
  - Data about data ;
  - Data that enables different exploitations of AV contents.

Metadata :

information that makes data useful

Metadata: key issues

- Creating metadata:
  - Manual and automatic indexing ;
- Representing metadata:
  - Document format ;
- Using metadata:
  - Document management.

Creating metadata

- Manual Annotation:
  - Abstract and conceptual interpretation;
- Automatic indexing:
  - Dealing with large amount of data;
- Working environment:
  - Helping the manual work by automatic assistants.

Creating metadata: the fundamental problem

- Objective:
  - Define descriptors of the AV content.
- Fundamental problem:
  - AV documents are not alphabetic.

Descriptors are not given with the document: they should be *extracted* and *interpreted*.

Automatic indexing: the semantic gap

- Objective:
  - Automatically extracting descriptors from AV contents.
- Problem:
  - Determining a relevant descriptor depends on the context;
  - Extracting algorithms are too close to the physical nature of the AV content to be useful in real contexts.
- Issue:
  - Mapping extracted descriptors to useful and relevant descriptors.

Automatic indexing: main approaches

- Temporal Segmentation: shots, scenes ;
- Spatial Segmentation: face detection, face recognition;
- Speech recognition and transcription ;
- Speech / transcription alignment;
- Close caption extraction;
- Etc.

Manual indexing: use the right word

- Goal :
  - Determining what is meaningful, and explicitating the associated meaning ;
- Problem:
  - There is no apriori meaning in AV documents: AV documents show perceptual reality and no conceptual meaning;
- Solution:
  - Paraphrasing AV contents with words.
- Issue:
  - Controlling words that are used for indexing.

Manual and automatic indexing

- Building a co-operation:
  - Mixing different automatic indexing methods.
  - Parameterising algorithms by high level knowledge on context;
  - Preparing manual indexing by automatic indexing.

Mutual reinforcements

- *THE* audio-visual problem:
  - No global apprehension, no means for browsing.
- Automatic indexing contribution:
  - Tools for browsing AV documents;
- Manual indexing contribution:
  - Meanings, associated with AV units, to help the AV document management and use.

Three domains for AV indexing

- *Automatic indexing* :
  - Defining units that help browsing;
- *Conceptual indexing*:
  - Interpreting content to define relevant units and their meaning;
- *Structural indexing* :
  - Structuring index to manage their use.

Why is it difficult ?

Interpretation

- Queries are abstract and conceptual :
  - Videos showing "unemployment", "sadness", etc.
- Audio-visual content is complex:
  - Cooperation between sounds, speech, images;
- Meaning of an audiovisual document is only partially conveyed by the video part:
  - Television = Radio + image ? (M. Chion)
  - What is the role of the video part in an AV document:
    - Illustration ?
    - Make the eye busy ?
    - ….
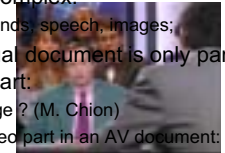
Interpretation is necessary

Problem

- Information for interpreting is not in the AV materials but:
  - In context of use;
  - In context of production
  - …

- Relation between what is shown and what is interpreted is "very" indirect !
  - Descriptors provided by automatic tools are too far from semantics to be useful.

Consequences

- Two levels :
  - Extracted physical descriptors
    - Exploit the information conveyed by the audiovisual form ;
  - Interpreted semantic indexes;
    - Re-formulate in terms of possible use or meaning for end users the audiovisual content
- But :
  - Descriptors are not index
  - Irreducible gap between the two.

But we really need:

- Dealing with huge amounts of video :
  - Needs for automatic tools ;

- Semantically and conceptually interpreting video:
  - Needs for relating analysis tools with conceptual interpretations of content.

What can be done ?

What can be done ? 1

- Patho-gnomonic signs of concepts in audiovisual manifestations
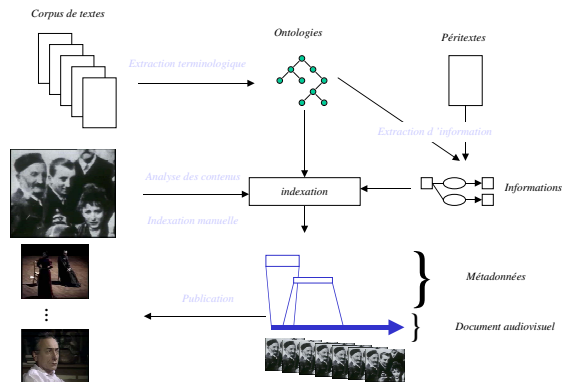 - For example: blue colour in background denoting stage shots:

What can be done ? -2

- Exploiting a priori information on video content:
  - Programming types : news, fiction;
  - Associating every type with the audiovisual information that can be extracted from content;
  - Building extraction scenarios that specify what can be extracted, how, and what it means.
- Making several technics cooperate:
  - Inscrusted text recognition & face detection;
  - Logos recognition & backgrounds colors;
  - Speech recognition & text analysis & macro-segmentation;
  - Etc.

Vue globale Indexation

Corpus de textes

Extraction terminologique

Ontologies

Péritextes

Extraction d'information

Analyse des contenus

indexation

Informations

Indexation manuelle

Publication

Métadonnées

Document audiovisuel

What can be done ? -3

- Exploiting a priori information on concepts used for interpreting :
  - concepts are related to socio-cultural schemata that specify how to represent them in audiovisual programming ;
  - These schemata can be related with extracting scenarios to detect when they are manifested by AV content.

65

Conclusion

- Situation:
  - Extracted information is still useless for us ;
  - Useful information relies upon semantic interpretation;
- Perspectives:
  - Formalising the regularities observed in semantic interpretation ;
  - Translating them in extraction scenarios;
  - Building tools to control these scenarios.

66