Semantically Enhanced Audio-Visual Repository

http://sorebellious.files.wordpress.com/2009/05/russeantrekket.jpg

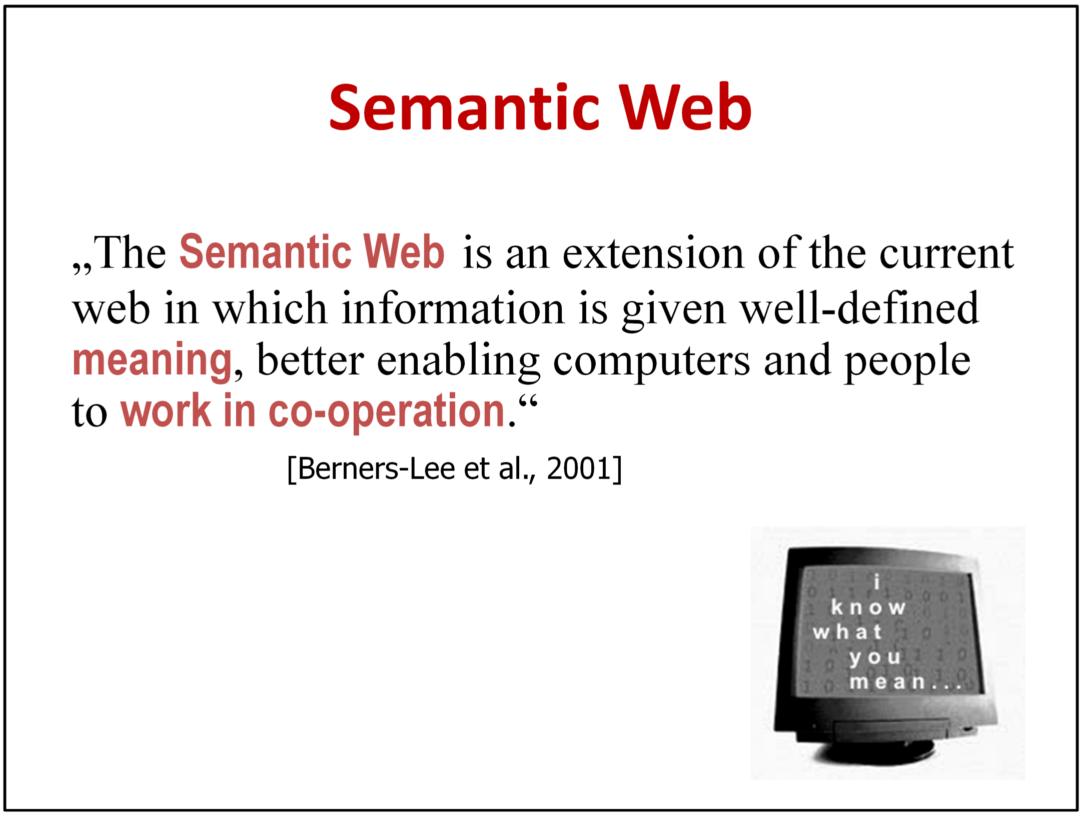
Svetla Boytcheva

State University of Library Studies and Information Technologies, Bulgaria

svetla.boytcheva@gmail.com

Hello,
My name is Svetla Boytcheva,
I am from the *State University of Library Studies and Information Technologies, Bulgaria*

I am going to present you work in progress for a research project aiming development of tools supporting semantic information for audio-visual repository .
At this initial stage we just finished with the system prototype implementation and currently we are in stage of testing the system.



Semantic Web

„The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning**, better enabling computers and people to **work in co-operation**.“

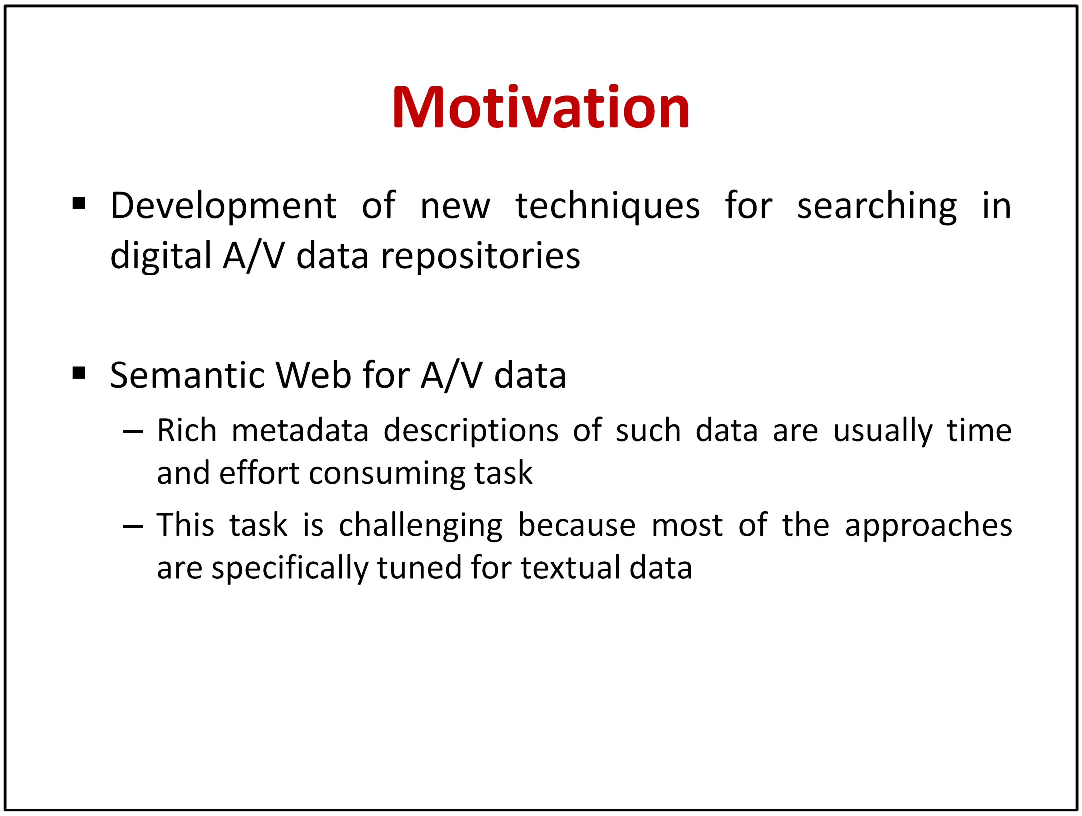
[Berners-Lee et al., 2001]

i
know
what
you
mean...

„The **Semantic Web** is an extension of the current web in which information is given well-defined **meaning**, better enabling computers and people to **work in co-operation**.“
[Berners-Lee et al., 2001]

CAN'T YOU DO ANYTHING RIGHT?
GLASBERGEN
Copyright © Randy Glasbergen. www.glasbergen.com

Usually we are relying on “intelligence” of computers searching information, but in order to solve this problem, computer scientists needs to do a lot of efforts in order to help the computers to “understand” human interpreted information and as a first step at least to add some additional information to help them searching and making inferences from data.



Motivation

- Development of new techniques for searching in digital A/V data repositories

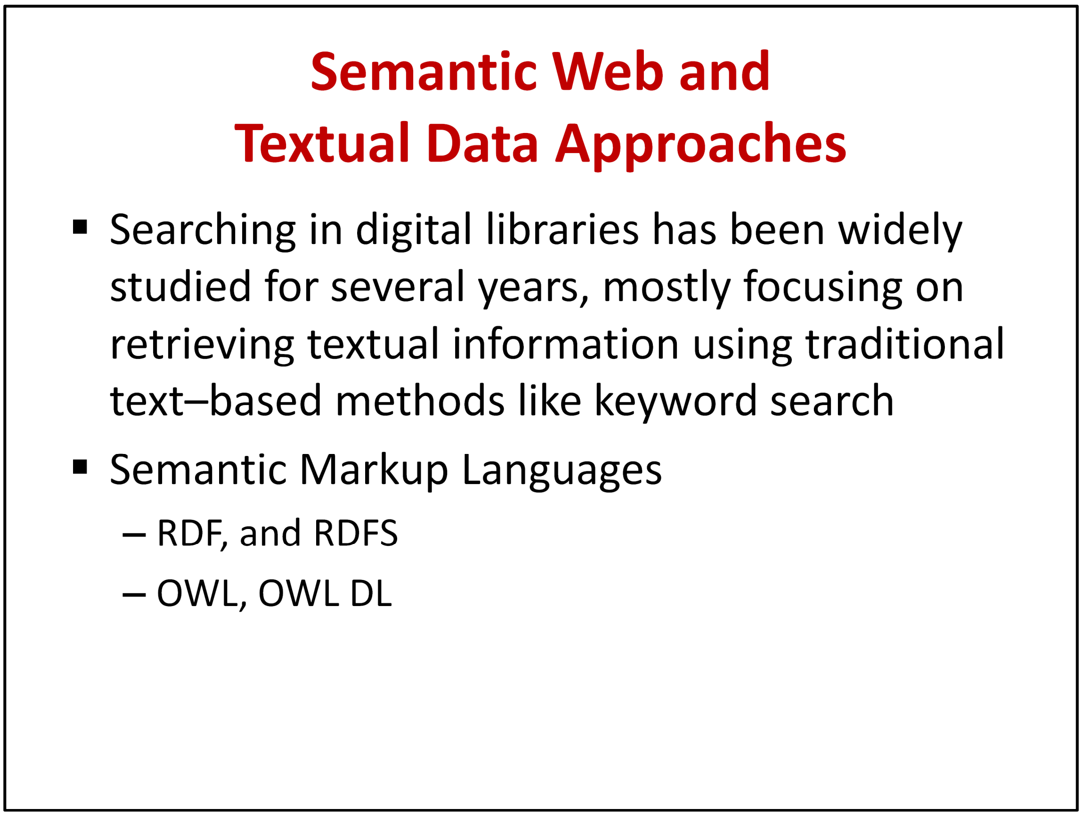
- Semantic Web for A/V data
  - Rich metadata descriptions of such data are usually time and effort consuming task
  - This task is challenging because most of the approaches are specifically tuned for textual data

Growing amount of stored digital audio and video data requires development of new techniques for semantically enhanced maintenance of such data

Semantic web, information extraction and information retrieval from texts are widely used nowadays, while technologies for capturing semantic information from audio and video data are making their first steps

Rich metadata descriptions of such data are usually time and effort consuming task

This task is challenging because most of the approaches are specifically tuned for textual data



Semantic Web and Textual Data Approaches

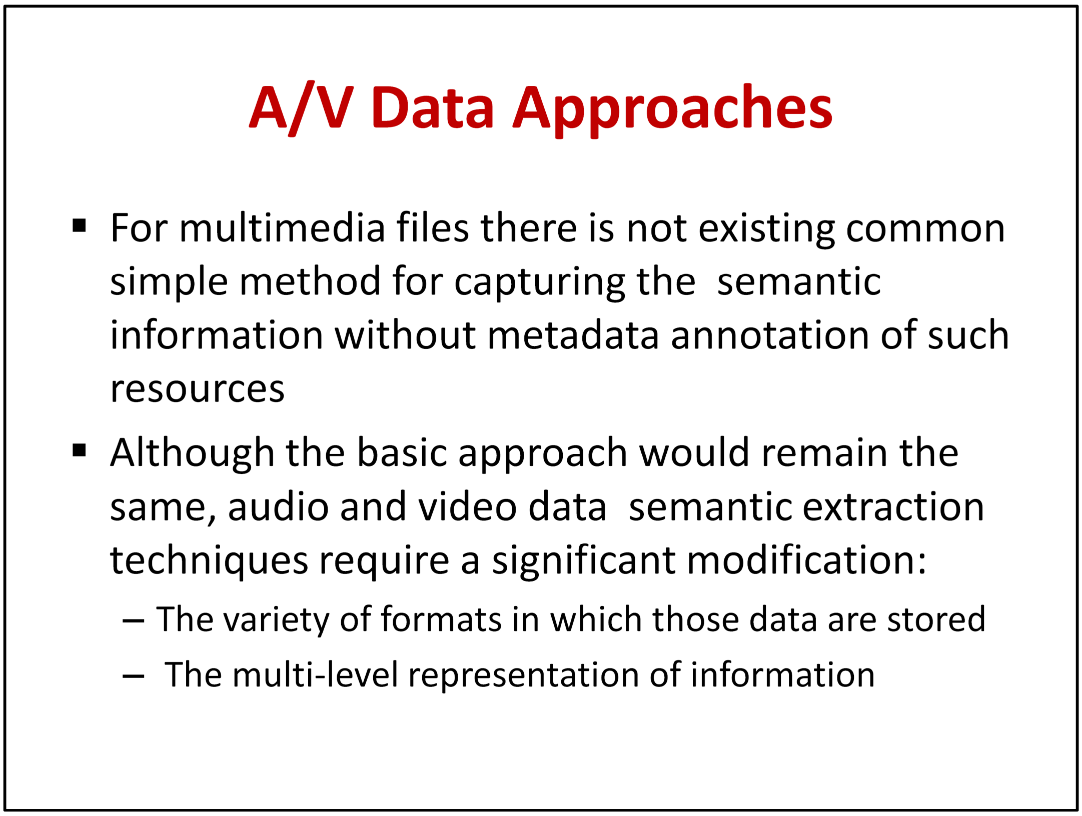
- Searching in digital libraries has been widely studied for several years, mostly focusing on retrieving textual information using traditional text–based methods like keyword search
- Semantic Markup Languages
  - RDF, and RDFS
  - OWL, OWL DL

Searching in digital libraries has been widely studied for several years, mostly focusing on retrieving textual information using traditional text–based methods like keyword search

Semantic Markup Languages

RDF, and RDFS

OWL, OWL DL



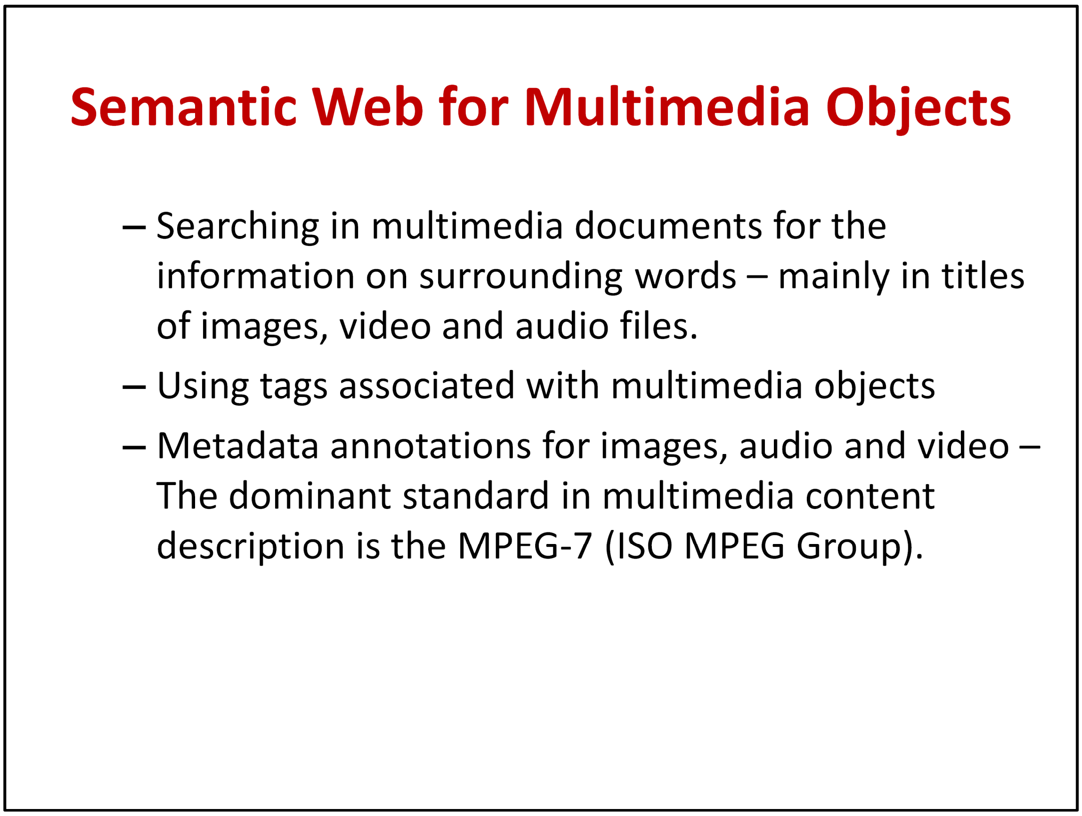
A/V Data Approaches

- For multimedia files there is not existing common simple method for capturing the semantic information without metadata annotation of such resources
- Although the basic approach would remain the same, audio and video data semantic extraction techniques require a significant modification:
  - The variety of formats in which those data are stored
  - The multi-level representation of information

For multimedia files there is not existing common simple method for capturing the semantic information without metadata annotation of such resources
Although the basic approach would remain the same, audio and video data semantic extraction techniques require a significant modification:

The variety of formats in which those data are stored

The multi-level representation of information



Semantic Web for Multimedia Objects

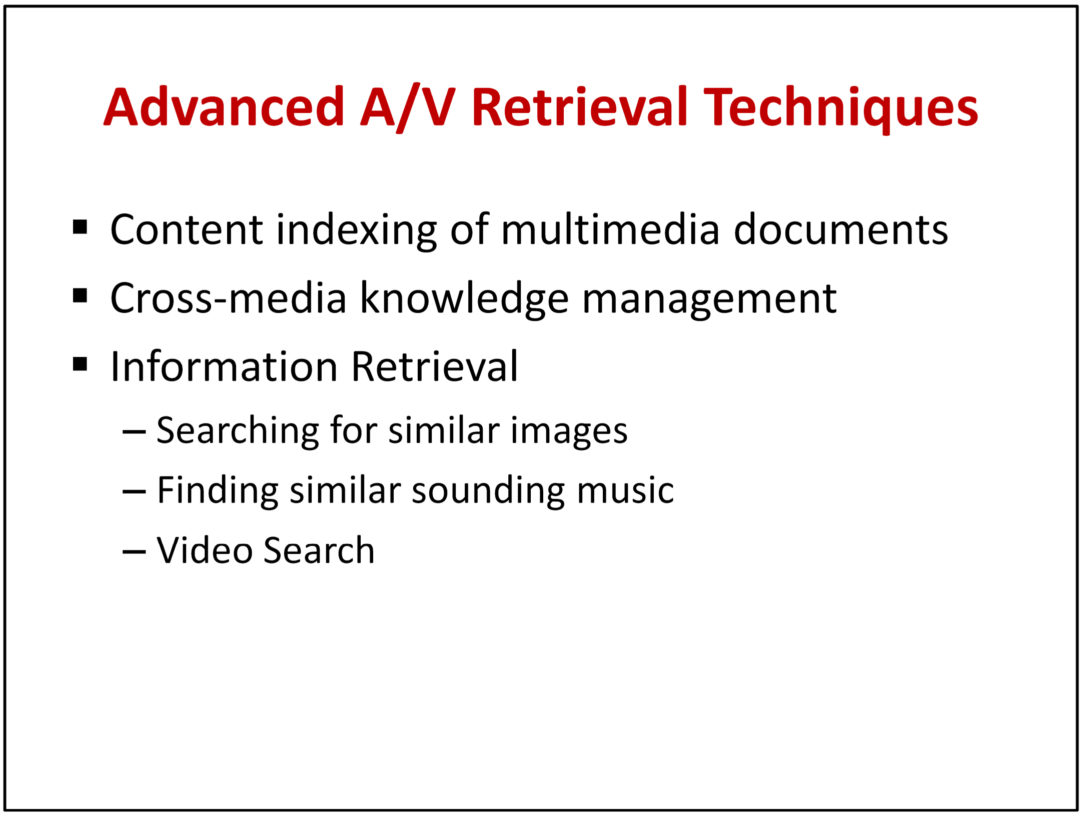
- Searching in multimedia documents for the information on surrounding words – mainly in titles of images, video and audio files.
- Using tags associated with multimedia objects
- Metadata annotations for images, audio and video – The dominant standard in multimedia content description is the MPEG-7 (ISO MPEG Group).

Last decade was proposed several methods for maintaining semantic information for multimedia files like:

Searching in multimedia documents for the information on surrounding words – mainly in titles of images, video and audio files. Unfortunately in this approach we receive not sound and correct answers due to metaphoric and abstract annotation of multimedia objects which some times not corresponds to their content.

Using tags associated with multimedia objects – there are also a lot of wrong tags and sometimes this leads to misconceptions

Metadata annotations for images, audio and video – The dominant standard in multimedia content description is the MPEG-7 (ISO MPEG Group). This standard provides rich general purpose multimedia content description capabilities. It includes both low-level features and high-level semantic description constructs.



Advanced A/V Retrieval Techniques

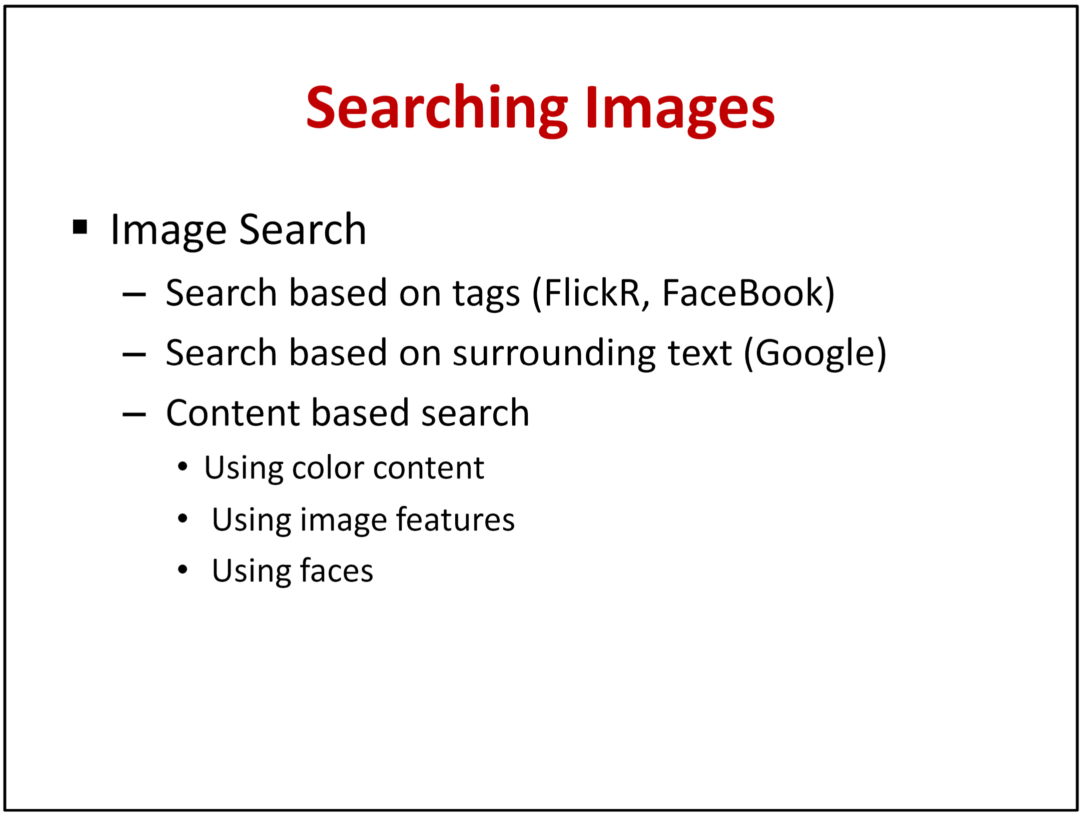
- Content indexing of multimedia documents
- Cross-media knowledge management
- Information Retrieval
  - Searching for similar images
  - Finding similar sounding music
  - Video Search

Content indexing of multimedia documents – There are used 3 layers for annotating multimedia documents - audio, video i text(subtitles). There usually are used natural language processing (NLP) techniques for annotating such documents and for context information extraction. Unfortunately the speech recognition can not be widely used for such purposes, it is restricted only for some languages and specific domains.

Cross-media knowledge management – mainly based on events monitoring

Information Retrieval

Searching for similar images
Finding similar sounding music
Video Search



Searching Images

- Image Search
  - Search based on tags (FlickR, FaceBook)
  - Search based on surrounding text (Google)
  - Content based search
    - Using color content
    - Using image features
    - Using faces

Search over tags associated with images
Users manually add Tags to images (FlickR, FaceBook)
Find images with tags that match the query keyword
Limitations -
Tags require human effort to create
Tags may be wrong

Use text associated with images for search
Search web for images
Use surrounding text
Text in URL for image filename
Text in HTML on page
Same as text search

Query can be an image and searching for similar images
Similarity is defined by features of the image
Color Content

Color Histogram - Distribution of pixel colors in image. No spatial information. Similarity based on histogram Distance.

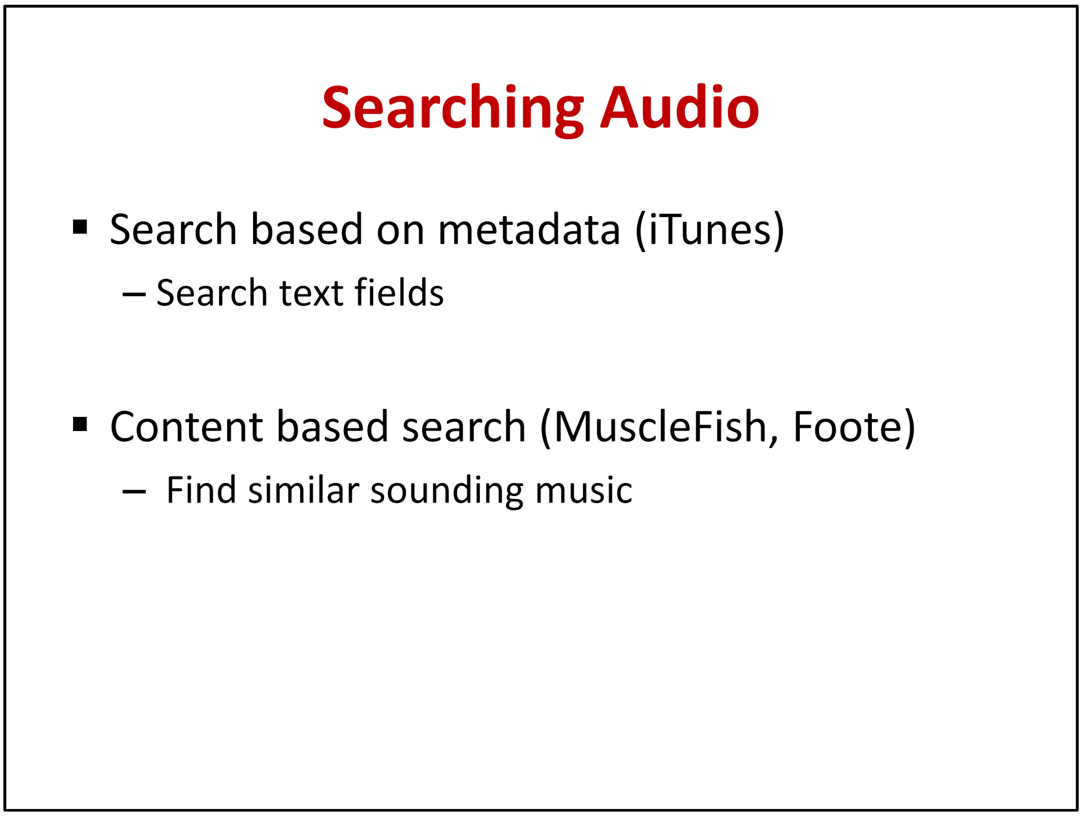
Color Corellogram - Color histogram as a function of distance between pixels. Multiple color histograms – one for each distance. Distribution of pixel color plus spatial information. Similarity based on correlogram difference

Image descriptors

Gradients at image keypoints - SIFT (Scale Invariant Feature Transform) Features (2004: David Lowe, UBC) . Select keypoints regions in image from extrema in scale space. Different images have different numbers of keypoints. Compute feature vectors X for each keypoint region. Feature vectors from histogram of gradient directions near the keypoint. SIFT features X are 128-dimensional vectors. Image described by N SIFT features. Features are X1,….XN, N is different for different images

Quantize for “Visual words” - Quantize SIFT features to create “visual words” to represent images (2006: Lienhart, University of Augsburg & Slaney, Yahoo!) Cluster SIFT features of representative images. Features X are in 128-dimensional space. Generate W clusters. Clusters define “visual words” . All features in same cluster are the same “visual word”. To compute visual words describing an image - Compute N SIFT X1,….XN features for the image,; Find nearest cluster center (codeword) to each features Xj. These clusters define the visual words for the image. Image is described by it’s visual words. Just like a document is described by the text words. Create image index ; Compute visual words for all images; Create a visual word index into the images; Compute visual words for query image; Use query words for retrieval; Just like text! Except the visual words aren’t quite as meaningful

Faces

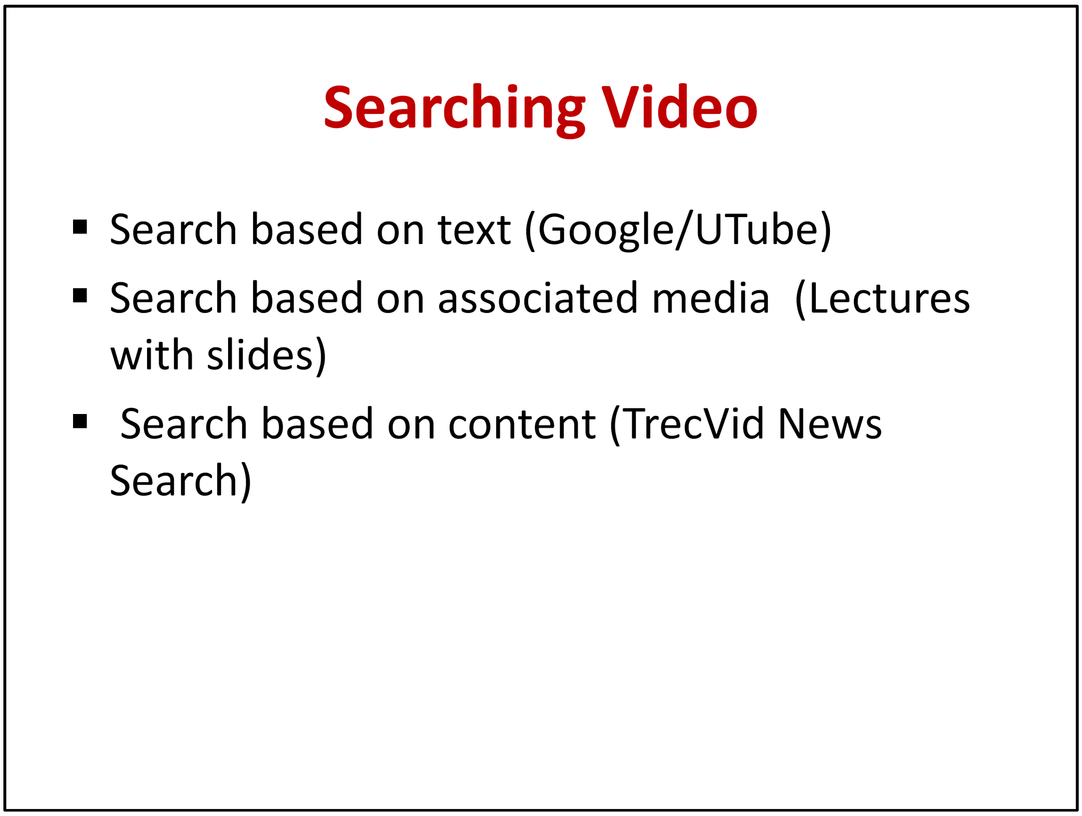


Searching Audio

- Search based on metadata (iTunes)
  - Search text fields
- Content based search (MuscleFish, Foote)
  - Find similar sounding music

Search based on metadata (iTunes)
Search text fields
Title
Artist
Album
Genre
Example 1997: Jon Foote, FXPAL; Similarity of Nat King Cole and Gregorian Chant

Content based search (MuscleFish, Foote)
Find similar sounding music
Compute spectral feature vectors (MFCC)
Quantize features to create audio histogram - Audio histogram describes sounds; Order of sounds is lost

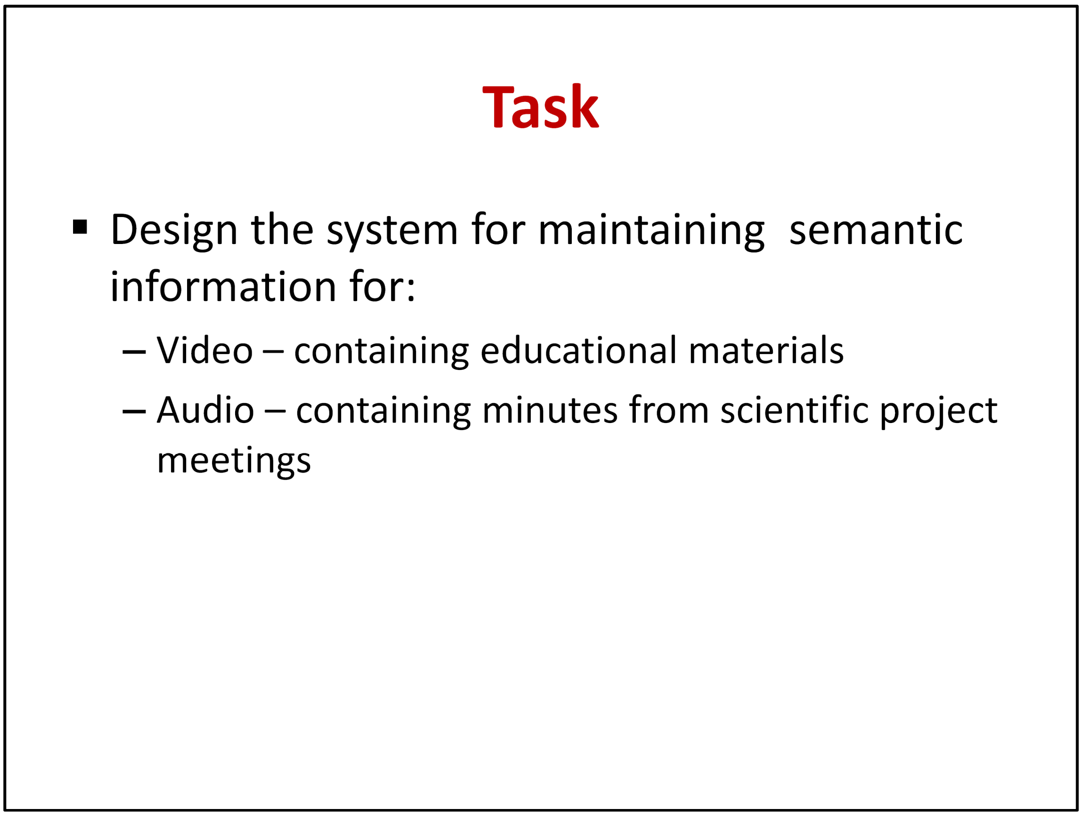


Searching Video

- Search based on text (Google/UTube)
- Search based on associated media (Lectures with slides)
- Search based on content (TrecVid News Search)

Search based on text (Google/UTube)
Search based on associated media (Lectures with slides)
Search based on content (TrecVid News Search)

Search for an entire video
    Search using surrounding text



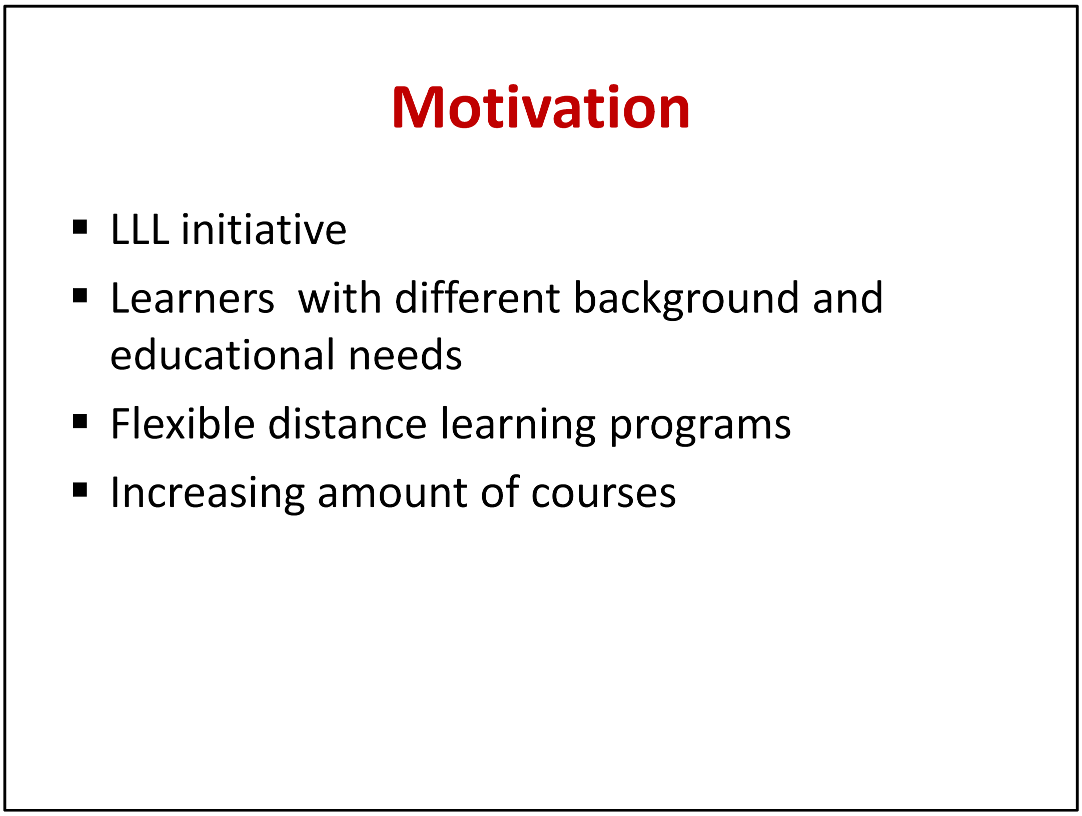
Task

- Design the system for maintaining semantic information for:
  - Video – containing educational materials
  - Audio – containing minutes from scientific project meetings

The main goal is to design the system for maintaining semantic information for:

Video – containing educational materials
Audio – containing minutes from scientific project meetings

In this presentation I will focus mainly on the first task



Motivation

- LLL initiative
- Learners with different background and educational needs
- Flexible distance learning programs
- Increasing amount of courses

LLL (long-life learning) needs increasing
A lot of courses for qualification are provided for several organizations – universities, companies, ...

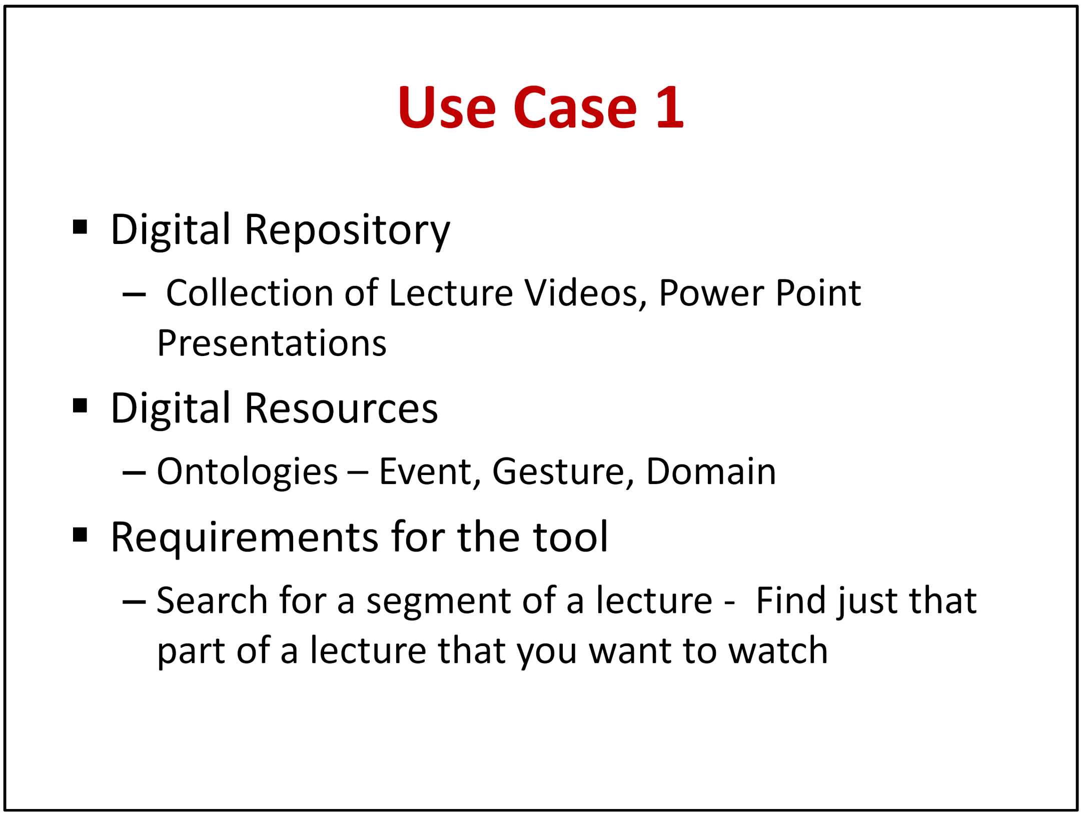
Training students at undergraduate and graduate university programs, as well as training employees in companies for additional qualification improvement
The trained people are coming with different background and learning needs

We need to provide flexible solutions and to present the users personalized view to the resources

For such group of users the more important functionality which needs to be provided by our system in order to support the education is not only to be able to play those video resources but also to search the appropriate information among them.

In order to provide such functionality we need an approach for semantic annotation of such video materials and tools for maintaining semantic metadata.

Unfortunately this challenging task is efforts, time consuming and requires not only additional resources to be available, but also a lot of theoretical issues regarding semantic information capturing and maintained.

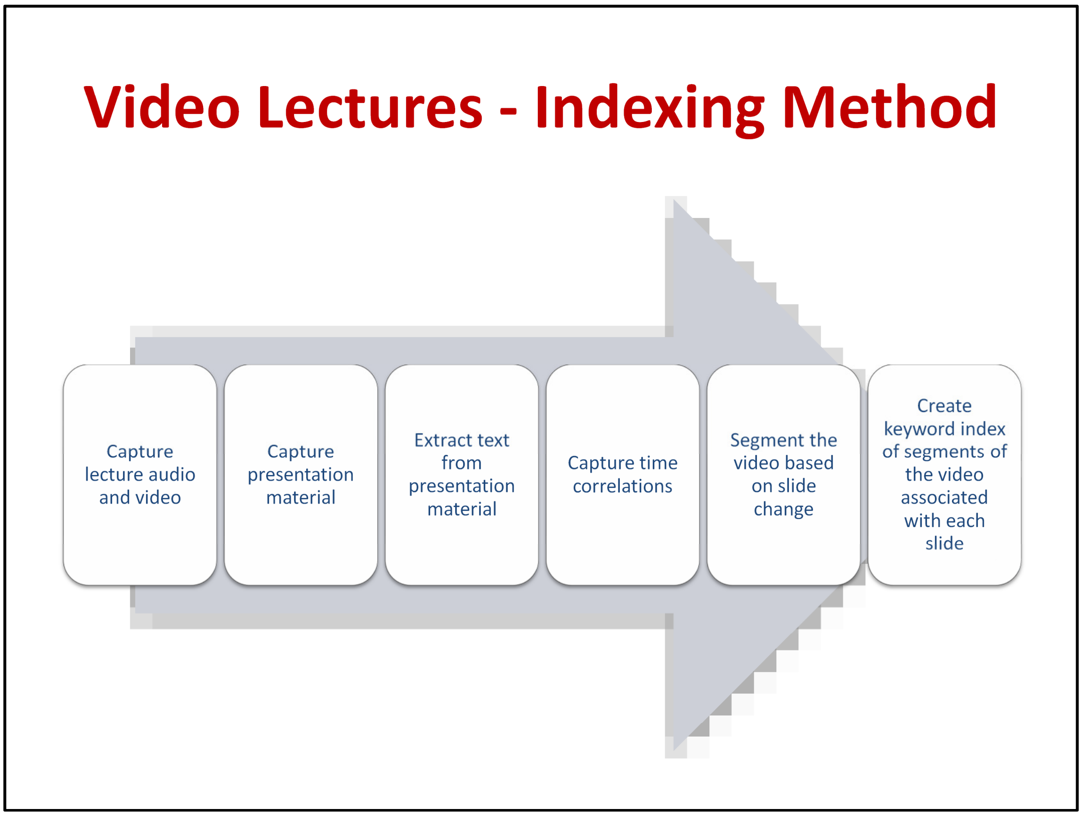


Use Case 1

- Digital Repository
  - Collection of Lecture Videos, Power Point Presentations
- Digital Resources
  - Ontologies – Event, Gesture, Domain
- Requirements for the tool
  - Search for a segment of a lecture - Find just that part of a lecture that you want to watch

The our video collection contains mainly filmed lectures stored as video files, accompanied with power point slides and other educational materials uploaded at e-learning platform for distance learning.
The lectures are with varying duration – from about 10 up to 45 minutes.

Digital Resources
  Ontologies – Event, Gesture, Domain
Requirements for the tool
  Search for a segment of a lecture - Find just that part of a lecture that you want to watch



Video Lectures - Indexing Method

Capture lecture audio and video

Capture presentation material

Extract text from presentation material

Capture time correlations

Segment the video based on slide change

Create keyword index of segments of the video associated with each slide

Indexing Method

Capture lecture audio and video

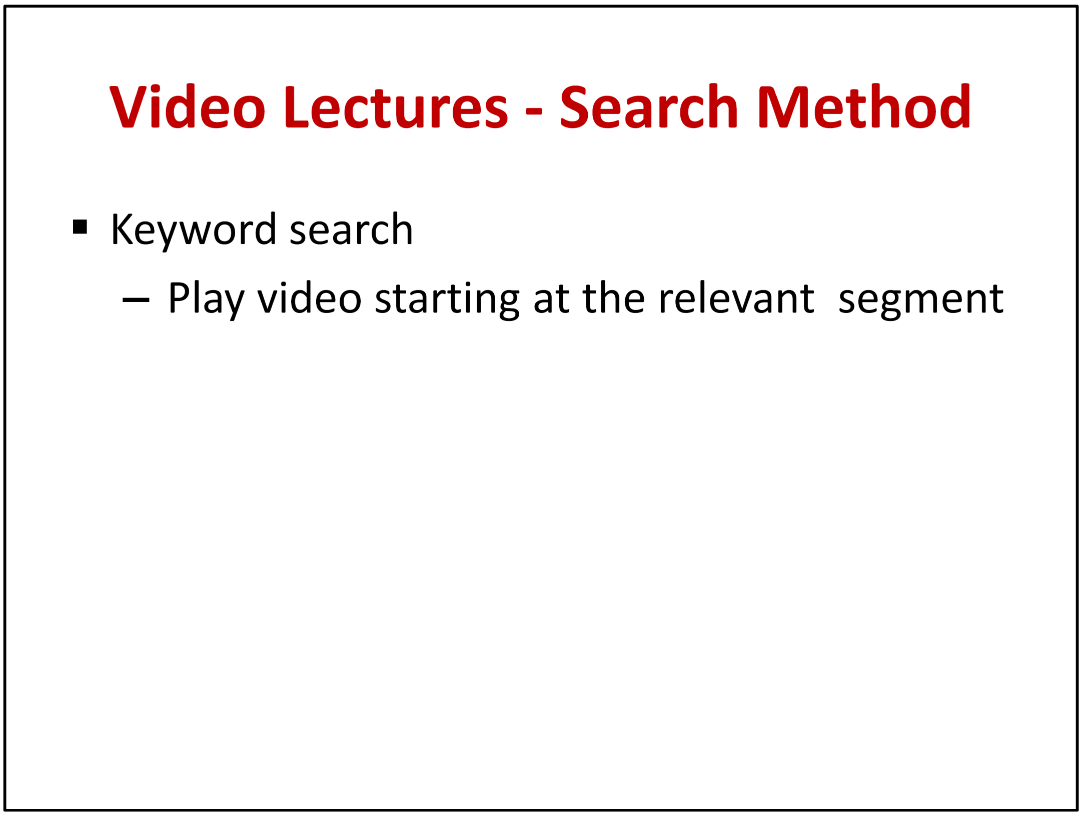
Capture presentation material

Extract text from presentation material

Capture time correlations

Segment the video based on slide change

Create keyword index of segments of the video associated with each slide

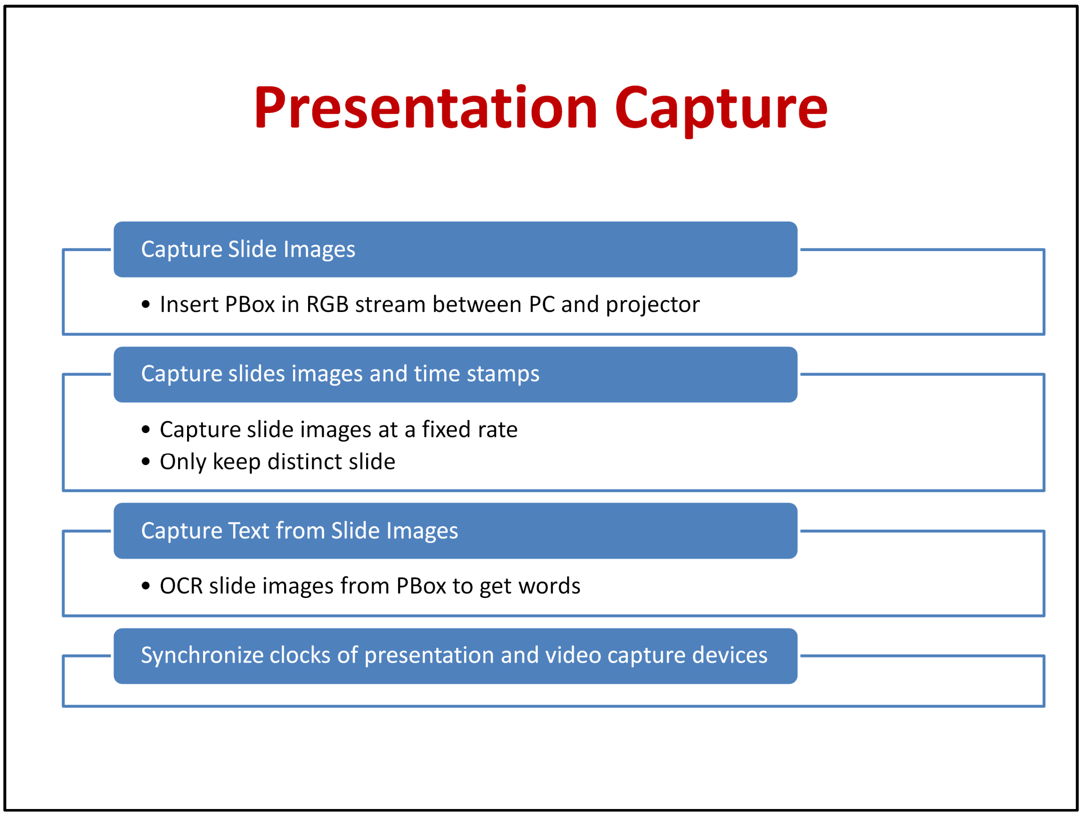


Video Lectures - Search Method

- Keyword search
  - Play video starting at the relevant segment

Keyword search

Play video starting at the relevant segment



Presentation Capture

Capture Slide Images

- Insert PBox in RGB stream between PC and projector

Capture slides images and time stamps

- Capture slide images at a fixed rate
- Only keep distinct slide

Capture Text from Slide Images

- OCR slide images from PBox to get words

Synchronize clocks of presentation and video capture devices

Capture Slide Images

ProjectorBox (PBox): Denoue and Hilbert FXPAL

Insert PBox in RGB stream between PC and projector

Capture slides images and time stamps

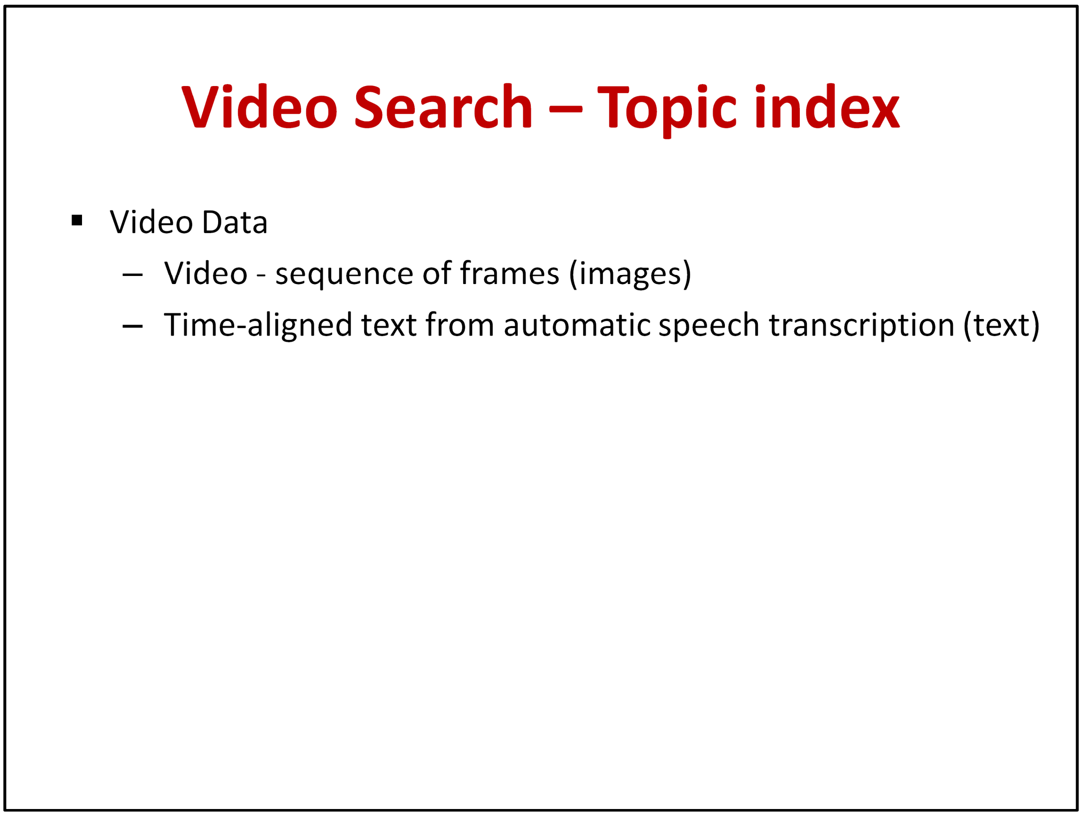
Capture slide images at a fixed rate

Only keep distinct slide

Capture Text from Slide Images

OCR slide images from PBox to get words (Optical Character Recognition (OCR) to convert text image to electronic text)

Synchronize clocks of presentation and video capture devices



Video Search – Topic index

- Video Data
  - Video - sequence of frames (images)
  - Time-aligned text from automatic speech transcription (text)

Video Data

Video - sequence of frames (images)

Time-aligned text from automatic speech transcription (text)

Pre-processing

Segment video into shots using image features

Compute pairwise similarity between frames of video

Similarity is based on image features

Segment when similarity is low

Select a representative keyframe for each shot

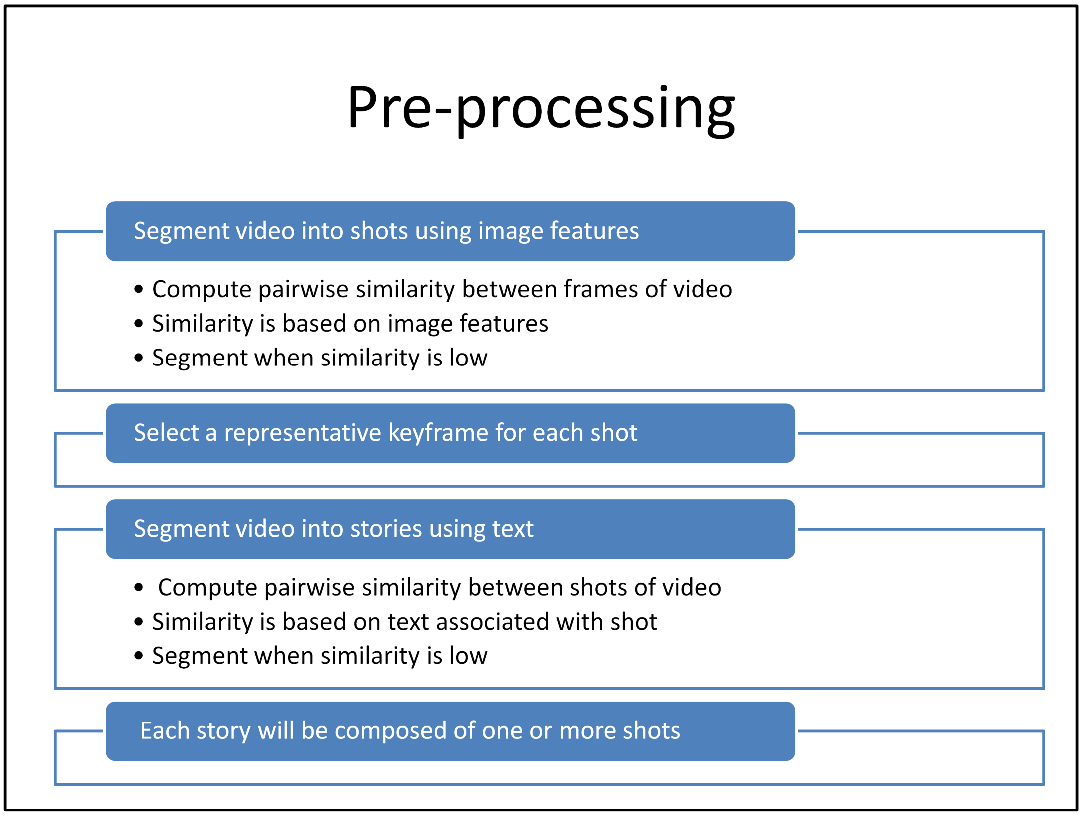
Segment video into stories using text

Compute pairwise similarity between shots of video

Similarity is based on text associated with shot

Segment when similarity is low

Each story will be composed of one or more shots



Pre-processing

Segment video into shots using image features

- Compute pairwise similarity between frames of video
- Similarity is based on image features
- Segment when similarity is low

Select a representative keyframe for each shot

Segment video into stories using text

- Compute pairwise similarity between shots of video
- Similarity is based on text associated with shot
- Segment when similarity is low

Each story will be composed of one or more shots

Pre-processing

Segment video into shots using image features

Compute pairwise similarity between frames of video

Similarity is based on image features

Segment when similarity is low

Select a representative keyframe for each shot

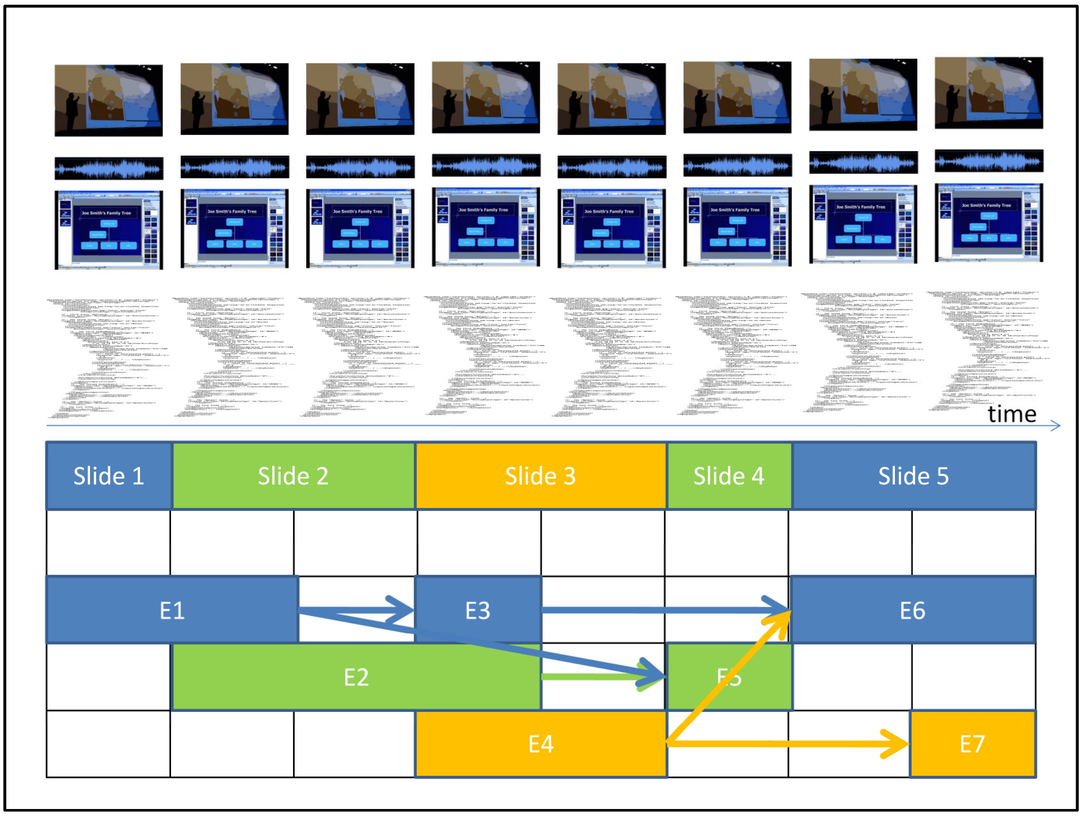
Segment video into stories using text

Compute pairwise similarity between shots of video

Similarity is based on text associated with shot

Segment when similarity is low

Each story will be composed of one or more shots



Segment Video

New segment when slide changes

Video associated with a slide

Create index into video segments associated with each slide

Index each slide in video based on text

Search

Keyword search locates relevant slide

Play video at starting time for that segment

Video

Sequence of frames (images), typically with audio 30 frames/second

Text Transcript of Audio

Time-correlated with video

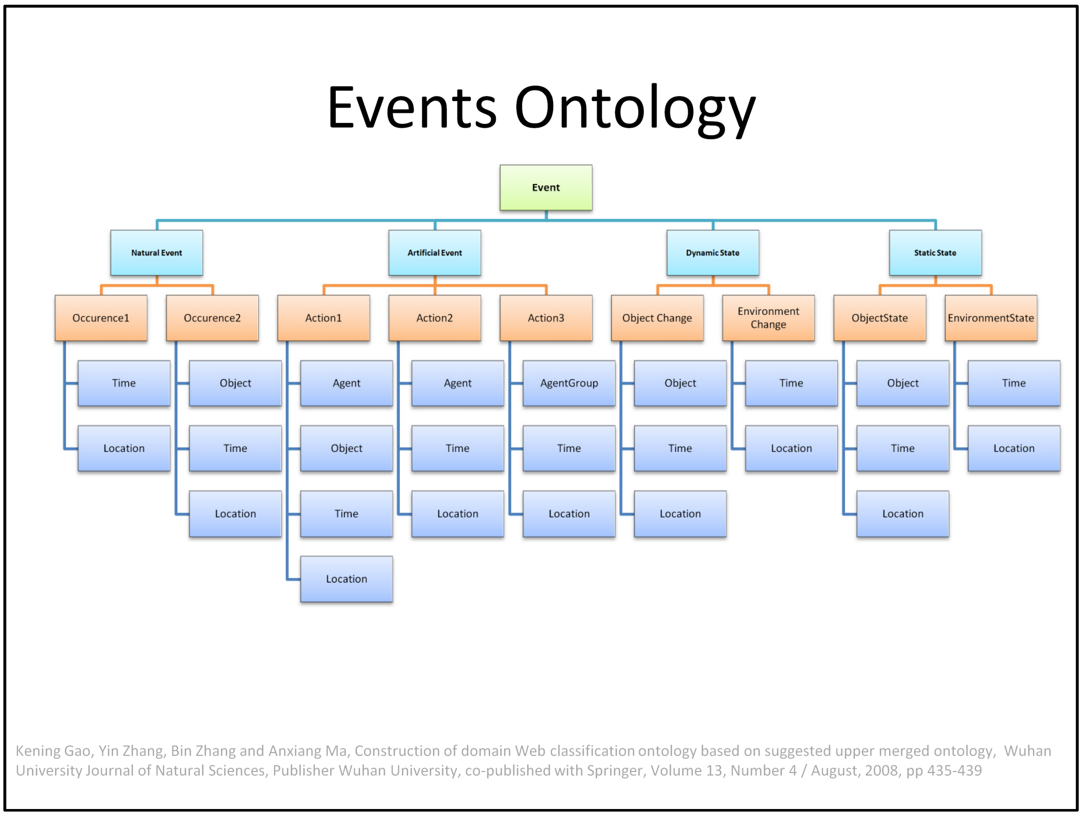
Segments of Video

Shot: Unbroken segment of video from a single camera

Story : Sequence of shots from the same lecture

Keyframe

Events Graph generations

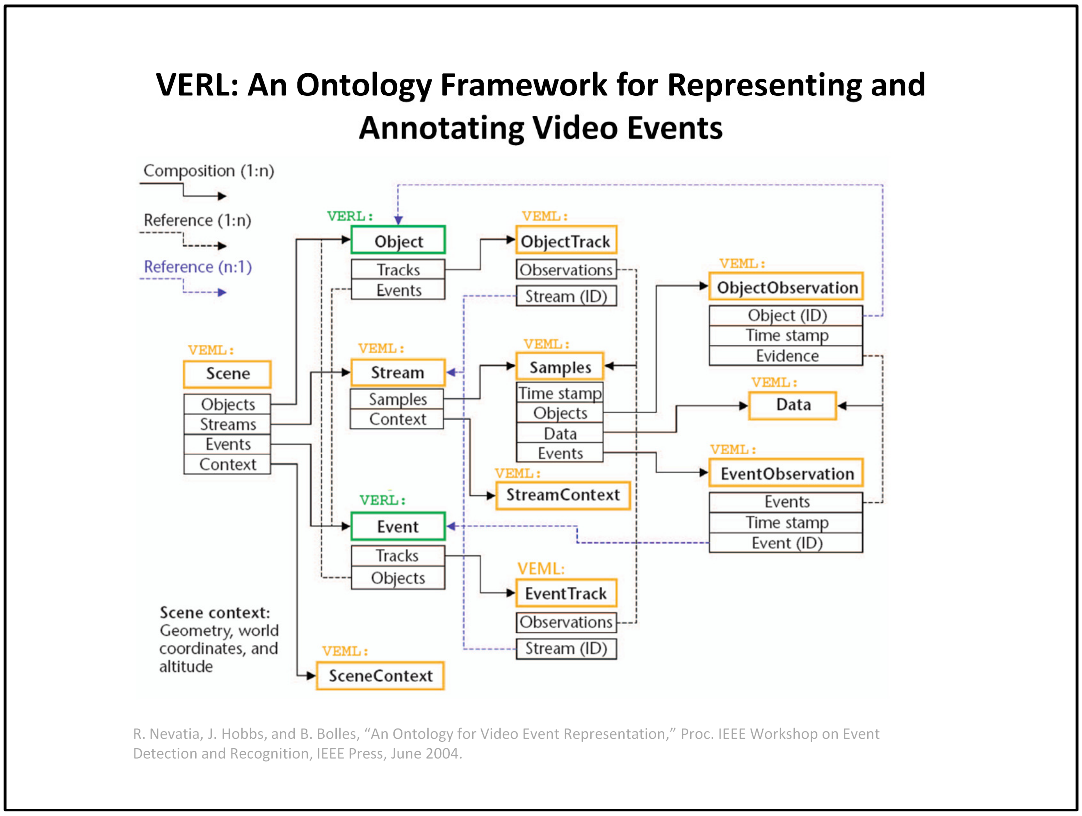


Events Ontology

- Event
  - Natural Event
    - Occurence1
      - Time
      - Location
    - Occurence2
      - Object
      - Time
      - Location
  - Artificial Event
    - Action1
      - Agent
      - Object
      - Time
      - Location
    - Action2
      - Agent
      - Time
      - Location
    - Action3
      - AgentGroup
      - Time
      - Location
  - Dynamic State
    - Object Change
      - Object
      - Time
      - Location
    - Environment Change
      - Time
      - Location
  - Static State
    - ObjectState
      - Object
      - Time
      - Location
    - EnvironmentState
      - Time
      - Location

Kening Gao, Yin Zhang, Bin Zhang and Anxiang Ma, Construction of domain Web classification ontology based on suggested upper merged ontology, Wuhan University Journal of Natural Sciences, Publisher Wuhan University, co-published with Springer, Volume 13, Number 4 / August, 2008, pp 435-439

We are using Event ontology

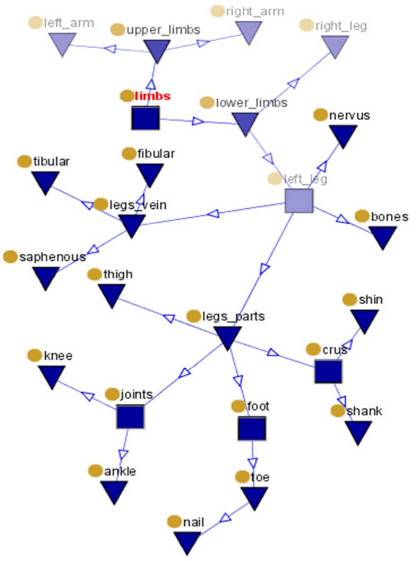


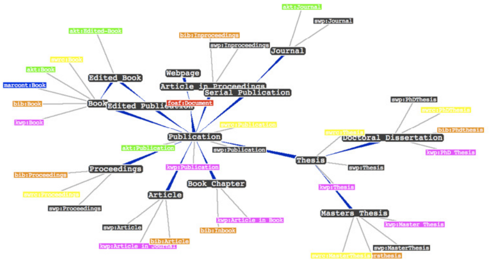
VERL: An Ontology Framework for Representing and Annotating Video Events

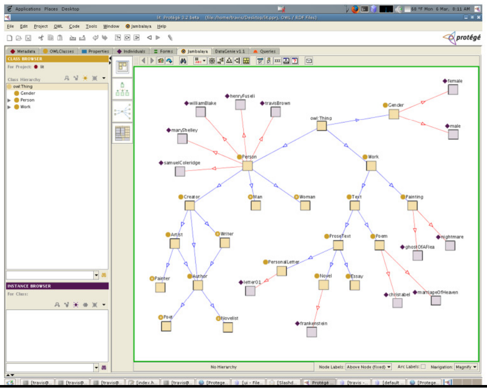
R. Nevatia, J. Hobbs, and B. Bolles, "An Ontology for Video Event Representation," Proc. IEEE Workshop on Event Detection and Recognition, IEEE Press, June 2004.

The metadata are based on standard - VERL: An Ontology Framework for Representing and Annotating Video Events

Domain Specific Ontologies







lit Protégé 3.2 beta (file:/home/travis/Desktop/lit.pprj, OWL / RDF Files)
File Edit Project OWL Code Tools Window Jambalaya Help
protégé
Metadata OWLClasses Properties Individuals Forms Jambalaya DataGenie v1.1 Queries
CLASS BROWSER
INSTANCE BROWSER
owl:Thing
Gender
Person
Work
female
male
Creator
Woman
Text
Painting
Artist
Writer
ProseText
Poem
Painter
Author
Novel
Essay
Poet
Novelist
frankenstein

Domain Specific Ontologies was used both in metadata generation and for searching queries – currently we tested the system for Computer Science, Medicine and Library Studies domains.

The ontologies are designed using OWL language and Protégé System

Queries Processing

- Multi-documents search
- Ranking the results
- Types of queries:
  - Simple - containing just keywords
  - Relations
    - Events relations
    - Time relations
    - Concepts relations

Multi-documents search
Types of queries:
Simple - containing just keywords
Relations
Events relations
Time relations
Concepts relations

Conclusion and Further Work

- In this challenging task we try to automate the process of metadata generation indexing video lecture materials
- We tried to add events structure and to use ontologies in order to be able to answer more complicated queries
- Some of the procedures was solved with different success and as further work we will try to find better solutions.

Thank you for Your attention!

Svetla Boytcheva
State University of Library Studies and Information Technologies, Bulgaria
svetla.boytcheva@gmail.com