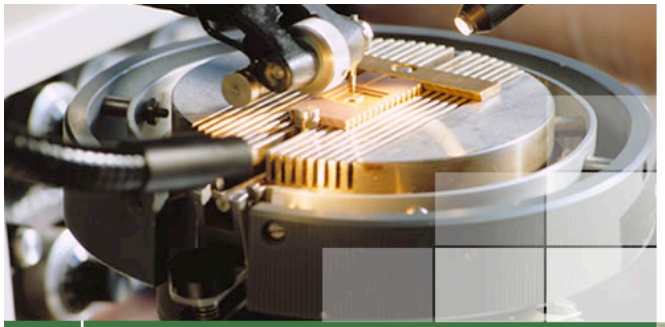


POLITECNICO DI MILANO

Dipartimento di
Elettronica e Informazione

Search Computing

Multimedia Information Retrieval and Search Engines

Alessandro Bozzon

Agenda

- Introduction and Motivation for Multimedia Information Retrieval (MIR)
- Content sources and challenges
- The MIR architecture
- Metadata
- Media Processing: (text), image, audio, video
- Multimodal content processing
- Query processing
- Current trends and open challenges
- References

Introduction and motivation

- Information society & (multimedia) information overload
- 161 exabytes of information was created or replicated worldwide in 2006
- IDC estimates 6X growth by 2010 to 988 exabytes (a zetabyte) / year
- That's more than in the previous 5,000 years.

  - **DATA from: Dr. Michael L. Brodie - Chief Scientist Verizon**

Where does content come from

- The largest source of data? → USERS
- YouTube Videos
  - 1.7 billion served / month
  - 1 million streams / day = 75 billion e-mails
- Facebook had [in 2007] …
  - 1.8 billion photos
  - 31 million active users
  - 100.000 new users / day
  - 1,800 applications
- MySpace, 185+ million registered users (Apr 2007), has…
  - Images:
    - 1+ billion - Millions uploaded / day- 150,000 requests / sec
  - Songs:
    - 25 million - 250,000 concurrent streams
  - Videos:
    - 60 TB - 60,000 uploaded / day - 15,000 concurrent streams

Challenges

- How to make multimedia content available to search engines and search based applications?

- Exploiting multimedia content requires:
  - Acquiring it
  - (Re) Formatting it
  - Indexing it
  - Querying it
  - Transmitting it
  - Browsing it

MIR: Query Examples [from MPEG overview]

- Play a few notes on a keyboard and retrieve a list of musical pieces similar to the required tune, or images matching the notes in a certain way, e.g., in terms of emotions
- Draw a few lines on a screen and find a set of images containing similar graphics, logos, ideograms,...
- Define objects, including color patches or textures and retrieve examples among which you select the interesting objects to compose your design
- On a given set of multimedia objects, describe movements and relations between objects and so search for animations fulfilling the described temporal and spatial relations
- Describe actions and get a list of scenarios containing such actions
- Using an excerpt of Pavarotti's voice, obtaining a list of Pavarotti's records, video clips where Pavarotti is singing and photographic material portraying Pavarotti

Some Terminology

- Raw Content: the media element in native format, before any processing
- Metadata: data about data
- Global metadata: metadata that refers to an entire media element (e.g., movie title and director)
- Local metadata: metadata pertaining a specific media segment (e.g., a scene in a movie)
- Manual metadata: media descriptions edited manually
- Automatic metadata (annotations): media descriptions extracted by the software
- Derived artifacts: secondary content elements extracted (manually or automatically), e.g., video key-frames, summarizations, freebies

More about metadata

- Metadata are data about data
  - They describe in a **structured** way properties of the data
    - E.g.: creator, owner, creation date, **description**
- Some metadata are explicitly provided with the data
  - E.g.: size, file name, etc.
- Others are **implicit**, and they must be extracted by means of algorithms and data analysis

Content acquisition

- In text or Web search engines, content is a closed or open collection of documents
- Textual Web content is acquired by crawlers, who exploit link navigation
- In MIR, content is acquired from many sources and in in multiple ways:
  - By crawling
  - By user's contribution
  - By syndicated contribution from content aggregators
  - Via broadcast capture (e.g., from air/cable/satellite broadcast, IPTV, Internet TV multicast, ..)

Content (re)formatting

- In textual search applications, the acquired content is ready to be indexed
- In MIR, content is on different media, has different formats, is described with different metadata
- Options range from low-quality, undescribed user-generated video to feature films with multi-language captions and extensive description
- Prior to any processing, multimedia content must be subject to a normalization procedure, to lower the degree of heterogeneity or to prepare alternative versions for different delivery channels (e.g., by increasing compression for mobile network delivery)

Content indexing

- In textual search engines, content need little (lexical) analysis before indexing
  - Index elements (words) are part of the content

- In MIR, content cannot be indexed directly
  - Computers **cannot understand** the meaning of a multimedia content
  - **Pixels** and **audio samples** just bring binary information

- Indexable features must be extracted from the input data
  - **Low level features**: concisely describe physical or perceptual properties of a media element (e.g., feature vectors)
  - **High level features**: domain concepts characterizing the content (e.g., extracted objects and their properties, content categorizations, etc)

Content querying

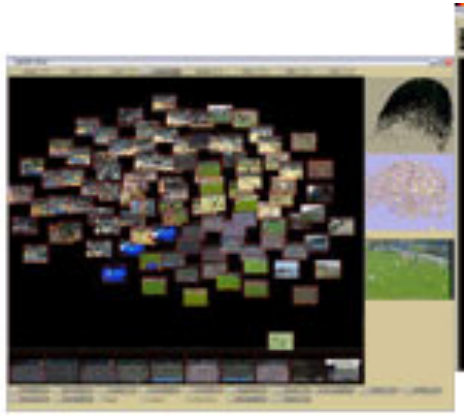
- In textual search applications, queries are keywords or expressions thereof
- In MIR, search can take place
  - By keyword
  - By (mono-media) example (e.g., query by image, query by humming, query by song similarity)
  - By (multi-media) example (e.g., query by video similarity)
- Query by example entails real time content processing
- MIR query processing naturally requires the interaction of multiple search engines (e.g., a text search engine for textual metadata and a content-based search engine for feature vectors)

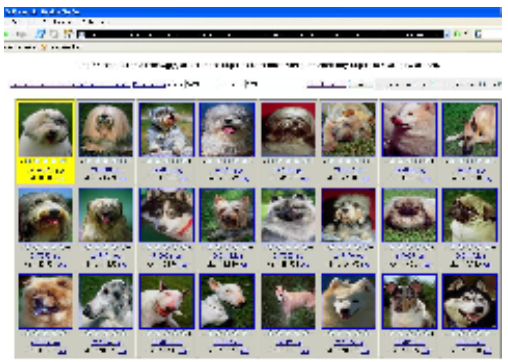
Content transmission

- In textual search applications, results are summarized in results list and then the original resource is accessed
- In MIR, result summarization is an open problem, which can be addressed in several ways
  - Textual summarization of relevant metadata
  - Key frames of a video
  - Preview fragment of a media element (e.g., first 30 seconds of a song)
  - By a distinct, yet correlated, media element (e.g., a movie trailer, a freebie of a music video, etc)
- Different delivery channels may also require different version of the media element (e.g., transcoding into 3GP for mobile terminals)
- VCR-like access to video may also require transcoding to a format that can be streamed with a protocol supporting VCR functionality

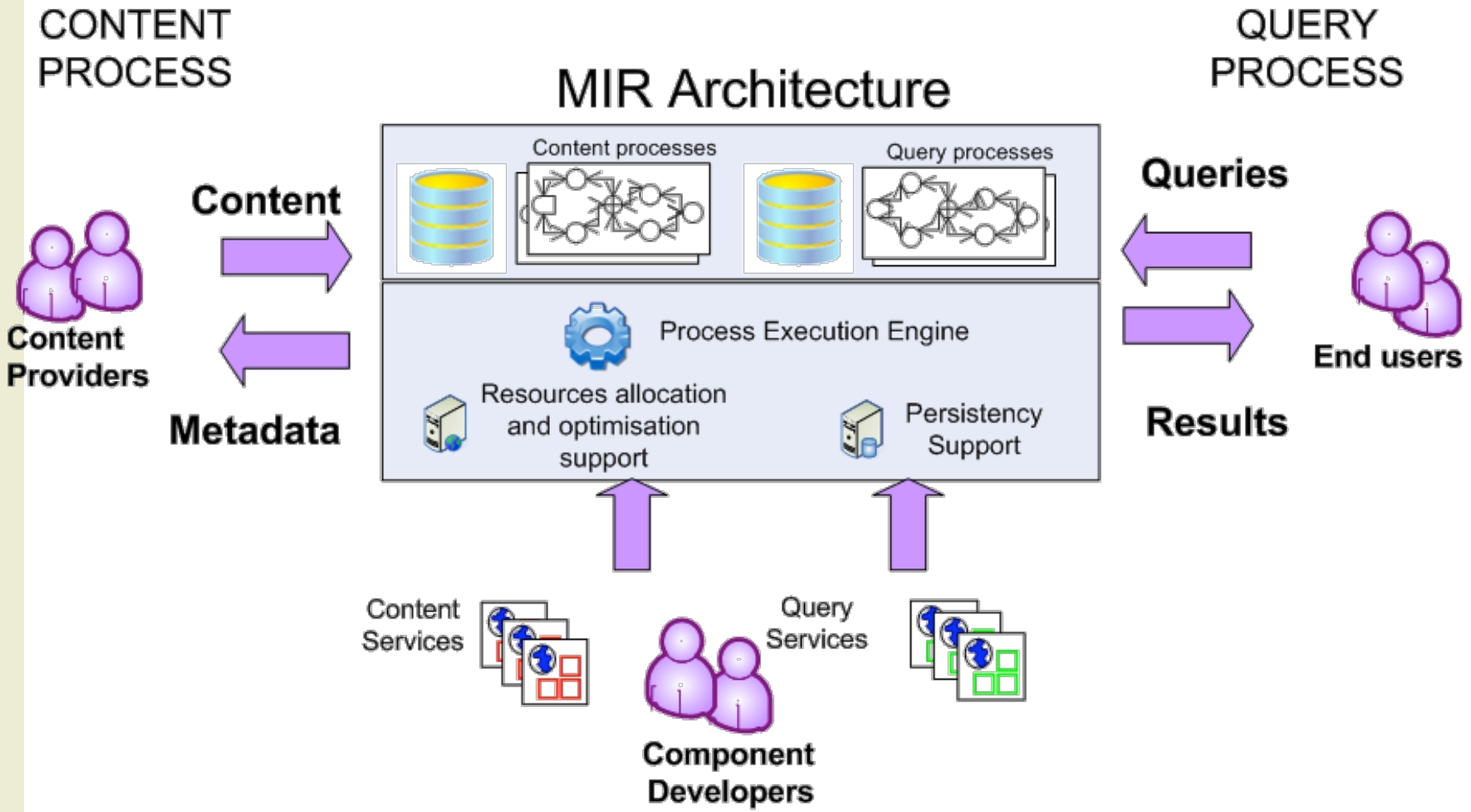
Content browsing

- In textual search engines, results are ranked linearly, browsed by navigating links, and read at a glance
- In MIR and similarity-based search applications, browsing results must consider multiple dimensions
  - **Relevance**: where the result appears in the sequence of retrieved media elements
  - **Space**: where the search has matched inside a spatially organized media element (e.g., an image)
  - **Time**: when a match occurs in a linear media element

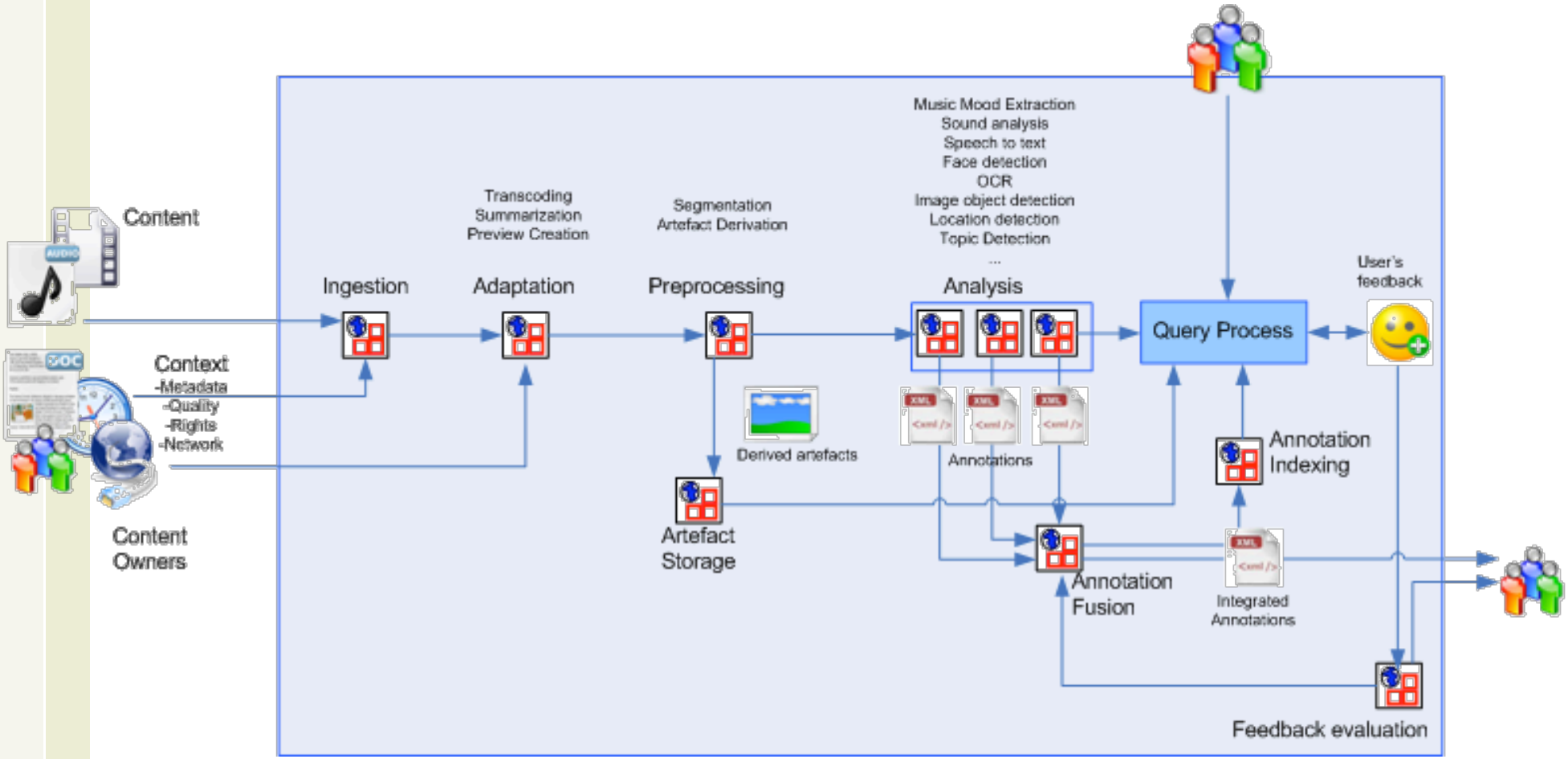




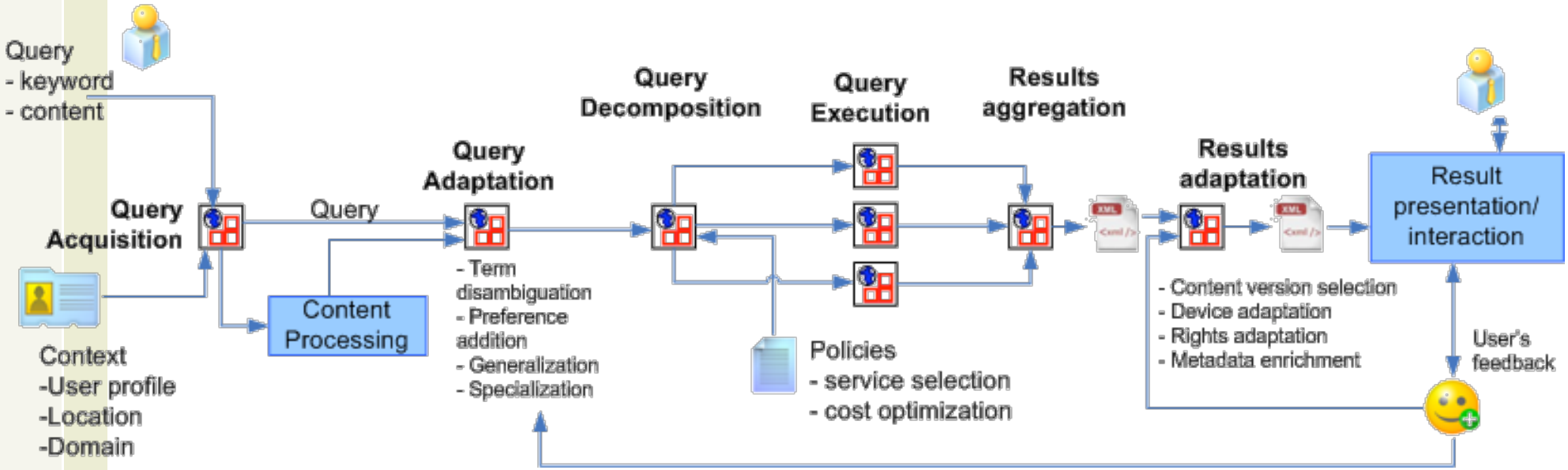
The MIR reference architecture



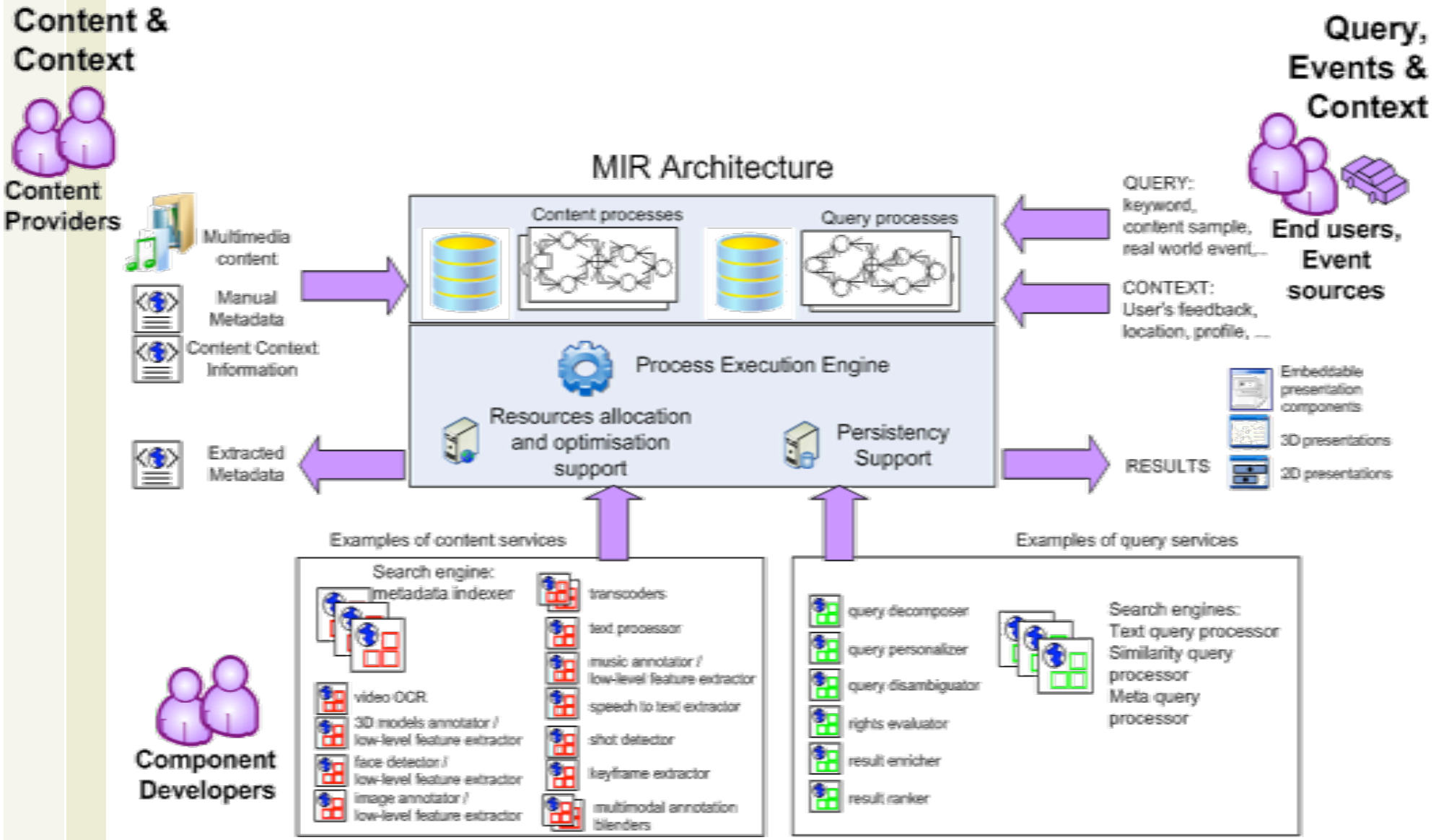
Overview of the content process



Overview of the query process

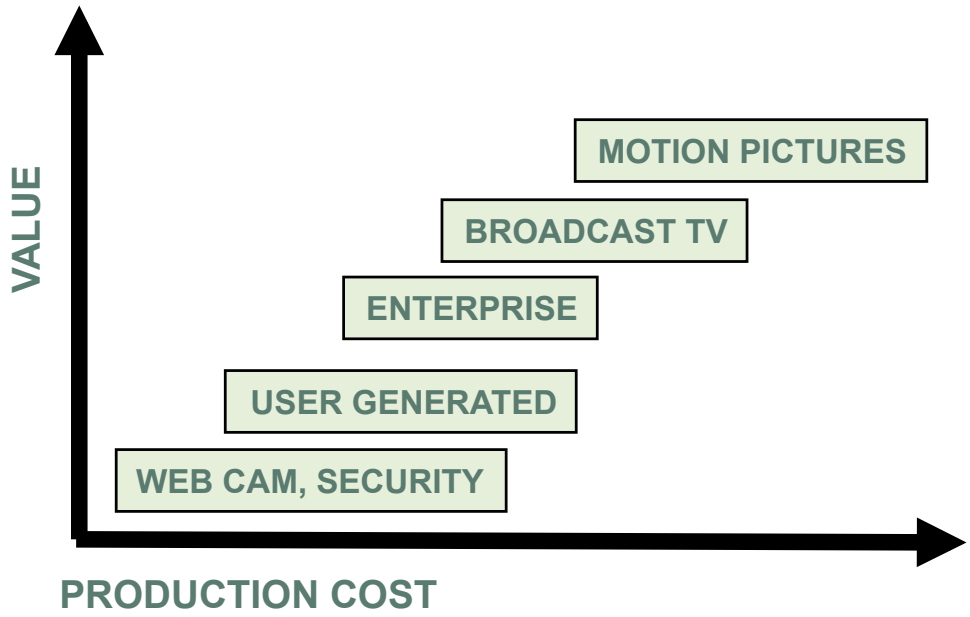


MIR Architecture: expanded view



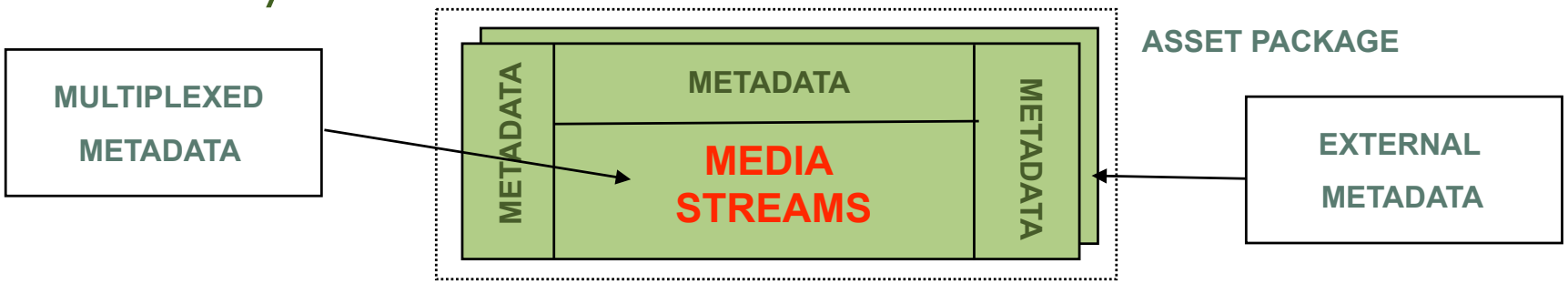
Acquisition: (video) content providers

- (Video) content comes from multiple sources, in a range of quality and value
  - Web cams, security apps
  - (Video/Audio) Telephony and teleconferencing
  - Industrial/Academic/Medical
  - User Generated Content
  - Public Access and Government Access
  - Rushes, Raw Footage
  - News
  - Advertising
  - TV Programming
  - Feature Films



Acquisition: (video) metadata sources & formats

- Content element may be accompanied by textual descriptions, which range in quantity and quality, from no description (e.g., web cam content) to multilingual high value data (closed captions and production metadata of motion pictures)
- Metadata may reside:
  - Embedded within content (e.g., close captions)
  - In surrounding Web pages or links (e.g., HTML content, link anchors, etc)
  - In domain-specific databases (e.g., IMDB for feature films)
  - In ontologies: http://www.daml.org/ontologies/keyword.html



Acquisition: (video) representative metadata standards

Standard	Body
MPEG-7, MPEG-21	ISO/IEC Int. Electrotechnical Comm., Motion Picture Expert Group
UPnP	Universal Plug and Play forum
MXF, MDD	SMPTE Society of Motion Picture and Television Engineers
AAF	AMWA Advanced Media Workflow Association
TV Anytime	ETSI European Telecommunication Standards Institute
Timed Text	W3C, 3GPP
RSS	Harward
Podcast	Apple
Media RSS	Yahoo

Acquisition: MPEG, MPEG-7, MPEG-21

- MPEG (Moving Picture Experts Group) is an ISO group devoted to set standards for audio and video compression and transmission.
- The MPEG standards consist of different Parts, that covers specific aspects of the specification.
- The standards also specifies Profiles and Levels.
  - Profiles define a set of implementation requirements for specific application
  - Levels define the range of appropriate values for the properties associated with them.
- MPEG has primarily standardized compression formats, but also media content descriptions

MPEG Main standards

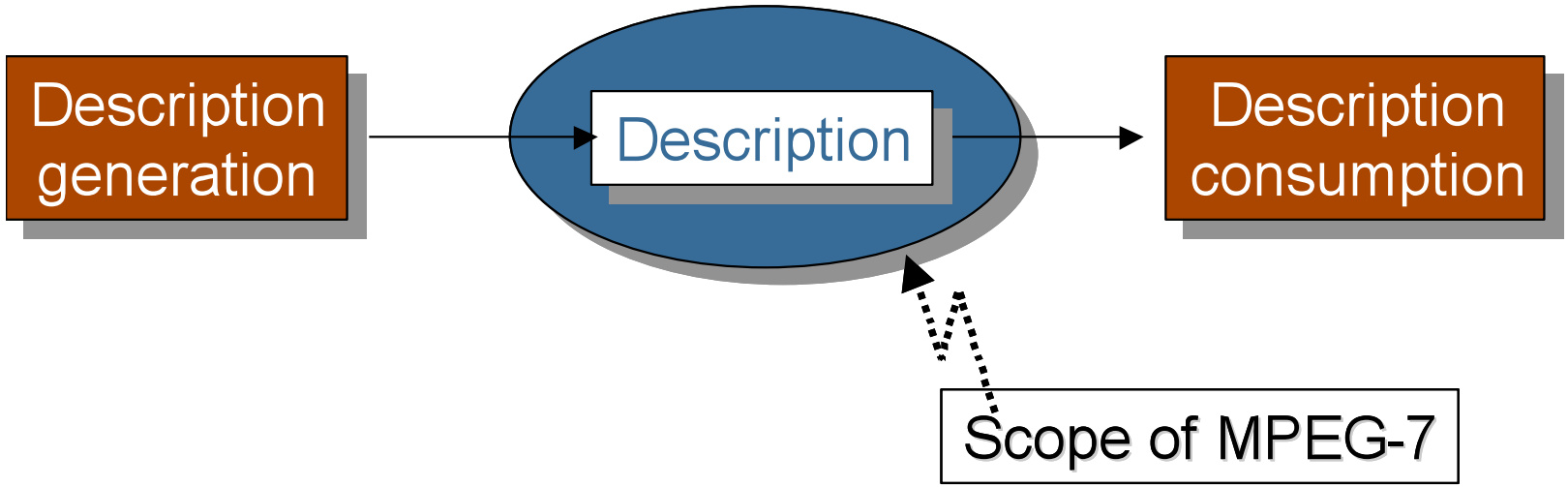
- **MPEG-1**: first compression standard for audio and video. Designed to allow moving pictures and sound to be encoded into the bitrate of a Compact Disc. Includes the popular Layer 3 (MP3) audio compression format.
- **MPEG-2**: Transport, video and audio standards for broadcast-quality television. Chosen for over-the-air digital television ATSC, DVB and ISDB, digital satellite TV services, digital cable television, SVCD, and DVD.
- **MPEG-3**: there is no MPEG-3 standard. MPEG-3 is not to be confused with MP3, which is MPEG-1 Audio Layer 3.
- **MPEG-4:** higher compression factors than MPEG-2. In addition, MPEG-4 moves closer to computer graphics applications. In more complex profiles, the MPEG-4 decoder becomes a rendering processor and the compressed bitstream describes three-dimensional shapes and surface texture. MPEG-4 also provides Intellectual Property Management and Protection (IPMP) to support digital rights management.

MPEG Standards for content description

- MPEG-7
  - 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
- MPEG-21
  - A normative open framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain

MPEG-7 Motivation and Scope

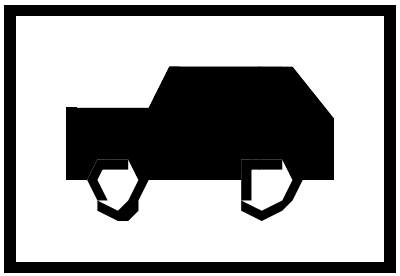
- Create standardized multimedia description framework
- Enable content-based access to and processing of multimedia information on the basis of descriptions of multimedia content and structure (metadata)
- Support range of abstraction levels for metadata from low-level signal characteristics to high-level semantic information



An Example MPEG-7 Media Description

- The following example gives an MPEG-7 description of a car that is depicted in an image:

```
<Mpeg7>
  <Description xsi:type="SemanticDescriptionType">
    <Semantics>
      <Label>
        <Name> Car </Name>
      </Label>
      <Definition>
        <FreeTextAnnotation>
          Four wheel motorized vehicle
        </FreeTextAnnotation>
      </Definition>
      <MediaOccurrence>
        <MediaLocator>
          <MediaUri> image.jpg </MediaUri>
        </MediaLocator>
      </MediaOccurrence>
    </Semantics>
  </Description>
</Mpeg7>
```



Acquisition: MPEG 21

- he MPEG-21 (aka ISO/IEC 21000) standard, from the Moving Picture Experts Group aims at defining an open framework for multimedia applications.
- MPEG-21 is based on two essential concepts: the definition of a fundamental unit of distribution and transaction, which is the Digital Item, and the concept of users interacting with them.
- MPEG-21 defines a "Rights Expression Language" standard as means of sharing digital rights/ permissions/restrictions for digital content from content creator to content consumer.

Acquisition: RSS and Media RSS

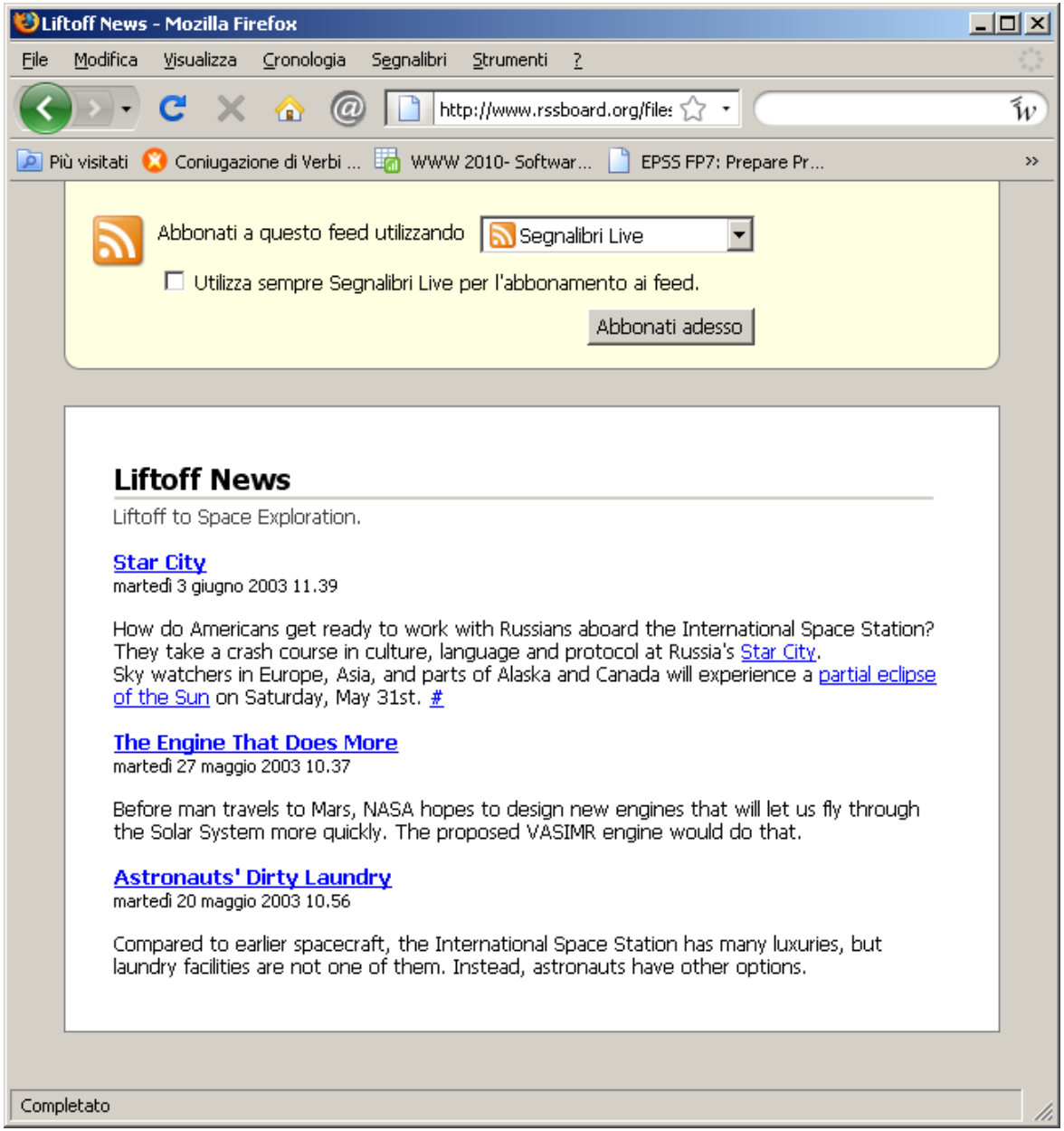
- RSS (Really Simple Syndication) describes a family of web feed formats used to publish frequently updated web resources (e.g., news)
- An RSS feed includes full or summarized text, plus metadata such as publishing dates and authorship
- RSS formats are specified using XML
- RSS 2.0 now "frozen"
- Media RSS proposed by Yahoo as an RSS module that supplements the <enclosure> element capabilities of RSS 2.0 to allow for more robust media syndication.

Acquisition: Example of RSS 2.0

```
<?xml version="1.0"?>
<rss version="2.0">
  <channel>
    <title>Lift Off News</title>
    <link>http://liftoff.msfc.nasa.gov/</link>
    <description>Liftoff to Space Exploration.</description>
    <language>en-us</language>
    <pubDate>Tue, 10 Jun 2003 04:00:00 GMT</pubDate>
    <lastBuildDate>Tue, 10 Jun 2003 09:41:01 GMT</lastBuildDate>
    <docs>http://blogs.law.harvard.edu/tech/rss</docs>
    <generator>Weblog Editor 2.0</generator>
    <managingEditor>editor@example.com</managingEditor>
    <webMaster>webmaster@example.com</webMaster>
    <ttl>5</ttl>

    <item>
      <title>Star City</title>
      <link>http://liftoff.msfc.nasa.gov/news/2003/news-starcity.asp</link>
      <description>How do Americans get ready to work with Russians aboard the
        International Space Station? They take a crash course in culture, language
        and protocol at Russia's Star City.</description>
      <pubDate>Tue, 03 Jun 2003 09:39:21 GMT</pubDate>
      <guid>http://liftoff.msfc.nasa.gov/2003/06/03.html#item573</guid>
    </item>
```

Acquisition: Browser rendition of RSS



Acquisition: an example of Media RSS

A music video with a link to a player window, and additional metadata about the video, including expiration date.

```
<rss version="2.0" xmlns:media="http://search.yahoo.com/mrss/"
xmlns:dcterms="http://purl.org/dc/terms/">
<channel>
<title>Music Videos 101</title>
<link>http://www.foo.com</link>
<description>Discussions of great videos</description>
        <item>
                <title>The latest video from an artist</title>
                <link>http://www.foo.com/item1.htm</link>
                <media:content url="http://www.foo.com/movie.mov" fileSize="12216320"
                type="video/quicktime" expression="full">
                <media:player url="http://www.foo.com/player?id=1111"
                height="200" width="400"/>
                <media:hash algo="md5">dfdec888b72151965a34b4b59031290a</media:hash>
                <media:credit role="producer">producer's name</media:credit>
                <media:credit role="artist">artist's name</media:credit>
                <media:category scheme="http://blah.com/scheme">music/artist
                name/album/song</media:category>
                <media:text type="plain">
                Oh, say, can you see, by the dawn's early light
                </media:text>
                <media:rating>nonadult</media:rating>
                <dcterms:valid>
                        start=2002-10-13T09:00+01:00;
                        end=2002-10-17T17:00+01:00;
                        scheme=W3C-DTF
                </dcterms:valid>
                </media:content>
        </item>
</channel>
</rss>
```

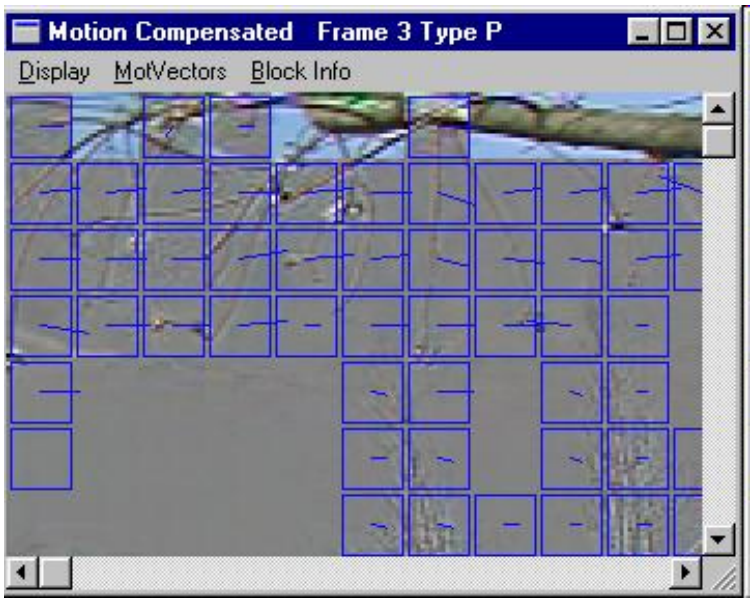
(Re)formatting: Digital video formats

- A digital video is a sequence of frames, recorded in interlaced or progressive scan format
- The Frame Aspect Ratio (FAR) defines the shape of each image (width divided by heigh), with 4:3 and 16:9 being the currently adopted values

- Pixel aspect ratio (PAR) describes how the width of pixels in a digital image compares to their height (rectangular pixels format exist for analog TV compatibility).
- Frame rate: number of frames per second (24 and 25 are common, but also lower and higher values are used)

(Re)formatting: compression

- Web media must be compressed, with lossy (but perceptually acceptable) transformations
- In video, compression works in two ways
  - Intra-Frame: an image is divided in blocks, whose content is "averaged"
  - Inter-frame: a frame is represented differentially with respect to the preceding one, by encoding only block that "have moved" and their motion vector
  - Example (MPEG compression)



Example.. continued

- Compression affects video search applications :
  - Content analysis algorithm must be either able to operate on the transformed domain, or must decode the video
  - Analysis in the compressed domain may be hampered by artificial image block (artifacts)
  - Different compressed formats need adequate playout capability (e.g., different plugins in the browser)
  - Some compressed formats may not be used for streaming or random access to content

Re(formatting): popular compression standards

Standard	Typical bitrates	Applications
M-JPEG, JPEG2000	Up to 60 Mbit/ sec	Consumer electronics, video editing systems
DVCAM	25M	Consumer
MPEG-1	1.5M	CD-ROM Multimedia
MPEG-2	4-20M	Broadcast TV, DVD
MPEG-4 H. 264	300K-12M	Mobile video, Podcast, IPTV
H.261 H.263	64k-1M	Video teleconferencing, telephony
Each standard has profiles, that balance latency, complexity, error resilience and bandwidth, specifically for a target application (e.g., file-based vs transport-based fruition)		

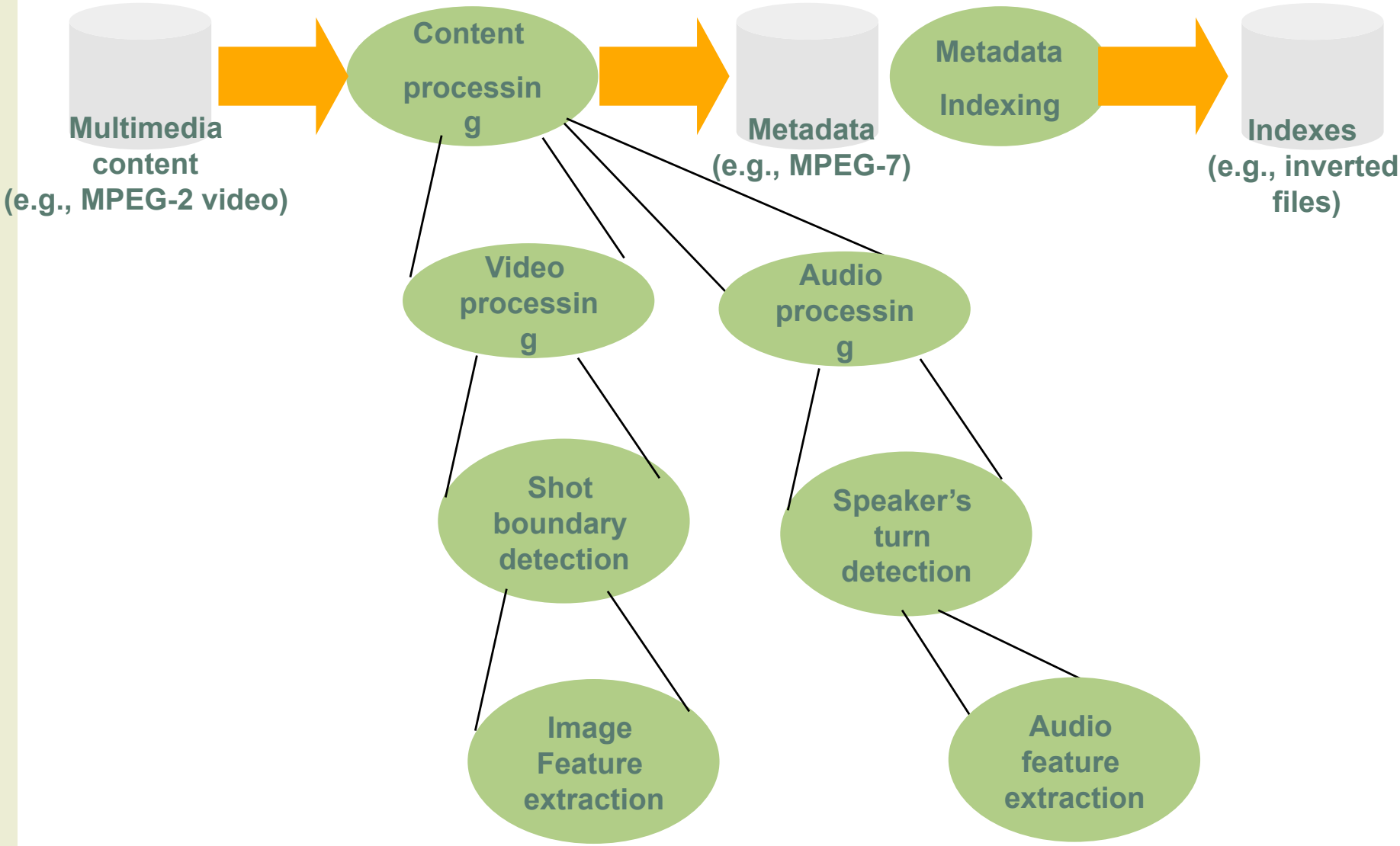
Re(formatting): transcoding

- Transcoding is the direct digital-to-digital conversion of one encoding to another
- Transcoding normally decodes the original data to an intermediate format (i.e. PCM for audio or YUV for video), in a way that still contains the content of the original, and then encodes the resulting file into the target format.
- Transrating is a process similar to transcoding in which files are coded to a lower bitrate without changing video formats
- Transsizing is changing the picture size of video, used if the output resolution differs from the media's resolution.
- Transrating compares to Bitrate Peeling, wherein a stream can be encoded at one bitrate but can be served at that or any lower bitrate (in theory)

Indexing: automatic annotation of multimedia contents

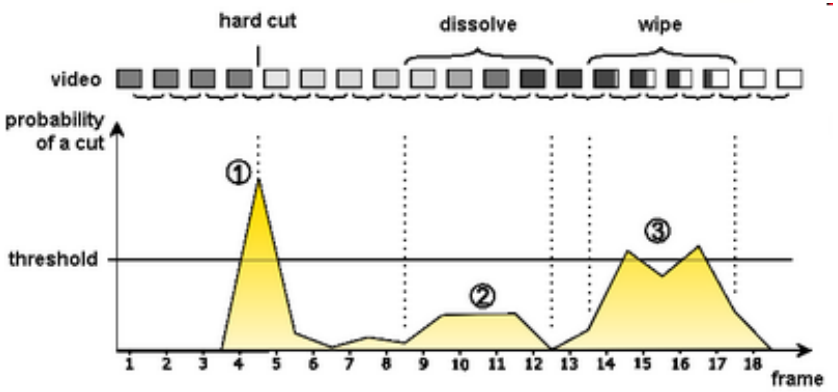
- The manual annotation of multimedia content is
  - **Expensive**
    - E.g.: manual segmentation of a video takes 10hs for each hour of a video
  - **Incomplete or inaccurate**
    - It is difficult for a user to grasp all the meanings associated with a multimedia object
  - **Difficult**
    - Some contents are difficult to describe with words
      - E.g.: a melody with not singing and irregular structure
- Automatica annotations provide **good quality** at a **"low"** cost

Indexing: the core pipeline



Indexing: media segmentation

- Media segmentation is to MIR what terms identification is to textual IR
- Audio or video are split in units for analysis and presentation purposes
  - **ANALYSIS**: a homogeneous region can be subjected to a specific analysis pattern (e.g., a speaker's turn is processed to get speaker identification, a music or silent scene is not); music mood extraction can be performed on a subsequence of the whole track
  - **PRESENTATION**: a (possibly long) video sequence can be represented by a single thumbnail or scene;
- **SHOT BOUNDARY DETERMINATION**: analyses consecutive frames for capturing typical camera motions (cut, fade in, fade out, dissolve, wipe, etc..)



Indexing: Media segmentation in MPEG-7

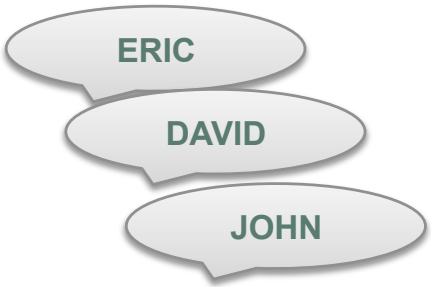
```
<Mpeg7 xsi:schemaLocation="urn:mpeg:mpeg7:schema:2001 ./davp-2005.xsd"
xmlns="urn:mpeg:mpeg7:schema:2001" xmlns:mpeg7="urn:mpeg:mpeg7:schema:2001"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
 <Description xsi:type="ContentEntityType">
  <MultimediaContent xsi:type="AudioVisualType">
   <AudioVisual>
    <StructuralUnit href="urn:x-mpeg-7-pharos:cs:AudioVisualSegmentationCS:root" />
    <MediaSourceDecomposition criteria="fast sample annotator segment">
     <AudioSegment>
      <StructuralUnit href="urn:x-mpeg-7-pharos:cs:SegmentationCS:audio" />
      <MediaTime>
       <MediaTimePoint>T00:00:00:0F100</MediaTimePoint>
       <MediaDuration>PT1M12S</MediaDuration>
      </MediaTime>
      <TemporalDecomposition criteria="audio:segmentation:fast:samplesegments">
       <AudioSegment>
        <StructuralUnit href="urn:x-mpeg-7-pharos:cs:AudioSegm:fast:sample " />
        <TextAnnotation type="segment_class" confidence="0.8765432">
         <FreeTextAnnotation>speech</FreeTextAnnotation>
        </TextAnnotation>
        <MediaTime>
         <MediaTimePoint>T00:00:00:0F100</MediaTimePoint>
         <MediaDuration>PT50S</MediaDuration>
        </MediaTime>
       </AudioSegment>
       <AudioSegment>
        <StructuralUnit href="urn:x-mpeg-7-pharos:cs:AudioSegmentationCS:fast:sample" />
        <TextAnnotation type="segment_class" confidence="0.7654321">
         <FreeTextAnnotation>music</FreeTextAnnotation>
        </TextAnnotation>
        <MediaTime>
         <MediaTimePoint>T00:00:50:0F100</MediaTimePoint>
         <MediaDuration>PT22S</MediaDuration>
        </MediaTime>
       </AudioSegment>
      </TemporalDecomposition>
     </AudioSegment>
    </MediaSourceDecomposition>
   </AudioVisual>
  </MultimediaContent>
 </Description>
</Mpeg7>
```

Indexing: feature extraction

- In pattern recognition and in image processing, Feature extraction is a special form of dimensionality reduction.
- It works by transforming input data into smaller output data, so that some "features" are retained in the output that serve for a given task (e.g., OCR, face detection, etc)
- Employed techniques are either general purpose (e.g., Principal Component Analysis) or task-specific (e.g., edge detection filters).

Indexing: audio

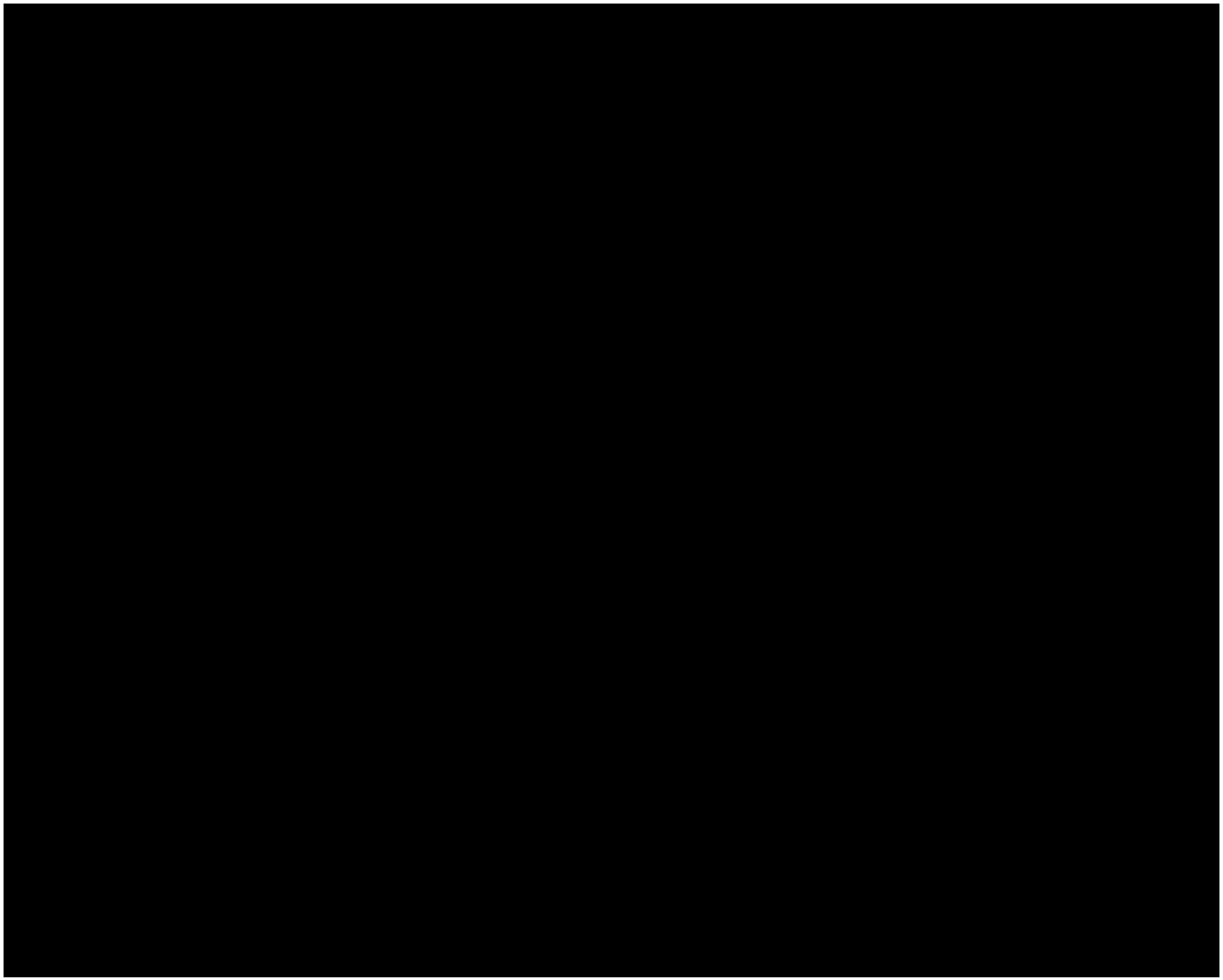
- Speaker Identification



- Word Spotting

- Speech to text

Indexing: example of speech to text



CREDITS: Thorsten Hermes@SSMT2006

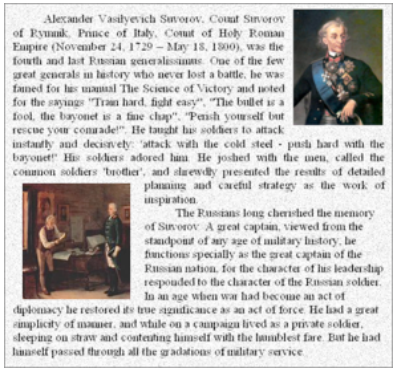
Indexing: audio

- Audio event identification

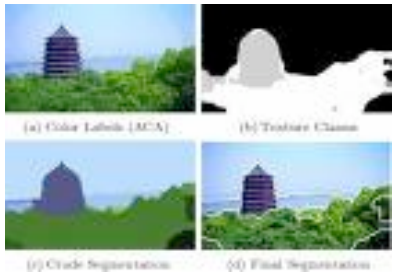
- Classification of music genres, mood, instruments, rhythmic features, structure

Indexing: image

- Text/Image segmentation


Alexander Vasilyevich Suvorov, Count Suvorov of Rymnik, Prince of Italy, Count of Holy Roman Empire (November 24, 1729 – May 18, 1800), was the fourth and last Russian generalissimus. One of the few great generals in history who never lost a battle, he was famed for his manual The Science of Victory and noted for the sayings "Train hard, fight easy", "The bullet is a fool, the bayonet is a fine chap", "Perish yourself but rescue your comrade!". He taught his soldiers to attack instantly and decisively: 'attack with the cold steel - push hard with the bayonet!' His soldiers adored him. He joshed with the men, called the common soldiers 'brother', and shrewdly presented the results of detailed planning and careful strategy as the work of inspiration.

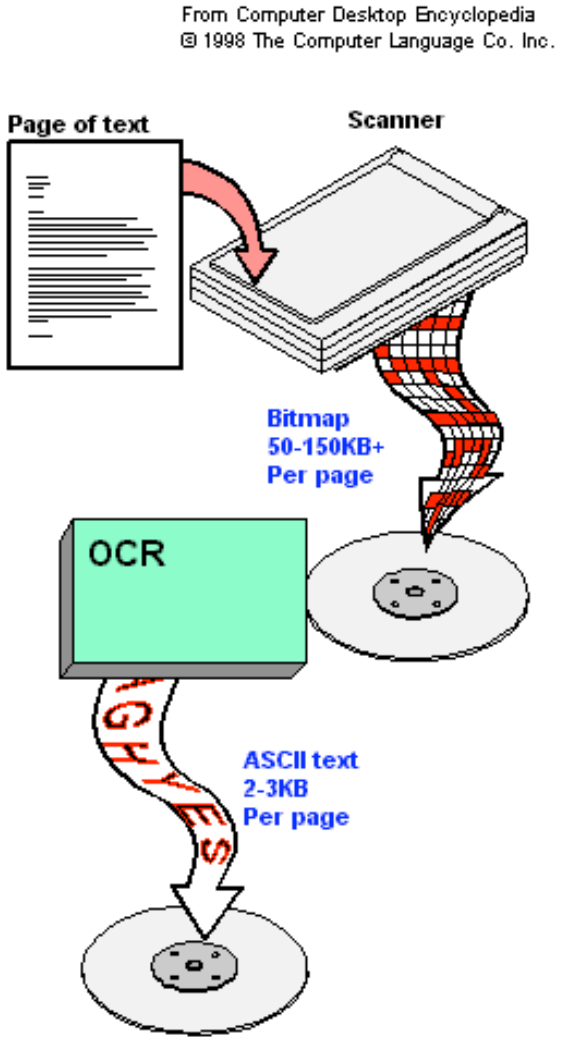
The Russians long cherished the memory of Suvorov. A great captain, viewed from the standpoint of any age of military history, he functions specially as the great captain of the Russian nation, for the character of his leadership responded to the character of the Russian soldier. In an age when war had become an act of diplomacy he restored its true significance as an act of force. He had a great simplicity of manner, and while on a campaign lived as a private soldier, sleeping on straw and contenting himself with the humblest fare. But he had himself passed through all the gradations of military service.



- Low-level feature extraction

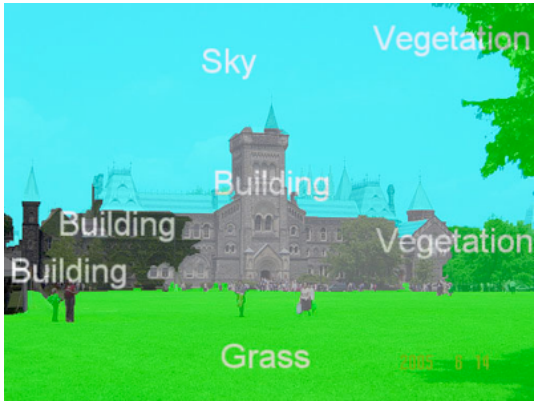
Indexing: Optical character recognition (OCR)

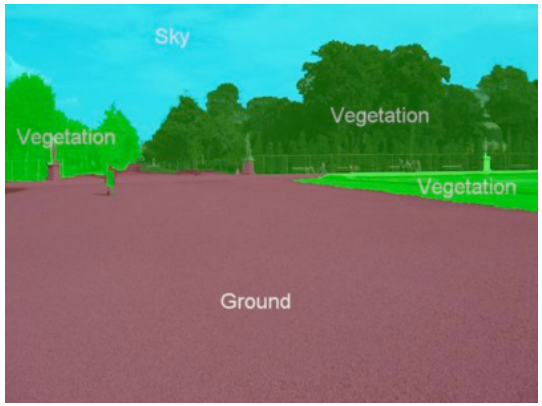
- OCR is a technique for translating images of typed or handwritten text into symbols
- Solved problem for typewritten text (99% accuracy)
- Commercial solutions for handwritten text (e.g, MS Tablet PC)
- Video OCR has specific problems, due to low resolution, small text size, and interference with background
- Detection is normally done on the most representative image of an entire shots, rather than frame by frame
- Approach: filter for enhancing resolution + pattern matching for character identification

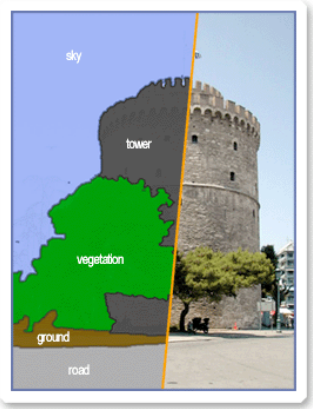


Indexing: concept detection

- Image analysis extract low level features from raw data (e.g., color histograms, color correlograms, color moments, co-occurrence texture matrices, edge direction histograms, etc..)
- Features can be used to build discrete classifiers, which may associate semantic concepts to images or regions thereof
- Concepts can be detected also from text (e.g., from manual or automatic metadata) using NLP techniques (FAST text search engine recognizes entities like geographical locations, professions, names of persons, domain-specific technical concepts, etc)





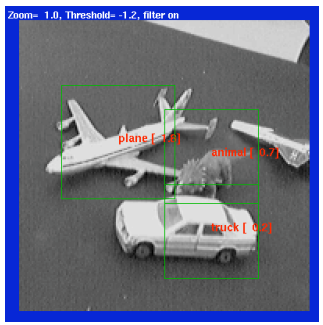


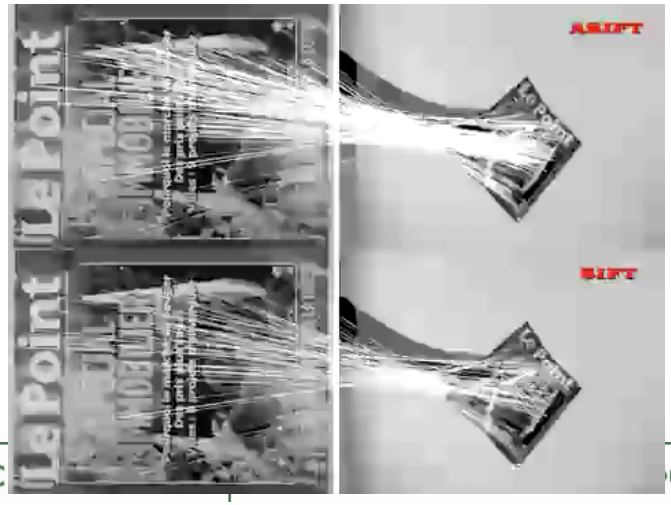
Indexing: image

- Face recognition and identification

CREDITS: Thorsten Hermes@SSMT2006

- Object Recognition



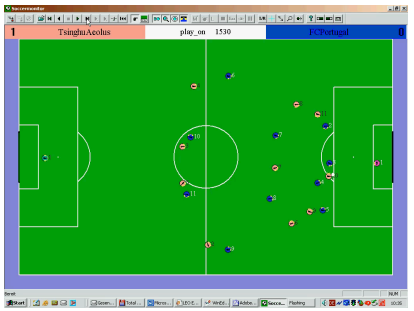


Indexing: video

- Motion detection and identification

CREDITS: Thorsten Hermes@SSMT2006

- Object Tracking

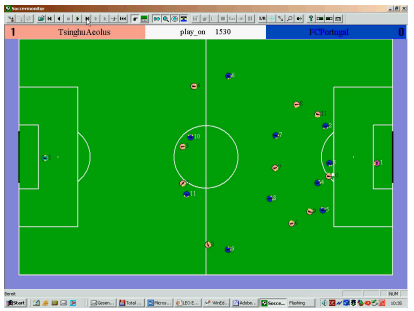


Indexing: video

- Motion detection and identification

CREDITS: Thorsten Hermes@SSMT2006

- Object Tracking



Indexing: video

- Shot/Scene detection

- Video OCR

Indexing: multimodal annotation fusion

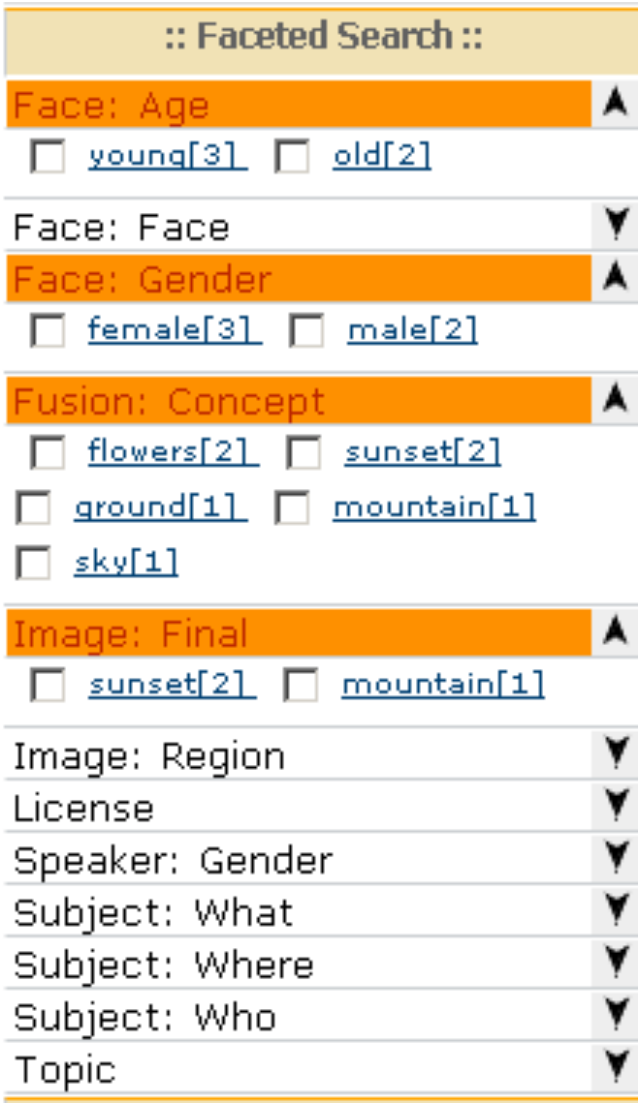
- Media segmentation and concept extraction are probabilistic processes
- The result is characterized by a confidence value
- Significance can be enhanced by comparing the output of distinct techniques applied to the same or similar problems
- Examples:
  - Media segmentation: shot detection + speaker's turn identification
  - Person recognition: voice identification + face detection
  - Concept detection: image based classification (e.g., "outdoor" & "water" + object extraction: "bird", "boat")

Querying: modalities

- In MIR applications, search keyword match the manual or automatic metadata
- A complementary approach is to provide an example of the desired content and look for similar media elements
- Similarity is a medium-dependent, domain-dependent, and subjective criterion
- Can be computed on low lever features (e.g., image color histograms, music bpm) or on high level concepts/categorization (e.g., melancholic images, party music)
- Can be multimodal (e.g., video similarity)
- Querying may also consider context information (e.g., the user's geographical position or the access device)

Faceted query

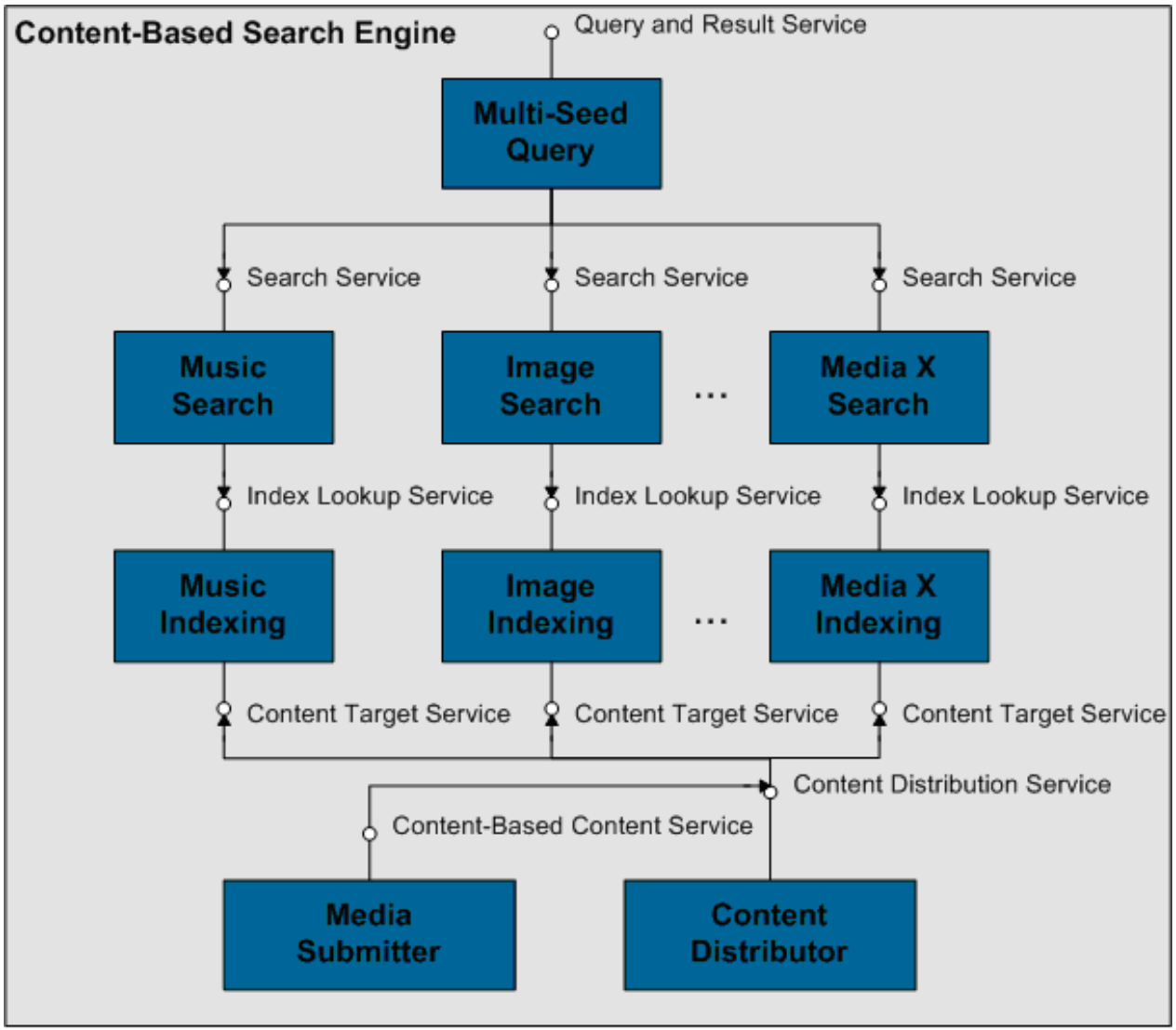
- When a media collection is large and its content unknown to the user, exposing part of the metadata can help
- This can be done by showing a compact representation of the categories of content (facets)
- A query can be restricted by selecting only the relevant facets



Querying: by keyword

- The keyword may match the manual metadata and/or the automatic metadata
- The match can be multimodal: in the audio, in a visual concept

Querying: by content/similarity - architecture



Search flickr Tags [One Tag Only Please] [Search] or show random images

Reference image for search: Show ONLY images from the reference photographer

1. Search mode: Theme

2. Find similar by Color / Texture

1. Find similar by Theme

OR

2. Find similar by Color / Texture

1. Find similar by Theme

OR

2. Find similar by Color / Texture

1. Find similar by Theme

OR

2. Find similar by Color / Texture

1. Find similar by Theme

OR

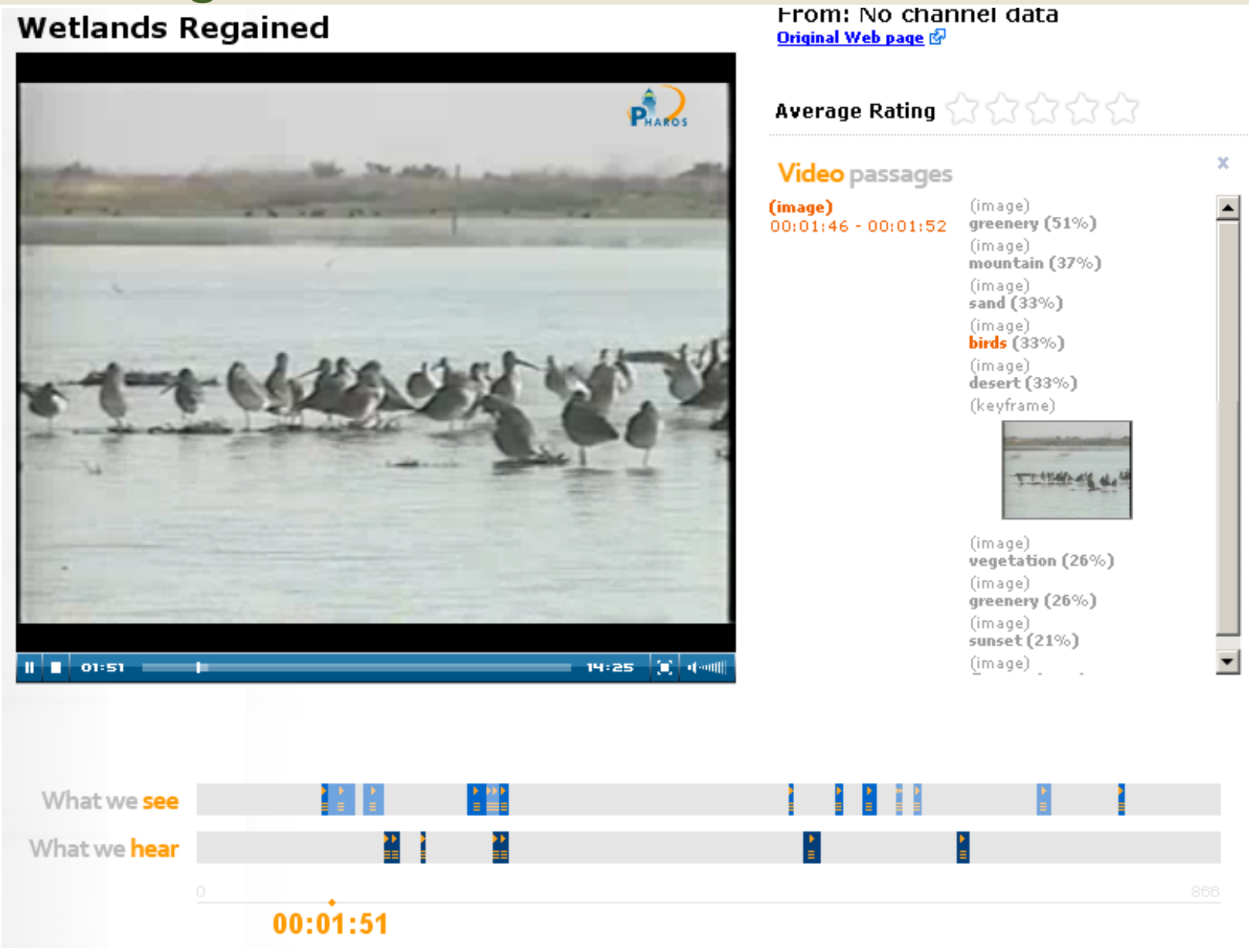
2. Find similar by Color / Texture

1. Find similar by Theme

OR

2. Find similar by Color / Texture

Browsing: timeline-based video access



State-of-the art of MSE

- Image search
  - www.tiltomo.com
  - www.tineye.com
  - www.pixsta.com
  - www.picsearch.com
- Music Search
  - www.midomi.com
  - www.audiobaba.com
  - http://www.bmat.com/
- Video Search
  - www.blinx.com
  - www.clipta.com
  - www.yovisto.com
- Entrerprise MIR search
  - www.autonomy.com
  - www.pictron.com
  - www.exalead.com
  - www.fastsearch.com

References

- MPEG-7:
  - MPEG-7 Overview http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm
  - Prof. Ray Larson & Prof. Marc Davis, UC Berkeley SIMS http://www.sims.berkeley.edu/academics/courses/is202/f03/
- RSS: http://www.rssboard.org/rss-specification
- MEDIA RSS: http://search.yahoo.com/mrss
- MPEG:http://en.wikipedia.org/wiki/MPEG
- Shot detection: http://en.wikipedia.org/wiki/Shot_boundary_detection

References

- MediaMill: http://www.science.uva.nl/research/mediamill
- Similarity search
  - www.midimi.com
  - www.tiltomo.com
  - http://tineye.com/
- PHAROS: http://www.pharos-audiovisual-search.eu/