INF5080 Multimedia Coding and Transmission
Vårsemester 2004, Ifi, UiO

Multimedia Metadata and MPEG-7

Wolfgang Leister

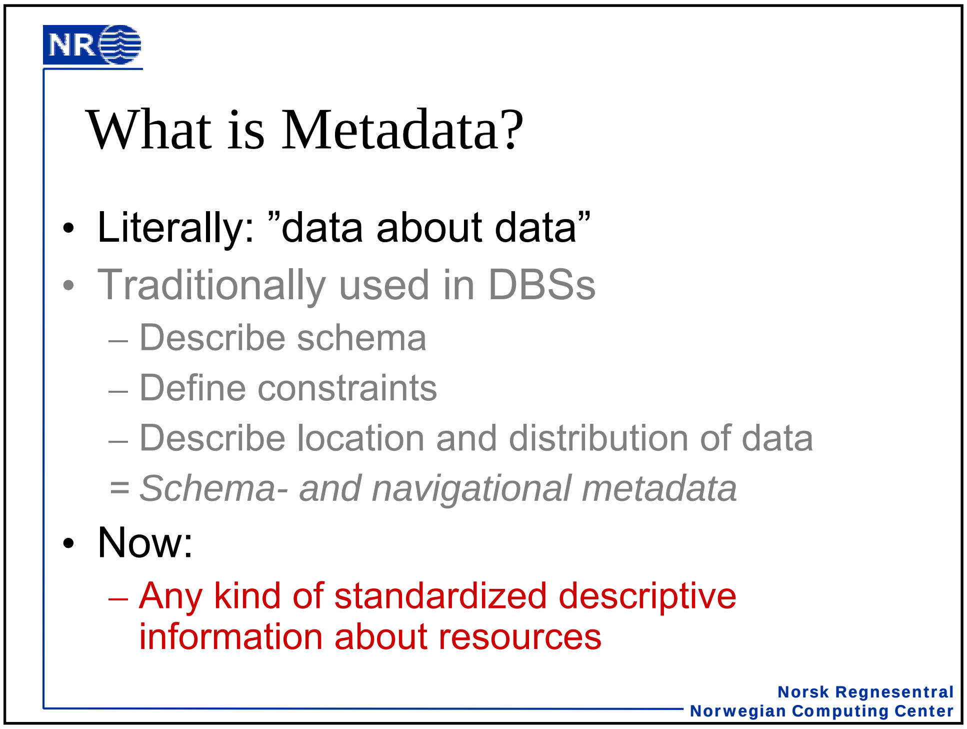
Knut Holmqvist

Ketil Lund

Halldór Matthías Sigurðsson

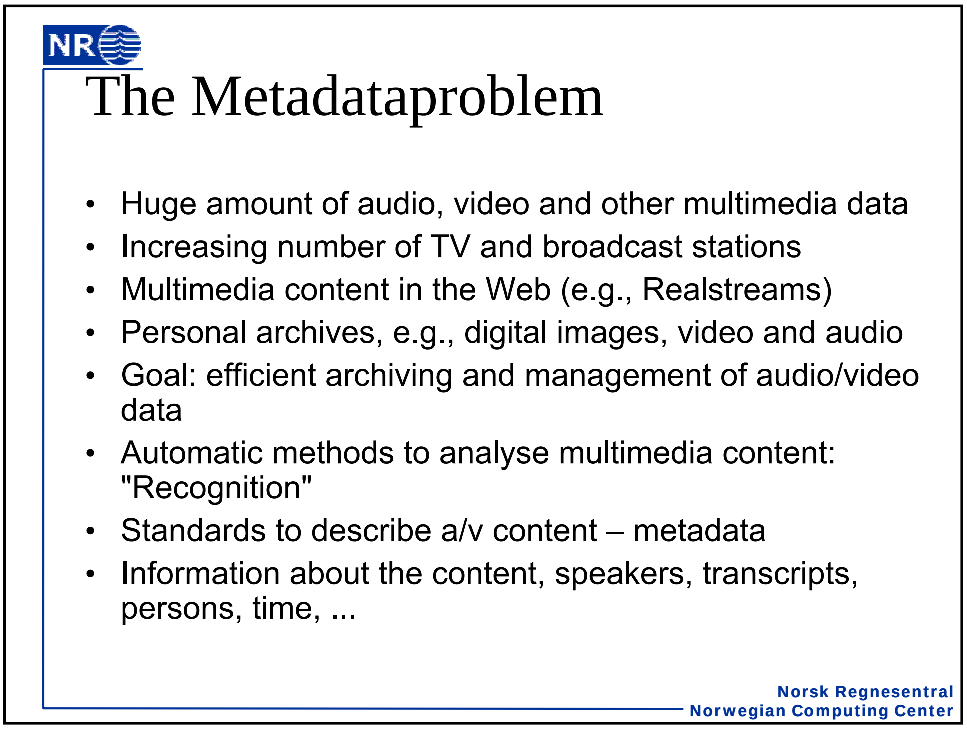
Today we'll talk about ...

- Metadata
- XML
- MPEG-7
- Dublin Core
- NewsML
- Application: AFR (Face Recognition)



What is Metadata?

- Literally: ”data about data”
- Traditionally used in DBSs
  - Describe schema
  - Define constraints
  - Describe location and distribution of data
  - = *Schema- and navigational metadata*
- Now:
  - Any kind of standardized descriptive information about resources



The Metadataproblem

- Huge amount of audio, video and other multimedia data
- Increasing number of TV and broadcast stations
- Multimedia content in the Web (e.g., Realstreams)
- Personal archives, e.g., digital images, video and audio
- Goal: efficient archiving and management of audio/video data
- Automatic methods to analyse multimedia content: "Recognition"
- Standards to describe a/v content – metadata
- Information about the content, speakers, transcripts, persons, time, ...

Multimedia Metadata

- What:
  - Metadata for multimedia resources
- Why:
  - Describe content
  - Index/random access to content
  - Describe formats
  - Management
  - ...
- Schema- and navigational metadata
  + associative metadata
- MM metadata can itself be multimedia data

Terminology

- Metadata
  - Structured data about data
  - Pre-defined elements with one or more values
- Metadata schema
  - Defines a set of metadata elements
  - Examples: Dublin Core, MPEG-7, ***TV Anytime, SMPTE Metadata Dictionary, EBU P/Meta***
- Encoding scheme
  - Defines the syntax
  - Examples: HTML, XML, RDF, MIME

Types of Multimedia Metadata

- Resource description
  - Content
  - Structural
  - Technical
- Resource management
  - Administrative
  - Preservation
  - Usage

Application Domains

- Different application domains have different metadata requirements:
  - File systems
  - Video servers
  - (MM)DBS
  - WWW
  - Digital libraries
  - e-Learning
  - Digital photography
  - ...
- No “one size fits all” – different solutions for different domains

Example:

Indexing of broadcast content

- Program, title
- Segment title into ”shots”, scenes and sequences
- describe metadata for each entity.

Title = Fredagskjøret
Creator = Radio Nova
Contributor.Presenter = DJ Kjøret
Date =……

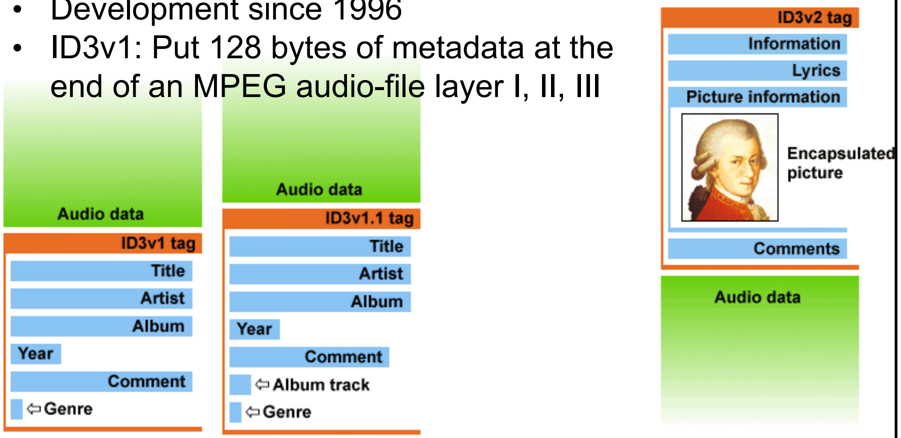
This solution is ...

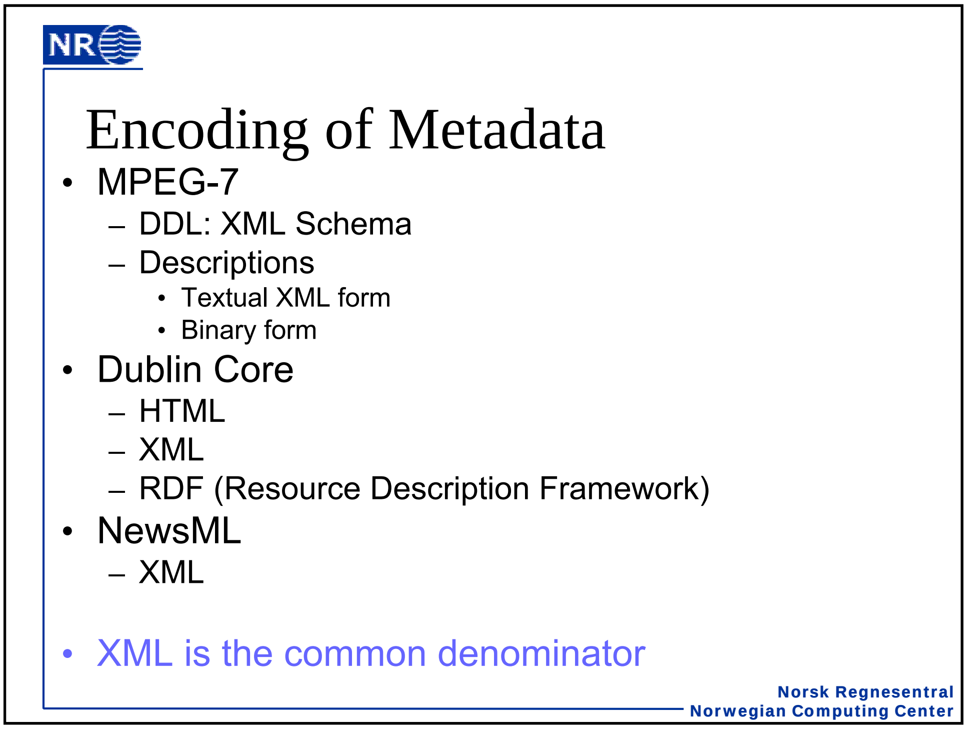
- ... not sufficient for multimedia
- ... not easy to use for others than librarians

http://www.id3.org

ID3

- Description of audio data
- Development since 1996
- ID3v1: Put 128 bytes of metadata at the end of an MPEG audio-file layer I, II, III

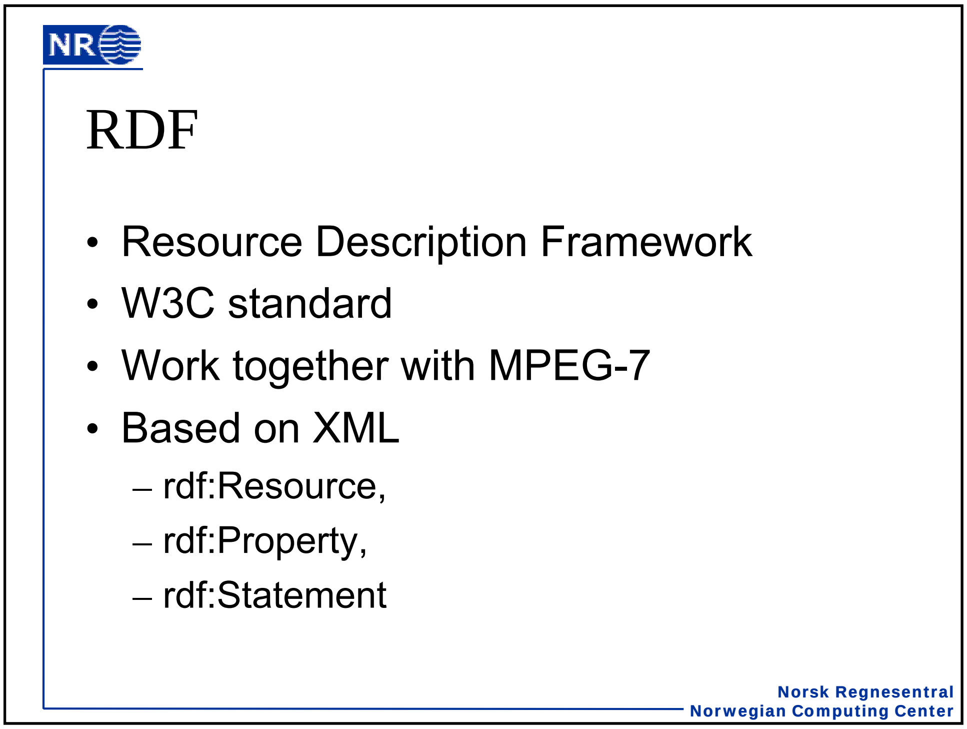




Encoding of Metadata

- MPEG-7
  - DDL: XML Schema
  - Descriptions
    - Textual XML form
    - Binary form
- Dublin Core
  - HTML
  - XML
  - RDF (Resource Description Framework)
- NewsML
  - XML

- XML is the common denominator



RDF

- Resource Description Framework
- W3C standard
- Work together with MPEG-7
- Based on XML
  - rdf:Resource,
  - rdf:Property,
  - rdf:Statement

XML 101

Syntax
- Basic syntax
  - Tags define elements
  - Attributtes describe elements
  - XML object must follow XML basic syntax
- DTD/XML schema define
  - which tags are allowed,
  - which combinations and
  - which attributes

```
<?xml version="1.0">
<notat>
  <fra>Knut</fra>
  <til>Knut</til>
  <melding>
    <overskrift>Husk!</overskrift>
    <tekst>
      Du må si noe om XML før du snakker om
MPEG-7 på kurset
    </tekst>
    <prioritet nivå="viktig"/>
  </melding>
</notat>
```

En takk til Arve Larsen for lån av XML slides

XML Schema

- Describes XML-syntax
- An XML schema-definition (.xsd) is an XML-document
- An XML schema-instance is an object that follows an XML schema.

Namespaces

- Used to separate attributes and elements defined in different standards.
- A document can use elements from several namespaces.
- A document can use a default namespace for all tags without explicit namespace.
- Example: replacing line 2 in example with

<notat xmlns="http://www.nr.no/imedia">
makes all a: optional.

```
<?xml version="1.0" encoding="UTF-8">
<a:notat xmlns:a="http://www.nr.no/imedia">
  <a:fra>Knut</a:fra>
  <a:til>Knut</a:til>
  <a:melding>
    <a:overskrift>Husk!</a:overskrift>
    <a:tekst>
      Du må si noe om XML før du snakker om
            MPEG-7 på kurset
    <a:tekst>
    <a:prioritet nivå="viktig"/>
  </a:melding>
</a:notat>
```

Schema Components

- Simple type definitions
- Complex type definitions
- Element declarations

} Primary components

- Attribute Group definitions
- Identity constraint definitions
- Model Group definitions
- Notation declarations
- Annotations

} Secondary components

- Attribute declarations
- Model groups
- Particles
- Wildcards

} Helper components

Example

```
<xs:schema
    xmlns:xs=“http://www.w3.org/1999/XMLSchema”
    targetNamespace=“http://www.mpeg.org/mpeg-7”
    version=“1.1”>
….
</xs:schema>
```

Example 2

Schema:

```
<complexType name=”VideoDoc">
      <element name=“Title”…./>
      <element name=“Producer”…./>
      <element name=“Date”…./>
   </complexType>
   <complexType name=”NewsDoc" base=”VideoDoc" derivedBy="extension">
      <element name=“Broadcaster” />
      <element name=“Time” />
   </type>
   <element name=”VideoCatalogue">
      <complexType>
         <element name="CatalogueEntry" minOccurs="0" maxOccurs="*" type=”VideoDoc"/>
      </complexType>
   </element>
```

Instance doc:

```
<CatalogueEntry type=”NewsDoc">
              <Title>”CNN 6 oclock News” </Title>
              <Producer>David James</ Producer >
              <Date>1999</Date>
              <Broadcaster>CNN</ Broadcaster >
</CatalogueEntry>
```

MPEG-7

- MPEG-7, formally named

"Multimedia Content Description Interface",

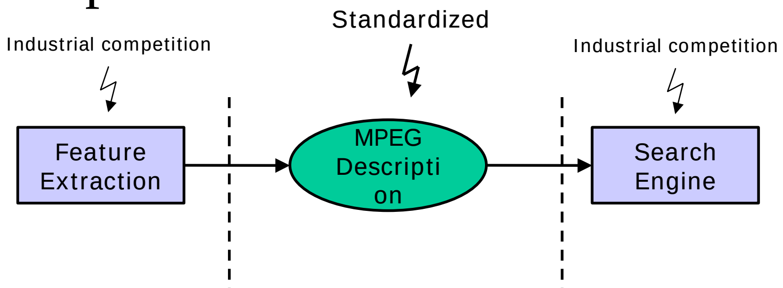
is *a standard for describing the multimedia content data that supports some degree of interpretation of the information’s meaning, which can be passed onto, or accessed by, a device or a computer code.*

- *ISO/IEC 15938*
- *Development since 1998, ISO standard 2001 (V1)*

MPEG-7

- A standard framework for describing all aspects of the content of a multimedia object
  - Low-level descriptions of individual objects in a scene
  - High-level abstract descriptions of scenes
  - Information related to content usage
  - Storage features
  - Structural information
  - ...
- Goal: Search, identify, filter, and browse audiovisual content

Scope



Extraction	MPEG-7 Scope	Use
Content analysis (D,DS) Feature extraction (D, DS) Annotation tools (DS) Authoring (DS)	Description Schemes (DSs) Descriptors (Ds) Language (DDL) Coding Schemes (CS)	Searching & filtering Classification Complex querying Indexing

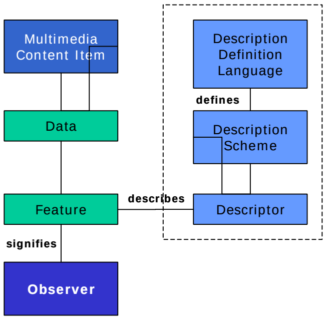
Source: Dr. John Smith
Multimedia Information Retrieval
and Management, Springer 2003

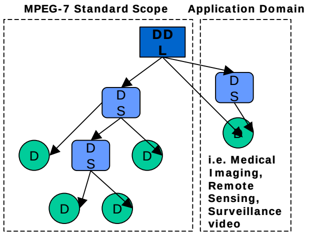
MPEG-7 main elements (I)

- Description Tools: Descriptors (D)
  - ***define the syntax and the semantics of metadata element***
  - ***Description Schemes (DS) specify the structure and semantics of the relationships between their components***
- A Description Definition Language (DDL)
  - define the syntax of the MPEG-7 Description Tools
- System tools, support binary coded representation for
  - ***efficient storage and transmission,***
  - ***transmission mechanisms,***
  - ***multiplexing of descriptions,***
  - ***synchronization of descriptions with content,***
  - ***management and protection of intellectual property***

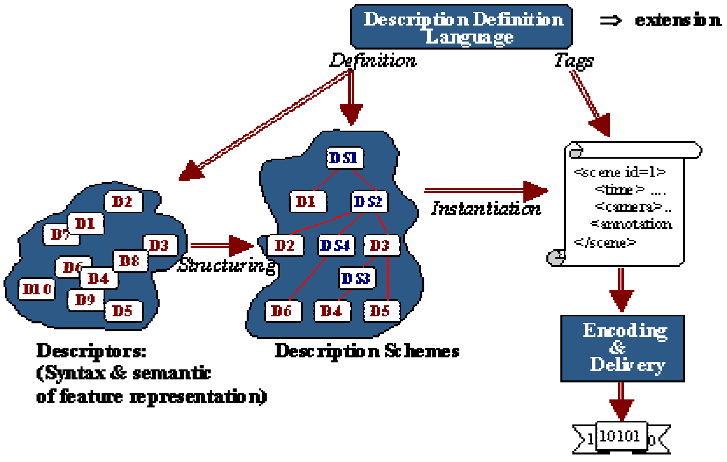
Main Elements (II)

- **Description Definition Language:**
  - The Description Definition Language (DDL) is the language specified in MPEG-7 for defining the syntax of Description Schemes and Descriptors. The DDL is based on the XML Schema Language
- **Description Schemes (DS):**
  - Description Schemes (DS) are description tools defines using DDL that describe entities or relationships pertaining to multimedia content
- **Descriptors(D):**
  - Descriptors are description tools defined using DDL that describe features, attributes, or groups of attributes of multimedia content.
- **Feature**
  - Features are defined as a distinctive characteristic of multimedia content that signifies something to a human observer, such as the “color” or “texture” of an image
- **Data**
  - Multimedia Data is defined as a representation of multimedia in a formalized manner suitable for communication, interpretation, or processing by automatic means (i.e. for example image or video)

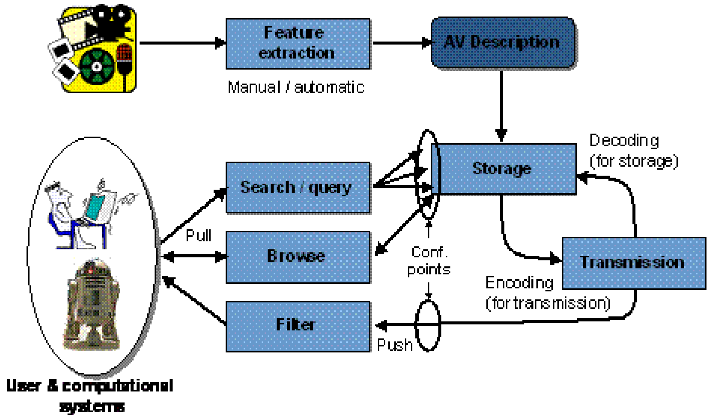




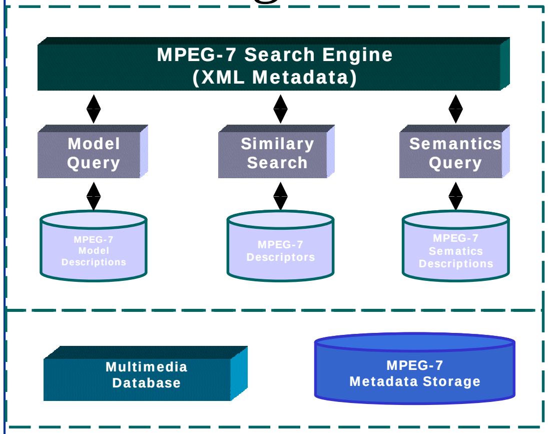
MPEG-7 Main Elements (III)



MPEG-7 Applications I



Searching and Indexing



- Semantics-based
  - people, places, events, objects, scenes
- Content-based
  - Color, texture, motion, melody, timbre
- Metadata
  - Title, author, dates
- Examples
  - Sketch up a logo and search to search for the company
  - Whistle a tune and search for the song
  - Search for information about a person from a picture of him

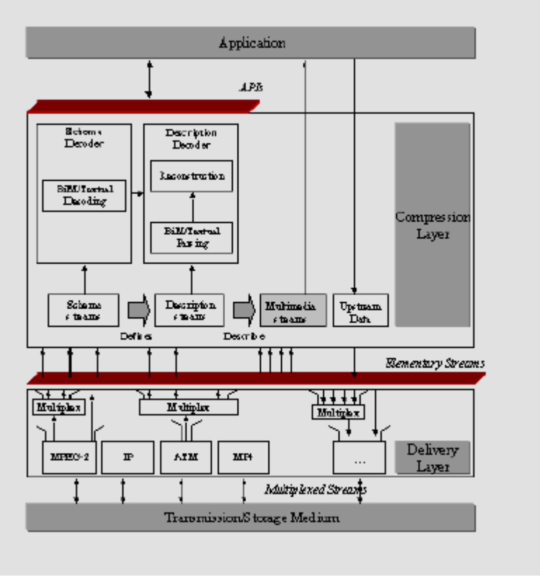
MPEG-7 Parts

- MPEG-7 Systems – binary encoding format, terminal architecture.
- MPEG-7 Description Definition Language
- MPEG-7 Visual – DT for Visual descriptions.
- MPEG-7 Audio – DT for Audio descriptions.
- MPEG-7 Multimedia Description Schemes
- MPEG-7 Reference Software
- MPEG-7 Conformance Testing
- MPEG-7 Extraction and use of descriptions – informative

MPEG-7 DDL

- **Description Definition Language**
- **The DDL is based on XML Schema Language.**
- **MPEG-7 adds extensions for audiovisual content**
- **DDL consists of:**
- **The XML Schema structural language components;**
- **The XML Schema datatype language components;**
- **The MPEG-7 specific extensions.**

MPEG-7 Terminal Architecture



Access units

MPEG-7 Visual Component

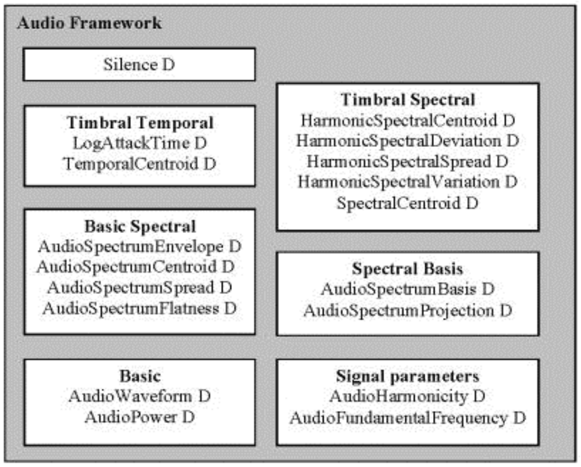
- Content-based image/video retrieval
- Six descriptor categories:
  - Color descriptors (7)
  - Texture descriptors (3)
  - Shape descriptors (3)
  - Motion descriptors (4)
  - Localization descriptors (2)
  - Face recognition (1)

- Grid Layout
- 2D-3D Multiple View
- Time Series
- Spatial 2D Coordinats
- Temporal Interpolation

MPEG-7 Audio Component

- Audio description framework
  - 17 low-level descriptors (temporal and spectral)
  - Generic (application independent)
- High-level audio description tools
  - Audio signature description scheme
  - Musical instrument timbre description tools
  - Melody description tools
  - General sound recognition and indexing description tools
  - Spoken content description tools

MPEG-7 Audio (II)



Audio Framework

Silence D

Timbral Temporal
LogAttackTime D
TemporalCentroid D

Timbral Spectral
HarmonicSpectralCentroid D
HarmonicSpectralDeviation D
HarmonicSpectralSpread D
HarmonicSpectralVariation D
SpectralCentroid D

Basic Spectral
AudioSpectrumEnvelope D
AudioSpectrumCentroid D
AudioSpectrumSpread D
AudioSpectrumFlatness D

Spectral Basis
AudioSpectrumBasis D
AudioSpectrumProjection D

Basic
AudioWaveform D
AudioPower D

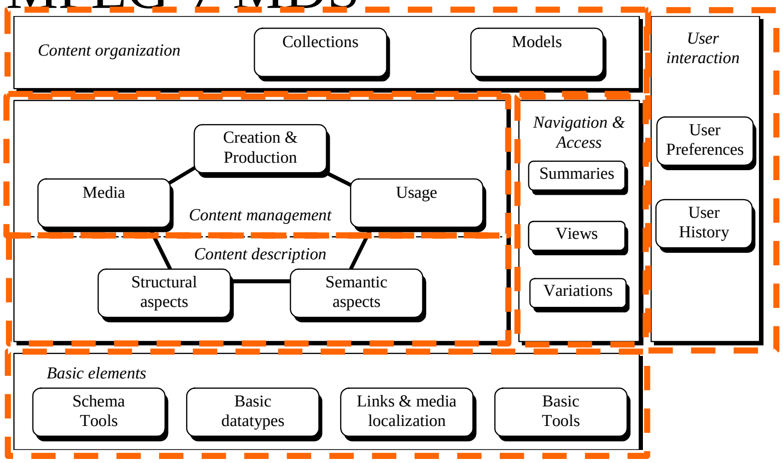
Signal parameters
AudioHarmonicity D
AudioFundamentalFrequency D

MPEG-7 MDS:

Multimedia Description Schemes

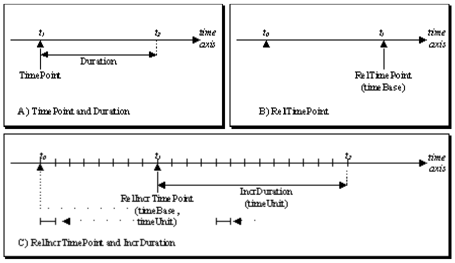
- MDS combine individual «descriptors»
- Content description: representation of perceivable information
- Content management: information about the media features, the creation and the usage of the AV content;
- Content organization: representation, the analysis and classification;
- Navigation and access: specification of summaries and variations;
- User interaction: description of user preferences and usage history

MPEG-7 MDS



(ISO/IEC 03)

MPEG-7 MDS (II)



A) TimePoint and Duration

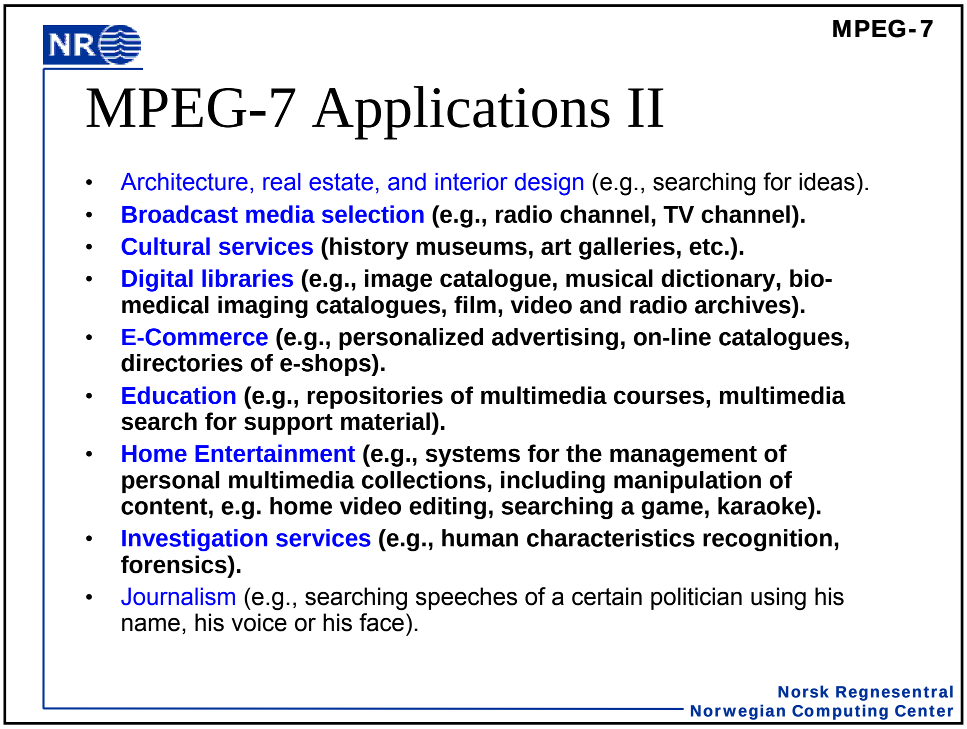
B) RelTimePoint

C) RelIncrTimePoint and IncrDuration

Overview of Time DSs

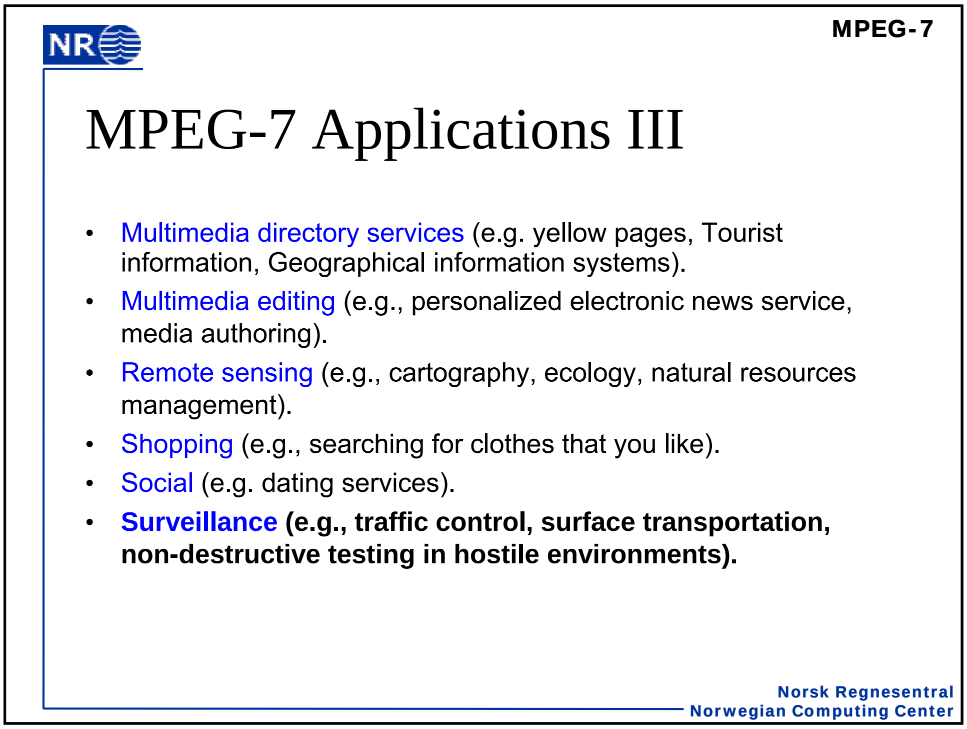
XM – the eXperimentation Model

- XM programs
- Framework for reference code
- Implements normative components (+)
  - Descriptors
  - Description Schemes
  - Coding schemes
  - DDL
  - BiM system components



MPEG-7 Applications II

- Architecture, real estate, and interior design (e.g., searching for ideas).
- **Broadcast media selection (e.g., radio channel, TV channel).**
- **Cultural services (history museums, art galleries, etc.).**
- **Digital libraries (e.g., image catalogue, musical dictionary, bio-medical imaging catalogues, film, video and radio archives).**
- **E-Commerce (e.g., personalized advertising, on-line catalogues, directories of e-shops).**
- **Education (e.g., repositories of multimedia courses, multimedia search for support material).**
- **Home Entertainment (e.g., systems for the management of personal multimedia collections, including manipulation of content, e.g. home video editing, searching a game, karaoke).**
- **Investigation services (e.g., human characteristics recognition, forensics).**
- Journalism (e.g., searching speeches of a certain politician using his name, his voice or his face).



MPEG-7 Applications III

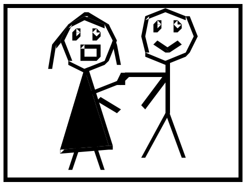
- Multimedia directory services (e.g. yellow pages, Tourist information, Geographical information systems).
- Multimedia editing (e.g., personalized electronic news service, media authoring).
- Remote sensing (e.g., cartography, ecology, natural resources management).
- Shopping (e.g., searching for clothes that you like).
- Social (e.g. dating services).
- **Surveillance (e.g., traffic control, surface transportation, non-destructive testing in hostile environments).**

MPEG-7 Applications IV

- MPEG-7 Camera
  - On-the-fly generation of content descriptions
- MPEG-7 Spoken Content Transcription service
  - Create an MPEG-7 Audio “SpokenContent” description file from an audio file in “wav” format
- Ricoh MovieTool
  - A tool for generating MPEG-7 descriptions based on the structure of a video
- Query-by-Humming (MusicLine)
  - Search for melodies by humming a tune

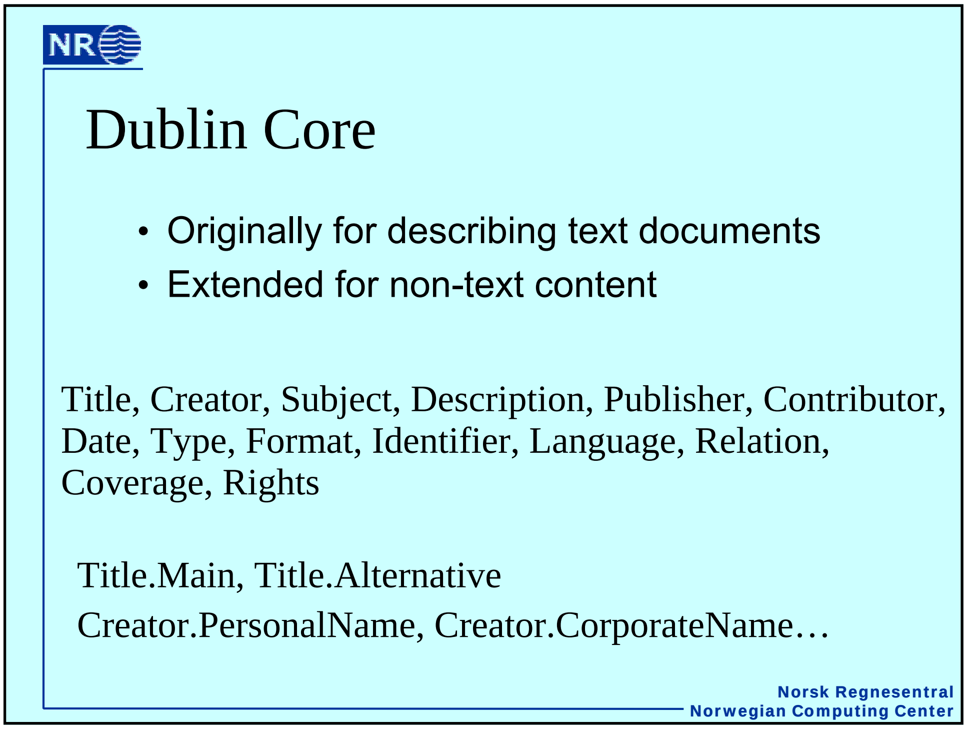
MPEG-7 Example

The following example gives an MPEG-7 description of the event of handshake between people:



```
<Mpeg7>
  <Description xsi:type="SemanticDescriptionType">
    <Semantics>
      <Label>
        <Name> Shake hands </Name>
      </Label>
      <SemanticBase xsi:type="AgentObjectType" id="A">
        <Label href="urn:example:acs">
         <Name> Person A </Name>
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="AgentObjectType" id="B">
        <Label href="urn:example:acs">
          <Name> Person B </Name>
        </Label>
      </SemanticBase>
      <SemanticBase xsi:type="EventType">
        <Label><Name> Handshake </Name></Label>
        <Definition>
          <FreeTextAnnotation> Clasping of right hands by two people </FreeTextAnnotation>
        </Definition>
        <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:agent" target="#A"/>
        <Relation type="urn:mpeg:mpeg7:cs:SemanticRelationCS:2001:accompanier“ target="#B"/>
      </SemanticBase>
    </Semantics>
  </Description>
</Mpeg7>
```

Source: Dr. John Smith
Multimedia Information Retrieval
and Management, Springer 2003



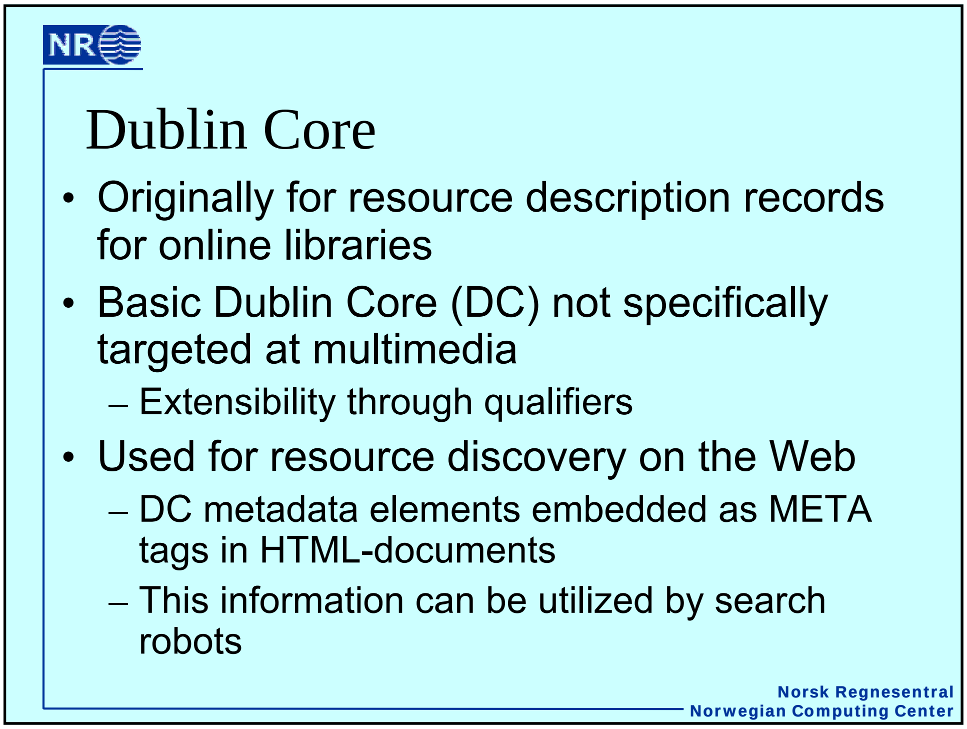
Dublin Core

- Originally for describing text documents
- Extended for non-text content

Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Language, Relation, Coverage, Rights

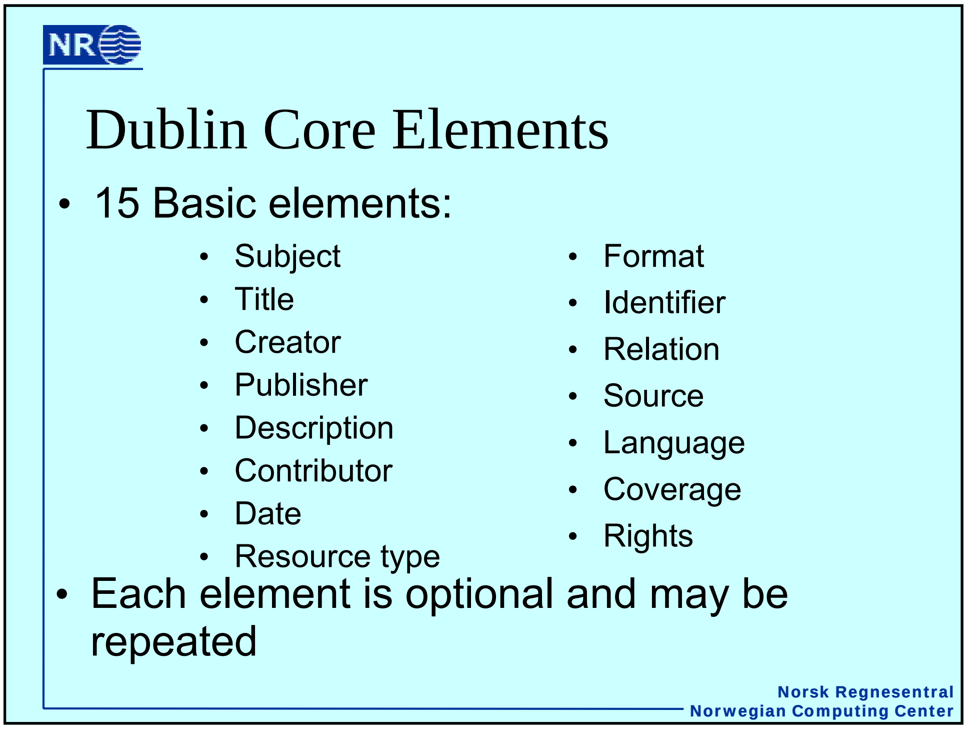
Title.Main, Title.Alternative

Creator.PersonalName, Creator.CorporateName…



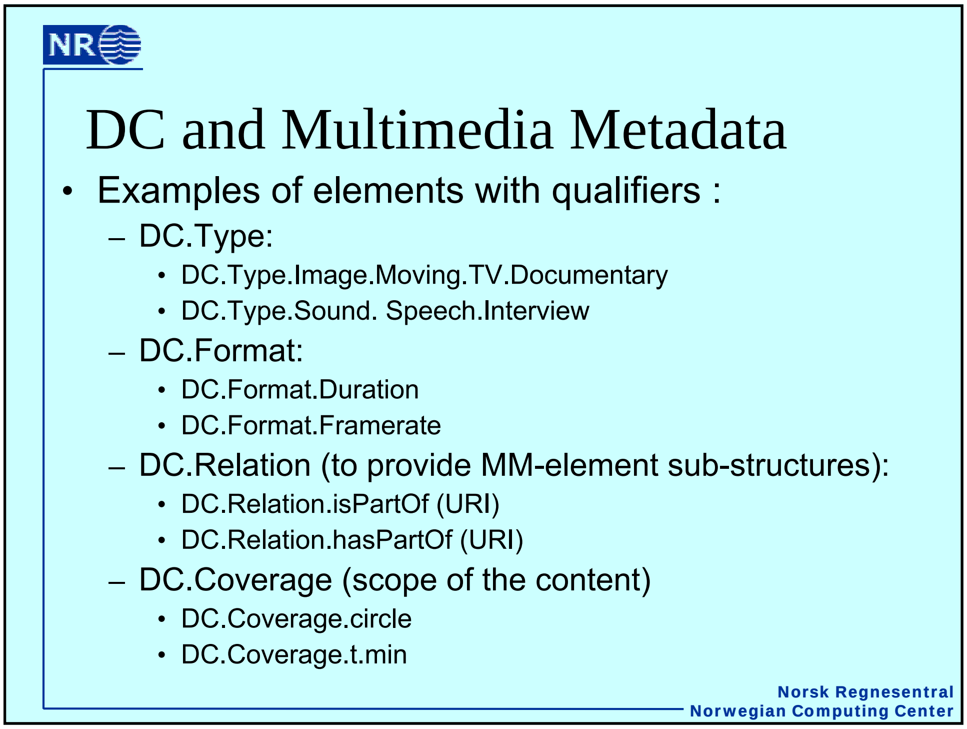
Dublin Core

- Originally for resource description records for online libraries
- Basic Dublin Core (DC) not specifically targeted at multimedia
  - Extensibility through qualifiers
- Used for resource discovery on the Web
  - DC metadata elements embedded as META tags in HTML-documents
  - This information can be utilized by search robots



Dublin Core Elements

- 15 Basic elements:
  - Subject
  - Title
  - Creator
  - Publisher
  - Description
  - Contributor
  - Date
  - Resource type
  - Format
  - Identifier
  - Relation
  - Source
  - Language
  - Coverage
  - Rights
- Each element is optional and may be repeated



DC and Multimedia Metadata

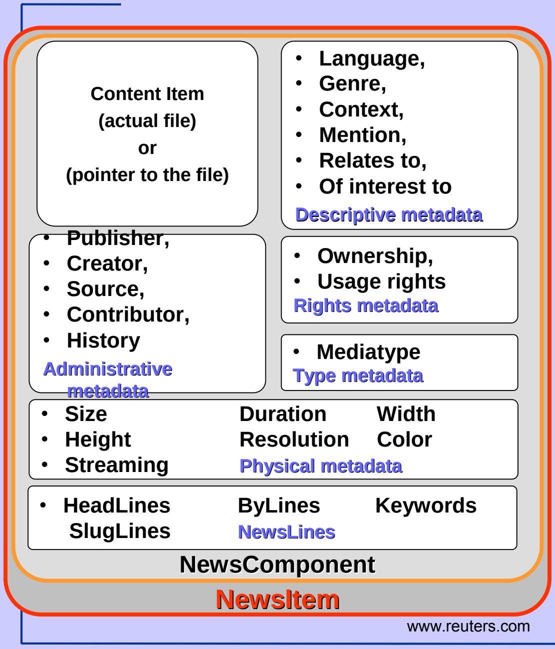
- Examples of elements with qualifiers :
  - – DC.Type:
    - DC.Type.Image.Moving.TV.Documentary
    - DC.Type.Sound. Speech.Interview
  - – DC.Format:
    - DC.Format.Duration
    - DC.Format.Framerate
  - – DC.Relation (to provide MM-element sub-structures):
    - DC.Relation.isPartOf (URI)
    - DC.Relation.hasPartOf (URI)
  - – DC.Coverage (scope of the content)
    - DC.Coverage.circle
    - DC.Coverage.t.min

DC - Examples

- Shared Online Media Archive (SOMA)
    - An online media archive for use by community radio stations
- Exchange Broadcast Binary and Metadata Format (XBMF)
    - File and metadata format for exchanging broadcasts between various broadcast content management systems
- EBU Core Metadata Set for Radio Archives
    - A simple set of metadata which is adapted for use in radio archives
- The Berkeley Digital Library SunSITE
    - Library containing text, images, video, and sound
- The Gateway to Educational Materials (GEM)
    - An effort to provide access to collections of Internet-based educational materials

NewsML

- News Markup Language for global news exchange
    - A media-independent, structural framework for interchange and management of multimedia news
    - Developed by IPCT (International Press Telecommunications Council)
    - Current version: 1.2 (October 2003)
- Typical areas of application
    - In and between editorial systems
    - Between news agencies and their customers
    - Between news service providers and end-users



Content Item
(actual file)
or
(pointer to the file)

- Language,
- Genre,
- Context,
- Mention,
- Relates to,
- Of interest to

Descriptive metadata

- Publisher,
- Creator,
- Source,
- Contributor,
- History

Administrative metadata

- Ownership,
- Usage rights

Rights metadata

- Mediatype

Type metadata

- Size
- Height
- Streaming

Duration
Resolution

Width
Color

Physical metadata

- HeadLines
SlugLines

ByLines
NewsLines

Keywords

NewsComponent

NewsItem

www.reuters.com

Structure

- NewsML
  - Root element
- NewsEnvelope
  - “Address label”
- NewsItem
  - A “piece of news”
- NewsComponent
  - Container for news objects

NewsML Features

- Arbitrary combination of media types and encodings
- Multiple languages
- Versioning
- Flexibility of display
- Efficient searching and filtering
- Generic infrastructure

NewsML Applications

- Reuter
- Business Wire
- Agence France Presse (AFP)
- United Press International (UPI)
- PR Newswire, UK
- Eidetica (Care4Cure)

Other Approaches

- Multimedia presentations
  - SMIL 2.0 (Meta-information module)
  - ***TV Anytime,***
  - ***SMPTE Metadata Dictionary,***
  - ***EBU P/Meta***
- Voice interaction
  - VoiceXML (Voice interaction with web-services)
- File systems
  - MXF (Material eXchange Format)
  - WinFS (File system in Windows Longhorn)

Domain (in-)dependency

- Three schemas for multimedia metadata:
  - Domain independent
    - MPEG-7
    - Dublin Core w/multimedia extensions
  - Domain specific
    - NewsML

Sakset fra Innlegg ved Håvard Hegna på seminar om
”Emerging Technologies” på E_SIG ved NR 22.januar 2004 om
”Face Recognition and MPEG-7” med utgangspunkt i
artikkelen nedenfor og MPEG-7-standarden:

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11
MPEG2003/M9801
July 2003, Trondheim, Norway

Title: Multi-view 3D-Face Descriptor: proposal for CE
Source: Samsung Advanced Institute of Technology
Author: Won-Sook LEE and KyungAh SOHN

A new face descriptor, which aims to contain multi-view 3D-information of a person to help for search, retrieval and browsing of images, videos and 3D-facial model databases.

MPEG-7 Face Recognition

- MPEG-7 Visual Description Tools:
  - color, texture, shape, motion, localization, and **face recognition**.
- MPEG-7 Visual – Basic Structures:
  - Grid layout, Time series, Multiple view, Spatial 2D coordinates,Temporal interpolation
- The 2D/3D (Multiple View) Descriptor
  - 2D Descriptors representing visual feature of 3D objects, seen from different view angles.
  - forms complete 3D view-based representation of object.

MPEG-7 Visual
face recognition

- The FaceRecognition descriptor can be used to retrieve face images which match a query face image.
- The descriptor represents the projection of a face vector onto a set of basis vectors which span the space of possible face vectors.
- The FaceRecognition feature set is extracted from a normalized face image, which contains 56 lines with 46 intensity values in each line. The centers of the two eyes in each face image are located on the 24th row and the 16th and 31st column for the right and left eye resp.
- This normalized image is then used to extract the one dimensional face vector which consists of the luminance pixel values from the normalized face image arranged into a one dimensional vector using a raster scan starting at the top-left corner of the image and finishing at the bottom-right corner of the image.
- The FaceRecogniton feature set is then calculated by projecting the one dimensional face vector onto the space defined by a set of basis vectors.

MPEG-7 Face-Recognition Press Release I

- **Tokyo,December 16, 2003** --- NEC Corporation (NEC) and Samsung Advanced Institute of Technology (SAIT) today announced that the MPEG (Moving Picture Experts Group) Committee has decided to adopt NEC and SAIT jointly proposed new face recognition technology for the upcoming MPEG-7 standard (*) to be published in "ISO/IEC 15938-3:2002/Amd.1." in the spring of 2004.
- The NEC/SAIT technology was chosen for facial recognition because it performed best in retrieval accuracy, speed and data size as benchmarked by MPEG-7.
- Referred to as MPEG-7 AFR (advanced face recognition descriptor), the technology is a description method that presents facial features in still or moving picture form for multimedia retrieval. It requires only 253 bits to accurately identify a face, NEC claimed.

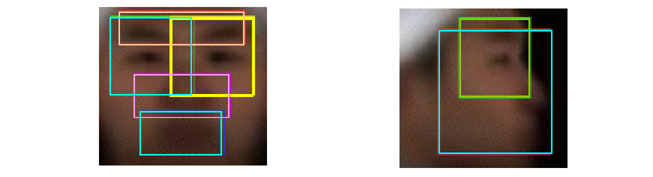
MPEG-7 Face-Recognition Press Release II

- In comparison to the previous standard, this technology achieves a reduction in the rate of retrieval error by one eighth (1/8) on average. In addition, it realizes a matching speed capability of one million times per second on a conventional PC thus making it **possible to retrieve a scene starring a specific person in approximately one second from a 24 hour video.**
- NEC developed "Cascaded Linear Discriminant Analysis", which selects features of human faces in order of performance within the cascading architecture and realizes **an accurate description of each face image in a minimum data size of 253 bits/face.**
- SAIT developed "Face Component Based Face Feature Representation Method" that **extracts facial features from each face component, such as the eyes and mouth**, and when applied to the NEC CLDA improves the level of accuracy of the technology.

Face Recognition by Principal Component Analysis (PCA)

- Method to extract features of a set of observations
  - ***Typical example: image analysis:***
  - ***1) in multispectral images the observations correlate***
  - ***2) project the original observations to a new coordinate system with fewer dimensions and minimal correlation,***
  - ***3) use the new «essential» set of observations for further processing.***
- Face recognition often uses PCA. *Eigenface*-method by Turk og Pentland (1991) is PCA. Uses observations for entire face.
- *Eigenfeature* is based on main features, nose, mouth eyes, and establishes *eigennose*, *eigenmouth* etc. Which are combined together.
- PCA needs training phase where «essential» coordinate systems for some objects are established.
- Note, NEC/SAIT uses something else: ”Linear Discriminant Analysis”

Feature extraction or holistic Description?



Five subregion definition on View (0°, 0°) Two subregion definition on View (80°, 0°)

Multi-view 3D-Face Descriptor

- Uses modified version of the current best algorithm, known as *Subregion-based LDA on Fourier space*
  A. Yamada and L. Cieplinski, “MPEG-7 Visual part of eXperimentation Model Version 17.1 ”, ISO/IEC JTC1/SC29/WG11 M9502 , Pattaya, Thailand, March 2003
- Reduced set of views for the training/characterization
  - typical poses in typical situations (i.e. frontal slightly angeled)
  - views with a large ”quasi-view”
  - views that appear a lot in practice
  - views whose ”quasi-views” together cover a large region
  - views that are easy to obtain
  - use just relevant regions for different poses
- The aim of the new descriptor is to find an optimization to build any-view-information by giving more information and by optimizing them.

Face Recognition

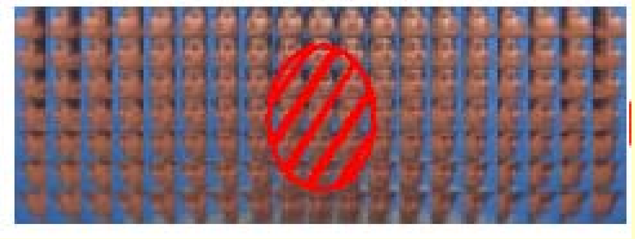
Face Representation

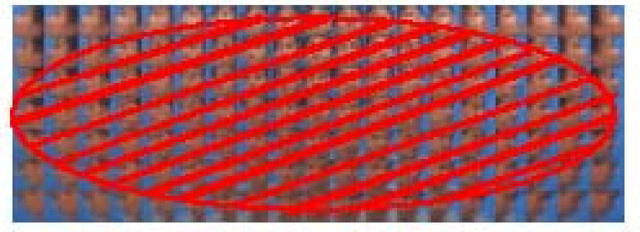
- Characterization of faces
- Compression, meta data
- Search and find

View-Mosaic

Area / quasi-view

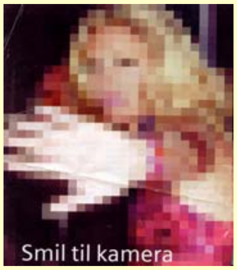
Front





10*7 poses = 133 images

Avisenes store piksler



Er avisenes bruk av ”store piksler” godt nok når de vil gi inntrykk av å være opptatt av ”i all anstendighet” å skjule identiteten til offeret på avisens forside?

End of Lecture

Thank you for your attention!