



THESEUS

Forschungsprogramm für eine
neue internetbasierte Wissensinfrastruktur

1st Int. Workshop on
Standards and Technologies in Multimedia Archives and Records (STAR)
Lausanne, 2010-04-26/27

Searching in video sequences - core technologies in THESEUS

26th April, 2010

Thomas Riegel

Siemens AG, Corporate Technology, Munich

Supported by

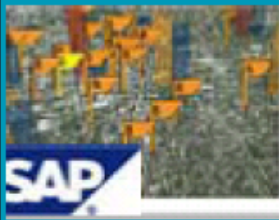


Federal Ministry
of Economics
and Technology



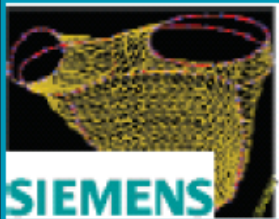
Overview

- » Introduction - THESEUS
- » Objectives and Challenges
- » System architecture
- » Sample Application “Wetten, Dass..?”
- » Live Demo



TEXO

Internet-based
service management



MEDICO

Medical diagnostics
information & knowledge
management



ALEXANDRIA

Community driven
semantic annotation of
information

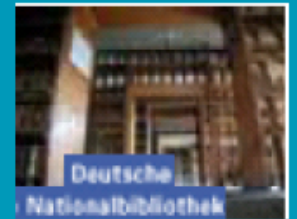
PROCESSUS

Knowledge based
Business Process
Management



CONTENTUS

Knowledge based access
to the cultural heritage



ORDO

Personalized Information &
Knowledge Management



Core Technology Cluster

Base Technologies for the Use Cases

Challenge:



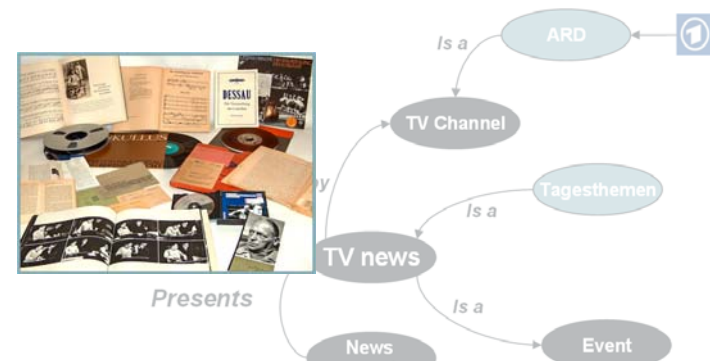
More than 30 Million hours
of audio-visual data stored in
European archives



How to navigate and search that overwhelming amount of data ?

Contentus develops an integrated system for the semantic provisioning of broadcast archives by

- » digitizing & restaurating
- » content analysis (metadata extraction, enrichment)
- » archiving & indexing the archival footage





**Image and Video Analysis
Video and Audio Codecs**

Fraunhofer Gesellschaft



**Multimodal
Dialogue Shell
for various devices**



SIEMENS

**Statistical
Machine Learning
Technologies**

**Ontology
Development &**



Image Recognition

Video Recognition

Video Codec

Standardization

**Metadata Generation,
Indexing, Retrieval**

Quality Assessment

Fingerprinting



Fraunhofer Gesellschaft



Fraunhofer Gesellschaft



Research on and development of a system and components for retrieving events, event courses and situations from media archives.

“How can a system support the search in large-scale image/video data stores, where meaningful results can only be retrieved by exploiting (inter-)relations between objects / events across multiple images, the situational context, and the application context ?”

Example search queries:

- » “Find scenes where celebrity A and politician B are approaching each other” (Media Domain)
- » “Find cases of patients with a similar lesion in the liver and a similar course of healing” (Medical Domain)
- » “Trace back a marked person in the video footage” (Surveillance Domain)



Efficient metadata usage:

- » Queries must be answered with the metadata generated by existing and available video analysis tools
- » Intermediate metadata (incl. confidence values) for the analysis tools are valuable information and shall be used when available

Exhaustive context usage:

- » Most queries can only be answered in a specific domain, including application and task context knowledge to restrict the search space and to add semantics
- » Any information cues should be used – but under consideration of their reliability

Query management:

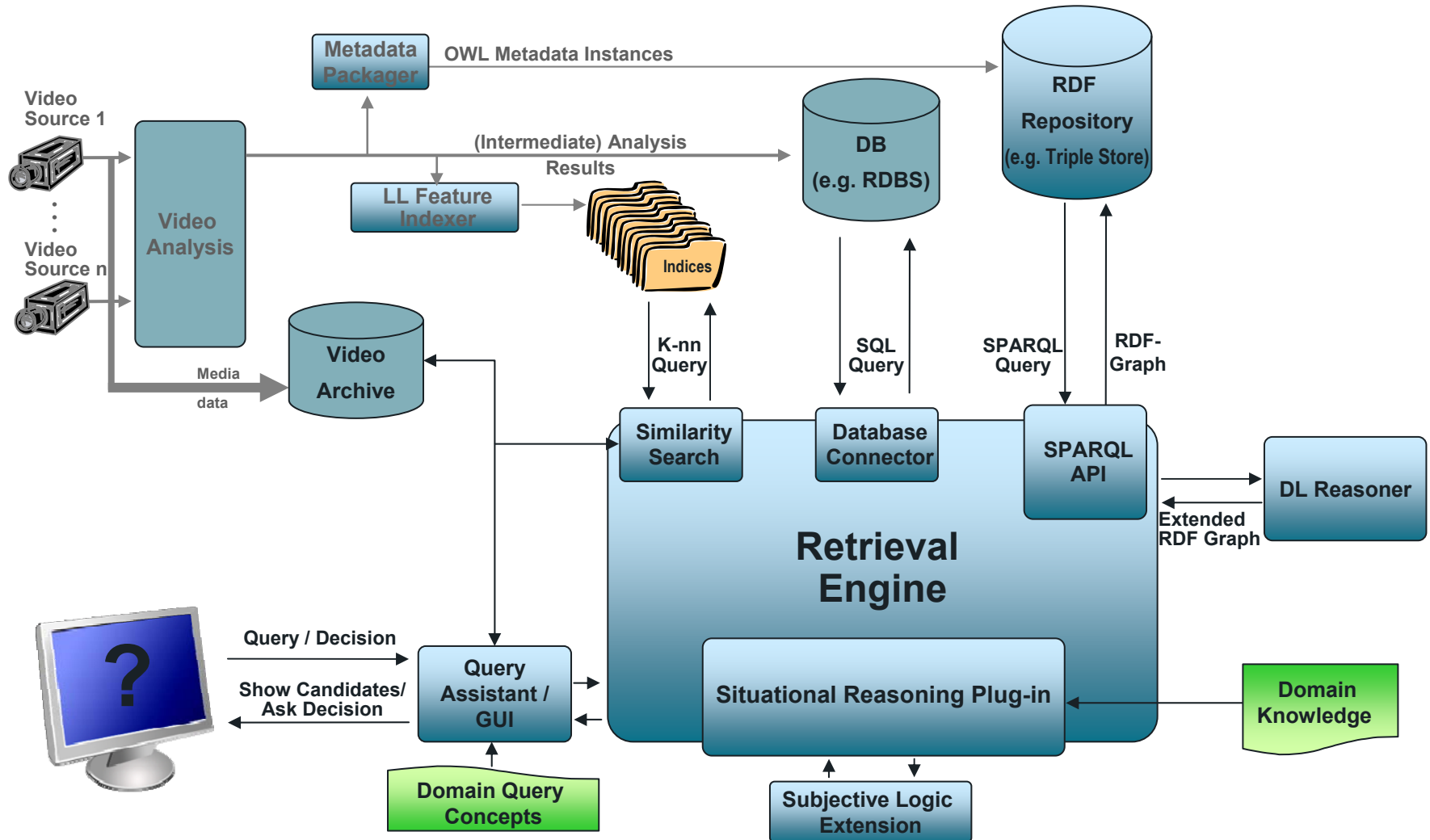
- » The required information may be distributed among different data base with different retrieval paradigms

System Architecture for Video Search



THESEUS

Forschungsprogramm für eine neue internetbasierte Wissensinfrastruktur



Show case “Wetten, dass ..?” TV programs



- » *Show me a picture of celebrity Anke Engelke ...*
- » *I'm interested in the most exciting bets of the “Wetten, dass ..?” TV programs. Please show me the Wettkönig-scores, a picture of each respective Wettkönig and her/his bet ...*

... based on the automatically extracted metadata and the summarizing annotation.

Available Video, Annotation and extracted Metadata:

- » 7 “Wetten, dass ..?” TV programs
 - » in total > 90 GB video data, ca. 18 hours
 - » approx. 150 guests and celebrities
 - » summarizing textual descriptions from ZDF archivists
- » Available metadata extractors
 - » Face-Detection (FhG – HHI)
 - » Shot-Detection (FhG – HHI)
 - » Logo-Detection (Siemens I MO IL)
- » Resulting in
 - » overall > 1,4 Mio detected faces, belonging to 47.500 Face-Id's,
 - in total ca. 400 MB metadata

```
<KAT>NF Reihe/Serie Wettbewerbsspiel/Quiz.</KAT>
<INHALTS>
<INHALT>XX:XX:XX:XX XX:XX:XX:XX XX:XX:XX</INHALT>
<INHALT>Live aus Freiburg mit Thomas #GOTTSCHALK.</INHALT>
</INHALTS>
<BILDINHS>
<BILDINH>20:18:05:00 XX:XX:XX:XX XX:XX:XX</BILDINH>
<BILDINH>Thomas #GOTTSCHALK begrüßt Dieter #THOMA (Ski Springer und</BILDINH>
<BILDINH>Stadtvetten-Repräsentant).</BILDINH>
<BILDINH>20:18:45:00 20:19:12:00 00:00:27</BILDINH>
<BILDINH>Zuspielteil: winkende Zuschauer auf dem Münsterplatz in Freiburg.</BILDINH>
<BILDINH>20:19:38:00 XX:XX:XX:XX XX:XX:XX</BILDINH>
<BILDINH>#GOTTSCHALK wettet, dass Freiburg es nicht schafft, 100</BILDINH>
<BILDINH>Toilettenüren aus studentischen Wohngemeinschaften auf den</BILDINH>
<BILDINH>Münsterplatz zu bringen (gelingt, Stadtvetten verloren).</BILDINH>
<BILDINH>20:21:53:00 20:22:25:00 00:00:32</BILDINH>
<BILDINH>Zuspielteil Schnittbilder Fußball-WM 2006 abwechselnd mit</BILDINH>
<BILDINH>Handball-WM 2006, deutsche Tore, Jubel Jürgen #KLINSMANN, Joachim</BILDINH>
<BILDINH>#LÖW, Heiner #BRAND.</BILDINH>
<BILDINH>20:22:36:00 XX:XX:XX:XX XX:XX:XX</BILDINH>
<BILDINH>Joachim #LÖW (Bundestrainer Fußball) und Heiner #BRAND</BILDINH>
<BILDINH>(Bundestrainer Handball) betreten Bühne.</BILDINH>
<BILDINH>20:23:12:00 20:28:27:00 00:05:15</BILDINH>
<BILDINH>Interview #GOTTSCHALK mit #BRAND und #LÖW über Umgang mit der</BILDINH>
<BILDINH>erhöhten öffentlichen Aufmerksamkeit, Handball-Euphorie in</BILDINH>
<BILDINH>Deutschland nach dem "Fußball-Sommer", Auswirkungen der Erfolge</BILDINH>
<BILDINH>auf die Nachwuchsarbeit, das Aussehen und die modische Kleidung von</BILDINH>
<BILDINH>LÖW, Bestreben der Fußballnationalmannschaft Europameister zu</BILDINH>
<BILDINH>werden, Kritik am Einsatz der B-Mannschaft beim Länderspiel gegen</BILDINH>
```





How to solve the sample queries ?

- » Interviews: Main persons (interviewer and interviewees/celebrities) are mentioned in textual summary
- » Appearance frequency of interviewee is higher than of interviewer (usually the answer is more detailed and prolonged than question)
- » Narrow down video footage to relevant shots (exploring textual summary)
- » Cluster similar faces and assign them the most probable person in accordance to their appearance frequency
- » Cascade this process to get person identities successively



Clustering of face-ID's according to visual similarity

Similarity measure:
Covariance descriptor on colour vector of pixels (hair and chest)

Face-ID's

469	492	494	...	
470	480	491	496	..
472	478	...		
:				

Ranking according to the number of contained frames

JAUCH

2171 Faces



Cluster 1

376 Faces



Cluster 2

107 Faces



Cluster 3

42 Faces



Cluster 4

Cluster 5

Cluster 6

Cluster 7

Cluster 8

Cluster 8

- » Annotated faces are stored
- » Identity suggestion for new/unknown faces
 - » Similarity ranking to stored faces



FID_496

0.18

0.68

1.35

Identity storage



Jauch



Steiner



Gottschalk



- » Identity model is refined by added faces



Bundesministerium
für Wirtschaft
und Technologie

Video Source View



Play Pause Replay Stop Demo Prop.

Browser Parameters

Show Query Scene Info

Learned Persons expected in this Scene



GOTTSCHALK



JAUCH



ROCK



STEINER



ENGELKE



KRIENER

Find my Star

[expand all](#) [load video overlay](#)

ENGELKE

1594 Faces | store cluster? ✓ ENGELKE



KRIENER

178 Faces | Matching Distance: 0,000 | store cluster? ✓ KRIENER



KRIENER

GOTTSCHALK

JAUCH

ENGELKE



- » Exploitation of (inter-)relations between “low-level” metadata, the situational context, and the application context necessary for answering semantic queries
- » Role-based identity examination in video sequences is a good example for this
- » Harmonized metadata description schemas desired (at least a core set) to enhance interoperability in media search (cf. JPSearch)
- » Standardized Query Language for querying distributed media archives (cf. MPQF / JPSearch)
- » Confidence values are necessary for image-based metadata (inherent uncertainty of image analysis)



Thank you !