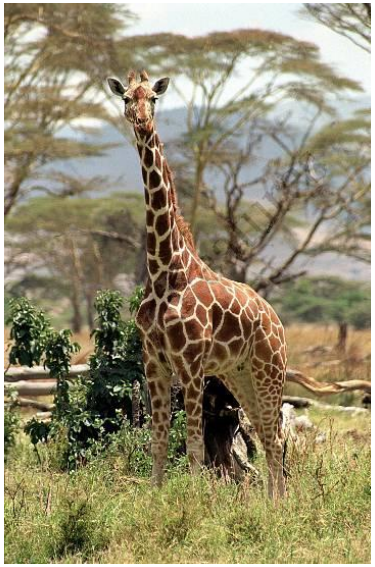
Multimedia Information Retrieval

http://morpheus.micc.unifi.it/learning/

State-of-the-art Multimedia Search Engines

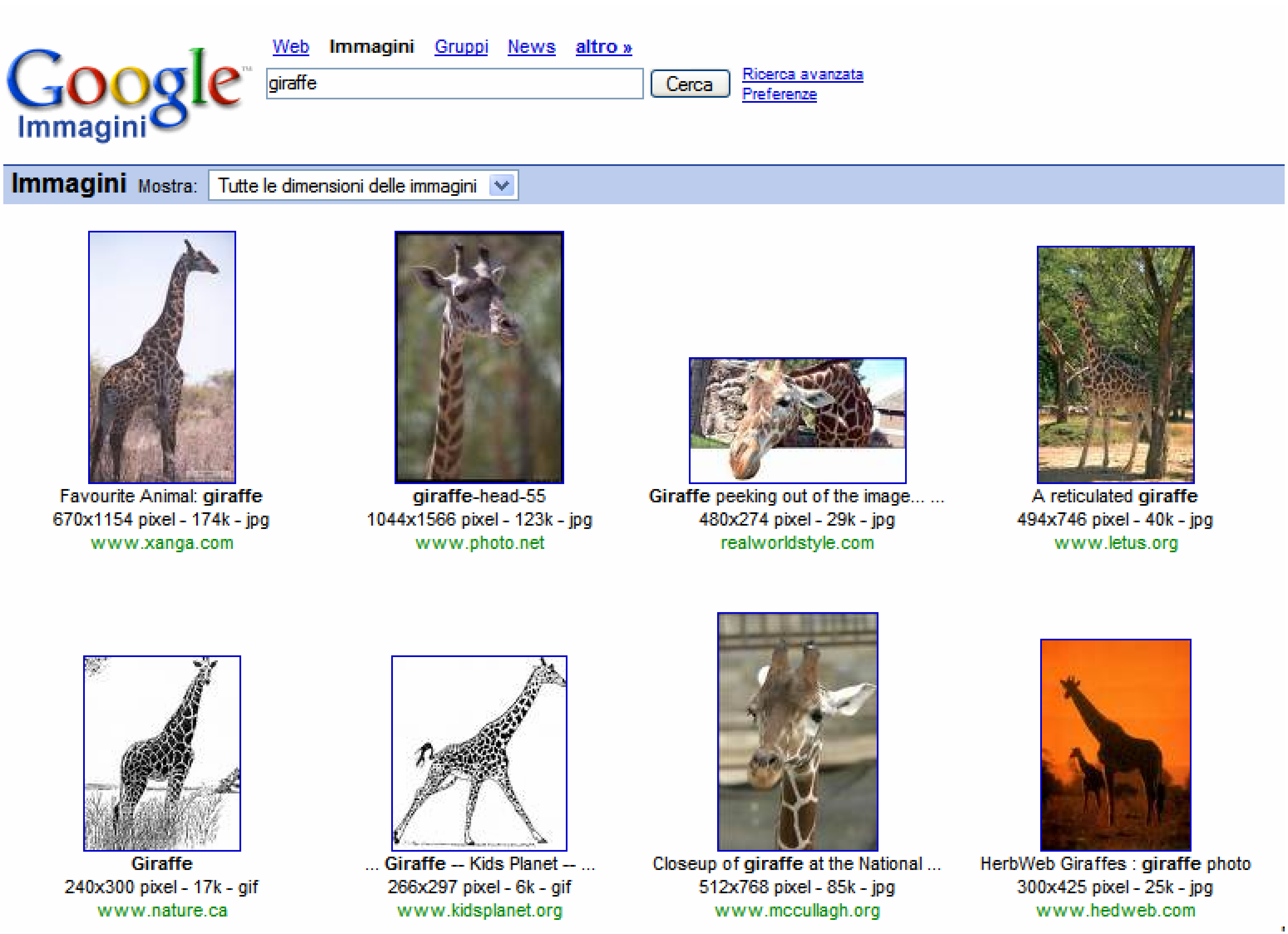
- Work better for simple concepts,
  e.g. Two people kissing, A picture of a giraffe
- Don’t work for complex queries
  e.g. A picture of a brick home with black shutters and white pillars, with a pickup truck in front of it (image)



Examples

- Find the pictures of giraffe
  - Keyword: giraffe
  - http://images.google.it/images?svnum=10&hl=it&lr=&rls=GGGL%2CGGGL%3A2006-19%2CGGGL%3Ait&q=giraffe&btnG=Cerca

- A picture of a brick home with black shutters and white pillars, with a pickup truck in front of it (image)
  - brick home shutters
  - http://images.google.it/images?sourceid=navclient-ff&ie=UTF-8&rls=GGGL,GGGL:2006-19,GGGL:it&q=brick+home+shutters+



giraffe

Immagini Mostra: Tutte le dimensioni delle immagini

Favourite Animal: **giraffe**
670x1154 pixel - 174k - jpg
www.xanga.com

giraffe-head-55
1044x1566 pixel - 123k - jpg
www.photo.net

Giraffe peeking out of the image... ...
480x274 pixel - 29k - jpg
realworldstyle.com

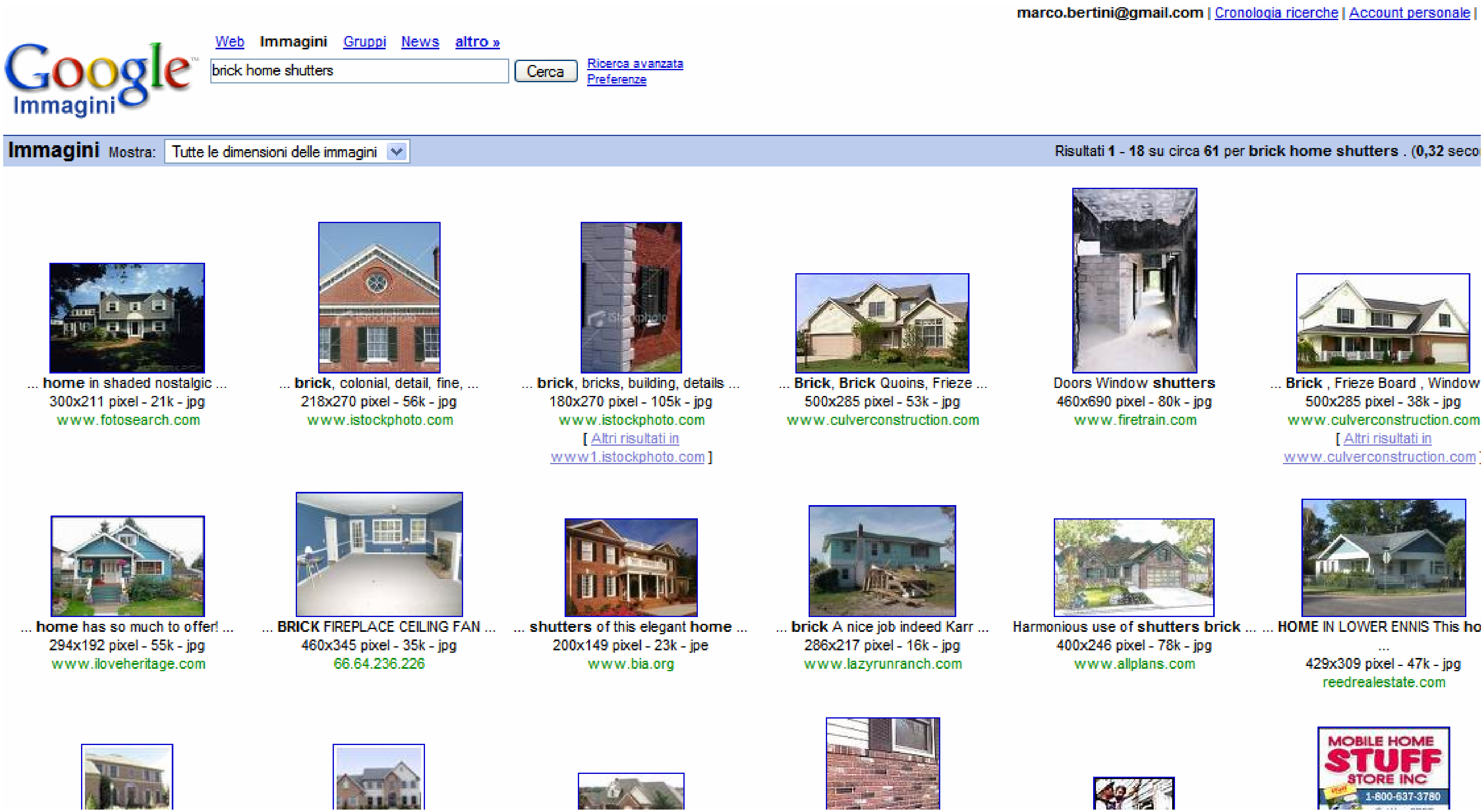
A reticulated **giraffe**
494x746 pixel - 40k - jpg
www.letus.org

Giraffe
240x300 pixel - 17k - gif
www.nature.ca

... **Giraffe** -- Kids Planet -- ...
266x297 pixel - 6k - gif
www.kidsplanet.org

Closeup of **giraffe** at the National ...
512x768 pixel - 85k - jpg
www.mccullagh.org

HerbWeb Giraffes : **giraffe** photo
300x425 pixel - 25k - jpg
www.hedweb.com



marco.bertini@gmail.com | Cronologia ricerche | Account personale |

Google Immagini

Web Immagini Gruppi News altro »

brick home shutters

Cerca

Ricerca avanzata
Preferenze

Immagini Mostra: Tutte le dimensioni delle immagini

Risultati **1** - **18** su circa **61** per **brick home shutters** . (**0,32** seco

... **home** in shaded nostalgic ...
300x211 pixel - 21k - jpg
www.fotosearch.com

... **brick**, colonial, detail, fine, ...
218x270 pixel - 56k - jpg
www.istockphoto.com

... **brick**, bricks, building, details ...
180x270 pixel - 105k - jpg
www.istockphoto.com
[Altri risultati in www1.istockphoto.com]

... **Brick**, **Brick** Quoins, Frieze ...
500x285 pixel - 53k - jpg
www.culverconstruction.com

Doors Window **shutters**
460x690 pixel - 80k - jpg
www.firetrain.com

... **Brick** , Frieze Board , Window
500x285 pixel - 38k - jpg
www.culverconstruction.com
[Altri risultati in www.culverconstruction.com

... **home** has so much to offer! ...
294x192 pixel - 55k - jpg
www.iloveheritage.com

... **BRICK** FIREPLACE CEILING FAN ...
460x345 pixel - 35k - jpg
66.64.236.226

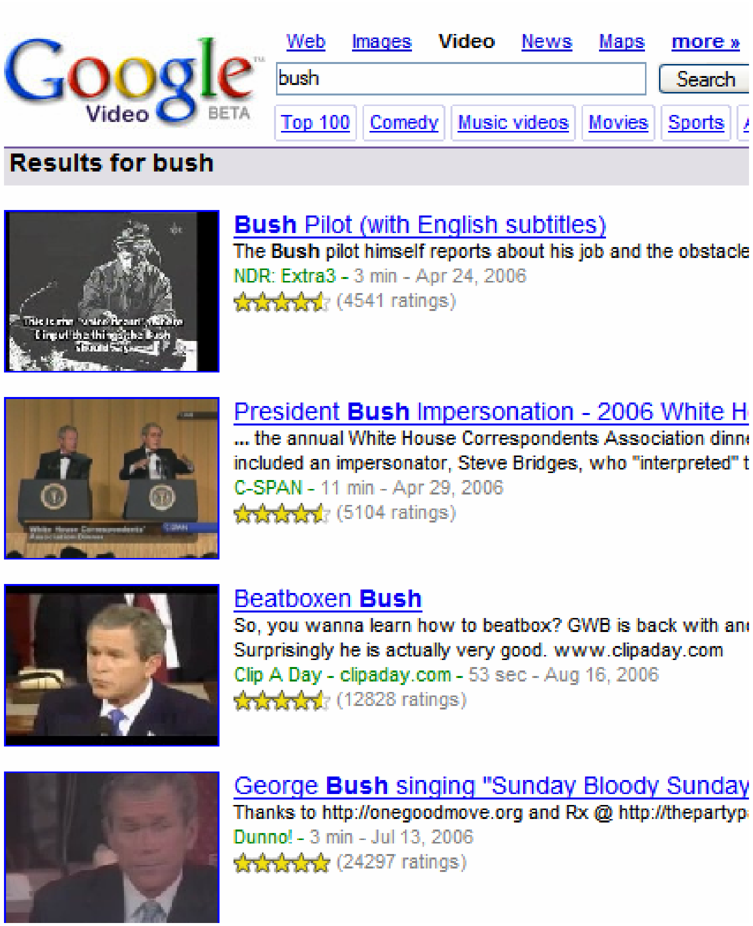
... **shutters** of this elegant **home** ...
200x149 pixel - 23k - jpe
www.bia.org

... **brick** A nice job indeed Karr ...
286x217 pixel - 16k - jpg
www.lazyrunranch.com

Harmonious use of **shutters brick** ...
400x246 pixel - 78k - jpg
www.allplans.com

... **HOME** IN LOWER ENNIS This ho
...
429x309 pixel - 47k - jpg
reedrealestate.com

- In Google Video try searching for “Bush” or “Bush speaking about Iraq”



bush

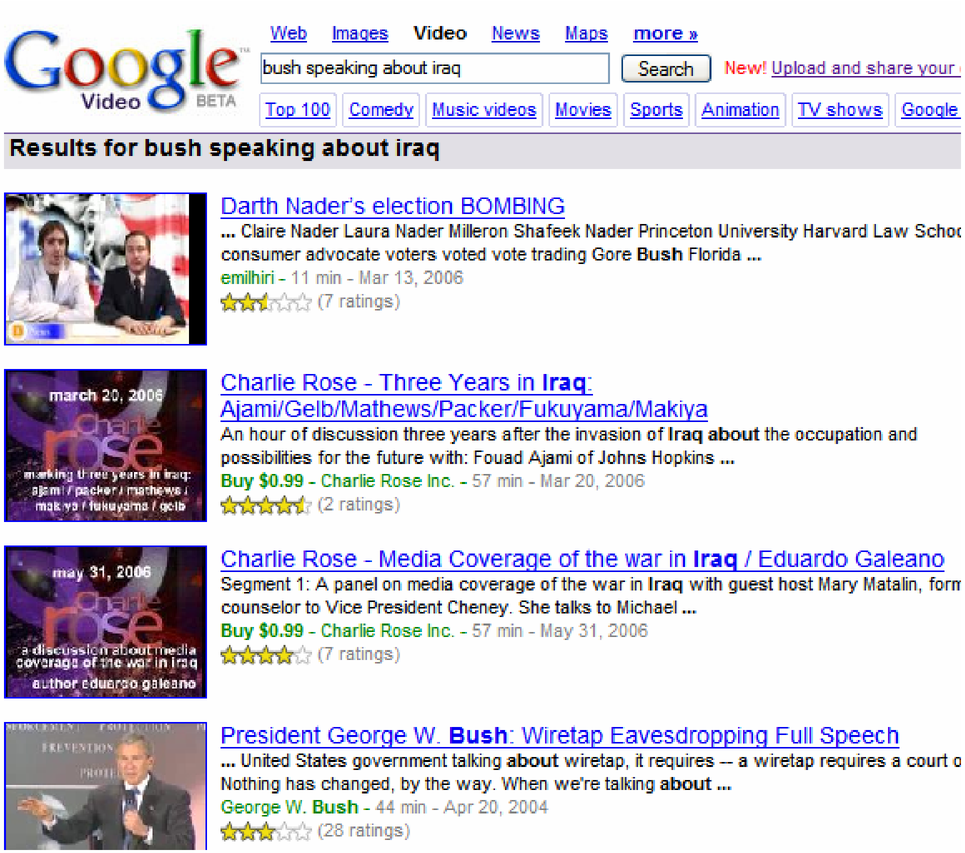
Results for bush

Bush Pilot (with English subtitles)
The **Bush** pilot himself reports about his job and the obstacle
NDR: Extra3 - 3 min - Apr 24, 2006
(4541 ratings)

President **Bush** Impersonation - 2006 White H
... the annual White House Correspondents Association dinn
included an impersonator, Steve Bridges, who "interpreted" t
C-SPAN - 11 min - Apr 29, 2006
(5104 ratings)

Beatboxen **Bush**
So, you wanna learn how to beatbox? GWB is back with an
Surprisingly he is actually very good. www.clipaday.com
Clip A Day - clipaday.com - 53 sec - Aug 16, 2006
(12828 ratings)

George **Bush** singing "Sunday Bloody Sunday
Thanks to http://onegoodmove.org and Rx @ http://thepartyp
Dunno! - 3 min - Jul 13, 2006
(24297 ratings)



bush speaking about iraq

New! Upload and share your

Results for bush speaking about iraq

Darth Nader's election BOMBING
... Claire Nader Laura Nader Milleron Shafeek Nader Princeton University Harvard Law Schoo
consumer advocate voters voted vote trading Gore **Bush** Florida ...
emilhiri - 11 min - Mar 13, 2006
(7 ratings)

Charlie Rose - Three Years in **Iraq**: Ajami/Gelb/Mathews/Packer/Fukuyama/Makiya
An hour of discussion three years after the invasion of **Iraq about** the occupation and
possibilities for the future with: Fouad Ajami of Johns Hopkins ...
Buy $0.99 - Charlie Rose Inc. - 57 min - Mar 20, 2006
(2 ratings)

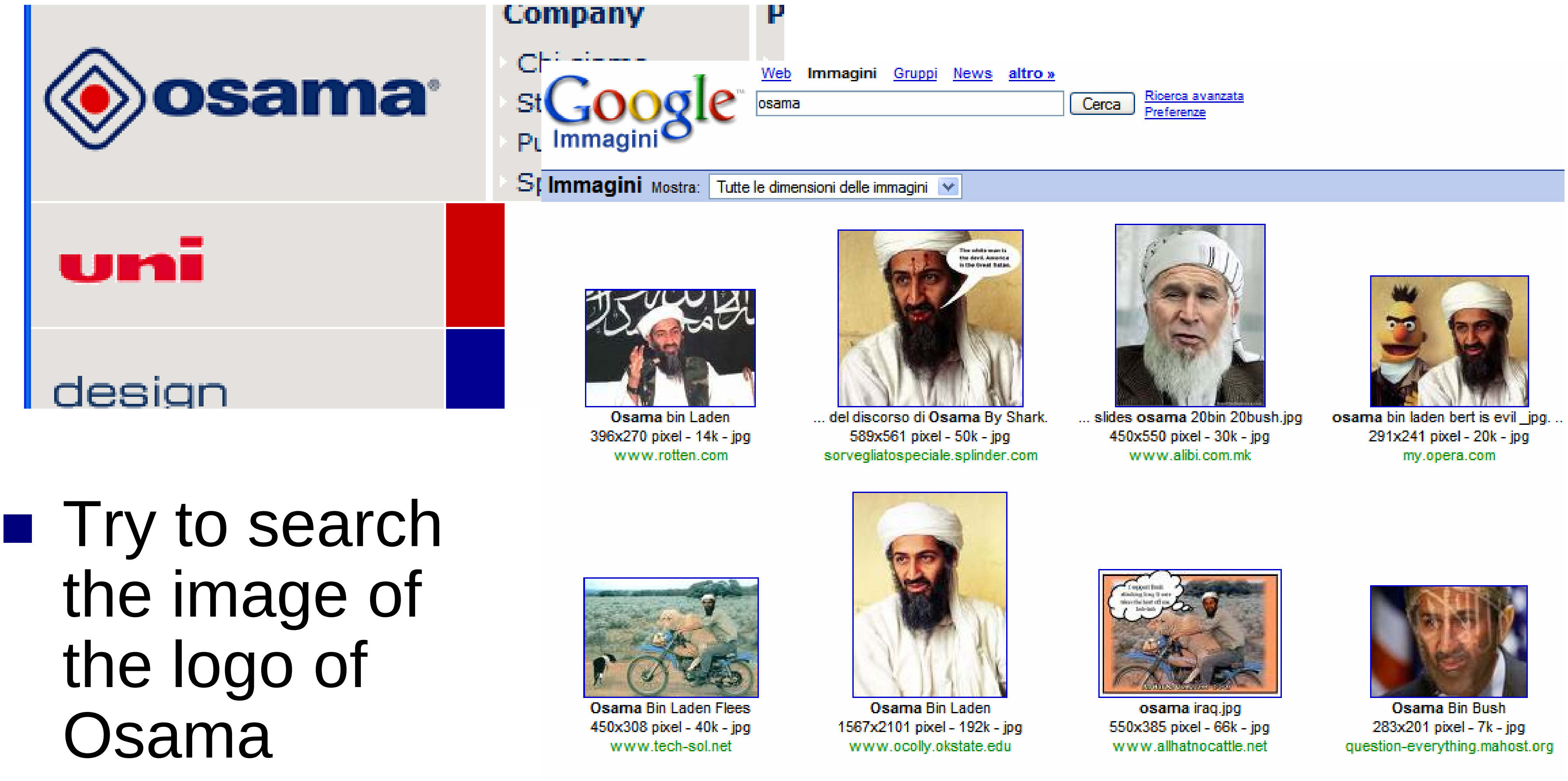
Charlie Rose - Media Coverage of the war in **Iraq** / Eduardo Galeano
Segment 1: A panel on media coverage of the war in **Iraq** with guest host Mary Matalin, form
counselor to Vice President Cheney. She talks to Michael ...
Buy $0.99 - Charlie Rose Inc. - 57 min - May 31, 2006
(7 ratings)

President George W. **Bush**: Wiretap Eavesdropping Full Speech
... United States government talking **about** wiretap, it requires -- a wiretap requires a court o
Nothing has changed, by the way. When we're talking **about** ...
George W. **Bush** - 44 min - Apr 20, 2004
(28 ratings)

Why this happens?

- Most of these search engines are keyword based
  - "False" multi-media search engine
  - Have to represent your idea in keywords
  - These keywords are expected to appear in the filename, or corresponding webpage
- Therefore……
  - Unable to handle semantic meaning of images
  - Unable to handle visual position
  - Unable to handle time information
  - Unable to use images as query
  - ……….

When I use a word," Humpty Dumpty said, in rather a scornful tone, "it means just what I choose it to mean—neither more nor less.



osama

Immagini Mostra: Tutte le dimensioni delle immagini

Osama bin Laden
396x270 pixel - 14k - jpg
www.rotten.com

... del discorso di Osama By Shark.
589x561 pixel - 50k - jpg
sorvegliatospeciale.splinder.com

... slides osama 20bin 20bush.jpg
450x550 pixel - 30k - jpg
www.alibi.com.mk

osama bin laden bert is evil _jpg. ..
291x241 pixel - 20k - jpg
my.opera.com

Osama Bin Laden Flees
450x308 pixel - 40k - jpg
www.tech-sol.net

Osama Bin Laden
1567x2101 pixel - 192k - jpg
www.ocolly.okstate.edu

osama iraq.jpg
550x385 pixel - 66k - jpg
www.allhatnocattle.net

Osama Bin Bush
283x201 pixel - 7k - jpg
question-everything.mahost.org

- Try to search the image of the logo of Osama

How Google does it?

- No image processing. Textual context!
    - In videos it uses closed captions and transcriptions
- File names, nearby words
- Distance from image to words
- “give me images with *flower* in the file name or near the image”

Solution

- ……it would be great to have multimedia search engine intelligent enough to associate its own keywords based on what's in the image.

- Content-based information retrieval (CBIR)

- Different from text IR:
  - Structure of data is more complex. Efficiency is an issue
  - Using of metadata
  - Characteristics of multimedia data
  - Operations to be performed
- Aspects:
  - Data modeling: Extract and maintain the features of objects
  - Data retrieval: based not only on description but on content

Retrieval process

- Query specification
  - fuzzy predicates: *similar to*
  - content predicates: *images containing an apple*
  - data type predicates: video, ...
- Query processing and optimization
  - Parsed, compiled, optimized for order of execution
  - Problem: many data types, different processing for each
- Answer
  - Relevance: similarity to query
- Iteration
  - Bad quality, so need to refine

- Multimedia Information Retrieval is quite big in scope:
- Data examples:
  - 2D/3D color/grayscale images: e.g., brain scans, scientific databases of vector fields
  - (2D) video,
  - (1D) voice/music; (1D) time series: e.g., financial/marketing time series; DNA/genomic databases
- Query examples:
  - find photographs with the same color distribution as this
  - find companies whose stock prices move as this one
  - find brain scans with a texture of a tumor
  - Find videos where something happens

Some solutions

- Reduce the problem to search for multi-dimensional points (feature vectors, but vector space is not used)
- Define a distance measure
  - for time series: e.g., Euclidean distance between vectors
  - for images: e.g., color distribution (Euclidean distance); another approach: *mathematical morphology*
  - Other features as vectors
- Often, for search within distance, the vectors are organized in R(*/+)-trees or other spatial trees
- Clustering plays important role

Query types

- All within given distance
  - Find all images that are within 0.05 distance from this one
- Nearest-neighbor
  - Find 5 stocks most similar to IBM
- All pairs within given distance
  - Further: clustering
- Whole object vs. sub-pattern match
  - Find parts of image that are...
  - E.g., in 512 × 512 brain scans, find pieces similar to the given 16 × 16 typical X-ray of a tumor
  - Like *passage retrieval* for text documents

Open problems

- How similarity function can be defined?
- What features of images (video, sound) there are?
- How to better specify the importance of individual features? (*Give me similar houses*: similar = size? color? structure? Architectural style?)
- How to determine the objects in an image?
- Integration with DBMSs and SQL for fast access and rich semantics
    - Integration with XML
    - Ranking: by similarity, taking into account history, profile

Open problems

- Object/event detection (computer vision and pattern recognition)
- Automatic feature selection
- Spatial indexing data structures (more than 1D)
- New types of data.
  - What features to select? How to determine them?
- Mixed-type data (e.g., web pages, or images with sound and description)
- What clustering/IR methods are better suited for what features? (What features for what methods?)
- Similar methods in data mining, ...

Content-based Video Retrieval

- Application
- Implementation
- Experience from TREC video track
  - Feature Extraction Task (High-level Semantics Feature)
  - Manual Retrieval Task (One-run Retrieval)
  - Interactive Retrieval Task (Multiple-run with Feedback)
  - Results & Demo (CMU and IBM)
- Conclusion

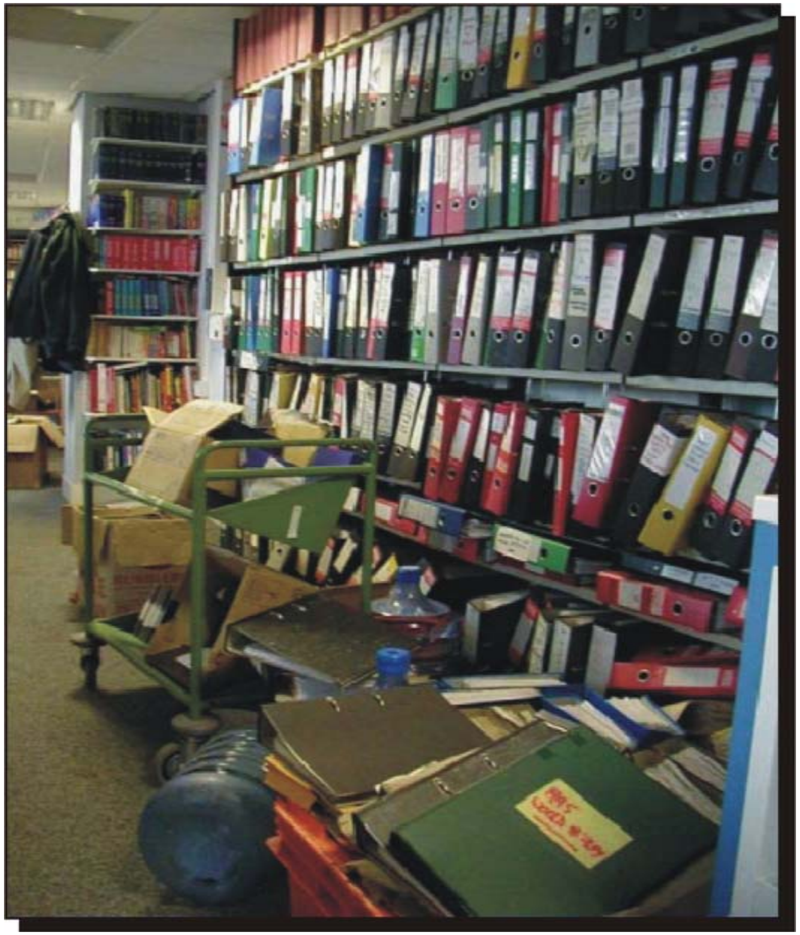
Application

- Increasing demand for visual information retrieval
    - Retrieve useful information from databases
    - Sharing and distributing video data through internet

- Example: BBC
    - BBC archive has +500k queries plus 1M new items … per year;
    - From the BBC …
        - Police car with blue light flashing
        - Government plan to improve reading standards
        - Two shot of Kenneth Clarke and William Hague

Application (Cont.)

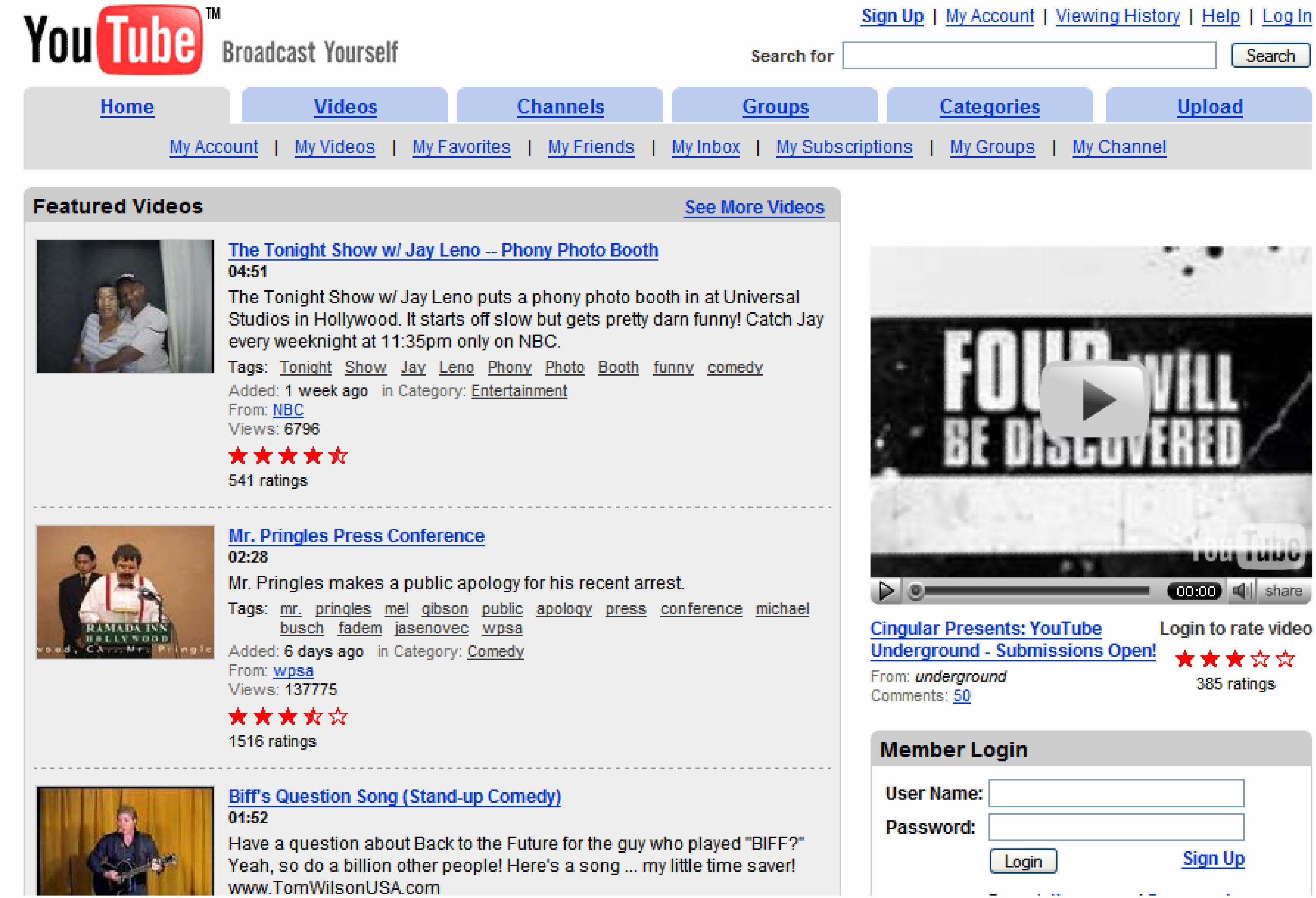
- Past project: ASSAVID in collaboration with BBC sports library:
- Develop automatic annotation systems for sports videos



Application (Cont.)

- Video Surveillance
  - Find where else the person appears
- Experience On-Demand
  - Help to remember previous events
- Provide useful information on traveling
  - Equipment on cars to retrieve useful multimedia information according to your location/preference
- ………
- Video content is plentiful … its now available digitally … we can work on it directly … so it follows

Application (Cont.)



Sign Up | My Account | Viewing History | Help | Log In

Search for [ ] Search

Home | Videos | Channels | Groups | Categories | Upload

My Account | My Videos | My Favorites | My Friends | My Inbox | My Subscriptions | My Groups | My Channel

Featured Videos See More Videos

The Tonight Show w/ Jay Leno -- Phony Photo Booth
04:51
The Tonight Show w/ Jay Leno puts a phony photo booth in at Universal Studios in Hollywood. It starts off slow but gets pretty darn funny! Catch Jay every weeknight at 11:35pm only on NBC.
Tags: Tonight Show Jay Leno Phony Photo Booth funny comedy
Added: 1 week ago in Category: Entertainment
From: NBC
Views: 6796
★★★★⯪
541 ratings

Mr. Pringles Press Conference
02:28
Mr. Pringles makes a public apology for his recent arrest.
Tags: mr. pringles mel gibson public apology press conference michael busch fadem jasenovec wpsa
Added: 6 days ago in Category: Comedy
From: wpsa
Views: 137775
★★★⯪☆
1516 ratings

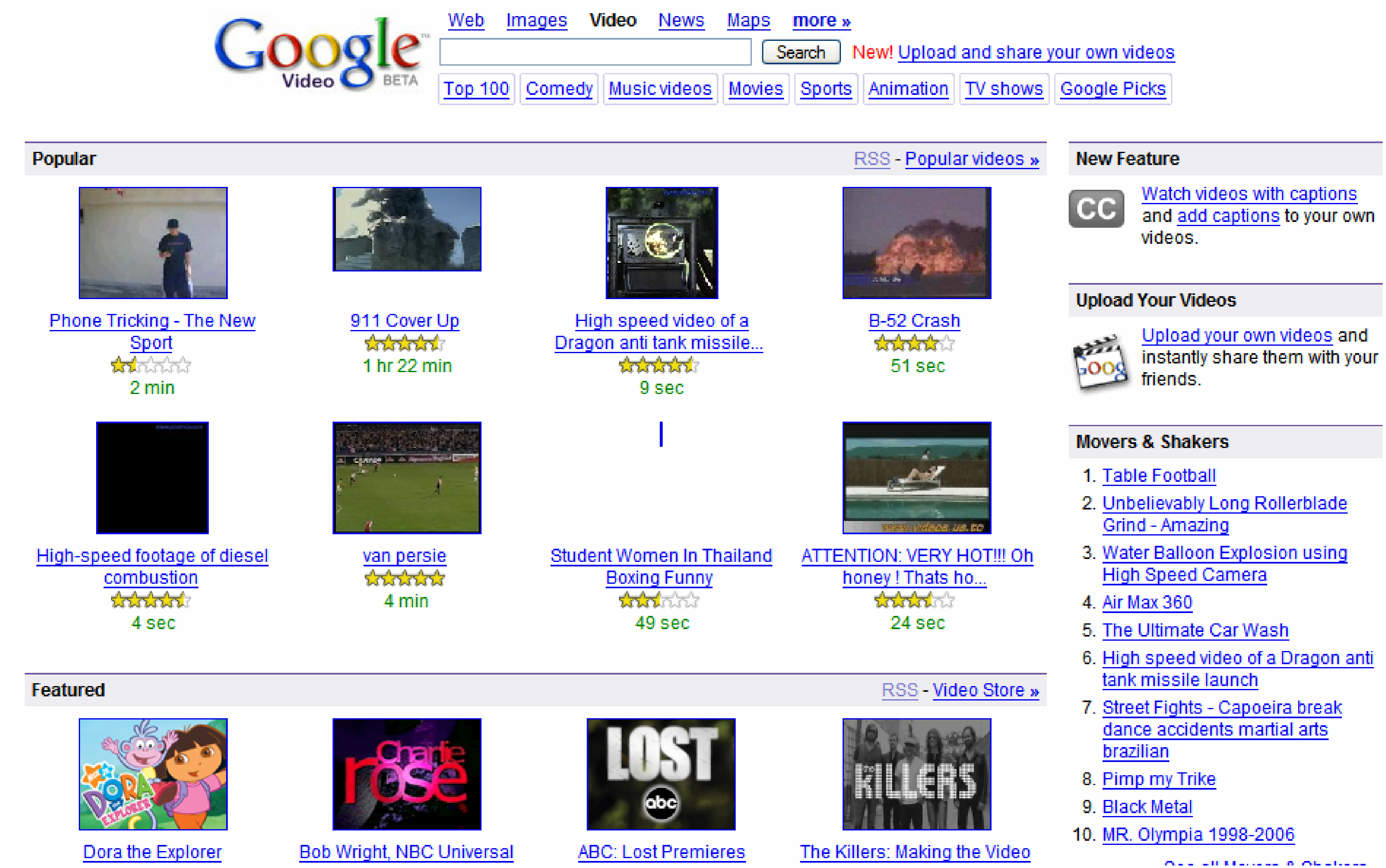
Biff's Question Song (Stand-up Comedy)
01:52
Have a question about Back to the Future for the guy who played "BIFF?" Yeah, so do a billion other people! Here's a song ... my little time saver! www.TomWilsonUSA.com

Cingular Presents: YouTube Underground - Submissions Open!
From: *underground*
Comments: 50

Login to rate video
★★★☆☆
385 ratings

Member Login
User Name: [ ]
Password: [ ]
Login Sign Up

Application (Cont.)



Web Images Video News Maps more »

Search New! Upload and share your own videos

Top 100 | Comedy | Music videos | Movies | Sports | Animation | TV shows | Google Picks

Popular RSS - Popular videos »

Phone Tricking - The New Sport
2 min

911 Cover Up
1 hr 22 min

High speed video of a Dragon anti tank missile...
9 sec

B-52 Crash
51 sec

High-speed footage of diesel combustion
4 sec

van persie
4 min

Student Women In Thailand Boxing Funny
49 sec

ATTENTION: VERY HOT!!! Oh honey ! Thats ho...
24 sec

Featured RSS - Video Store »

Dora the Explorer

Bob Wright, NBC Universal

ABC: Lost Premieres

The Killers: Making the Video

New Feature

CC Watch videos with captions and add captions to your own videos.

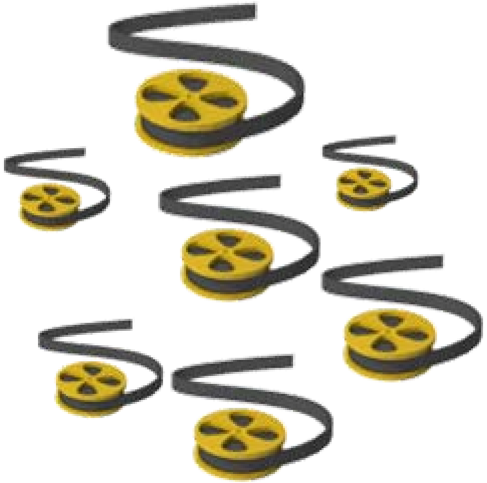
Upload Your Videos

Upload your own videos and instantly share them with your friends.

Movers & Shakers

1. Table Football
2. Unbelievably Long Rollerblade Grind - Amazing
3. Water Balloon Explosion using High Speed Camera
4. Air Max 360
5. The Ultimate Car Wash
6. High speed video of a Dragon anti tank missile launch
7. Street Fights - Capoeira break dance accidents martial arts brazilian
8. Pimp my Trike
9. Black Metal
10. MR. Olympia 1998-2006

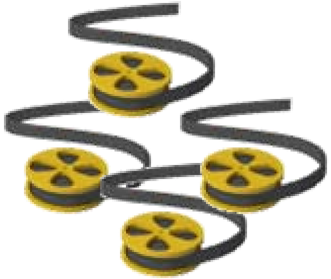
Typical Retrieval Framework



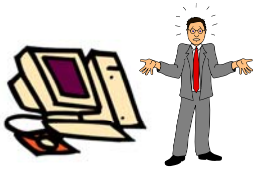
- User : provide query information that represents his information needs
- Database: store a large collection of video data
- Goal: Find the most relevant shots from the database
    - Shots: “paragraph” in video, typically 20 – 40 seconds, which is the basic unit of video retrieval

Bridging the Gap

Video Database



User



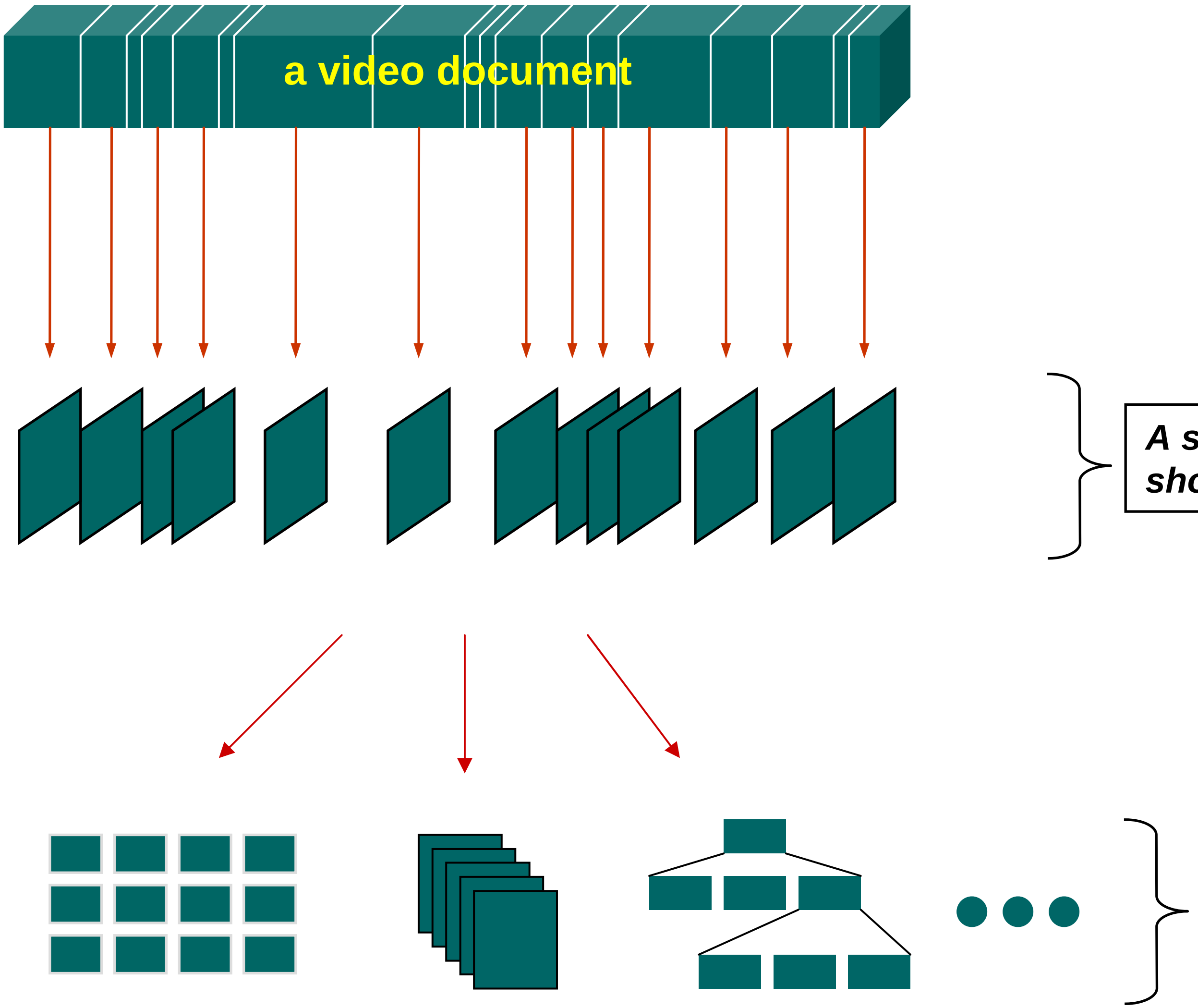
Result

Automatically Structure Video Data

- The first step for video retrieval: Video “programmes” are structured into logical scenes, and physical shots
- If dealing with text, then the structure is obvious:
    - paragraph, section, topic, page, etc.

- All text-based indexing, retrieval, linking, etc. builds upon this structure;
- Automatic shot boundary detection and selection of representative keyframes is usually the first step;

Typical automatic structuring of video

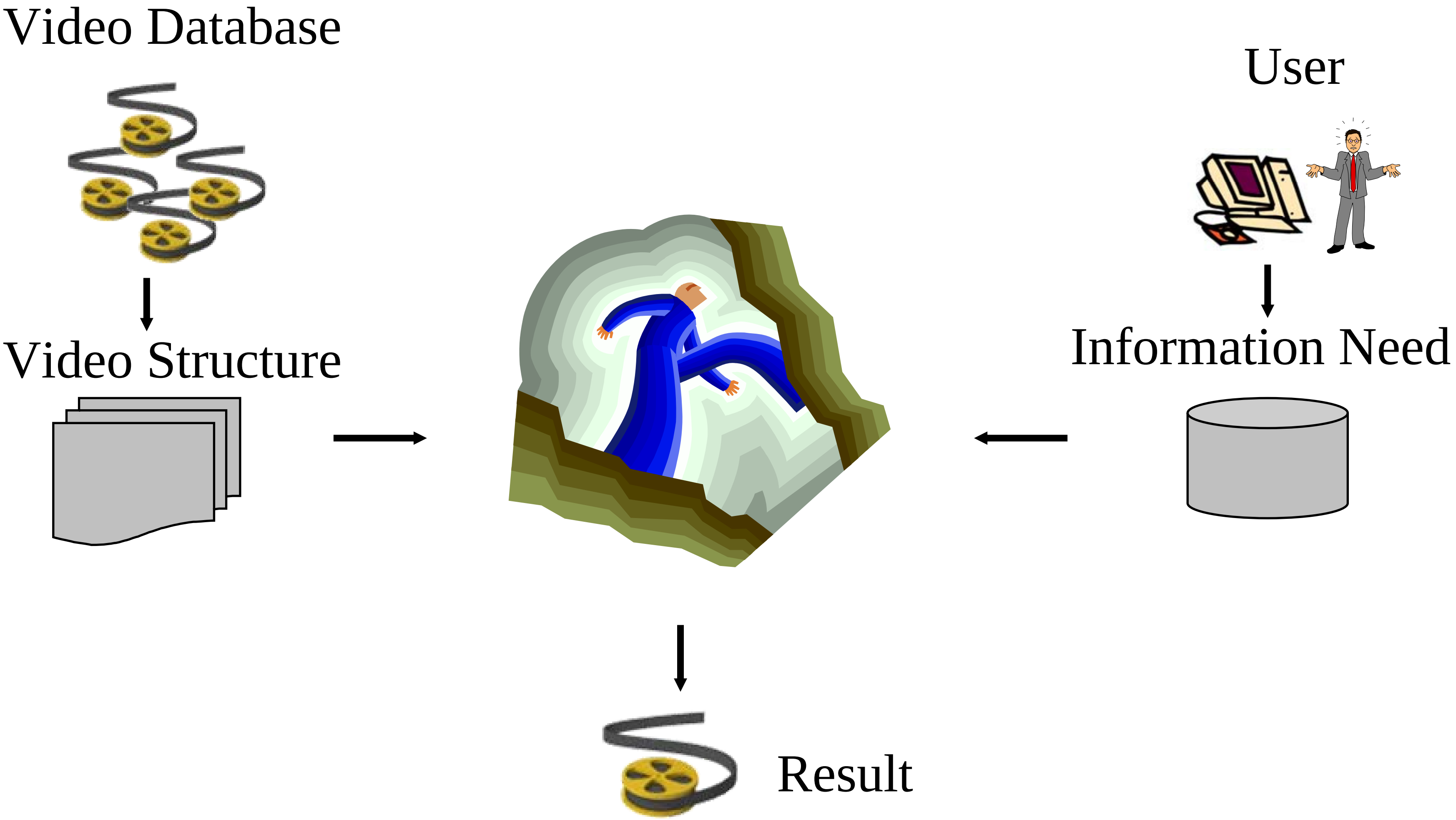


a video document

A set of shots

Keyframe browser combined with transcript or object-based search

Bridging the Gap



Video Database

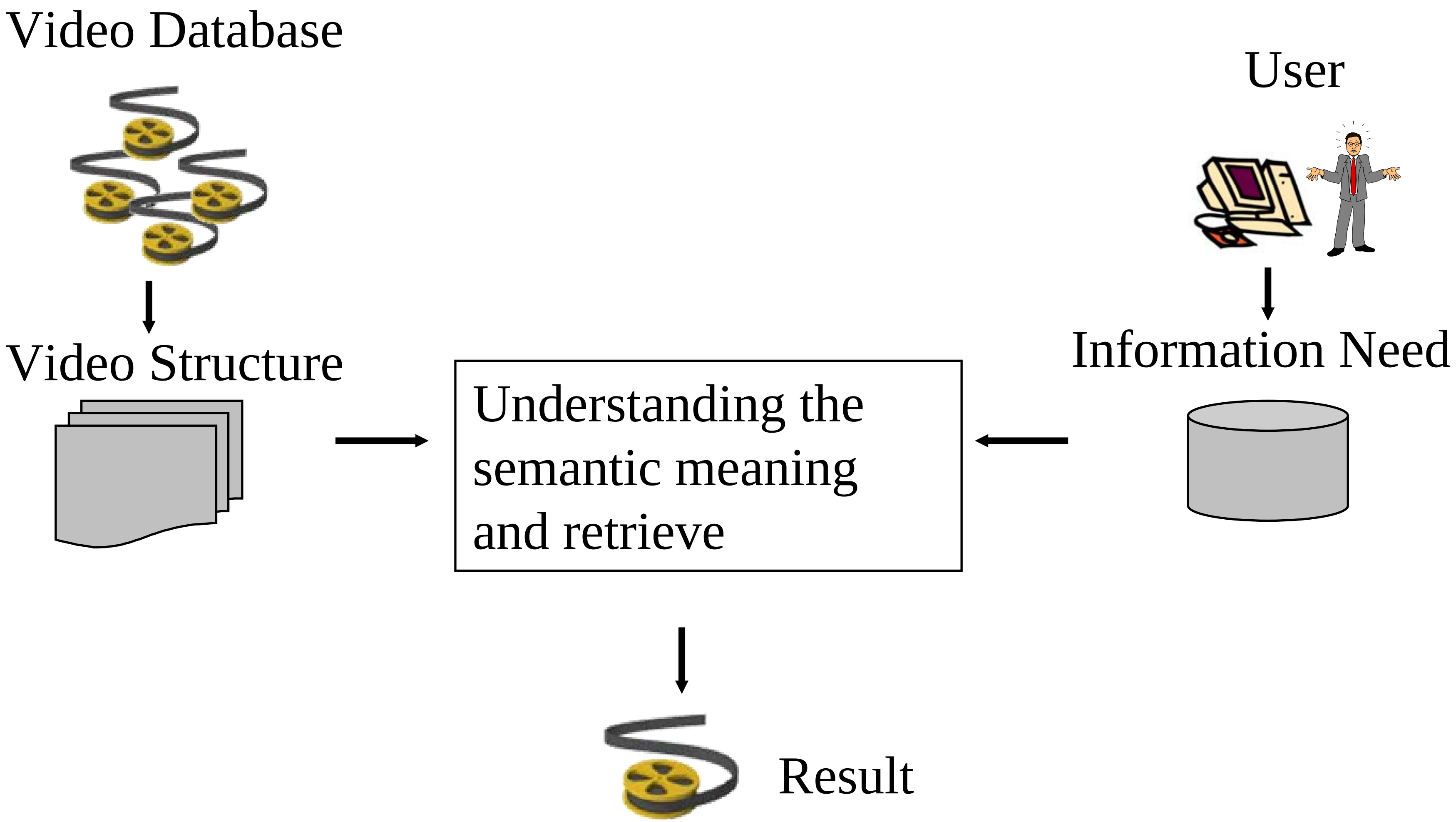
Video Structure

User

Information Need

Result

Ideal solution



Video Database

Video Structure

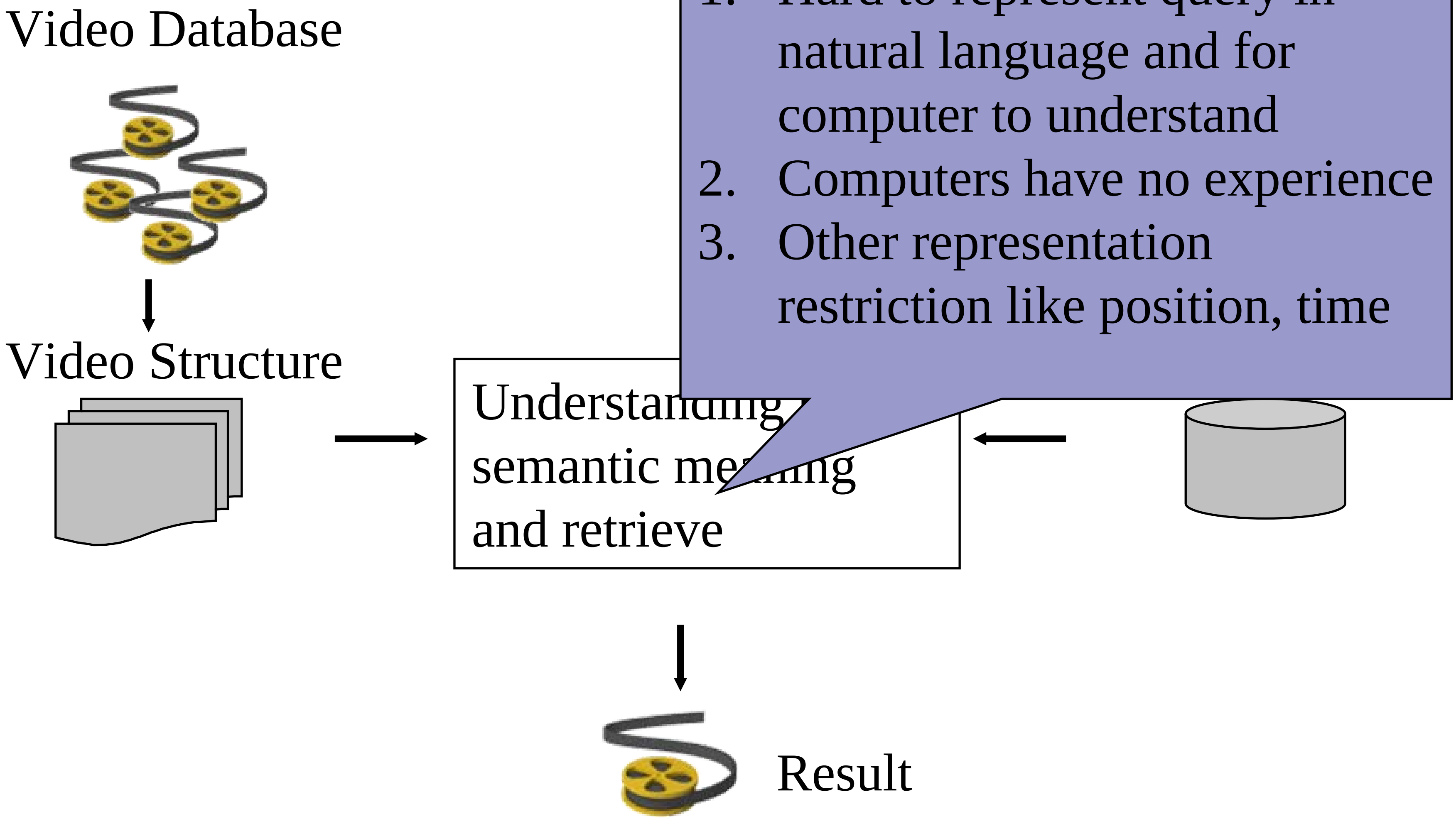
User

Information Need

Understanding the semantic meaning and retrieve

Result

Ideal solution



Video Database

Video Structure

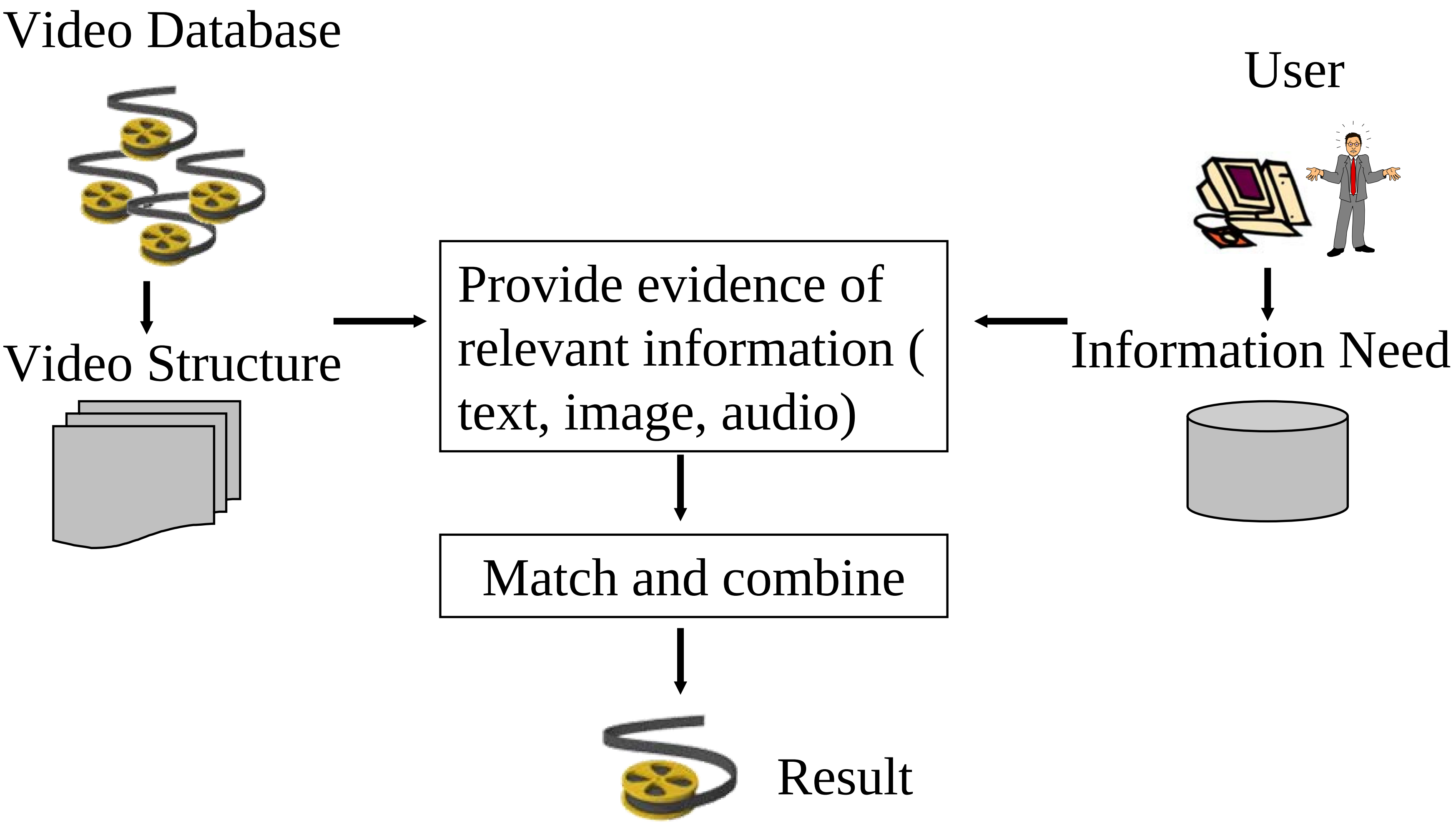
Understanding semantic meaning and retrieve

Result

However,

1. Hard to represent query in natural language and for computer to understand
2. Computers have no experience
3. Other representation restriction like position, time

Alternative Solution



Video Database

User

Video Structure

Provide evidence of relevant information (text, image, audio)

Information Need

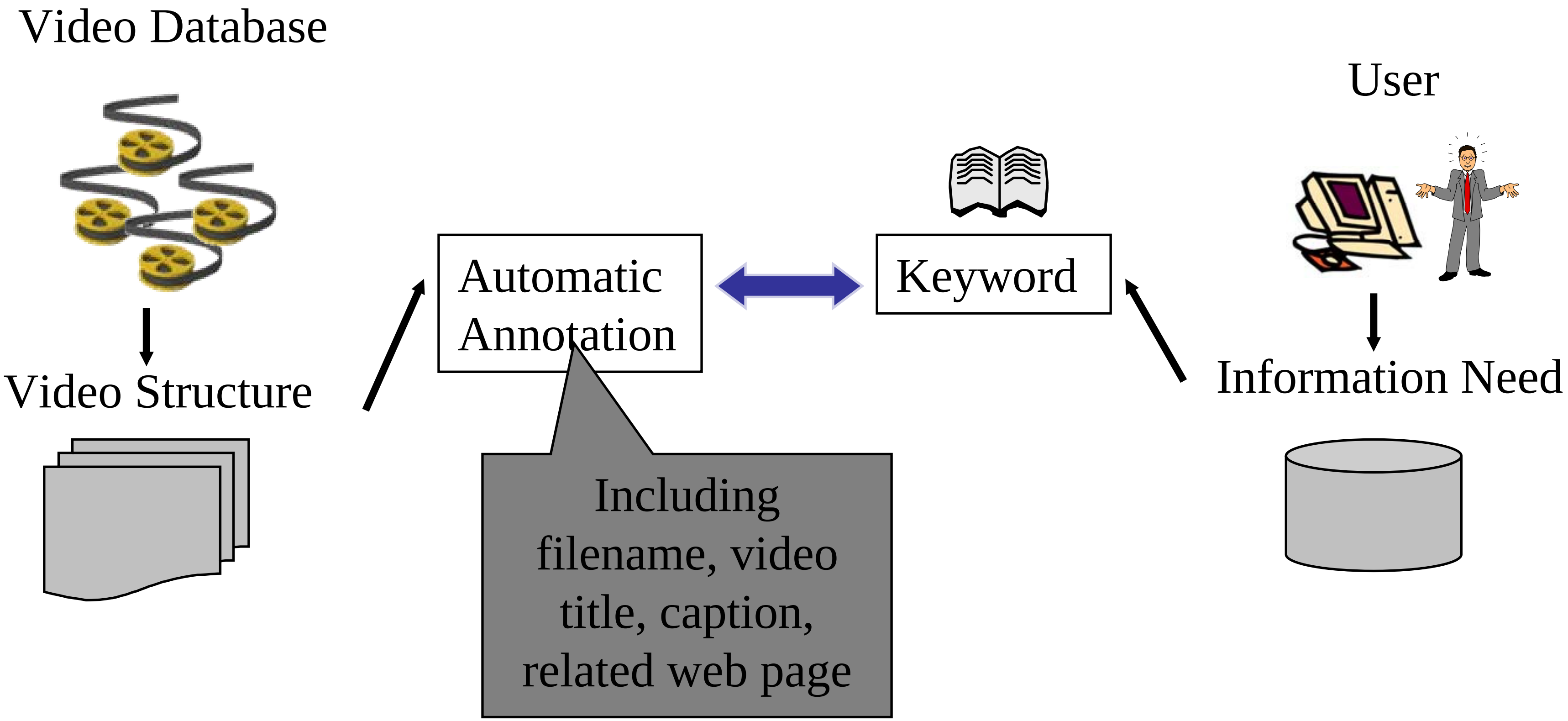
Match and combine

Result

Evidence-based Retrieval System

- General framework for current video retrieval system
- Video retrieval based on the evidence from both users and database, including
  - Text information
  - Image information
  - Motion information
  - Audio information
- Return a relevant score for each evidence
- Combination of the scores

Keyword-based System



Video Database

Video Structure

Automatic Annotation

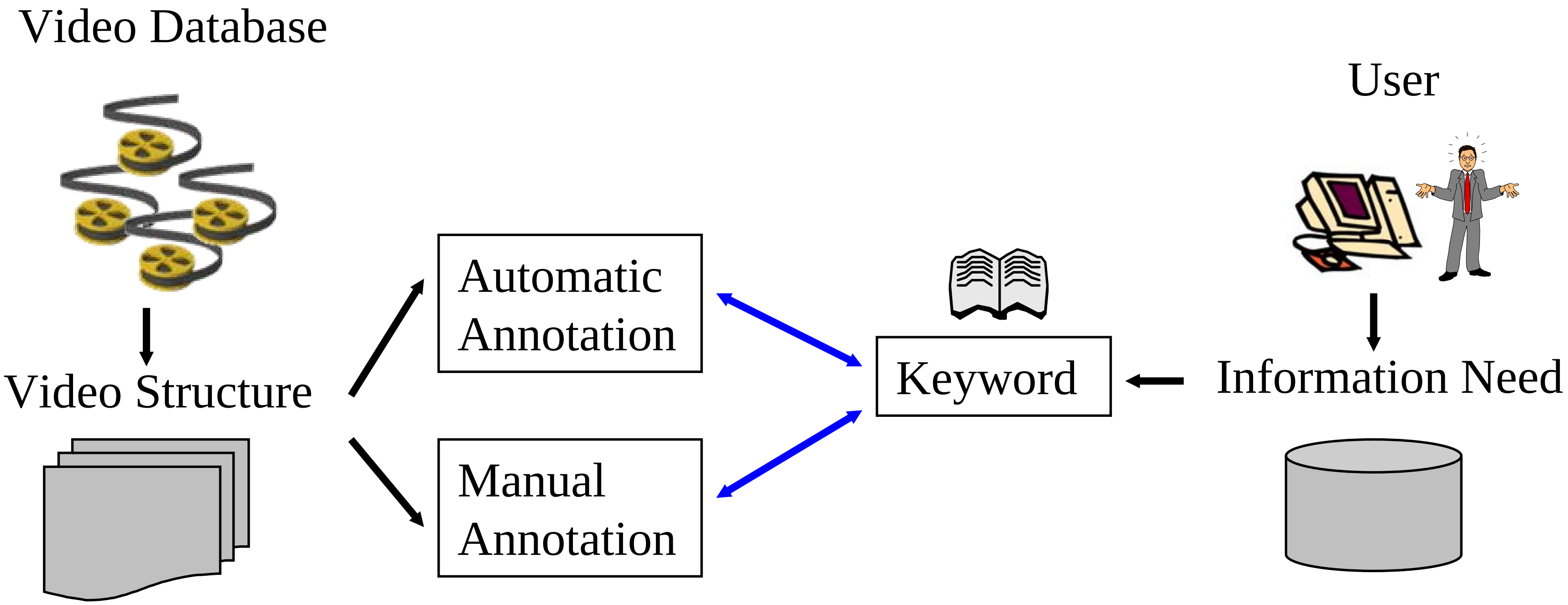
Including filename, video title, caption, related web page

Keyword

User

Information Need

Keyword-based System


Video Database

Video Structure

Automatic Annotation

Manual Annotation

Keyword

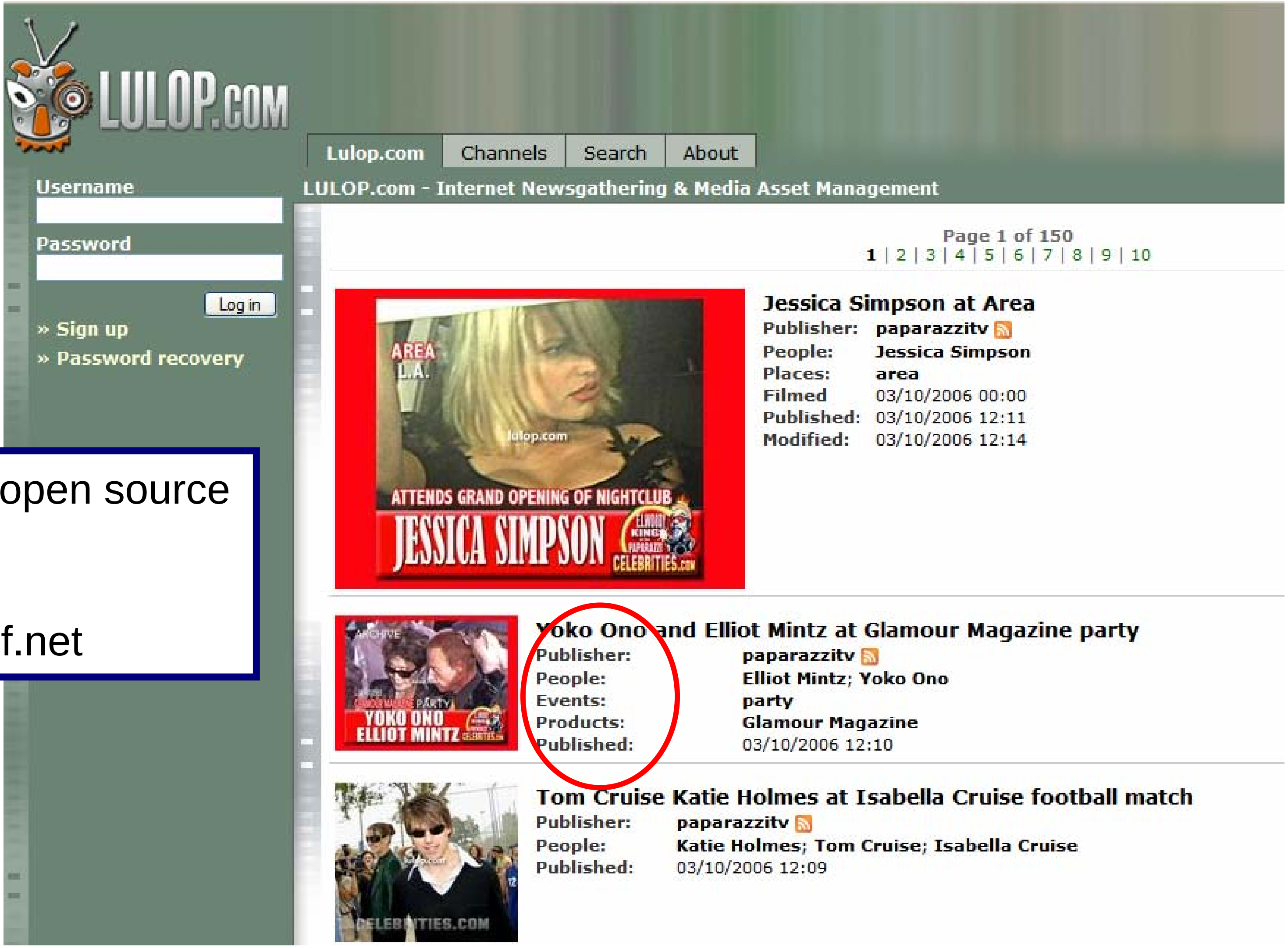
User

Information Need

Manual Annotation

- Manually creating annotation/keywords for image / video data
- Examples: Gettyimages.com (image retrieval)
- Pros:
    - Represent the semantic meaning of video
- Cons
    - Time-consuming, labor-intensive
    - Keyword is not enough to represent information need

Manual annotation using metadata



LULOP.com - Internet Newsgathering & Media Asset Management

Page 1 of 150
1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10

Jessica Simpson at Area

Publisher:	paparazzitv
People:	Jessica Simpson
Places:	area
Filmed	03/10/2006 00:00
Published:	03/10/2006 12:11
Modified:	03/10/2006 12:14

Yoko Ono and Elliot Mintz at Glamour Magazine party

Publisher:	paparazzitv
People:	Elliot Mintz; Yoko Ono
Events:	party
Products:	Glamour Magazine
Published:	03/10/2006 12:10

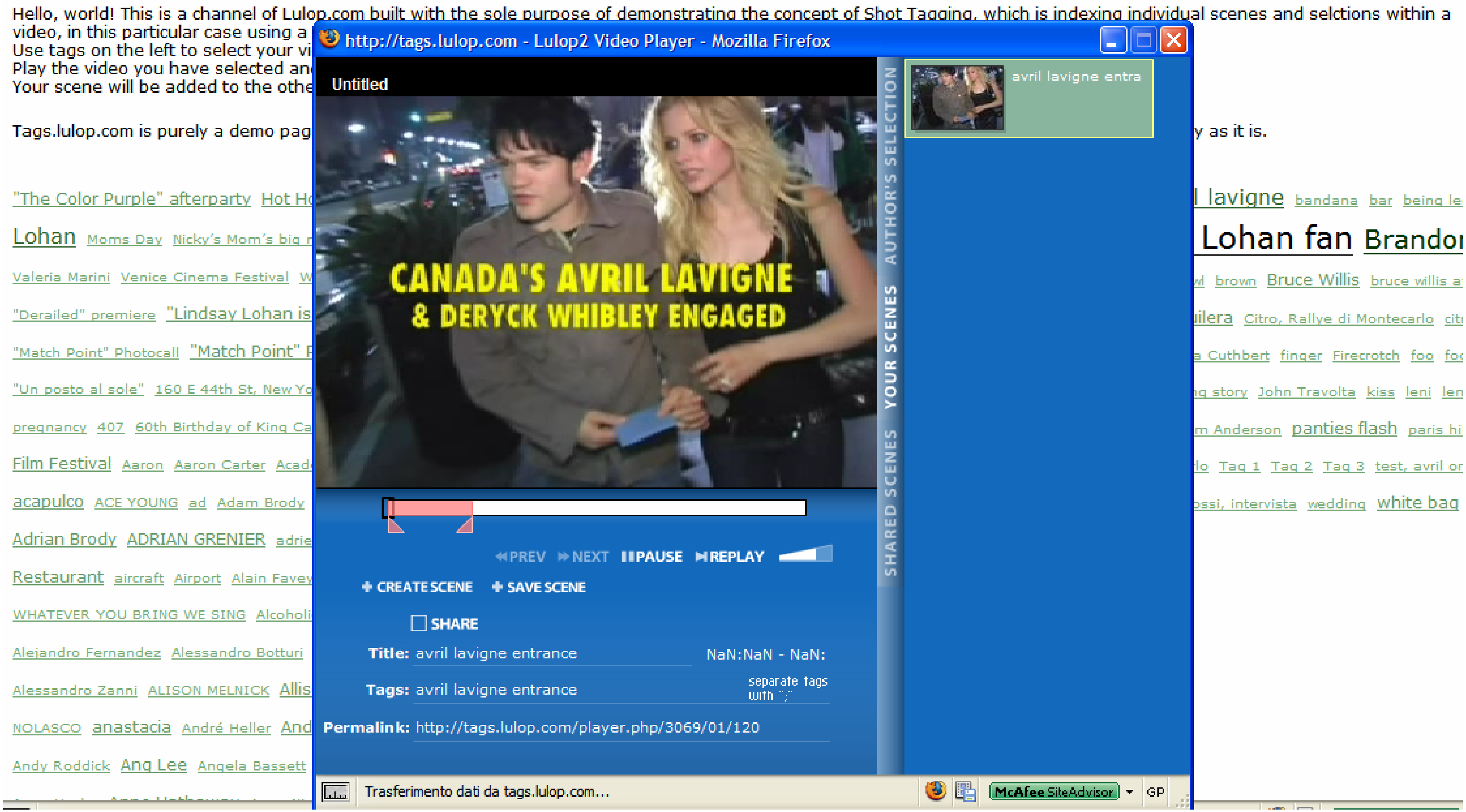
Tom Cruise Katie Holmes at Isabella Cruise football match

Publisher:	paparazzitv
People:	Katie Holmes; Tom Cruise; Isabella Cruise
Published:	03/10/2006 12:09

Try the open source version:

lulop2.sf.net

Tagging



Hello, world! This is a channel of Lulop.com built with the sole purpose of demonstrating the concept of Shot Tagging, which is indexing individual scenes and selctions within a
video, in this particular case using a
Use tags on the left to select your vi
Play the video you have selected an
Your scene will be added to the othe

Tags.lulop.com is purely a demo pag
y as it is.

"The Color Purple" afterparty Hot H
l lavigne bandana bar being le
Lohan Moms Day Nicky's Mom's big r
Lohan fan Brando
Valeria Marini Venice Cinema Festival W
wl brown Bruce Willis bruce willis a
"Derailed" premiere "Lindsay Lohan is
uilera Citro, Rallye di Montecarlo cit
"Match Point" Photocall "Match Point" F
a Cuthbert finger Firecrotch foo fo
"Un posto al sole" 160 E 44th St, New Yo
ng story John Travolta kiss leni len
pregnancy 407 60th Birthday of King Ca
m Anderson panties flash paris hi
Film Festival Aaron Aaron Carter Acad
lo Tag 1 Tag 2 Tag 3 test, avril or
acapulco ACE YOUNG ad Adam Brody
ossi, intervista wedding white bag
Adrian Brody ADRIAN GRENIER adrie
Restaurant aircraft Airport Alain Favey
WHATEVER YOU BRING WE SING Alcoholi
Alejandro Fernandez Alessandro Botturi
Alessandro Zanni ALISON MELNICK Allis
NOLASCO anastacia André Heller And
Andy Roddick Ang Lee Angela Bassett

http://tags.lulop.com - Lulop2 Video Player - Mozilla Firefox

Untitled

Title: avril lavigne entrance

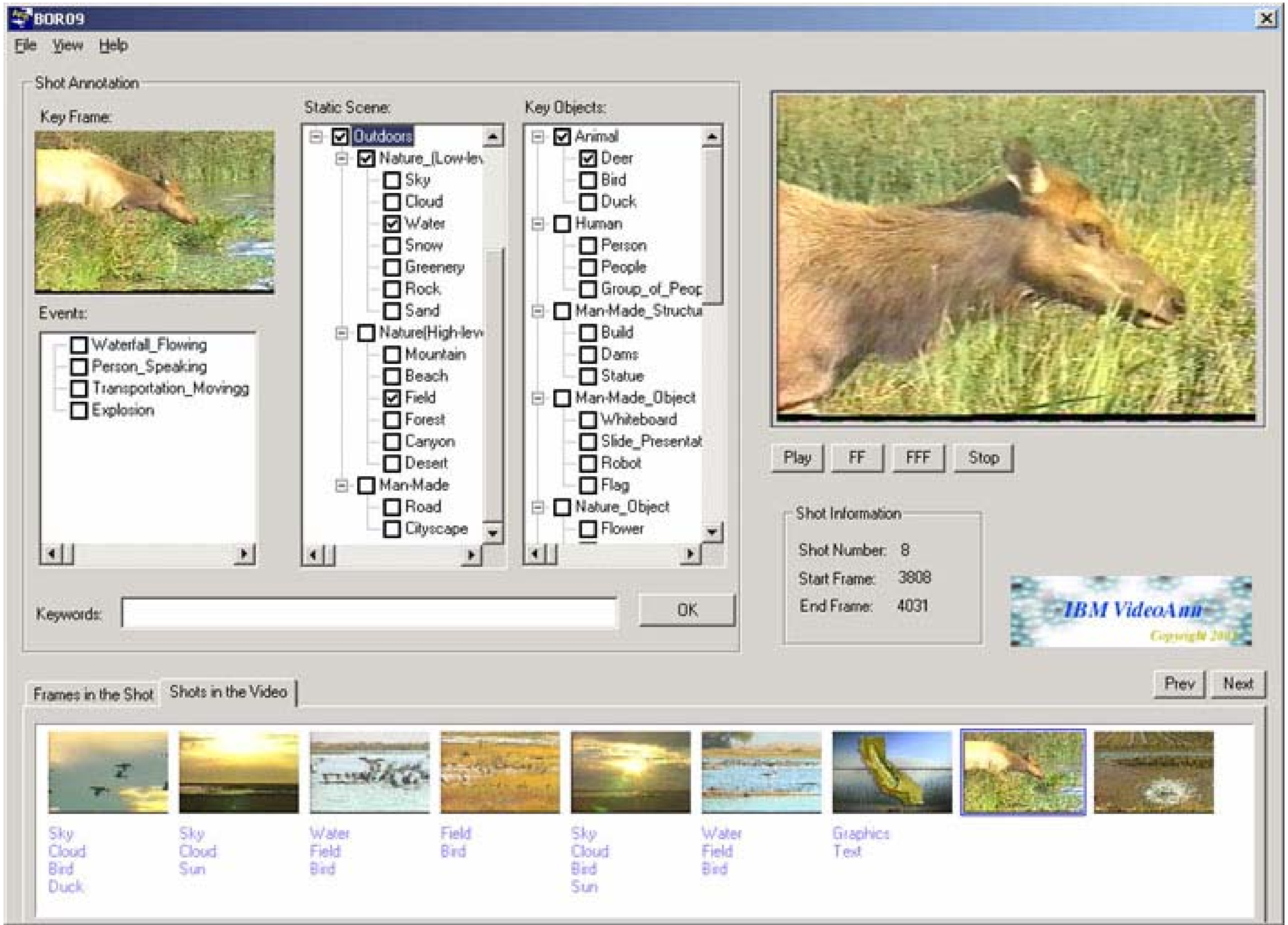
Tags: avril lavigne entrance

Permalink: http://tags.lulop.com/player.php/3069/01/120

Trasferimento dati da tags.lulop.com...

Manual annotation using taxonomy

(sort of…)



Shot Annotation

Key Frame:

Events:

- ☐ Waterfall_Flowing
- ☐ Person_Speaking
- ☐ Transportation_Movingg
- ☐ Explosion

Static Scene:

- ☑ Outdoors
  - ☑ Nature_(Low-lev
    - ☐ Sky
    - ☐ Cloud
    - ☑ Water
    - ☐ Snow
    - ☐ Greenery
    - ☐ Rock
    - ☐ Sand
  - ☐ Nature(High-lev
    - ☐ Mountain
    - ☐ Beach
    - ☑ Field
    - ☐ Forest
    - ☐ Canyon
    - ☐ Desert
  - ☐ Man-Made
    - ☐ Road
    - ☐ Cityscape

Key Objects:

- ☑ Animal
  - ☑ Deer
  - ☐ Bird
  - ☐ Duck
- ☐ Human
  - ☐ Person
  - ☐ People
  - ☐ Group_of_Peop
- ☐ Man-Made_Structu
  - ☐ Build
  - ☐ Dams
  - ☐ Statue
- ☐ Man-Made_Object
  - ☐ Whiteboard
  - ☐ Slide_Presentat
  - ☐ Robot
  - ☐ Flag
- ☐ Nature_Object
  - ☐ Flower

Keywords:

Shot Information

Shot Number: 8
Start Frame: 3808
End Frame: 4031

Sky
Cloud
Bird
Duck

Sky
Cloud
Sun

Water
Field
Bird

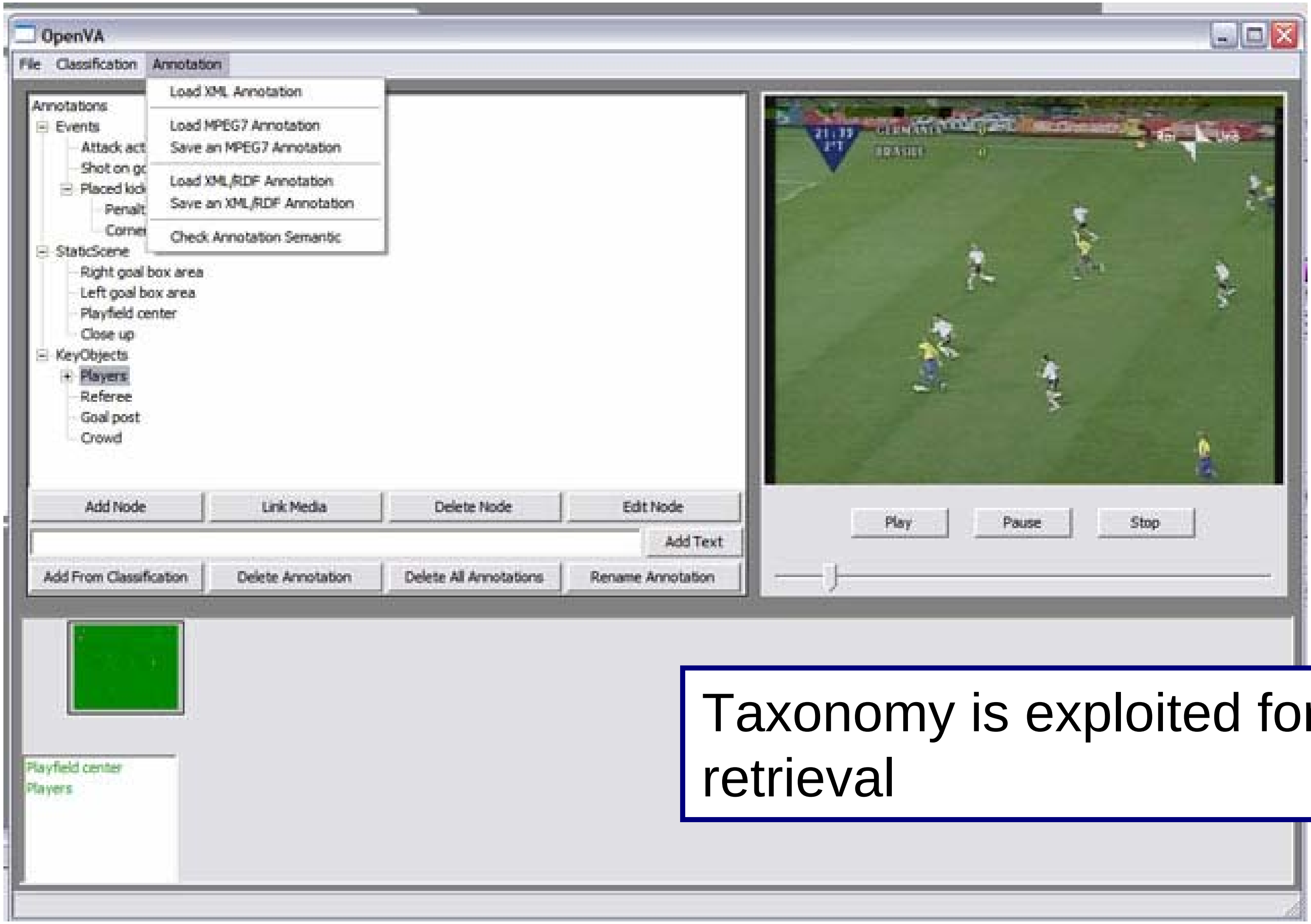
Field
Bird

Sky
Cloud
Bird
Sun

Water
Field
Bird

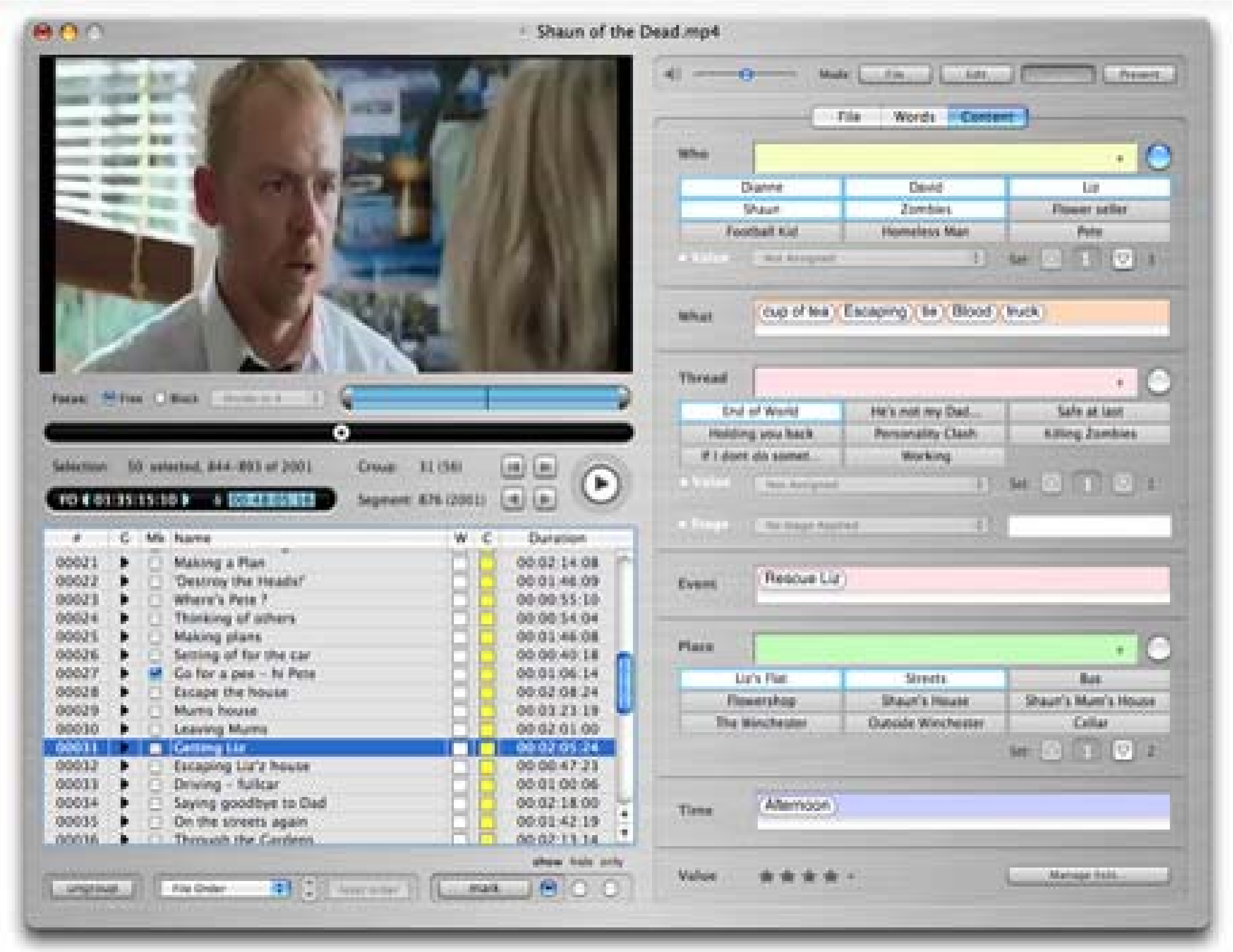
Graphics
Text

Manual annotation using taxonomy



Taxonomy is exploited for retrieval

Manual annotation using 4 Ws



Selection: 50 selected, 844-893 of 2001 Group: 11 (50)

Segment: 876 (2001)

#	C	Mk	Name	W	C	Duration
00021	▶	□	Making a Plan	□	🟨	00:02:14:08
00022	▶	□	'Destroy the Heads!'	□	🟨	00:01:46:09
00023	▶	□	Where's Pete ?	□	🟨	00:00:55:10
00024	▶	□	Thinking of others	□	🟨	00:00:54:04
00025	▶	□	Making plans	□	🟨	00:01:46:08
00026	▶	□	Setting of for the car	□	🟨	00:00:40:18
00027	▶	☑	Go for a piss - hi Pete	□	🟨	00:01:06:14
00028	▶	□	Escape the house	□	🟨	00:02:08:24
00029	▶	□	Mums house	□	🟨	00:03:23:19
00030	▶	□	Leaving Mums	□	🟨	00:02:01:00
00031	▶	□	Getting Liz	□	🟨	00:02:05:24
00032	▶	□	Escaping Liz's house	□	🟨	00:00:47:23
00033	▶	□	Driving - fullcar	□	🟨	00:01:00:06
00034	▶	□	Saying goodbye to Dad	□	🟨	00:02:18:00
00035	▶	□	On the streets again	□	🟨	00:01:42:19
00036	▶	□	Through the Gardens	□	🟨	00:02:13:14

Who: Dianne, David, Liz, Shaun, Zombies, Flower seller, Football Kid, Homeless Man, Pete

What: cup of tea, Escaping, lie, Blood, truck

Thread: End of World, He's not my Dad..., Safe at last, Holding you back, Personality Clash, Killing Zombies, If I dont do somet..., Working

Event: Rescue Liz

Place: Liz's Flat, Streets, Bus, Flowershop, Shaun's House, Shaun's Mum's House, The Winchester, Outside Winchester, Cellar

Time: Afternoon

Value: ★★★★

Who

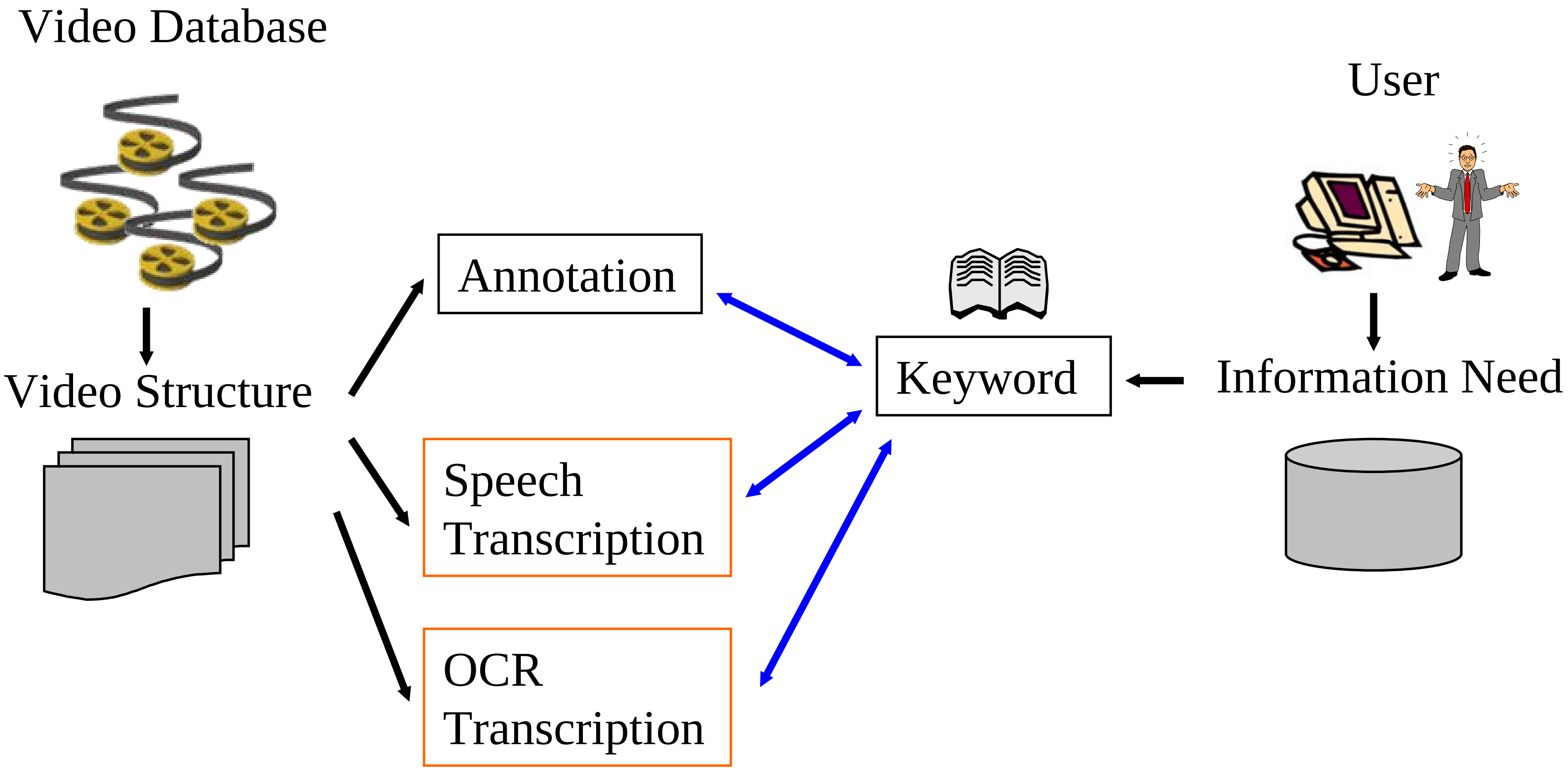
What

Frameline uses MPEG-47: i.e. MPEG-4 with MPEG-7 stream

Where

When

Speech and OCR transcription



Video Database

Video Structure

Annotation

Speech Transcription

OCR Transcription

Keyword

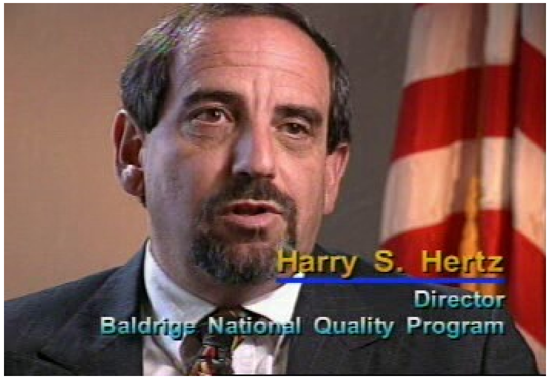
User

Information Need

Query using speech/OCR information

Query:

Find pictures of Harry Hertz, Director of the National Quality Program, NIST



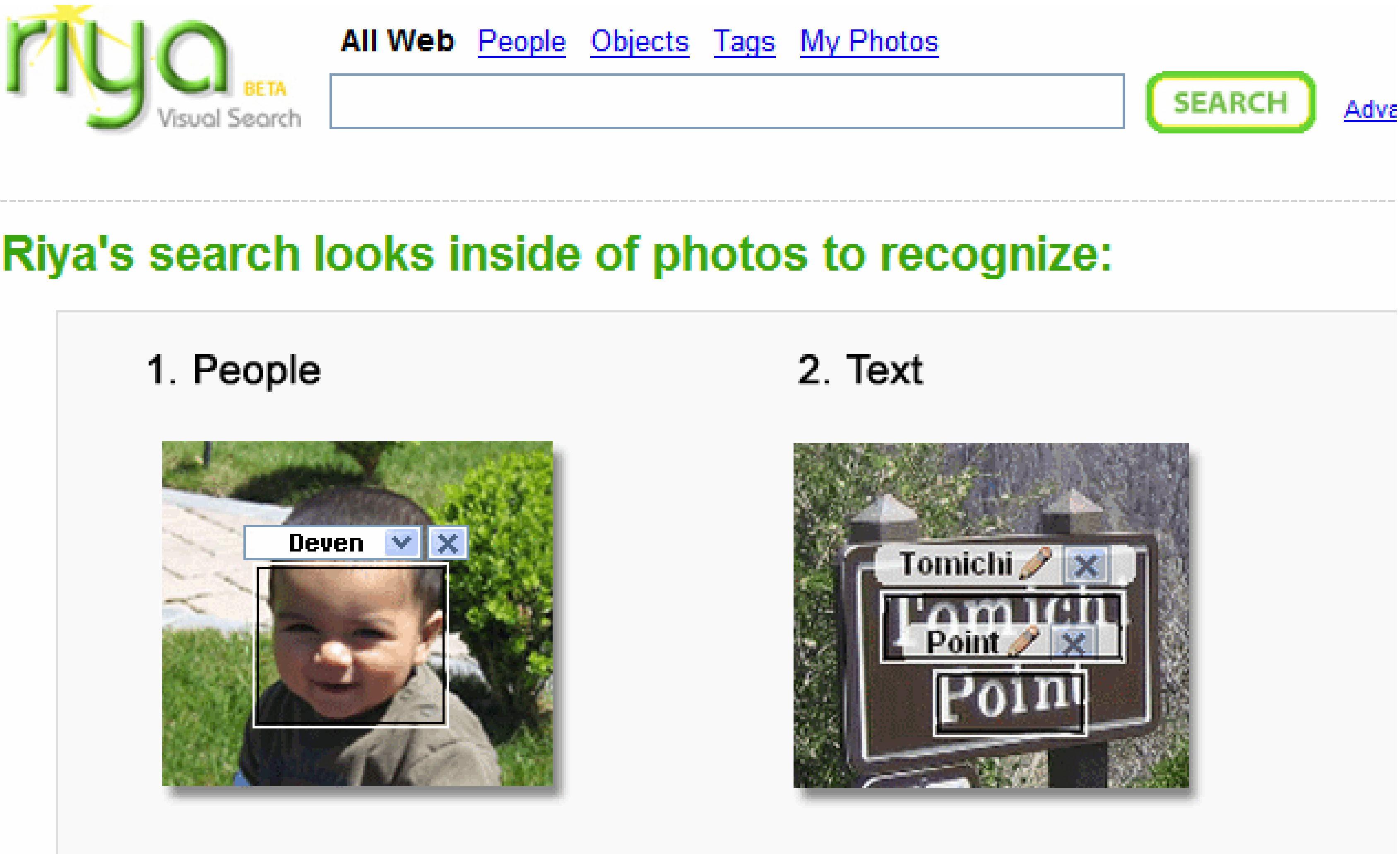
Speech:

We’re looking for people that have a broad range of expertise that have business knowledge that have knowledge on quality management on quality improvement and in particular …

OCR:

```
H,arry Hertz a Director aro 7 wa-
,i,,ty Program
,Harry Hertz a Director
```

Automatic face and OCR recognition

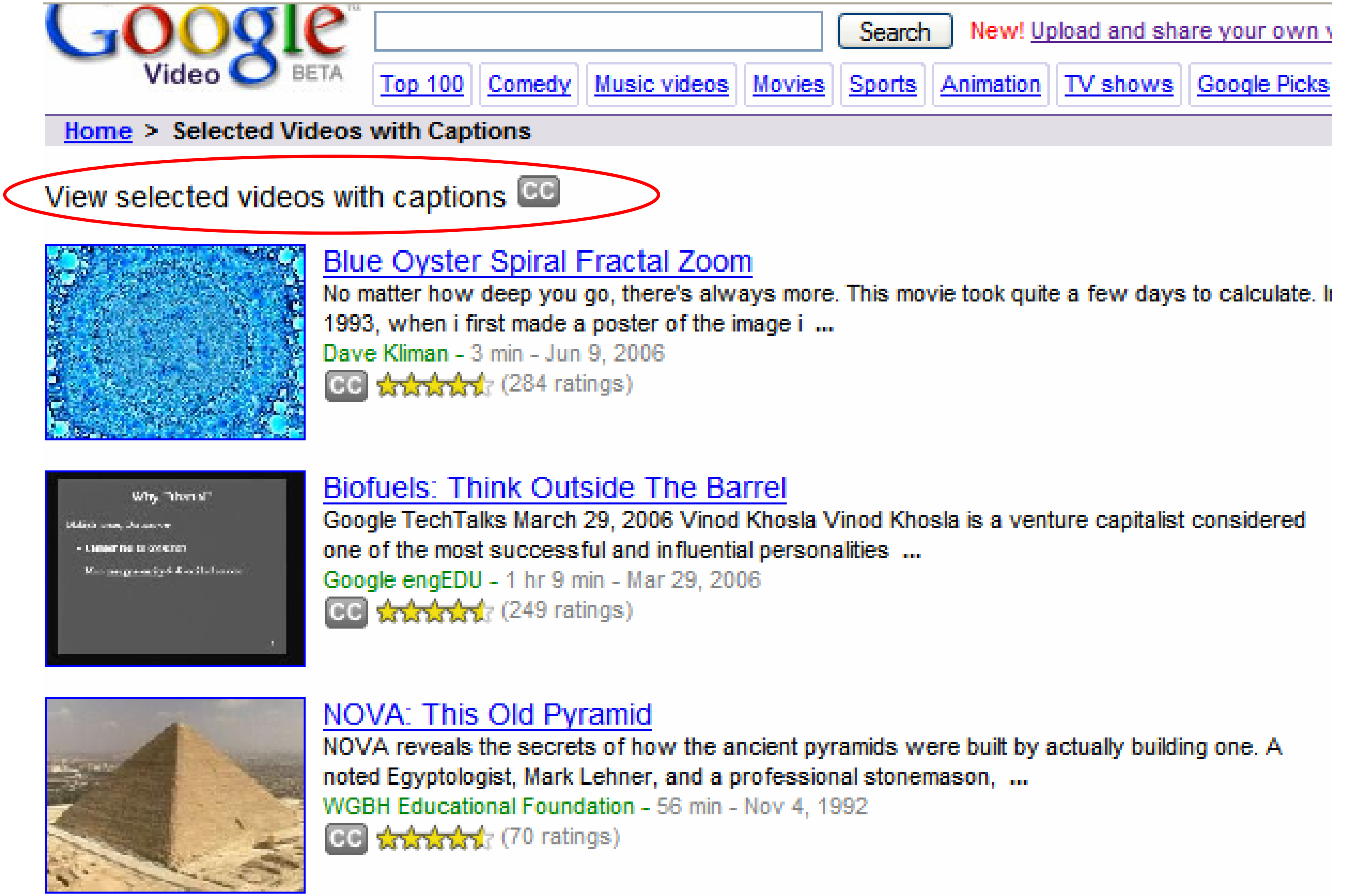


Riya's search looks inside of photos to recognize:

1. People

2. Text

Closed captions



Home > Selected Videos with Captions

View selected videos with captions CC

Blue Oyster Spiral Fractal Zoom

No matter how deep you go, there's always more. This movie took quite a few days to calculate. I 1993, when i first made a poster of the image i ...

Dave Kliman - 3 min - Jun 9, 2006

CC ★★★★☆ (284 ratings)

Biofuels: Think Outside The Barrel

Google TechTalks March 29, 2006 Vinod Khosla Vinod Khosla is a venture capitalist considered one of the most successful and influential personalities ...

Google engEDU - 1 hr 9 min - Mar 29, 2006

CC ★★★★☆ (249 ratings)

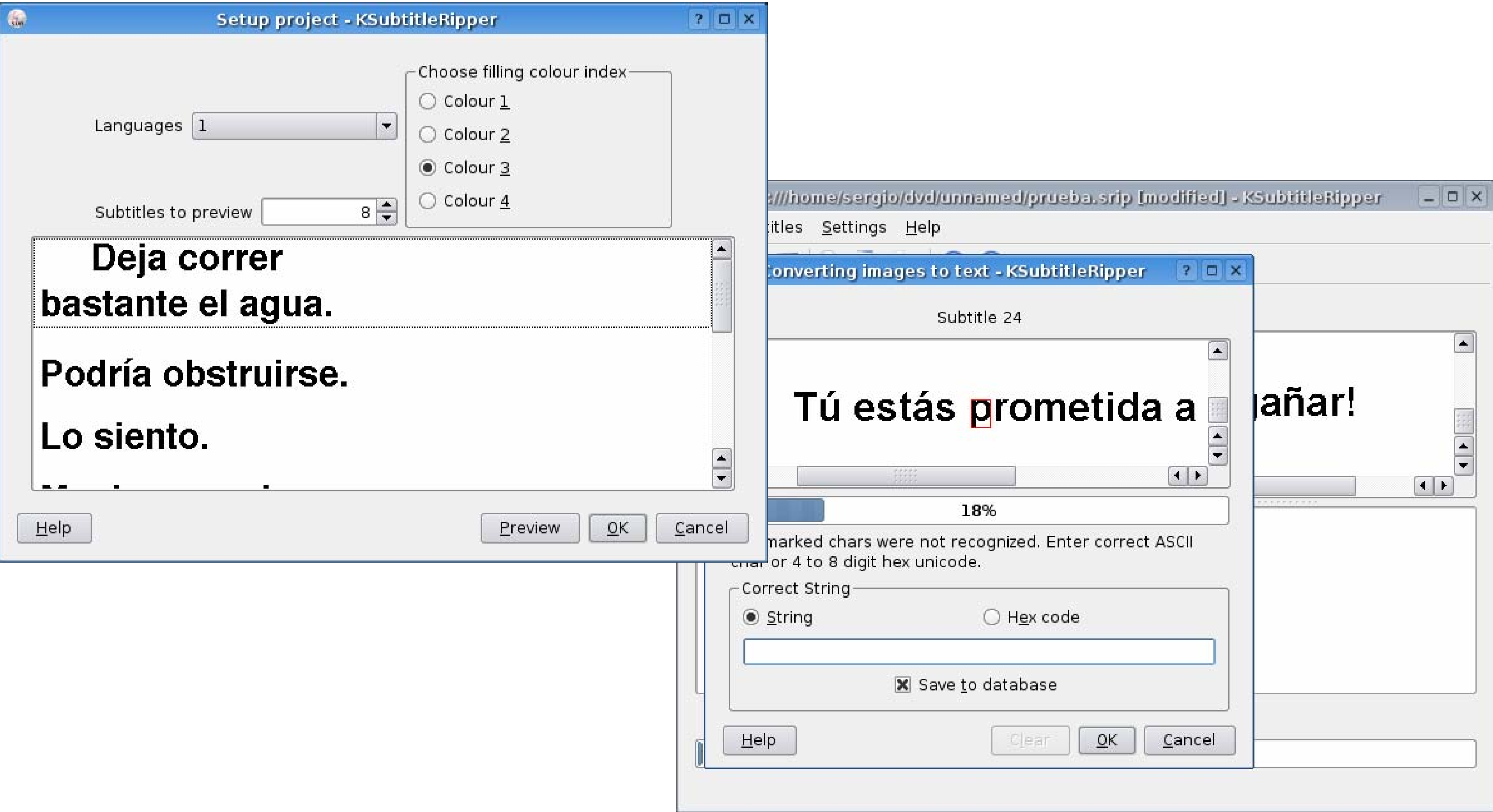
NOVA: This Old Pyramid

NOVA reveals the secrets of how the ancient pyramids were built by actually building one. A noted Egyptologist, Mark Lehner, and a professional stonemason, ...

WGBH Educational Foundation - 56 min - Nov 4, 1992

CC ★★★★☆ (70 ratings)

DVD subtitle ripping and OCR



Setup project - KSubtitleRipper

Languages: 1

Subtitles to preview: 8

Choose filling colour index
- ○ Colour 1
- ○ Colour 2
- ● Colour 3
- ○ Colour 4

Deja correr
bastante el agua.

Podría obstruirse.

Lo siento.

:///home/sergio/dvd/unnamed/prueba.srip [modified] - KSubtitleRipper

Converting images to text - KSubtitleRipper

Subtitle 24

Tú estás prometida a

añar!

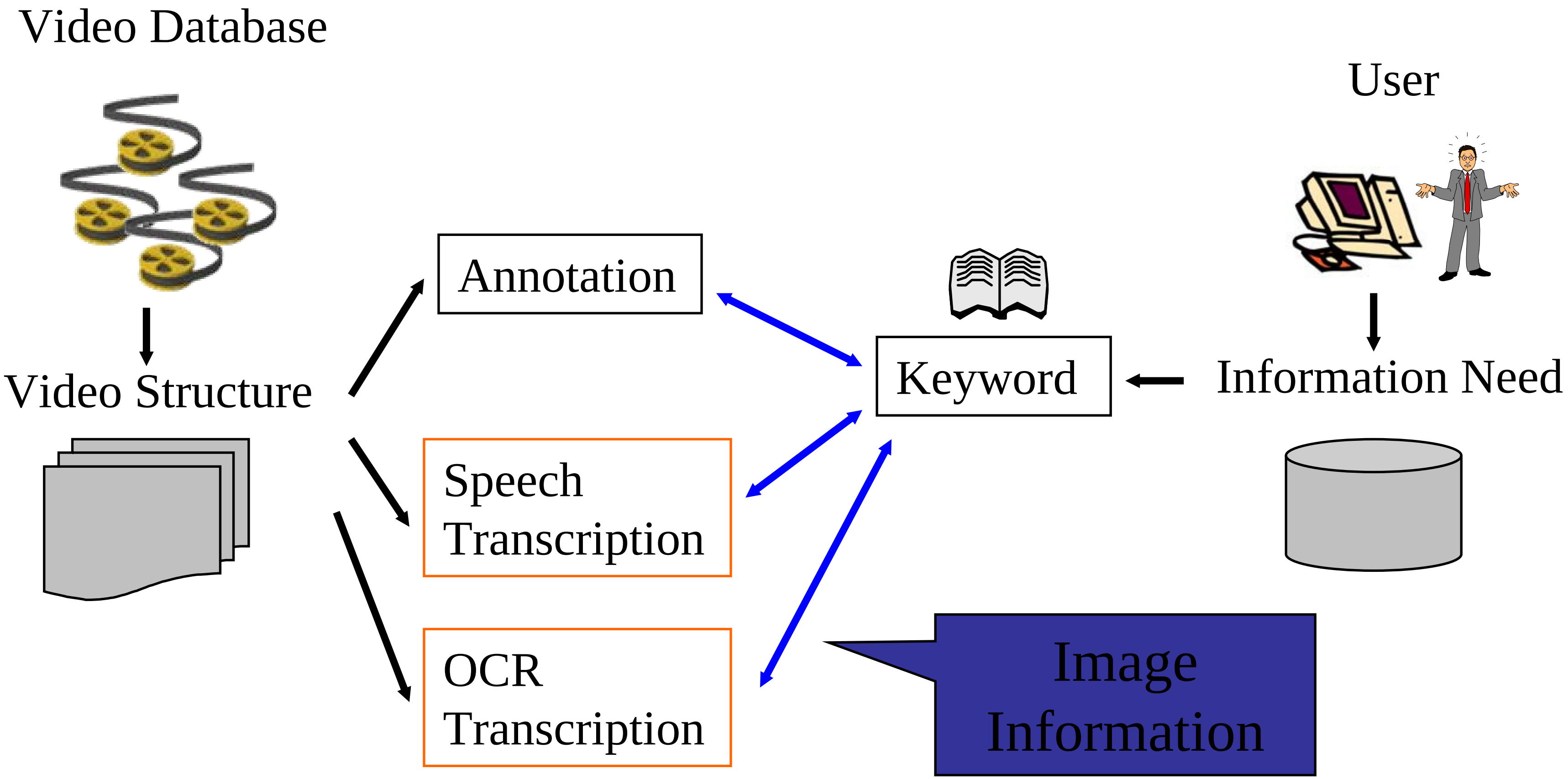
18%

marked chars were not recognized. Enter correct ASCII char or 4 to 8 digit hex unicode.

Correct String
- ● String
- ○ Hex code

☒ Save to database

What we lack?



Video Database

Video Structure

Annotation

Speech Transcription

OCR Transcription

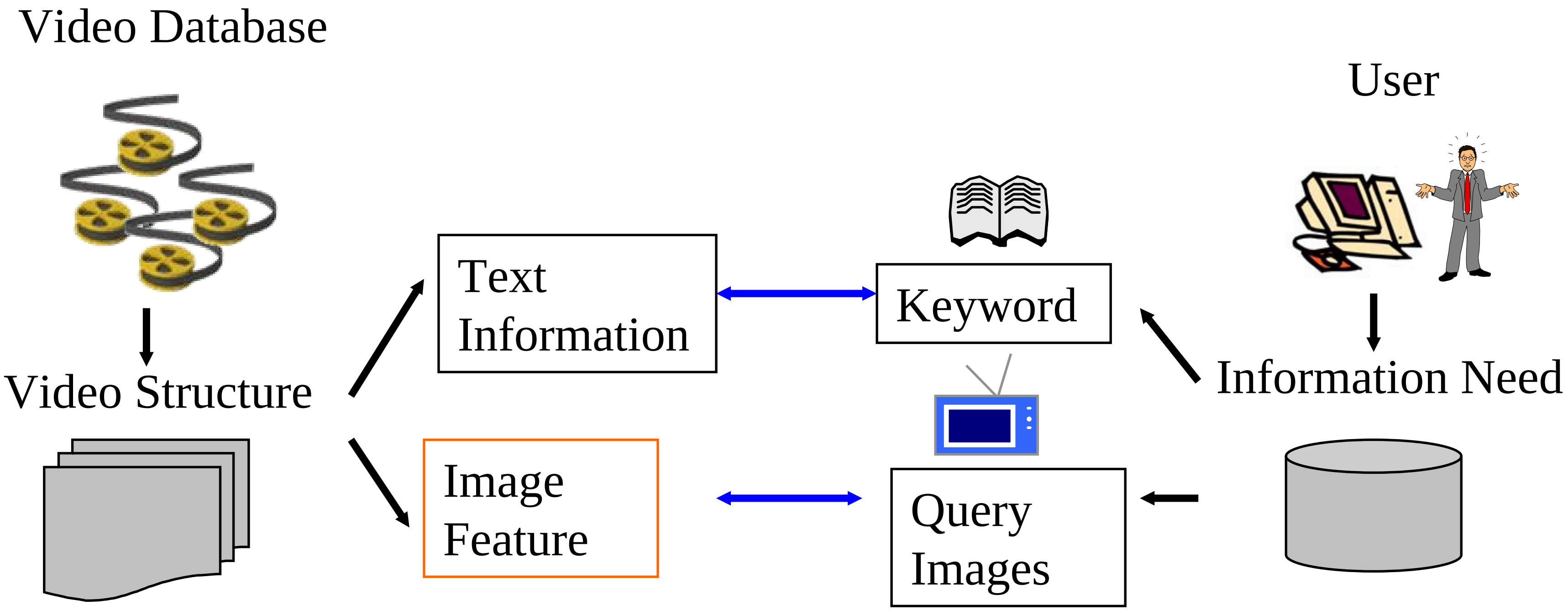
Keyword

User

Information Need

Image Information

Image-based Retrieval



Video Database

User

Text Information

Keyword

Video Structure

Information Need

Image Feature

Query Images

Global Low-level Image Feature

- Color-based Feature
  - Color Histogram
  - Color Percentage
  - Color Correlogram
  - Color Moments
- Texture-based Feature
  - Gabor Filter
  - Wavelet
- Shape/Structure Feature

Regional Low-level Image Feature

- Segmentation into objects – hard problem !
- Extract low-level features from each regions

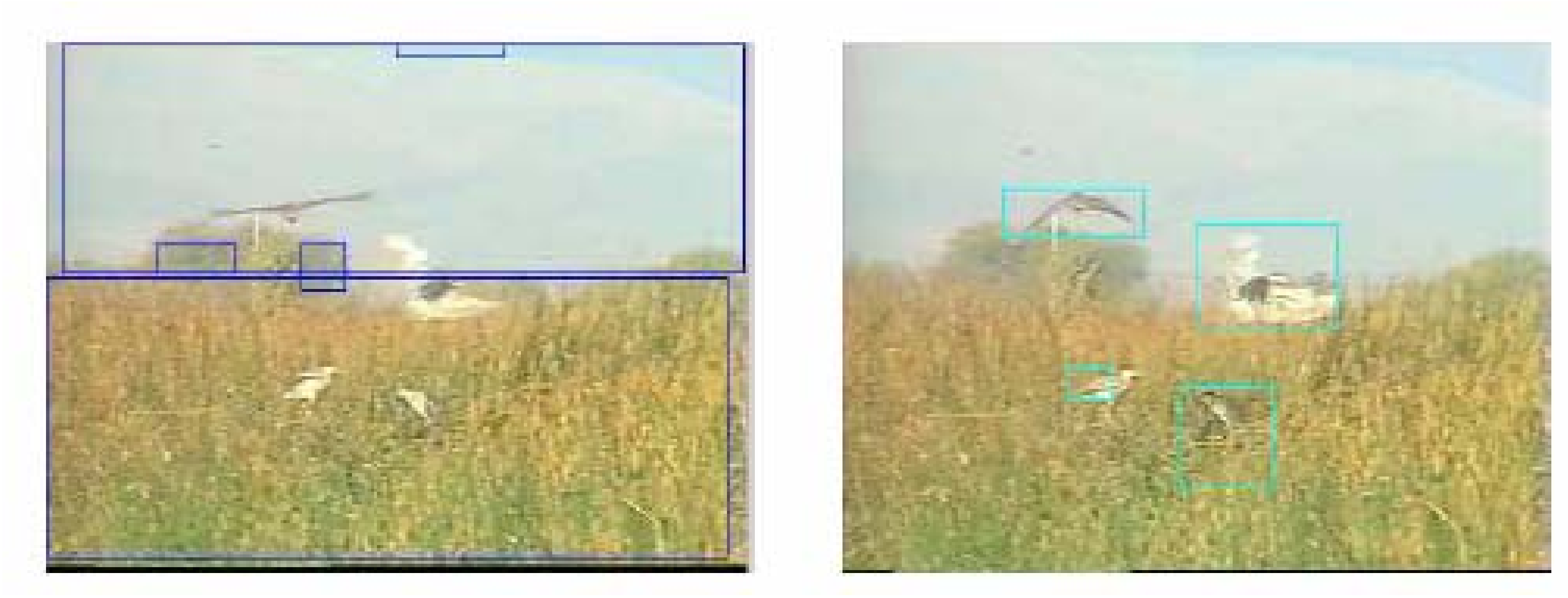
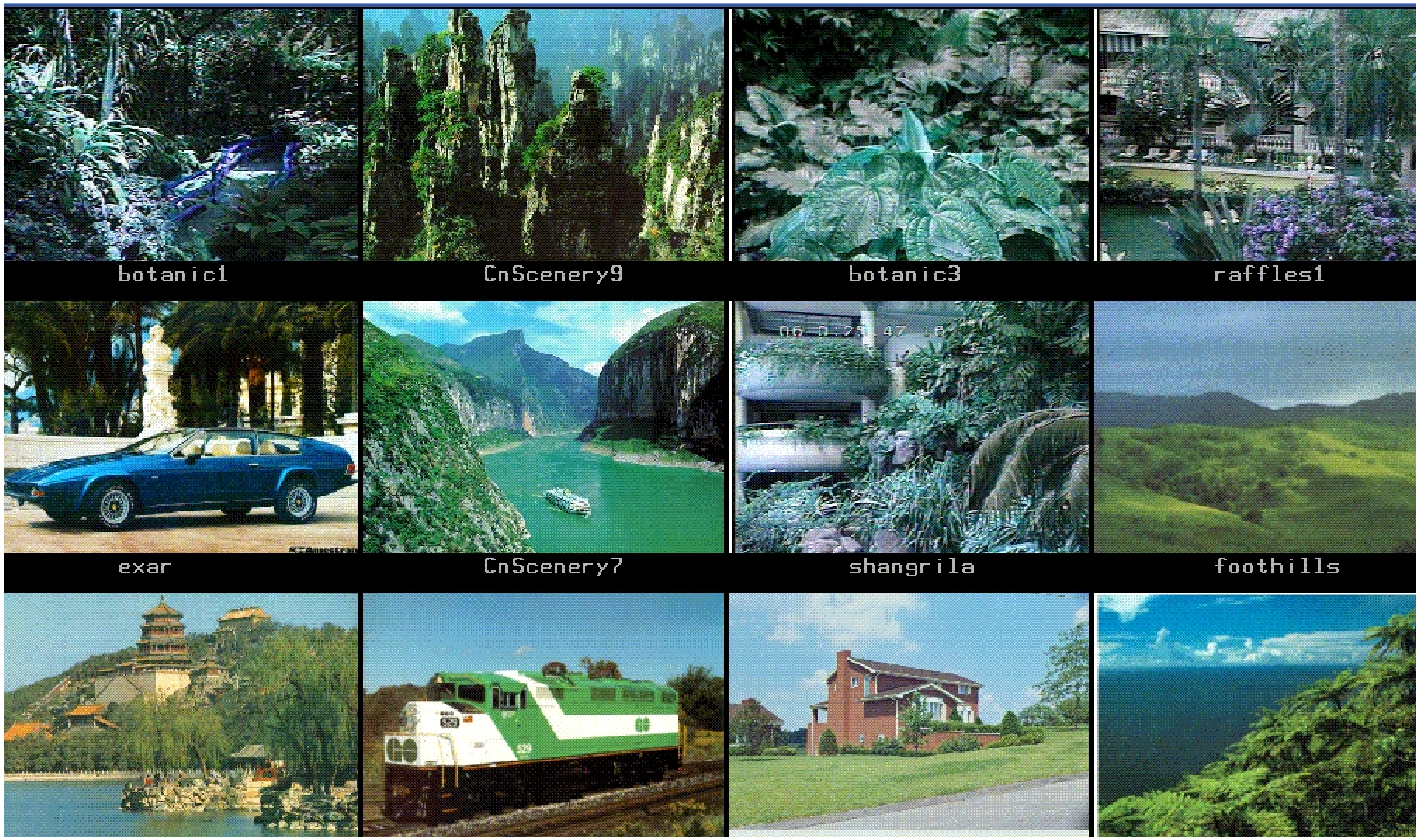


Image Search

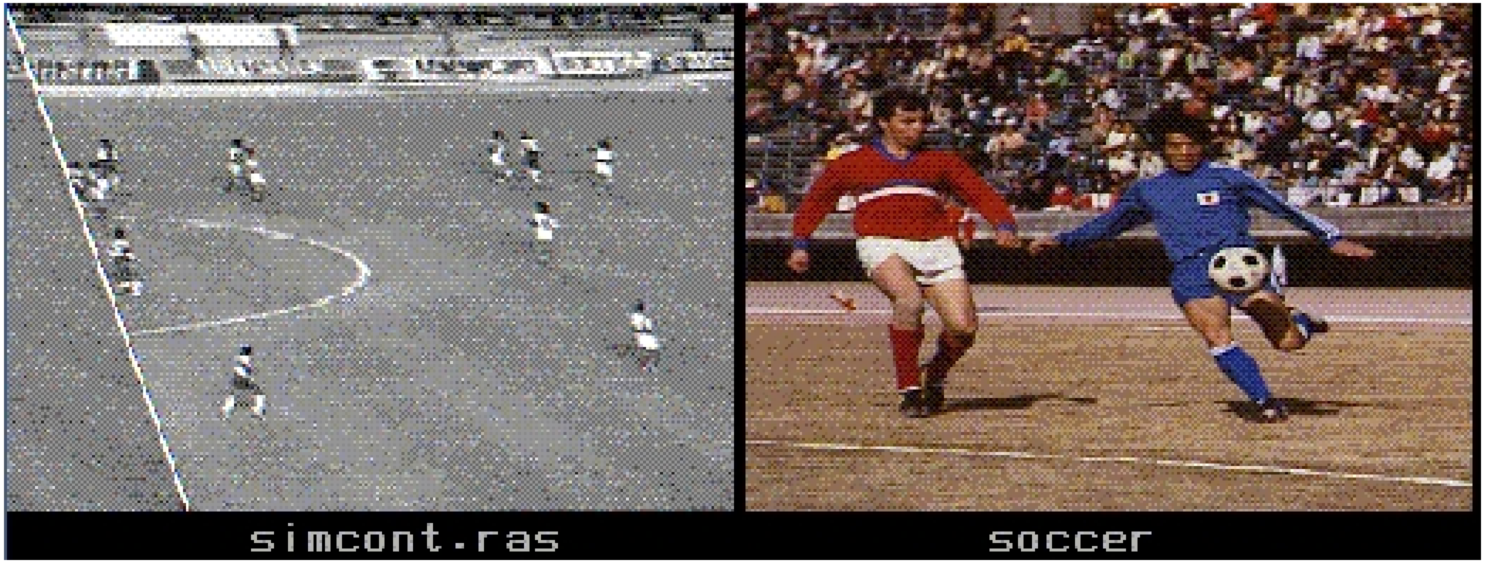
- Feature Representation
  - Image: represented as a series of real number, or a vector of features, *(f1, …., fn)*
  - Distance Function: The distance between two vectors, typically Euclidean Distance
  - Probably “Nearest is relevant”
    - The nearest images in the database is relevant to the query images.

Finding Similar Images

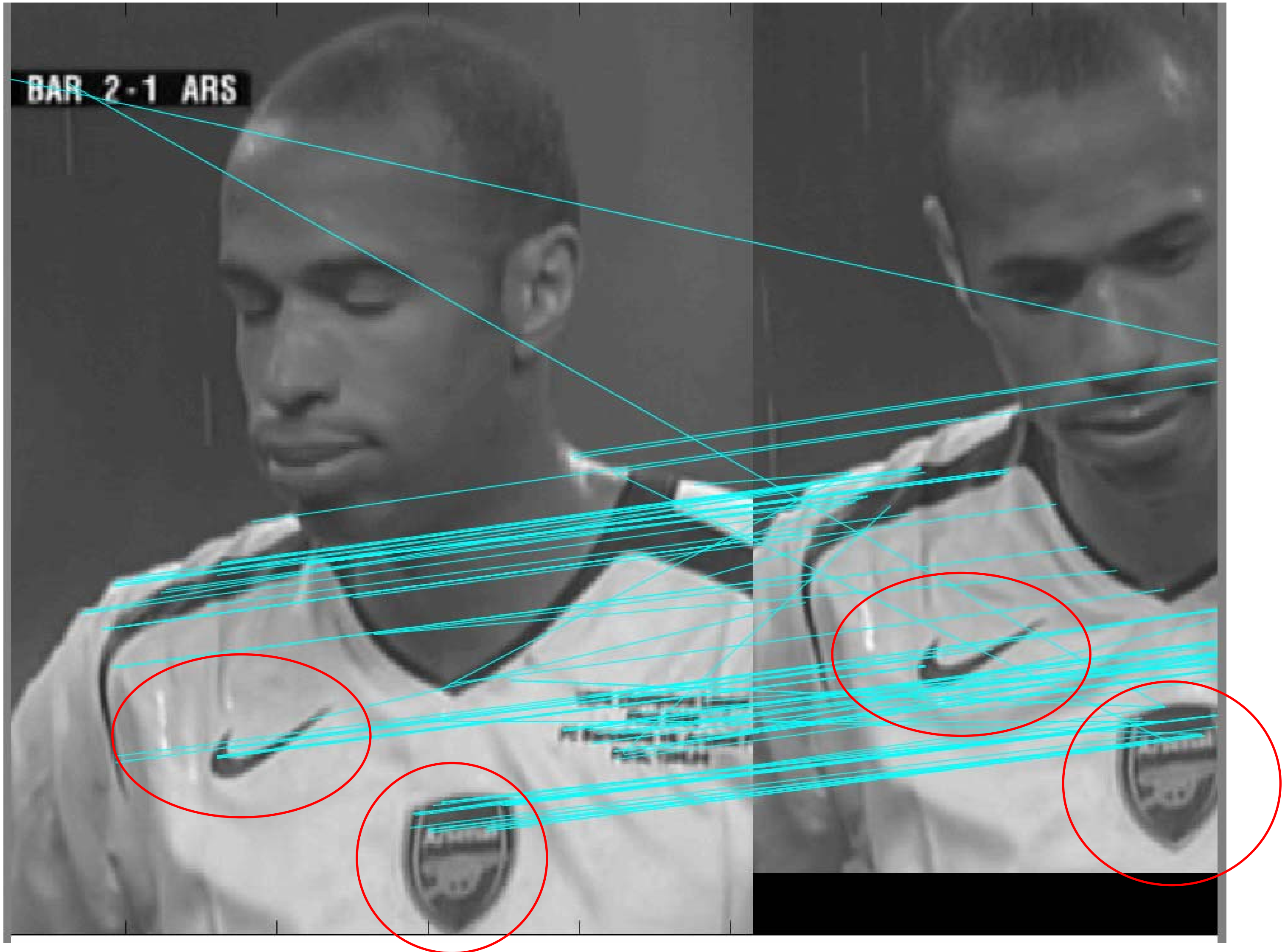


But…..

- Low-level feature doesn’t work in all the cases



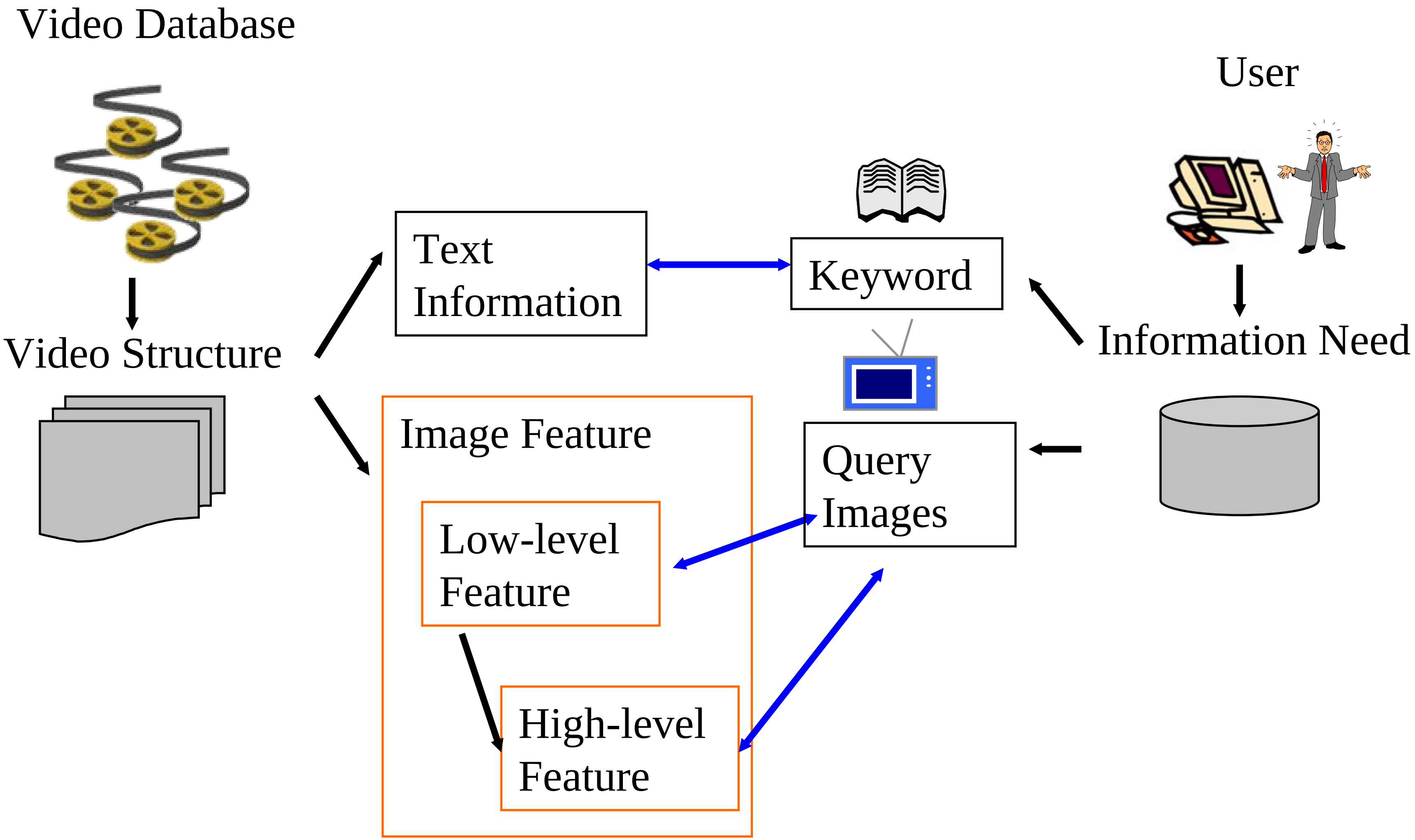
Find similar objects



High-level Image Feature

- Objects: Persons, Roads, Cars, Skies…
- Scenes: Indoors, Outdoors, Cityscape, Landscape, Water, Office, Factory…
- Event: Parade, Explosion, Picnic, Playing Soccer…
- Generated from low-level features

Image-based Retrieval



Video Database

Video Structure

Text Information

Keyword

User

Information Need

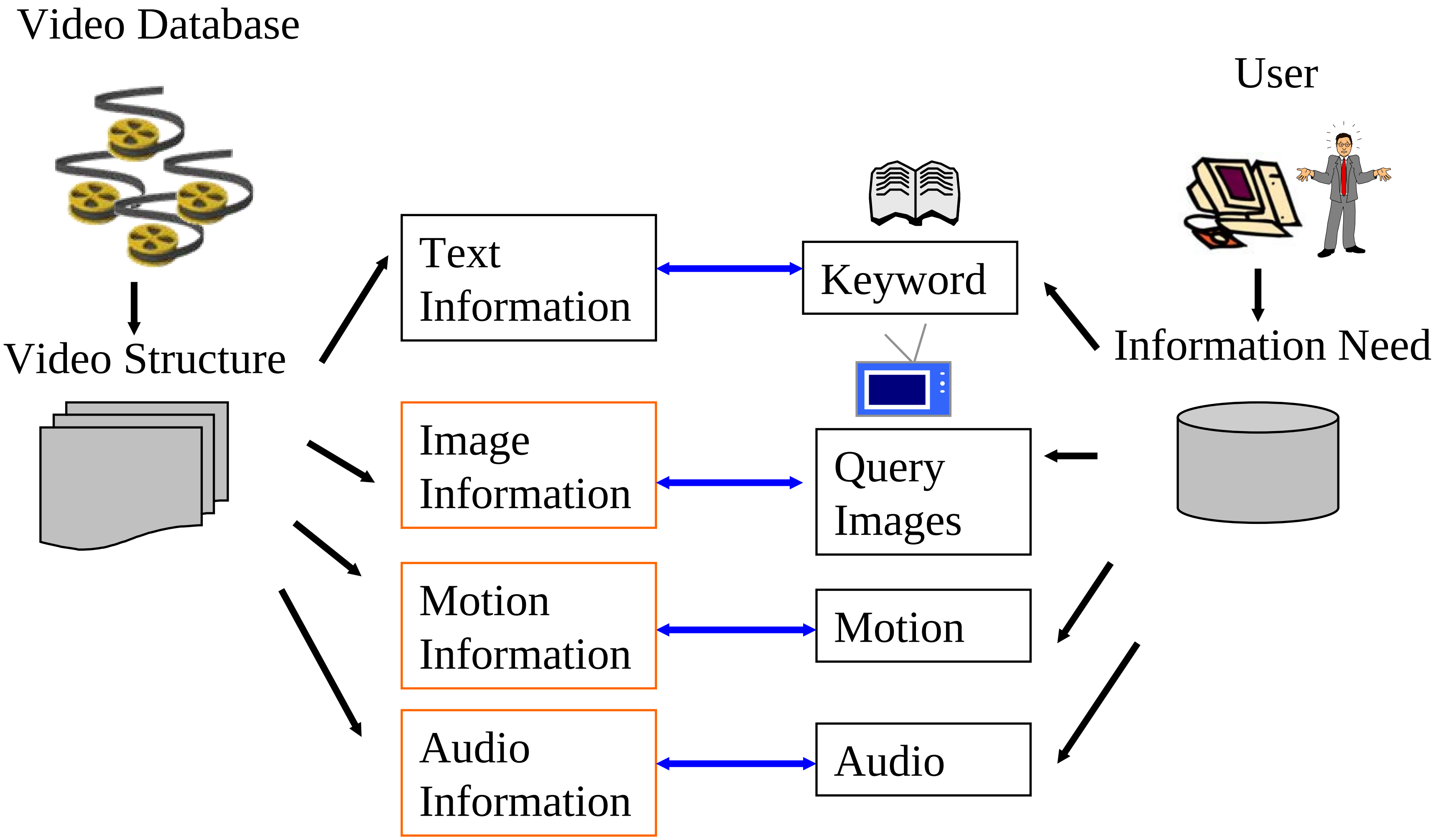
Image Feature

Low-level Feature

High-level Feature

Query Images

More Evidence in Video Retrieval



Video Database

Video Structure

Text Information

Image Information

Motion Information

Audio Information

Keyword

Query Images

Motion

Audio

User

Information Need

Combination of multi-modal results

- Difference characteristics between multi-modal information
  - Text-based Information: better for middle and high level queries
    - e.g. Find the video clip of dancing women wearing dresses
  - Image-based Information: better for low and middle level queries
    - e.g. Find the video clip of green trees
- Combination of multi-modal information

Other Useful Technique

- Query Expansion
- Cross-Modal Relation
- Relevance Feedback

Recap

- Video Retrieval is to bridge the gap between user information need and video database
- Multi-modal evidence
    - Text-based (most popular)
    - Image-based
    - Motion-based
    - Audio-based
- Combination of the evidence

Introduction to TREC Video Retrieval Track

- Full Name: Text REtrieval Conference
- TREC Video Track web site: http://www-nlpir.nist.gov/projects/trecvid/
- TREC series sponsored by the National Institute of Standards and Technology (NIST) with additional support from other U.S. government agencies
  - Goal is to encourage research in information retrieval

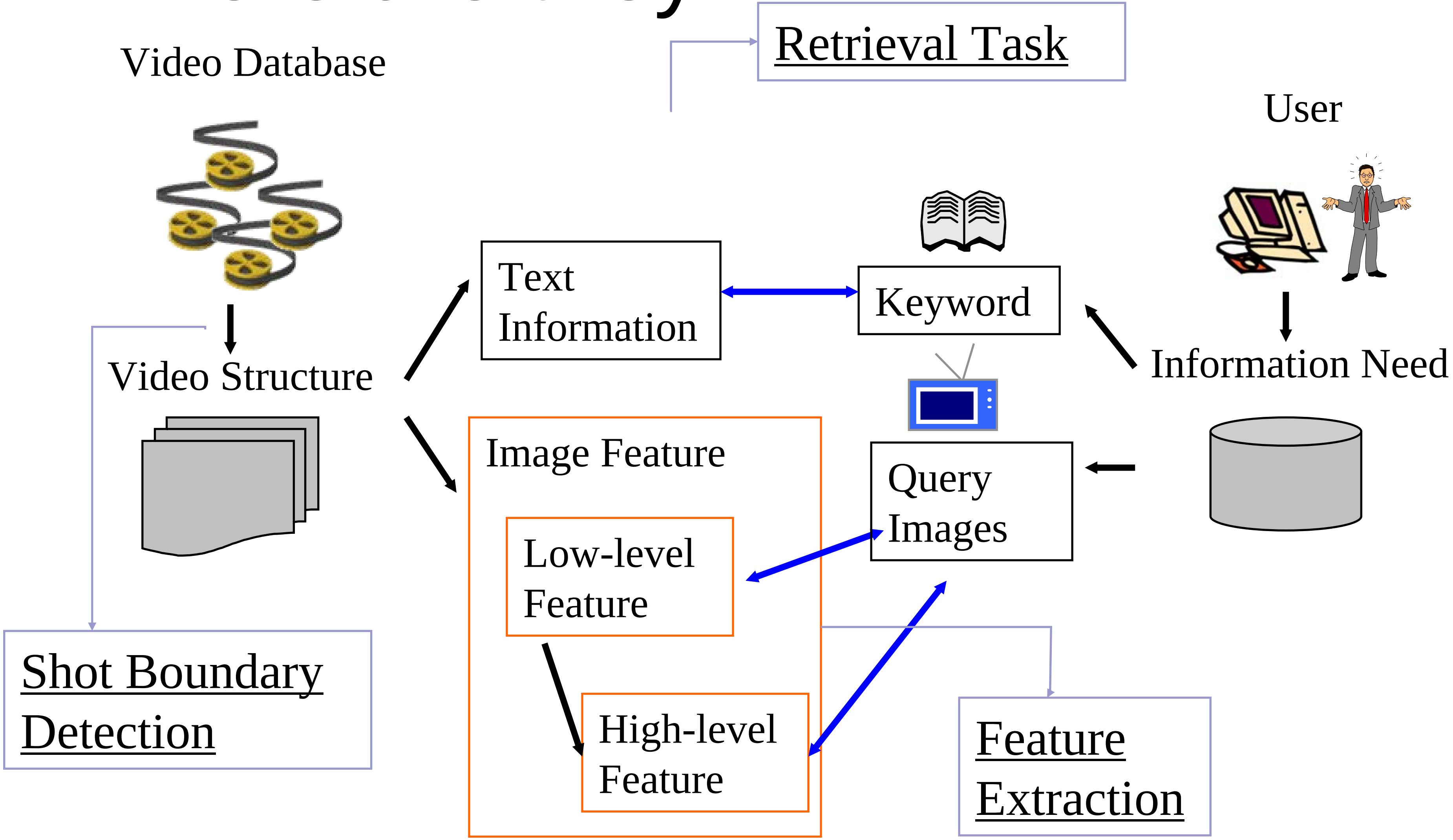
Introduction to TREC Video Retrieval Track

- Video Retrieval Track started in 2001
  - Goal is investigation of content-based retrieval from digital video
  - Focus on the *shot* as the unit of information retrieval rather than the scene or story/segment/clip
- Current state-of-the-art Video Retrieval Competition
  - 17 active participants, including groups from CMU, IBM Research, Microsoft Research Asia, MediaMill, LIMSI, Dublin City University.

Main tasks in TREC

- Shot boundary detection
- Semantic Feature Extraction Task
- Video Retrieval Task
  - Manual Retrieval: Human formulate a query and then automatically retrieve from collection
  - Interactive Retrieval: Full human access and feedback

Where are they?



Retrieval Task

Video Database

User

Text Information

Keyword

Video Structure

Information Need

Image Feature

Query Images

Low-level Feature

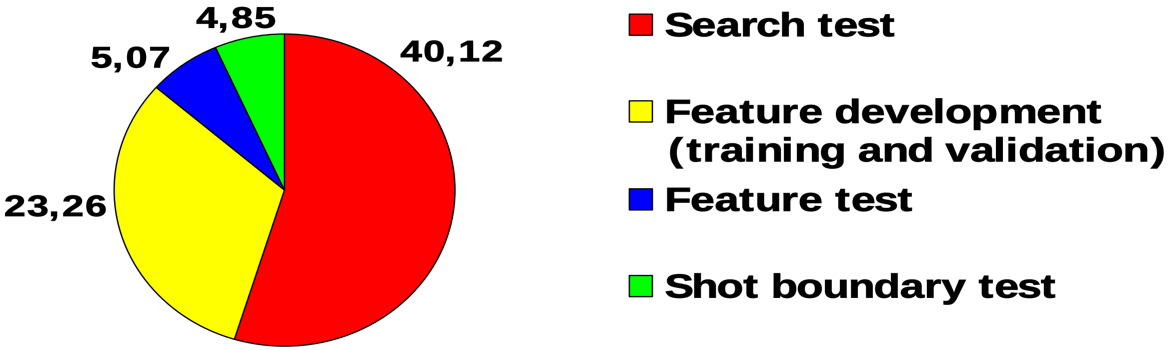
High-level Feature

Shot Boundary Detection

Feature Extraction

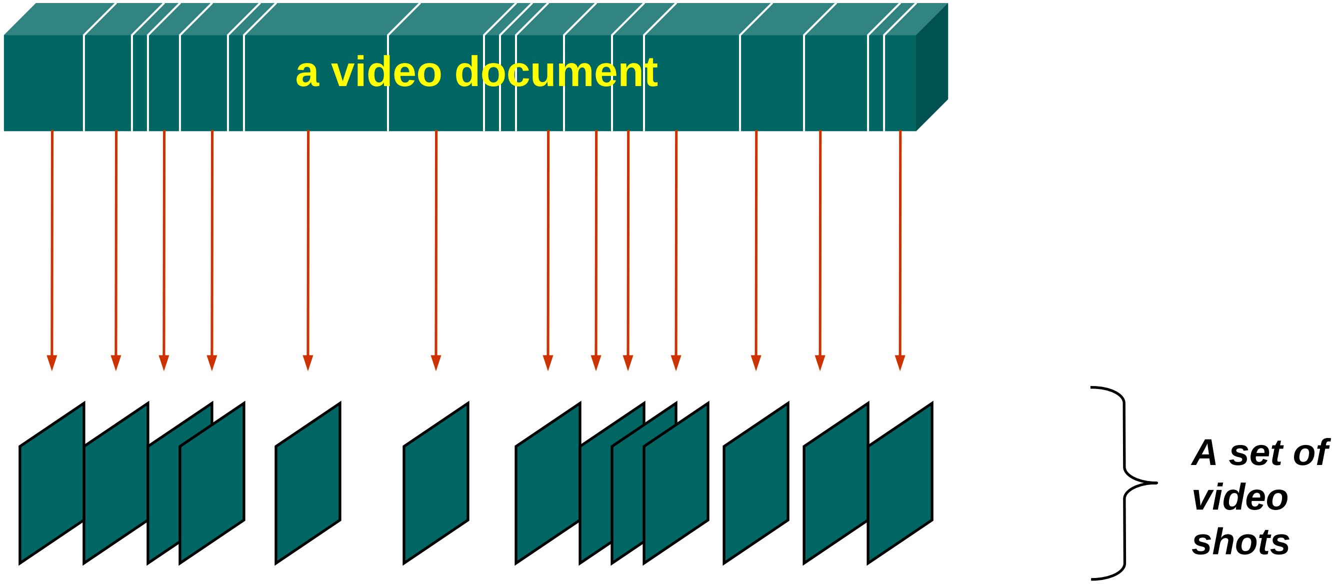
Video Data

- Difficult to get video data for use in TREC because ©
- Used mainly Internet Archive
  - advertising, educational, industrial, amateur films 1930-1970
  - produced by corporations, non-profit organisations, trade groups, etc.
  - Noisy, strange color, but real archive data
  - 73.3 hours partitioned as follows:



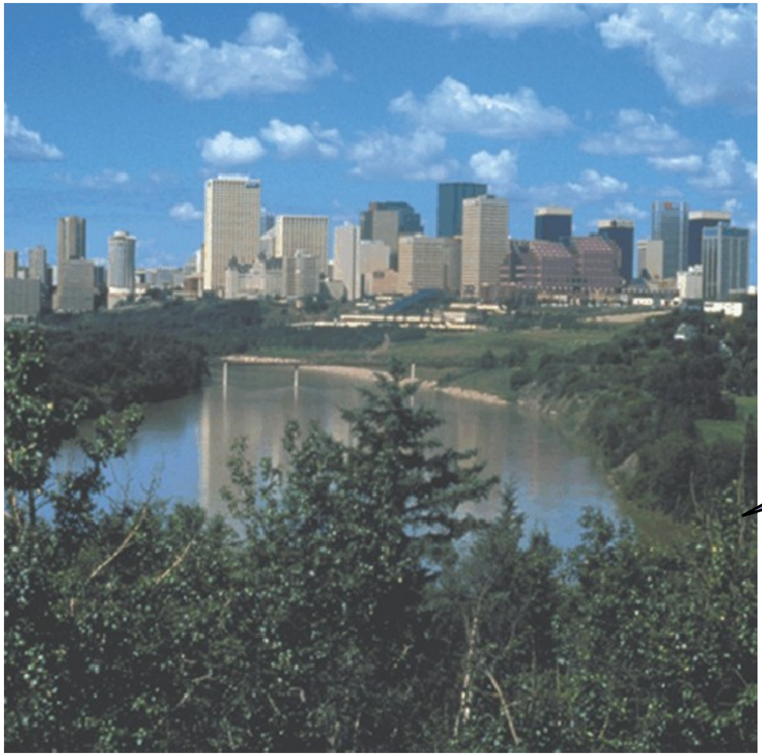
Shot Boundary Detection

- Fundamental primitive of most/all work in content-based video retrieval



Feature Extraction

- Extracted high-level semantic feature from video
- Assign a video clip to one or more of several categories of video



High-level features: Cityscape, Lake, Trees, Water, Sky

Feature Extraction

- Interesting itself but when it serves to help video navigation and search then its importance increases
- Benefits:
  - Retrieval - Find video from a particular class
  - Filtering - Remove irrelevant and distracting information categories from summaries and visualizations

The Features

- Face
  - Clip contains at least one human face with the nose, mouth, and both eyes visible. Pictures of a face meeting the above conditions count
- People
  - Clip contains a group of two more humans, each of which is at least partially visible and is recognizable as a human
- On-screen Text
  - Clip contains superimposed text large enough to be read

The Features

Indoor

Clip contains a recognizably indoor location, i.e., inside a building

Outdoor

Clip contains a recognizably outdoor location, i.e., one outside of buildings

Cityscape

Clip contains a recognizably city/urban/suburban setting

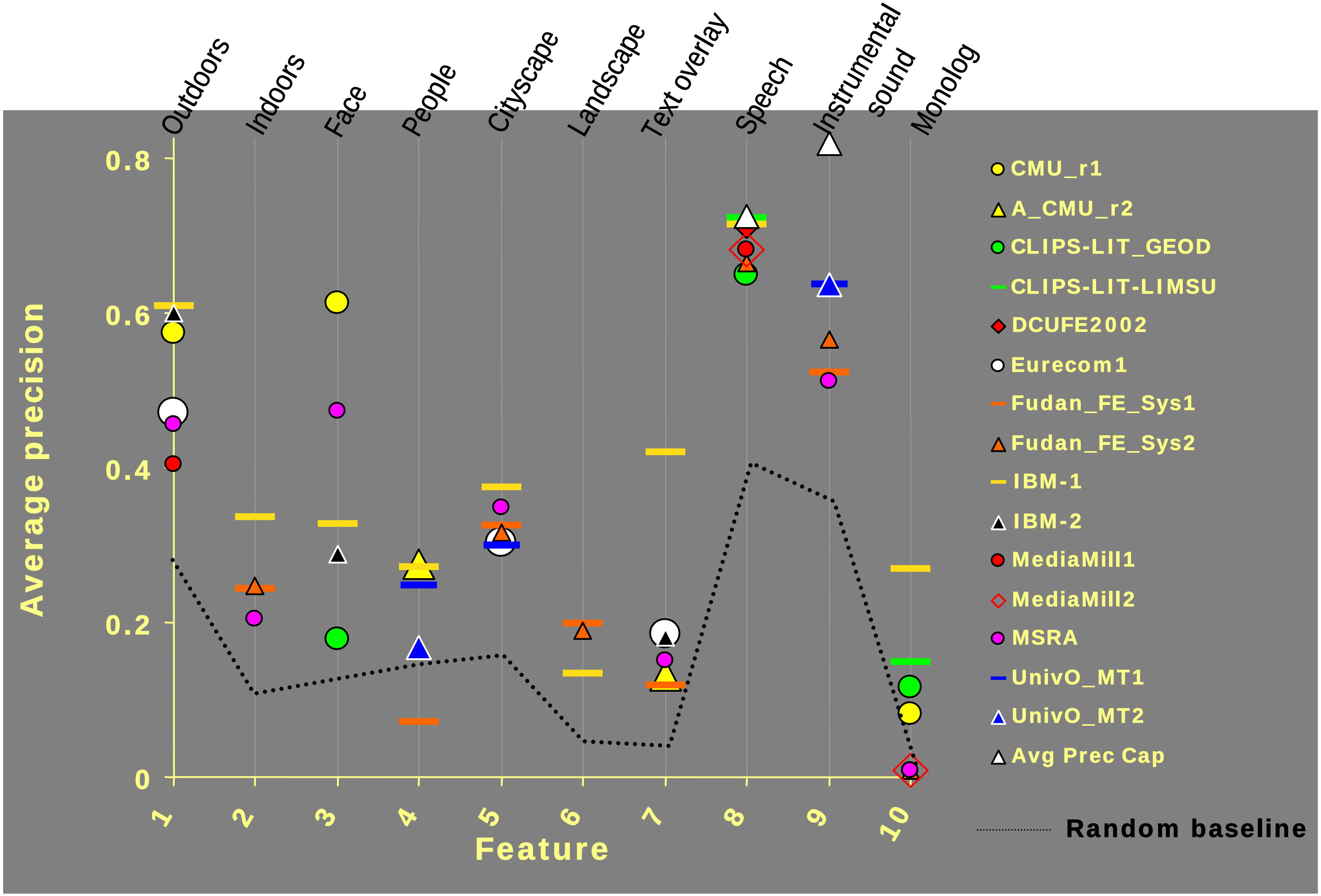
Landscape

Clip contains a predominantly natural inland setting, i.e., one with little or no evidence of development by humans. Scenes with bodies of water that are clearly inland may be included

Non-Video (Audio) Features

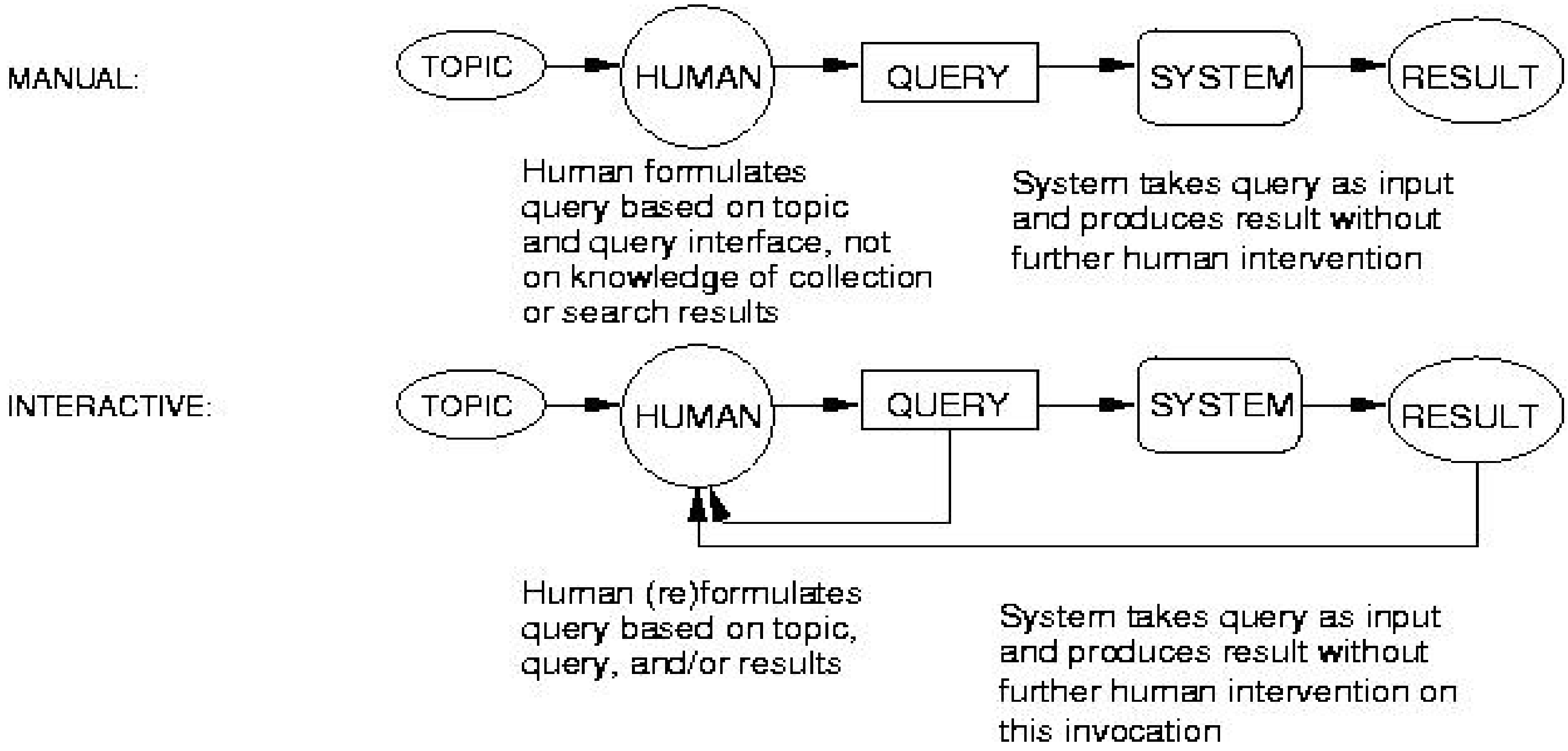
- Speech
  - A human voice uttering words is recognizable as such in this segment
- Instrumental Sound
  - Sound produced by one or more musical instruments is recognizable as such in this segment
- Monologues
  - Segment contains an event in which a single person is at least partially visible and speaks for a long time without interruption by another speaker. Pauses are ok if short

TREC02 Results



Video Search Task

- The most important task and final goal
- Manual & Interactive Search Task



Human formulates query based on topic and query interface, not on knowledge of collection or search results

System takes query as input and produces result without further human intervention

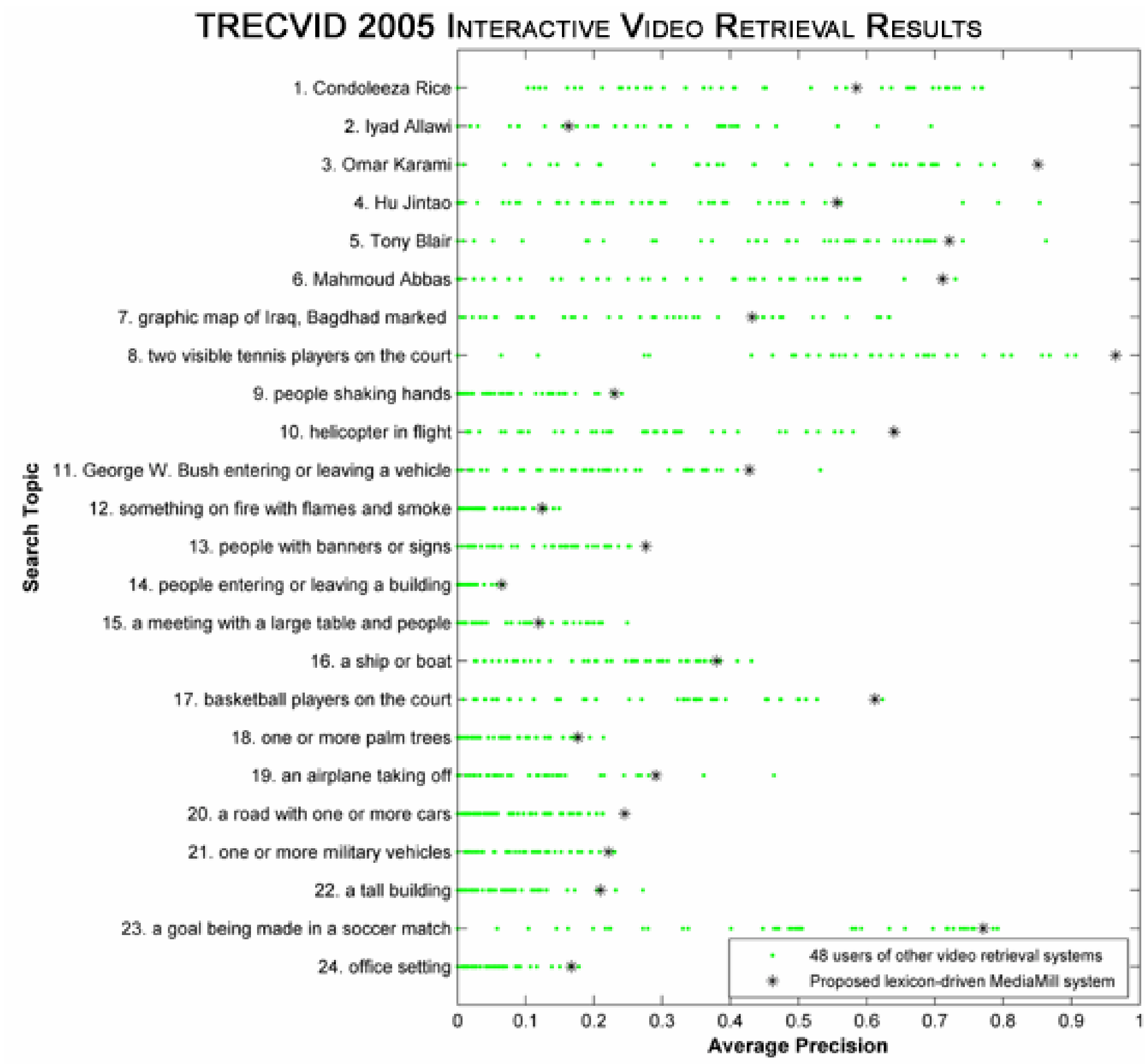
Human (re)formulates query based on topic, query, and/or results

System takes query as input and produces result without further human intervention on this invocation

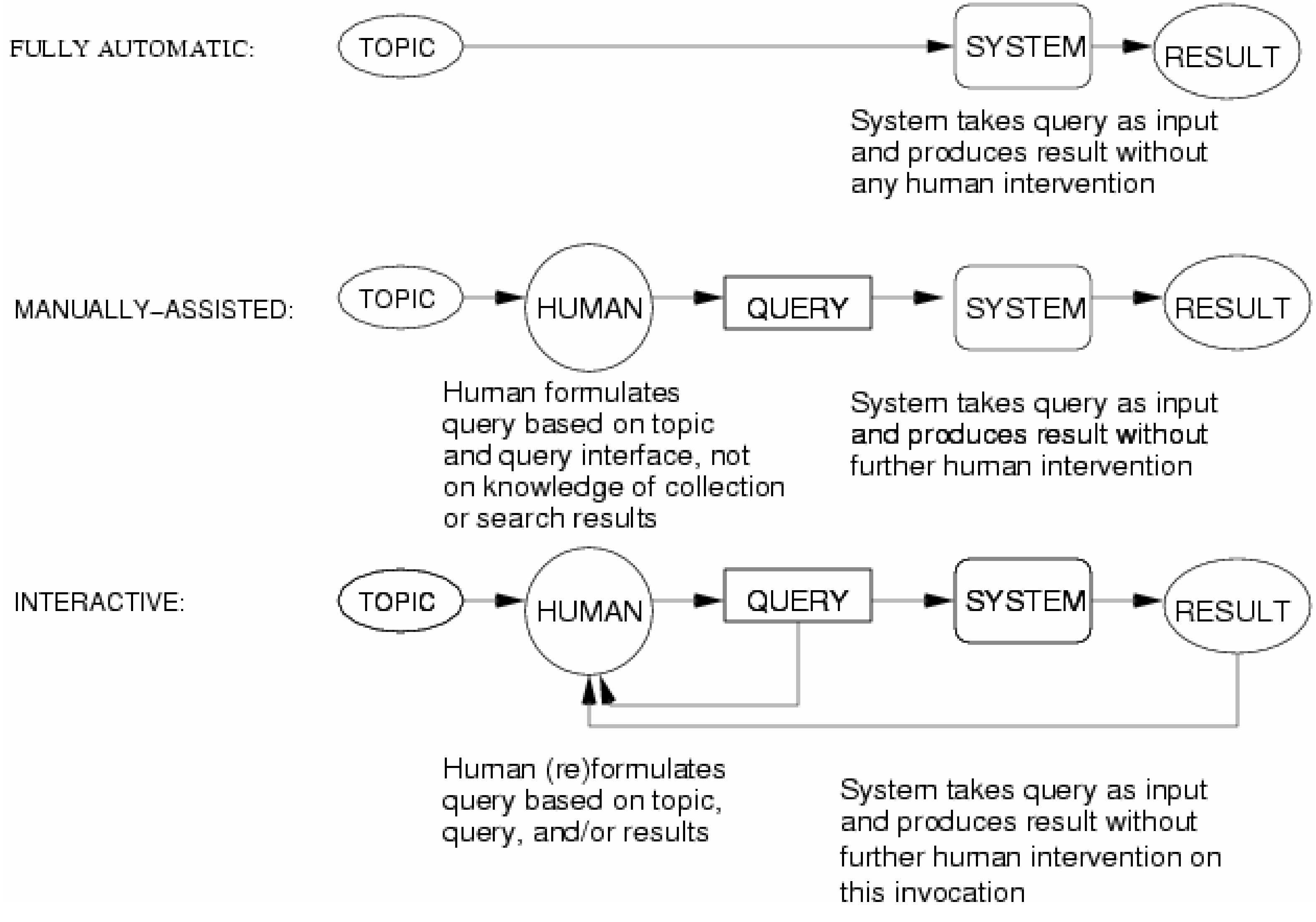
Queries for 2002 TREC Video Track

- Specific item or person
  - Eddie Rickenbacker, James Chandler, George Washington, Golden Gate Bridge, Price Tower in Bartlesville, OK
- Specific fact
  - Arch in Washington Square Park in NYC, map of continental US
- Instances of a category
  - football players, overhead views of cities, one or more women standing in long dresses
- Instances of events/activities
  - people spending leisure time at the beach, one or more musicians with audible music, crowd walking in an urban environment, locomotive approaching the viewer

TRECVid 2005 search topics



TRECvid 2006



System takes query as input and produces result without any human intervention

Human formulates query based on topic and query interface, not on knowledge of collection or search results

System takes query as input and produces result without further human intervention

Human (re)formulates query based on topic, query, and/or results

System takes query as input and produces result without further human intervention on this invocation

Queries for 2006 TREC Video Track

- Example types of video needs

I'm interested in video material containing:

  - □ a specific person
  - □ one or more instances of a category of people
  - □ a specific thing
  - □ one or more instances of a category of things
  - □ a specific event/activity
  - □ one or more instances of a category of events/activities
  - □ a specific location
  - □ one or more instances of a category of locations
  - □ combinations of the above

- Topics may target commercials as well as news content.

Some TRECVid 2006 high level features

- **Sports**: Shots depicting any sport in action
- **Entertainment**: Shots depicting any entertainment segment in action
- **Weather**: Shots depicting any weather related news or bulletin
- **Court**: Shots of the interior of a court-room location
- **Office**: Shots of the interior of an office setting
- **Meeting**: Shots of a Meeting taking place indoors
- **Studio**: Shots of the studio setting including anchors, interviews and all events that happen in a news room
- **Outdoor**: Shots of Outdoor locations
- **Building**: Shots of an exterior of a building
- **Desert**: Shots with the desert in the background
- **Vegetation**: Shots depicting natural or artificial greenery, vegetation woods, etc.
- **Mountain**: Shots depicting a mountain or mountain range with the slopes visible
- **Road**: Shots depicting a road
- …

Sample Query

- XML Representation

```
<!DOCTYPE videoTopic SYSTEM "videoTopics.dtd">
<videoTopic num="077">

<textDescription text="Find pictures of George Washington" />

<imageExample
src="http://www.cia.gov/csi/monograph/firstln/955pres2.gif"
desc="face" />

<videoExample src="01681.mpg"  start="09m25.938s"
stop="09m29.308s" desc="face" />

</videoTopic>
```

Evaluation Metric

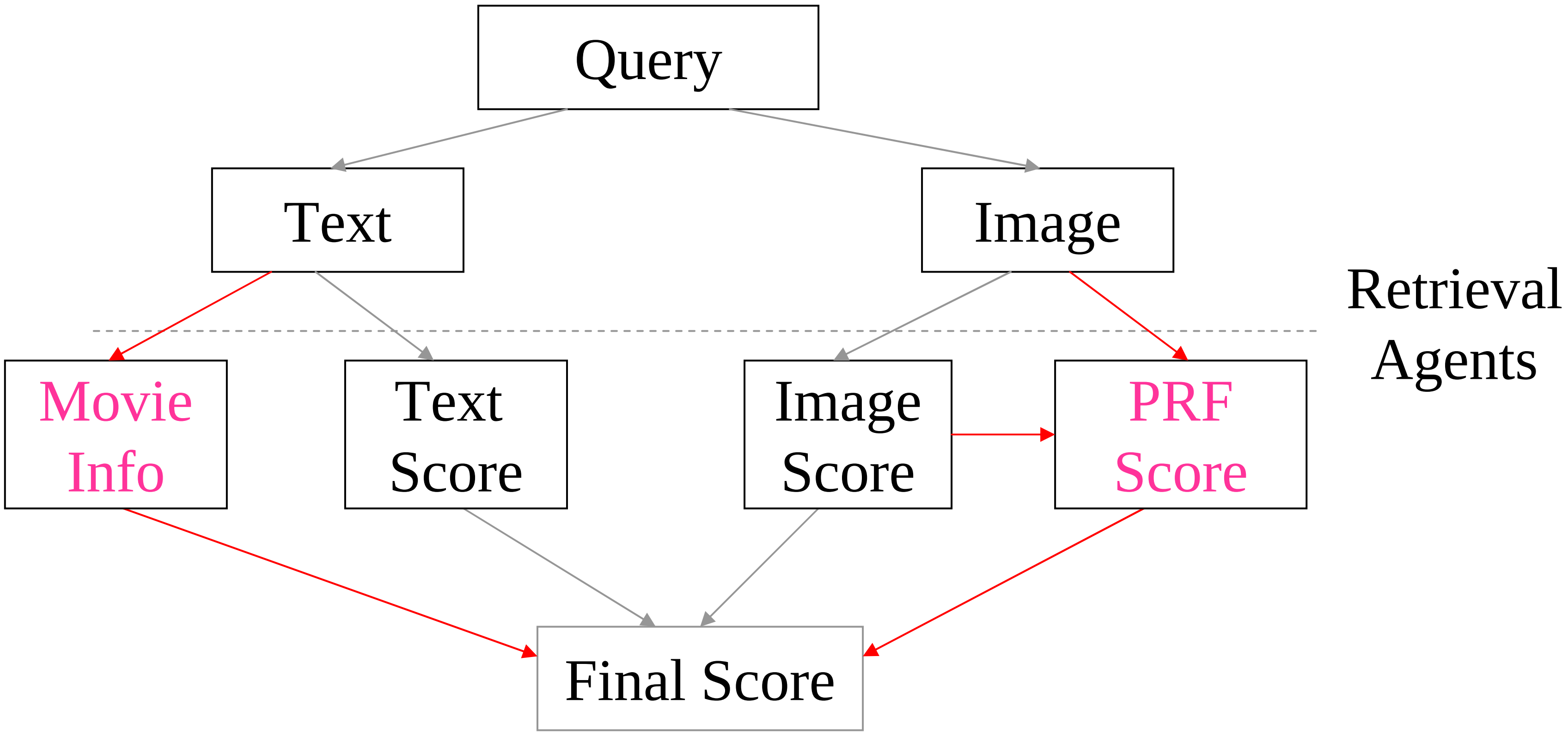
- Goal: Maximize the Mean Average Precision
  - Result set limited to 100 shots
  - Precision = (# relevant shots retrieved)/(total # shots retrieved)
  - Average precision: compute precision after each retrieved relevant shot and then average these precisions over the total number of retrieved relevant shots in the collection for that topic
  - Submitting the maximum number of shots per result set can never lower the average precision for that submission
  - Mean Average Precision = average of the average precision measures for each topic

Demo

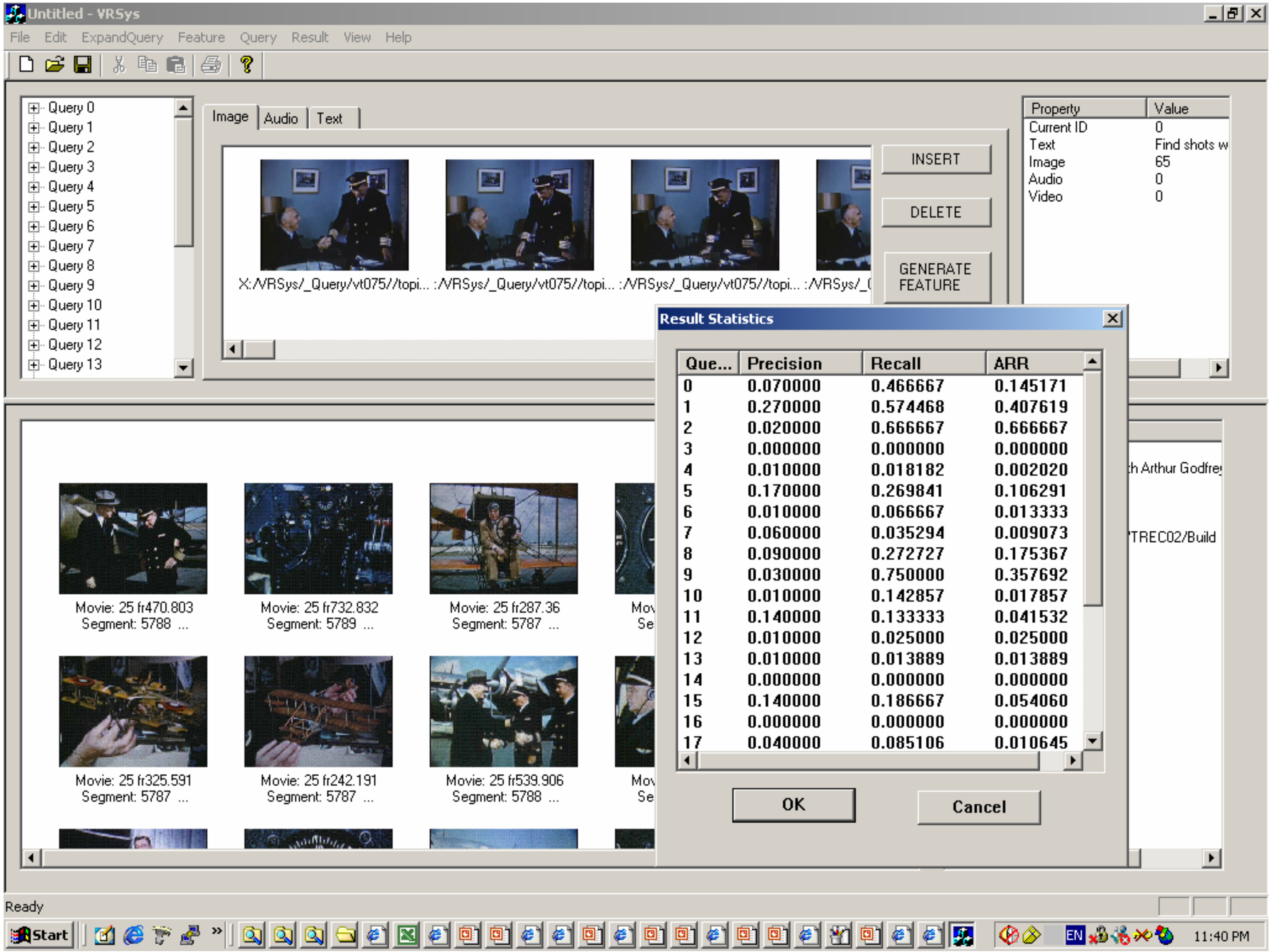
- CMU Interactive Search System
- IBM Video Retrieval System

  http://mp7.watson.ibm.com/marvel/
- UvA MediaMill

CMU Manual Retrieval System



Snapshot of the CMU system (2002)



Property	Value
Current ID	0
Text	Find shots w
Image	65
Audio	0
Video	0

Que...	Precision	Recall	ARR
0	0.070000	0.466667	0.145171
1	0.270000	0.574468	0.407619
2	0.020000	0.666667	0.666667
3	0.000000	0.000000	0.000000
4	0.010000	0.018182	0.002020
5	0.170000	0.269841	0.106291
6	0.010000	0.066667	0.013333
7	0.060000	0.035294	0.009073
8	0.090000	0.272727	0.175367
9	0.030000	0.750000	0.357692
10	0.010000	0.142857	0.017857
11	0.140000	0.133333	0.041532
12	0.010000	0.025000	0.025000
13	0.010000	0.013889	0.013889
14	0.000000	0.000000	0.000000
15	0.140000	0.186667	0.054060
16	0.000000	0.000000	0.000000
17	0.040000	0.085106	0.010645

Movie: 25 fr470.803 Segment: 5788 ...

Movie: 25 fr732.832 Segment: 5789 ...

Movie: 25 fr287.36 Segment: 5787 ...

Mov Se

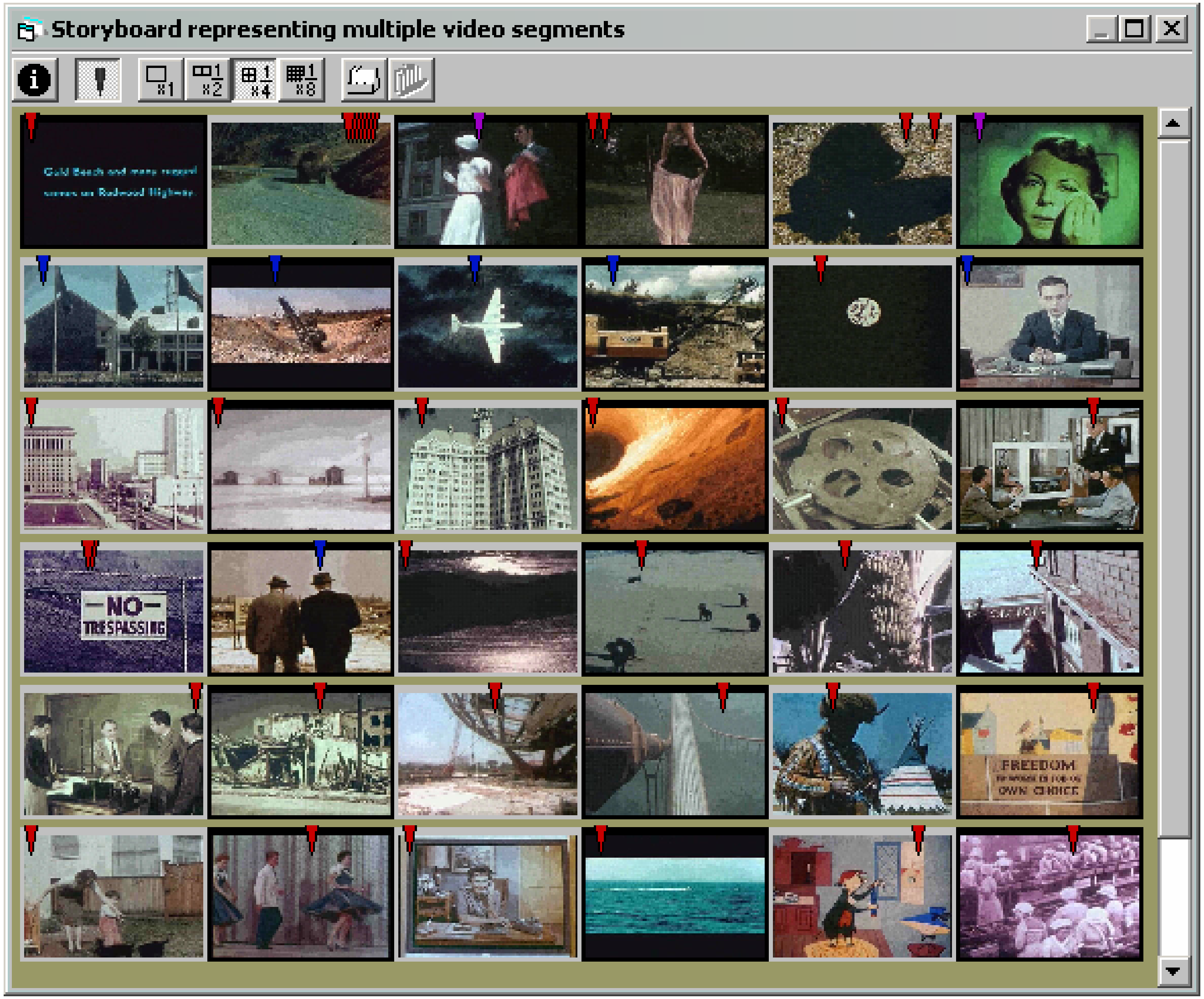
Movie: 25 fr325.591 Segment: 5787 ...

Movie: 25 fr242.191 Segment: 5787 ...

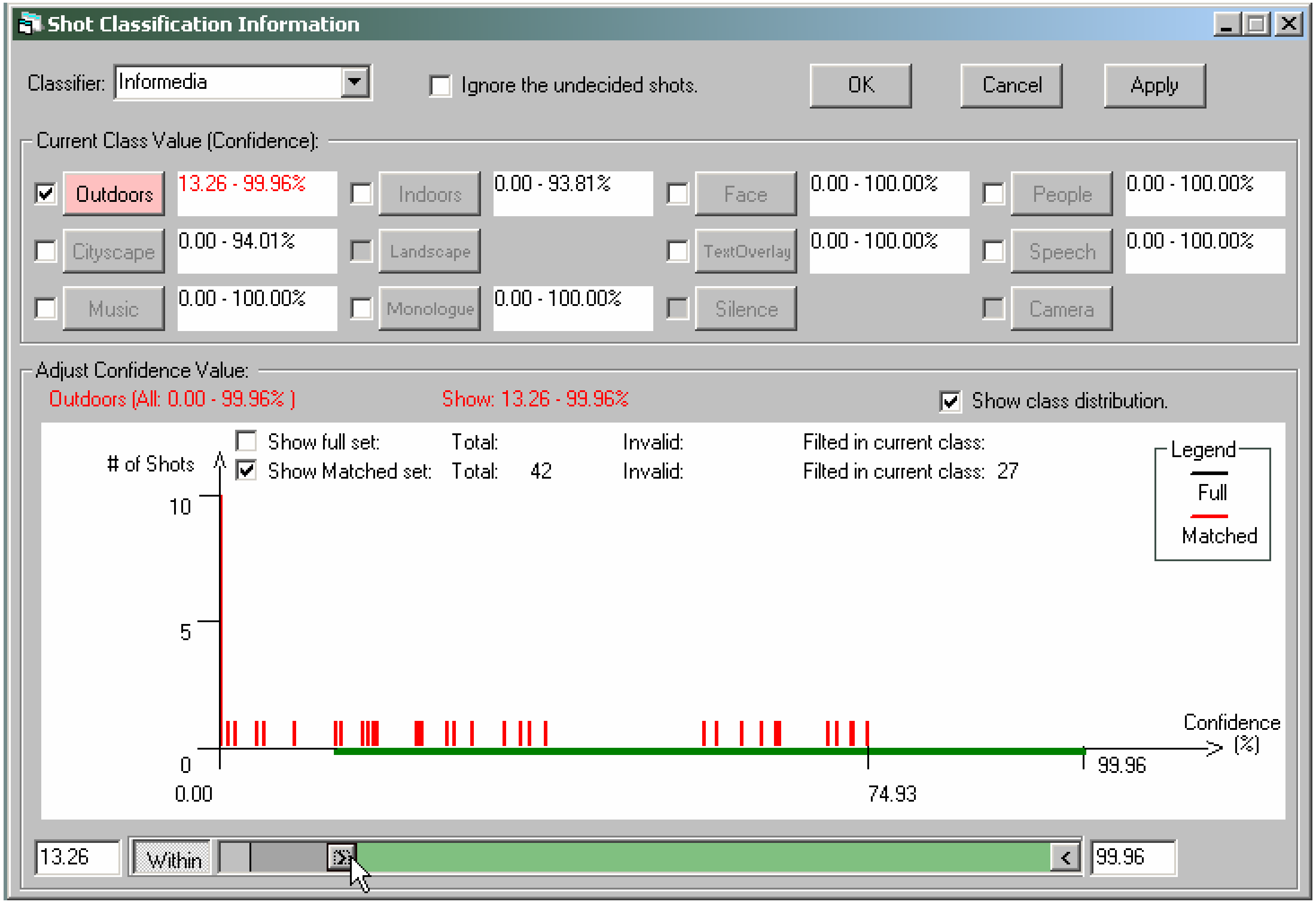
Movie: 25 fr539.906 Segment: 5788 ...

Mov Se

Snapshot of the CMU system (2002)


Storyboard representing multiple video segments
1 x1
1 x2
1 x4
1 x8
Gold Beach and many rugged scenes on Redwood Highway
NO TRESPASSING
FREEDOM
OWN CHOICE

CMU Filter Interface for using Image Features (2002)



Shot Classification Information

Classifier: Informedia

□ Ignore the undecided shots.

Current Class Value (Confidence):

☑ Outdoors 13.26 - 99.96%	□ Indoors 0.00 - 93.81%	□ Face 0.00 - 100.00%	□ People 0.00 - 100.00%
□ Cityscape 0.00 - 94.01%	□ Landscape	□ TextOverlay 0.00 - 100.00%	□ Speech 0.00 - 100.00%
□ Music 0.00 - 100.00%	□ Monologue 0.00 - 100.00%	□ Silence	□ Camera

Adjust Confidence Value:

Outdoors (All: 0.00 - 99.96%)

Show: 13.26 - 99.96%

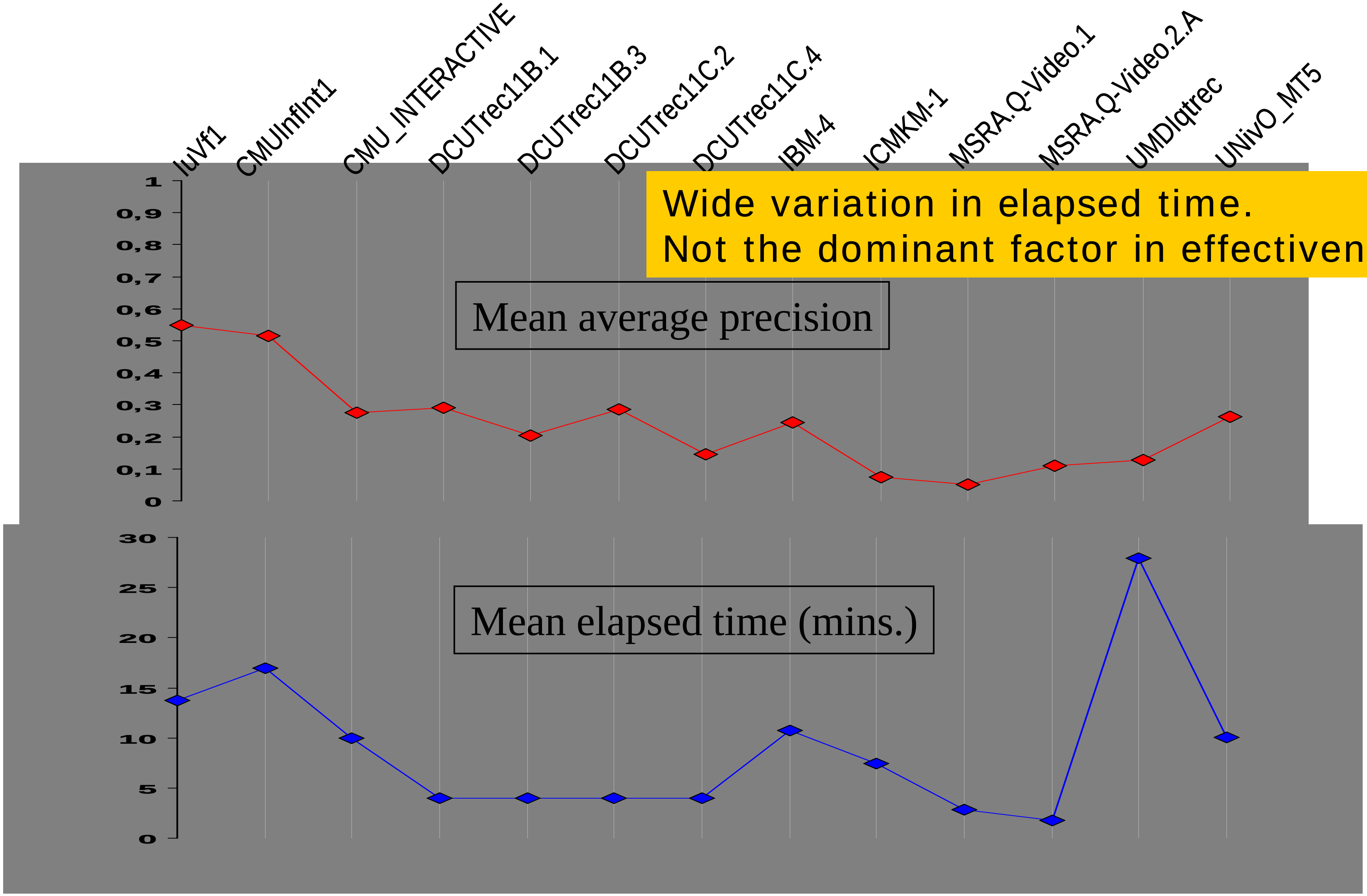
☑ Show class distribution.

□ Show full set: Total: Invalid: Filted in current class:

☑ Show Matched set: Total: 42 Invalid: Filted in current class: 27

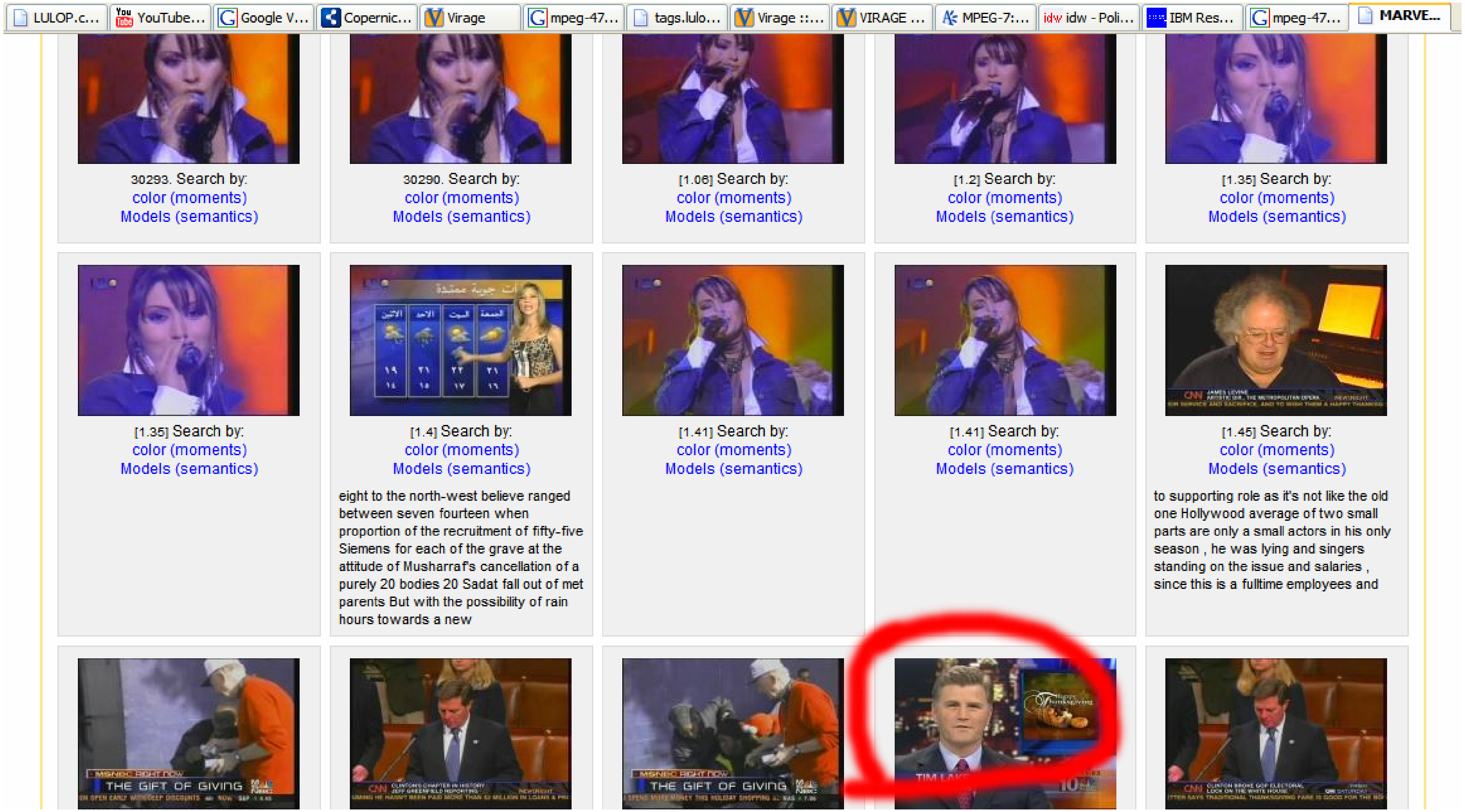
13.26 Within 99.96

Mean AvgP vs. mean elapsed time



Wide variation in elapsed time.
Not the dominant factor in effectiveness

IBM Marvel system



30293. Search by:
color (moments)
Models (semantics)

30290. Search by:
color (moments)
Models (semantics)

[1.06] Search by:
color (moments)
Models (semantics)

[1.2] Search by:
color (moments)
Models (semantics)

[1.35] Search by:
color (moments)
Models (semantics)

[1.35] Search by:
color (moments)
Models (semantics)

[1.4] Search by:
color (moments)
Models (semantics)

eight to the north-west believe ranged between seven fourteen when proportion of the recruitment of fifty-five Siemens for each of the grave at the attitude of Musharraf's cancellation of a purely 20 bodies 20 Sadat fall out of met parents But with the possibility of rain hours towards a new

[1.41] Search by:
color (moments)
Models (semantics)

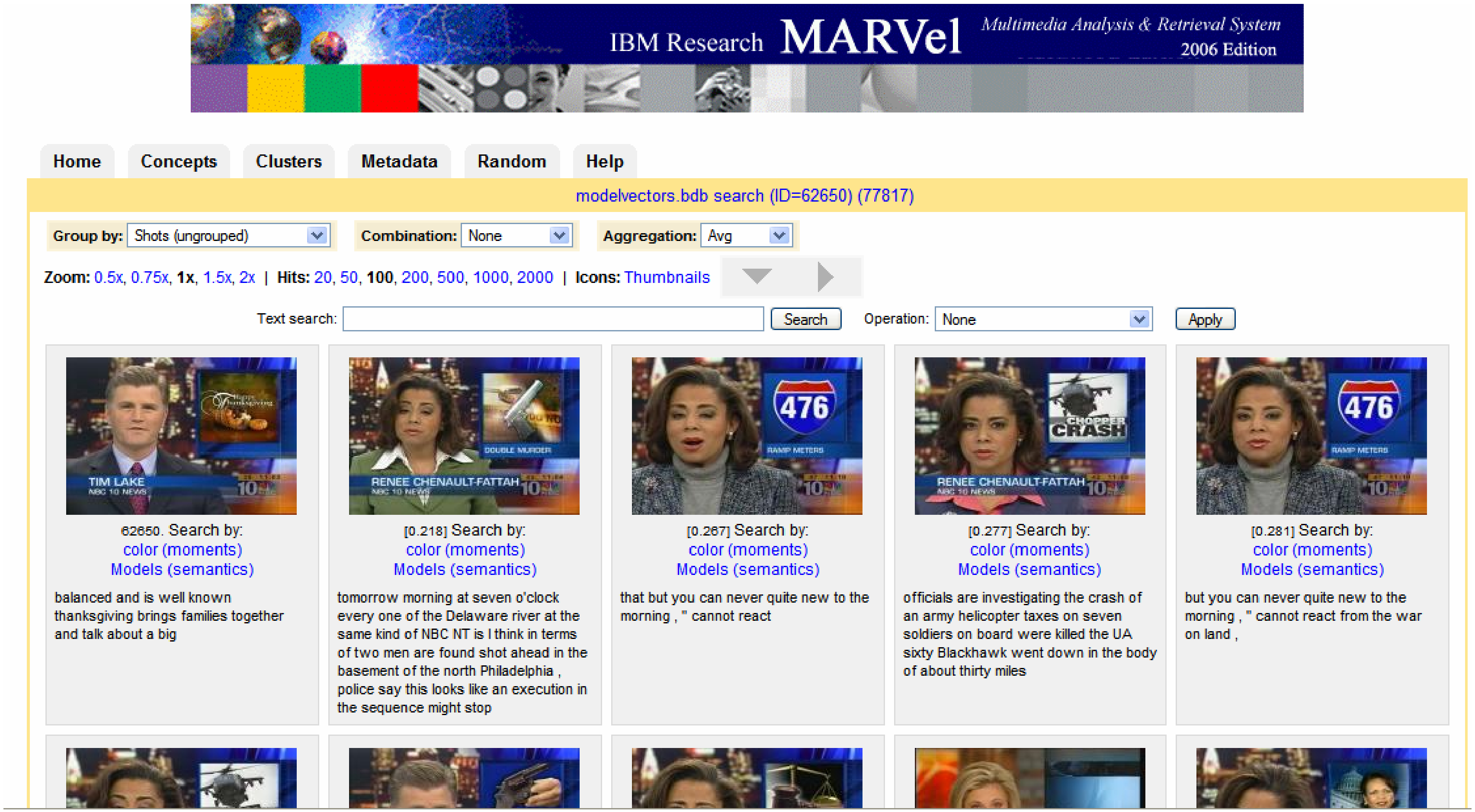
[1.41] Search by:
color (moments)
Models (semantics)

[1.45] Search by:
color (moments)
Models (semantics)

to supporting role as it's not like the old one Hollywood average of two small parts are only a small actors in his only season , he was lying and singers standing on the issue and salaries , since this is a fulltime employees and

Search using color similarity

IBM Marvel: search using semantics from previous results



modelvectors.bdb search (ID=62650) (77817)

Group by: Shots (ungrouped) | Combination: None | Aggregation: Avg

Zoom: 0.5x, 0.75x, **1x**, 1.5x, 2x | Hits: 20, 50, **100**, 200, 500, 1000, 2000 | Icons: Thumbnails

Text search: | Operation: None

62650. Search by:
color (moments)
Models (semantics)

balanced and is well known thanksgiving brings families together and talk about a big

[0.218] Search by:
color (moments)
Models (semantics)

tomorrow morning at seven o'clock every one of the Delaware river at the same kind of NBC NT is I think in terms of two men are found shot ahead in the basement of the north Philadelphia , police say this looks like an execution in the sequence might stop

[0.267] Search by:
color (moments)
Models (semantics)

that but you can never quite new to the morning , " cannot react

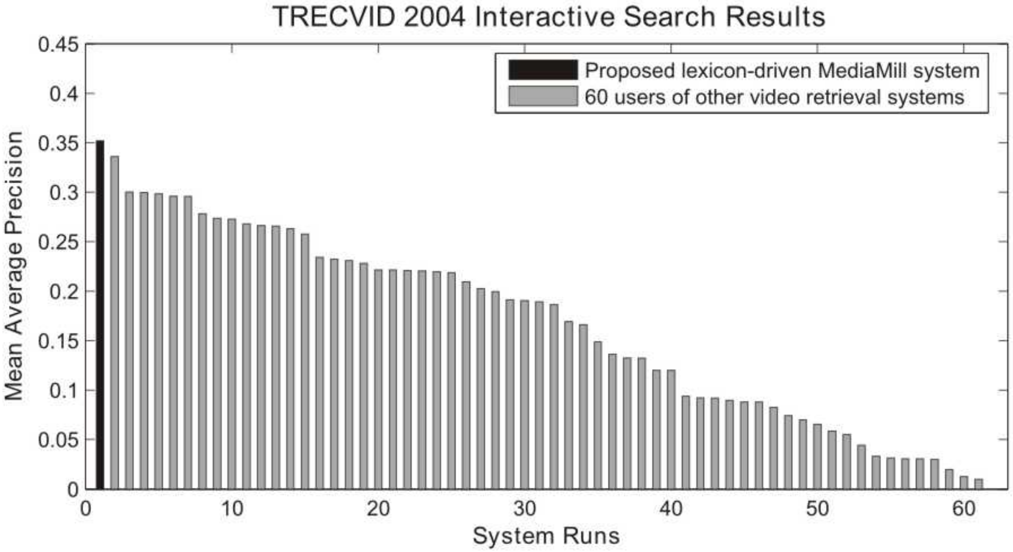
[0.277] Search by:
color (moments)
Models (semantics)

officials are investigating the crash of an army helicopter taxes on seven soldiers on board were killed the UA sixty Blackhawk went down in the body of about thirty miles

[0.281] Search by:
color (moments)
Models (semantics)

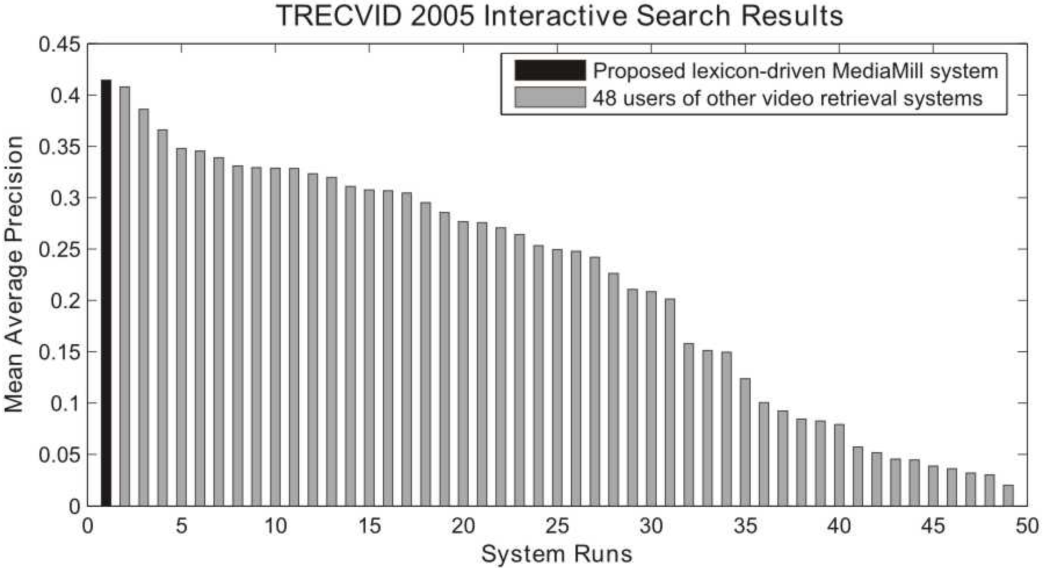
but you can never quite new to the morning , " cannot react from the war on land ,

UvA MediaMill



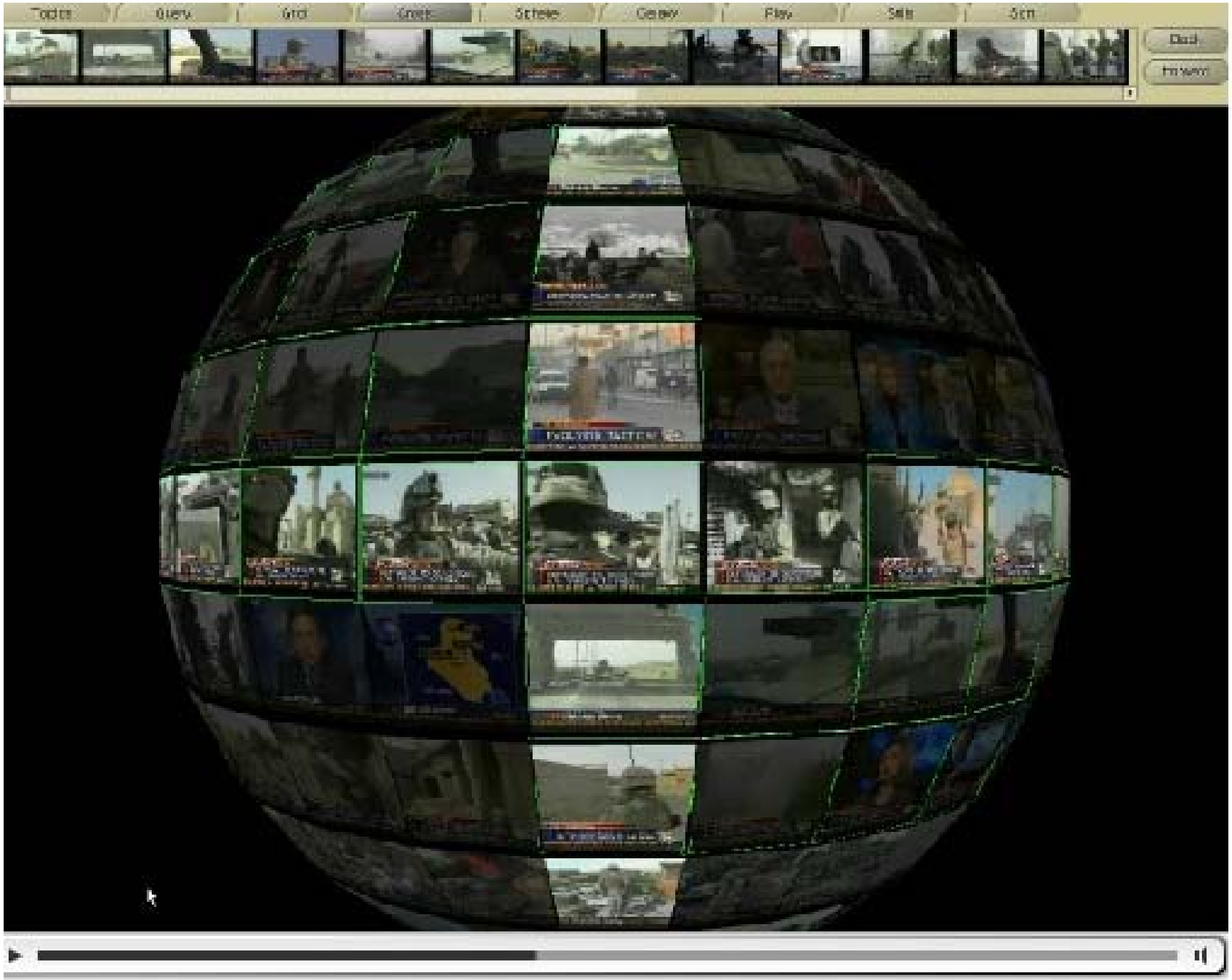
32 concept detectors

Interactive video search



101 concept detectors

UvA MediaMill – cross browser



Conclusion

- The goal of content-based video retrieval is to build more intelligent video retrieval engine via semantic meaning
- Many applications in daily life
- Combine evidence from different aspects
- Hot research topic, few business system
  - Check Techcrunch.com for info on business ventures
- State-of-the-art performance is still unacceptable for normal users, space to improve

Virage: a business that has survived

VIRAGE®

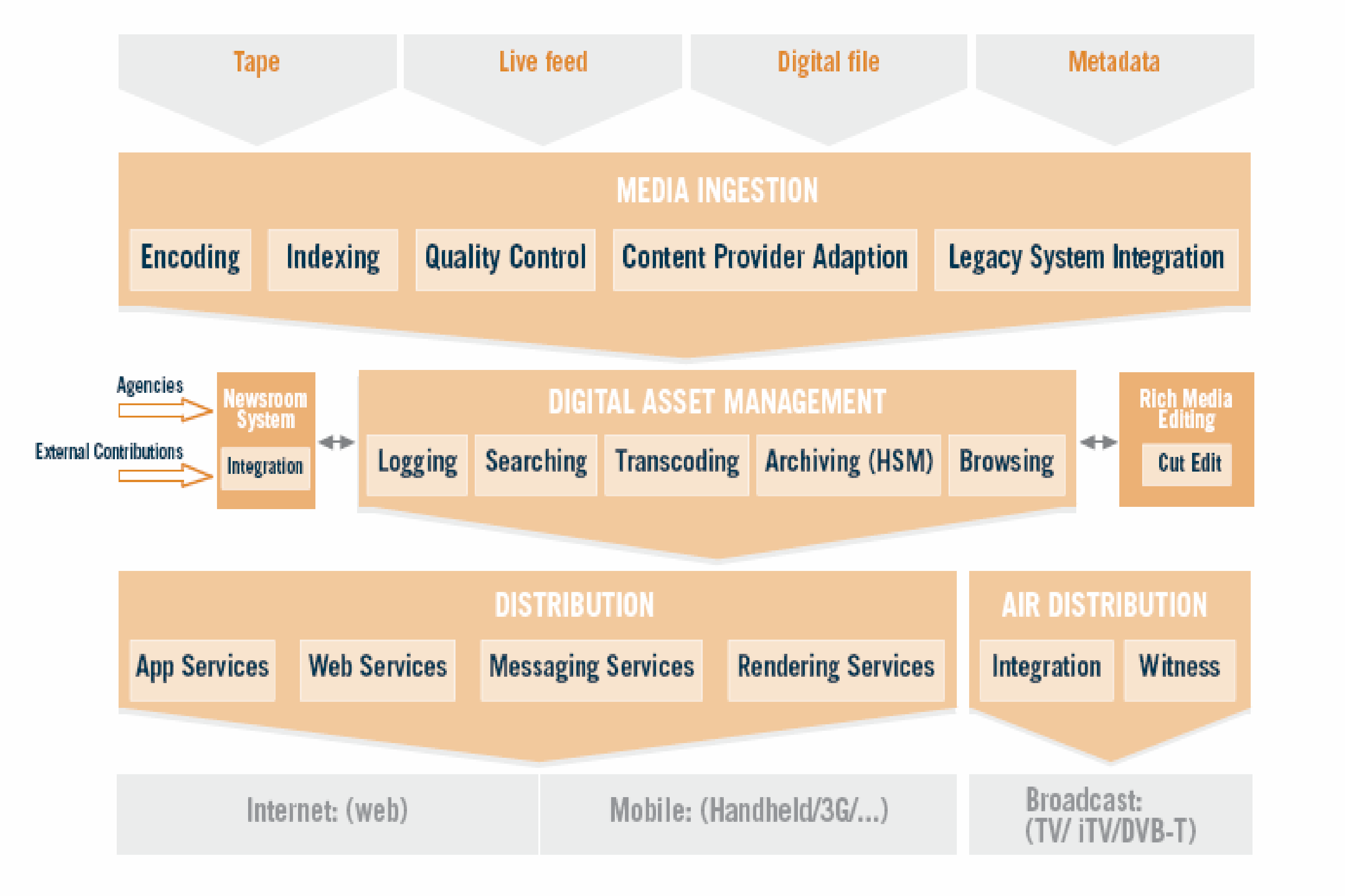
Rich Media Management

Security & Surveillance

Virage is the world leader in Rich Media Management software and Intelligent Video Analytics. Bringing gether complementary technologies from multimedia, security and infrastructure specialists, Virage offers unrivalled product set capable of television, video, audio and CCTV challenges of any kind. From making television content fully searchable and accessible via IPTV to supplying and managing complex security systems, Virage has unrivalled experience and expertise.

te for Japanese Language National Library of Medicine New York University Open University Oxford University Princeton University Stanford

Reply.it – Multimedia asset management



Credits

- Rong Yan – Carnegie Mellon
- Alexander Gelbukh