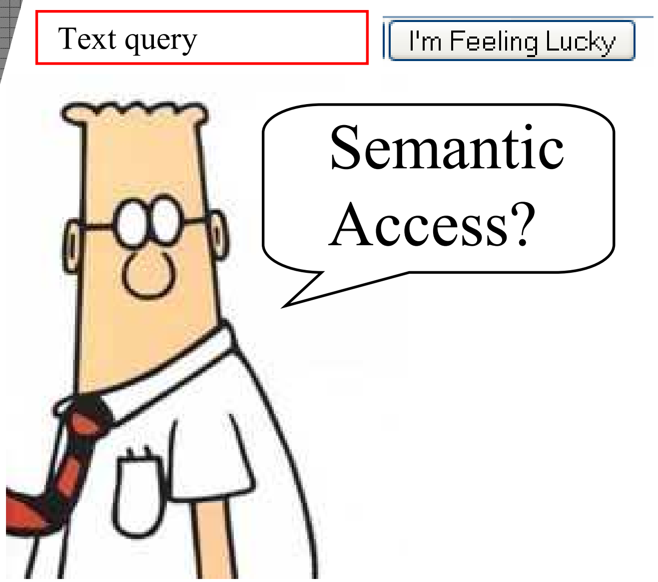
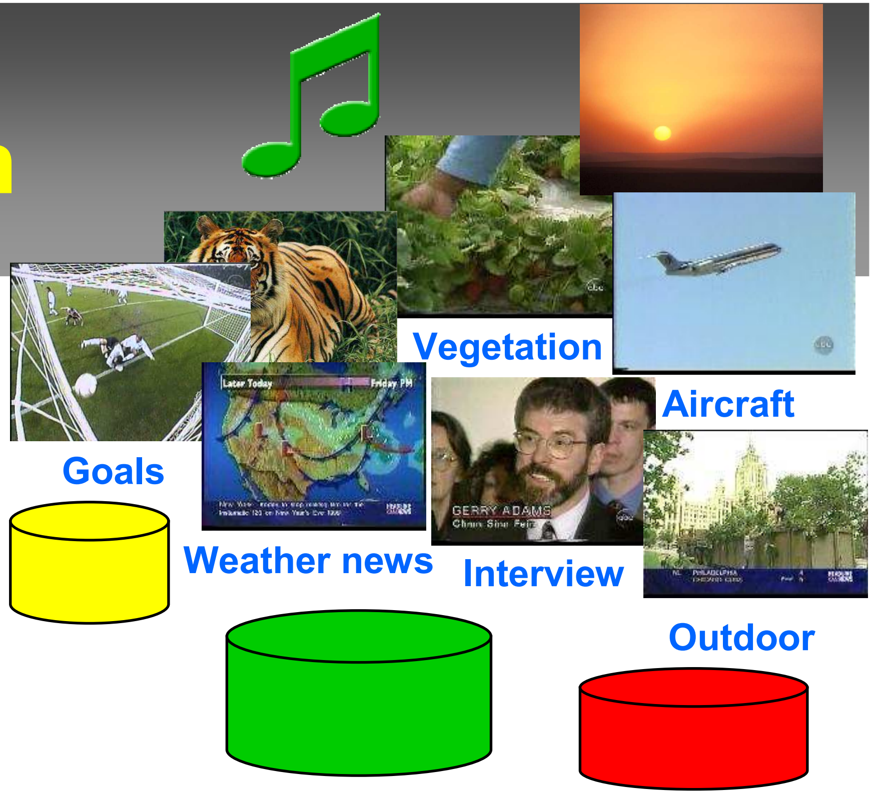
# Weak Retrieval
## from Multimedia Archives

**Cees Snoek, Marcel Worring, Jan-Mark Geusebroek, Dennis Koelma and Frank Seinstra**

MediaMill, University of Amsterdam, The Netherlands

# Introduction




Multimedia Archives

- Generic concept detection required
  - ✓ We propose: Semantic Value Chain

Semantic Gap

- Interactive retrieval required
  - ✓ We propose: Semantic Video Search Engine

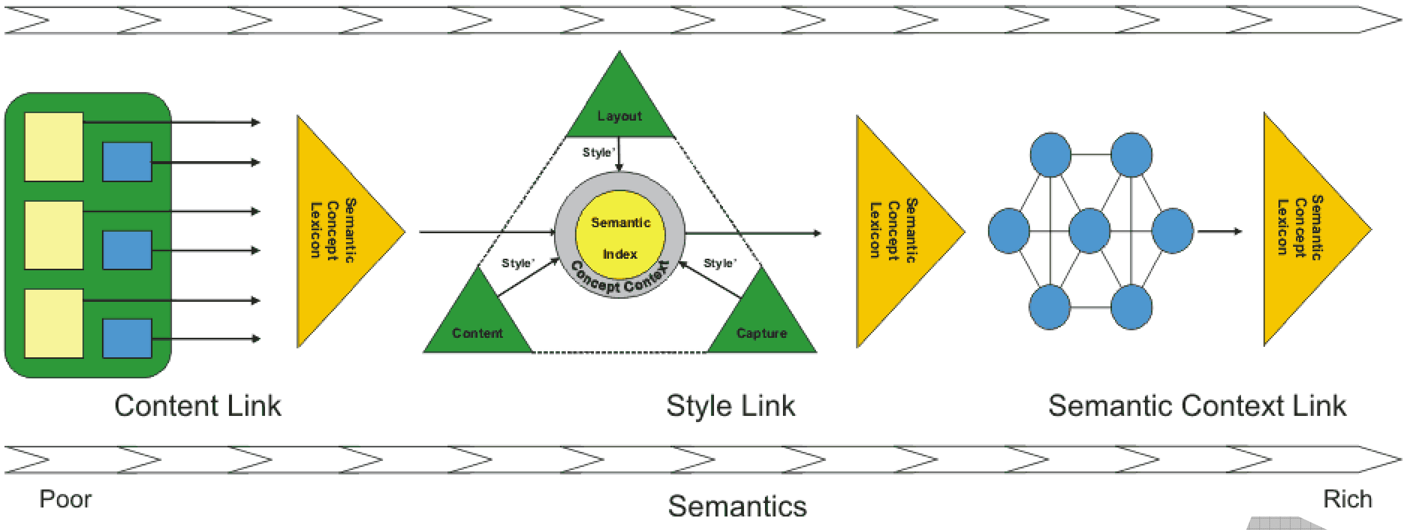
} Weak Retrieval

# Semantic Value Chain

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion

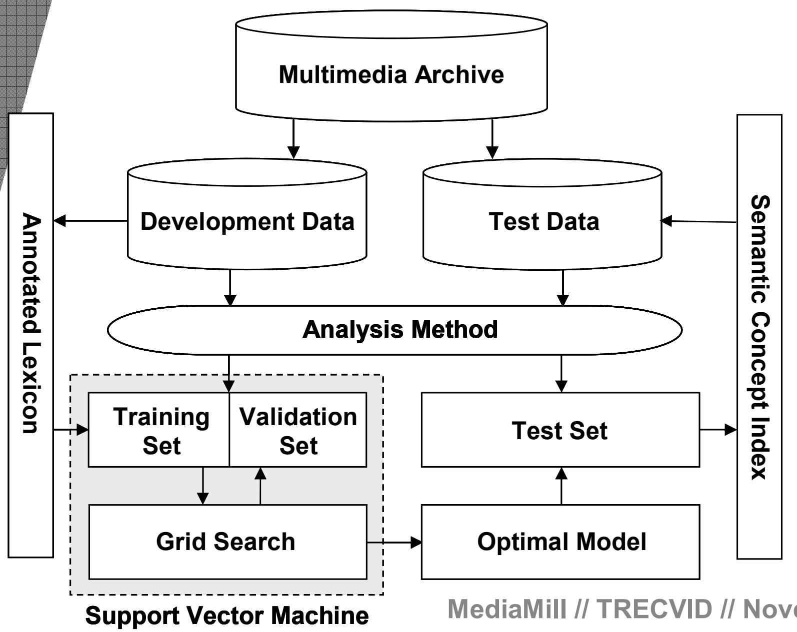
➢ Produced video is result of an authoring process
  - ✓ Author starts with semantic intention
  - ✓ Author uses style elements to convey semantic message
  - ✓ Author produces digital multimedia document

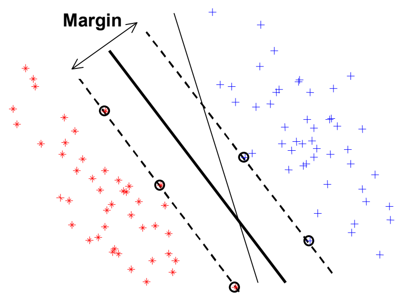
➢ Video analysis is inverse authoring process



# General Link Architecture

- Concept detection is pattern recognition problem
  - ✓ Each link obtains a vector pattern representation
  - ✓ Exploits annotated lexicon for supervised learning
- Use Support Vector Machine for classification
  - ✓ Grid search for optimal SVM parameters
  - ✓ Split development data in different sets




$$p(\omega|\vec{s}) = \frac{1}{1+\exp(\alpha\gamma(\vec{s})+\beta)}$$

# Annotated Lexicon

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion


Animal


Footbal


Road


Beach

Stock Quotes


Golf


Financial Anchor


Cartoon


Building


Airplane Take Off


Boat

Graphic


People


Car


Vegetation

Overlayed Text


Basket Scored


Bill Clinton


Sporting Event


Studio Setting


Physical Violence


Train


Baseball


News Subject Monologue


Anchor


Outdoor


Ice Hockey


People Walking


Madeleine Albright


Soccer


Bicycle


Weather News

# Content Link

References:
Geusebroek et al, PAMI '01
Seinstra & Koelma, CCPE, '04
Gauvain et al, Sp. Comm. '02


Das-2

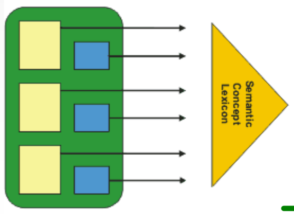
## Unimodal

- ✓ 1 out of 15 frames were reduced to vector *i*, of 18 visual concepts (sky,grass,…) using color invariant features
- ✓ Cluster reduced processing from +/- 250 days to 48 hours
- ✓ Learn word associations between speech and concept
- ✓ Construct frequency histogram *t*

## Multimodal

- ✓ Concatenate *t* with *i* into fused vector *m*
- ✓ Rank concepts based on validation set performance

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 1. | Weather News | 9. | People Walking | 17. | Golf | 25. | Road |
| 2. | Stock Quotes | 10. | Financial Anchor | 18. | People | 26. | Beach |
| 3. | Anchor | 11. | Ice Hockey | 19. | American Football | 27. | Train |
| 4. | Overlayed Text | 12. | Cartoon | 20. | Outdoor | 28. | M. Albright |
| 5. | Basketball | 13. | Studio Setting | 21. | Car | 29. | Building |
| 6. | Graphics | 14. | Physical Violence | 22. | B. Clinton | 30. | Aircraft |
| 7. | Baseball | 15. | Vegetation | 23. | News Subj. Monologue | 31. | Bicycle |
| 8. | Sporting Event | 16. | Boat | 24. | Animal | 32. | Soccer |

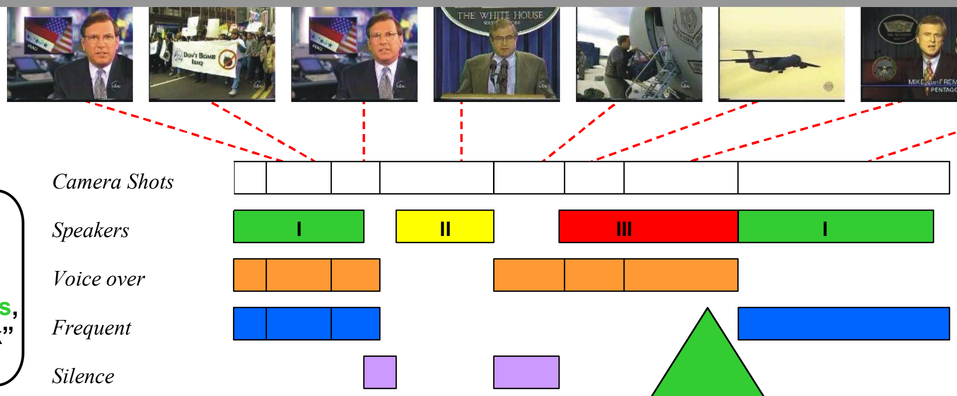



Content Link

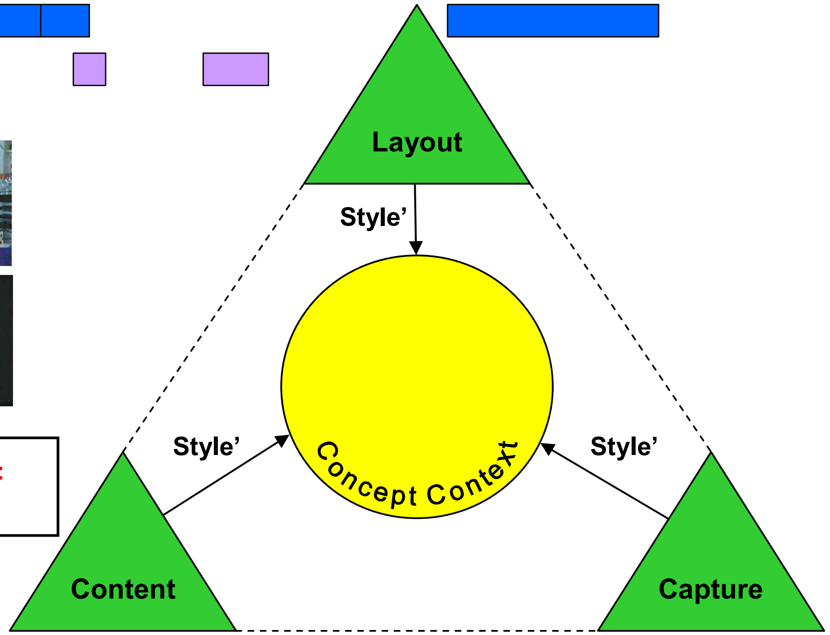
# Style Link

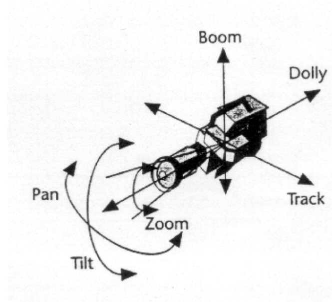
References:
Schneiderman et al, IJCV '04
Gauvain et al, Sp. Comm. '02
Informedia, CMU

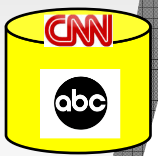
- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion




"Ann Compton ABC news, New York"



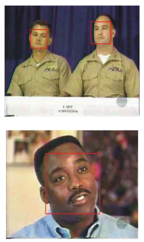



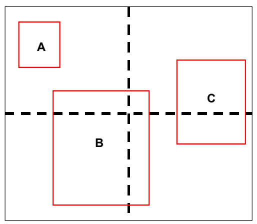
match?

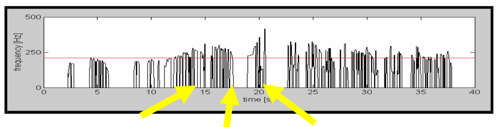
Jemmy Curter

isName?

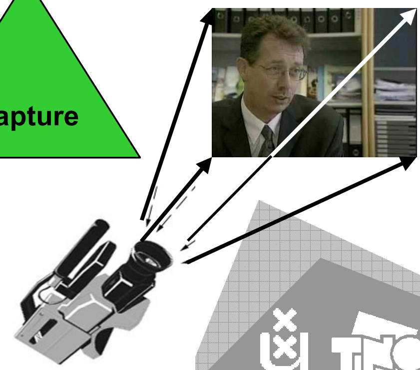
"Call now: 1800…"





…fast enough? Home run!



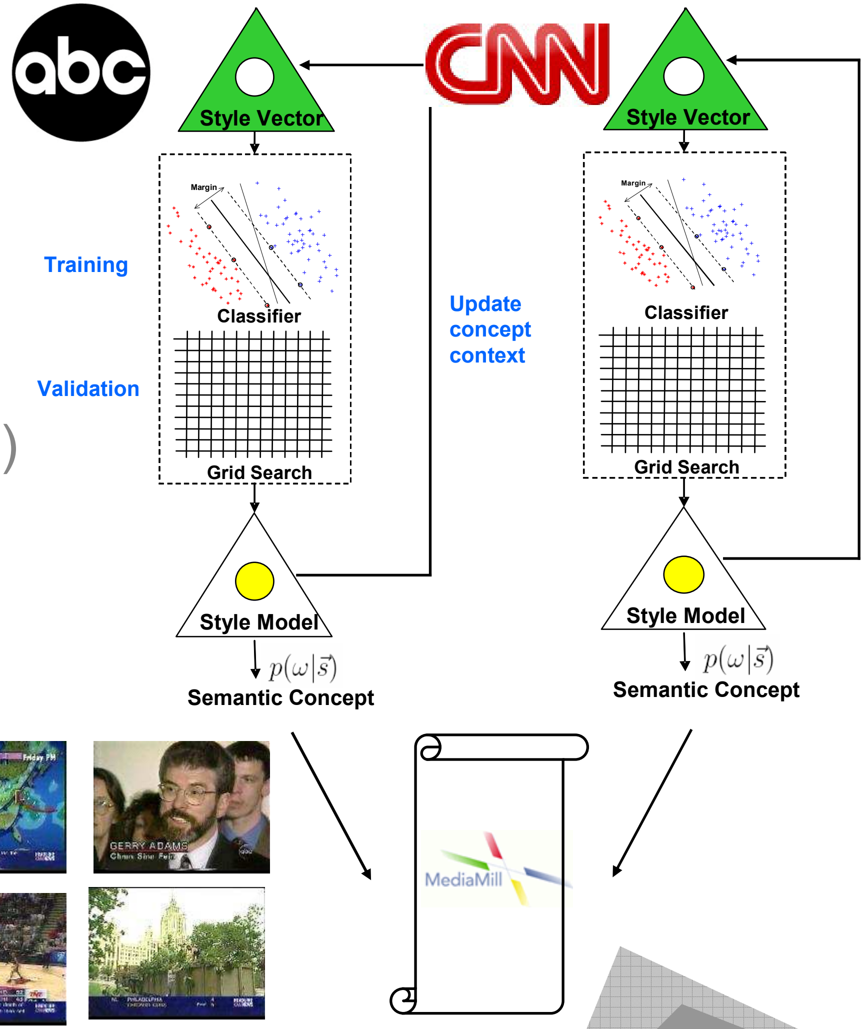
# Style Link Experiments

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion

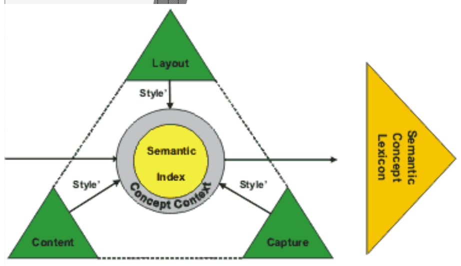
➢ **Concept Order**
  - ✓ AC1 vs AC2

➢ **Concept Threshold**
  - ✓ AC1 (0.5) vs AC3 (0.1)

➢ **Station Separation**
  - ✓ AC1 vs COM




Ranked list



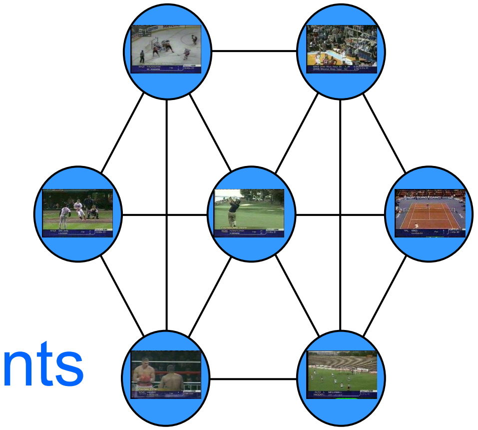

Style Link

References:
Iyengar et al, ACM MM '03
Wu et al, ICME '04

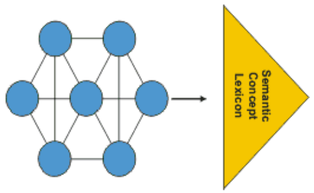
# Semantic Context Link


- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion


- ➢ Relies on semantic concept probabilities only
- ➢ We explore different context configurations
  - ✓ Stacked SVM: one 32 dimensional concept vector
  - ✓ Ontology: rescore concept probability based on co-occurrence



- ➢ Semantic Context Link Experiments
  - ✓ Influence of Style Link (CC vs R4)
  - ✓ Difference between two style runs (R4 vs R5)
  - ✓ Difference between context learning and rule based ontology (R5 vs OR5)



Semantic Context Link

- Introduction
- Semantic Value Chain
- **Results**
- Semantic Search Engine
- Results
- Conclusion

# Results - AP

| | Content Link | | Style Link | | | | Semantic Context Link | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | BU | VF | AC1 | AC2 | AC3 | COM | CC | R4 | R5 | OR5 |
| Boat | 0.108 | **0.117** | 0.096 | 0.094 | 0.098 | 0.101 | 0.084 | 0.070 | 0.096 | 0.042 |
| M. Albright | **0.238** | 0.136 | 0.023 | 0.027 | 0.021 | 0.035 | 0.000 | 0.015 | 0.018 | 0.021 |
| B. Clinton | 0.123 | 0.130 | 0.150 | **0.156** | 0.154 | 0.160 | 0.135 | 0.149 | 0.155 | 0.105 |
| Train | **0.083** | 0.054 | 0.062 | 0.050 | 0.074 | 0.072 | 0.041 | 0.005 | 0.004 | 0.023 |
| Beach | 0.008 | 0.020 | 0.017 | 0.011 | 0.010 | **0.021** | 0.010 | 0.012 | 0.010 | 0.006 |
| Basket scored | 0.017 | 0.118 | 0.180 | **0.214** | 0.200 | 0.141 | 0.174 | 0.209 | 0.193 | 0.194 |
| Airplane take off | 0.051 | 0.065 | 0.037 | 0.042 | **0.073** | 0.043 | 0.052 | 0.040 | 0.050 | 0.037 |
| People walking | 0.134 | 0.150 | 0.159 | 0.151 | 0.138 | 0.166 | 0.139 | **0.170** | 0.168 | 0.106 |
| Phys. violence | 0.062 | 0.064 | 0.071 | 0.052 | 0.076 | 0.067 | 0.080 | **0.086** | 0.069 | 0.064 |
| Road | 0.073 | 0.089 | 0.129 | 0.135 | 0.135 | 0.118 | 0.080 | 0.138 | **0.141** | 0.120 |
| MAP | 0.090 | 0.094 | 0.093 | 0.095 | **0.096** | 0.093 | 0.078 | 0.089 | 0.090 | 0.072 |

- Observations:
  - ✓ Fusion (VF) outperforms unimodal (BU)
  - ✓ Style Link: concept order, threshold, and combination matter
  - ✓ Influence of Style Link is evident: R4 14% better than CC
  - ✓ Learning improves upon rule based ontology with 25%
  - ✓ MAP fair measure for comparison?

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion

# Results – P@100

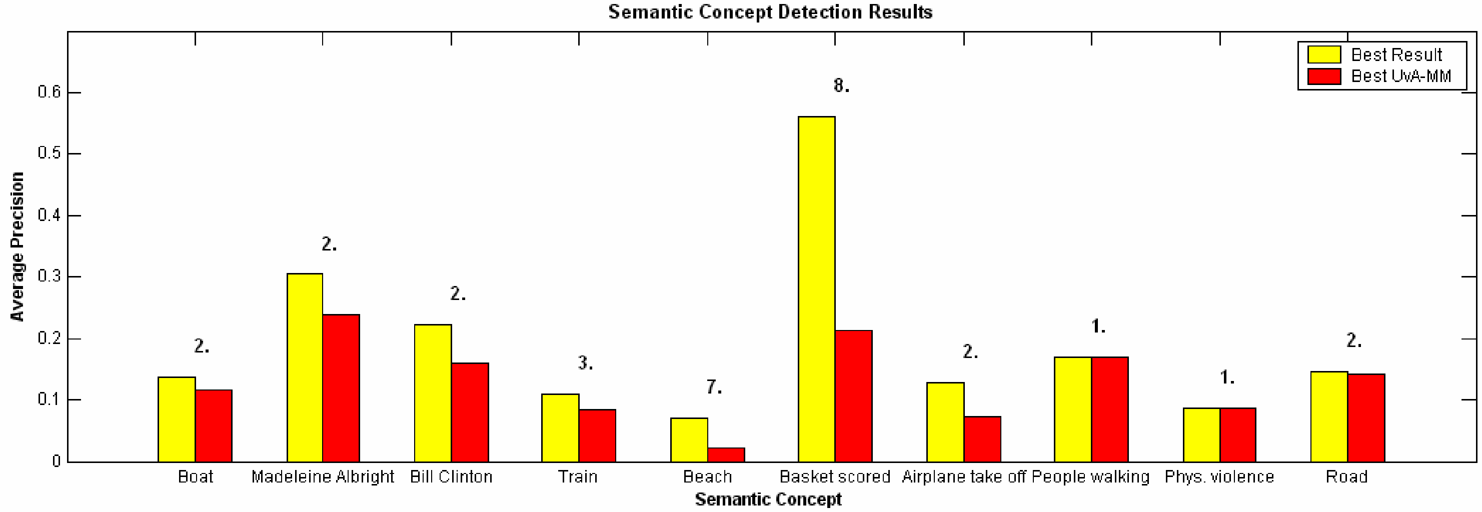
| | Content Link | | Style Link | | | | Semantic Context Link | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | BU | VF | AC1 | AC2 | AC3 | COM | CC | R4 | R5 | OR5 | Avg |
| Boat (441) | **44** | 42 | 38 | 36 | 40 | 39 | 34 | 37 | 39 | 37 | 38.6 |
| M. Albright (19) | 10 | **12** | 5 | 6 | 5 | 5 | 0 | 4 | 4 | 5 | 5.6 |
| B. Clinton (409) | 25 | 26 | 35 | 32 | 34 | 36 | 26 | 37 | **47** | 33 | 32.5 |
| Train (43) | **9** | 7 | 7 | 6 | 8 | 8 | 7 | 3 | 2 | 5 | 6.2 |
| Beach (374) | 10 | 13 | 12 | 11 | 9 | **14** | 9 | 12 | 11 | 9 | 11.0 |
| Basket scored (103) | 8 | 24 | 21 | **35** | 30 | 21 | 26 | 30 | 33 | 33 | 26.1 |
| Airplane take off (62) | 9 | 10 | 8 | **11** | 9 | 9 | 10 | 8 | 10 | 10 | 9.4 |
| People walking (1695) | 65 | 65 | 72 | 57 | 65 | 71 | 68 | **83** | 77 | 54 | 67.7 |
| Phys. violence (292) | 17 | 17 | 25 | 18 | 13 | 24 | 17 | **31** | 23 | 19 | 20.4 |
| Road (938) | 47 | 43 | 53 | 51 | 53 | 41 | 41 | 51 | **55** | 52 | 48.7 |
| Avg | 24.4 | $\boxed{25.9}$ | $\boxed{27.6}$ | 26.3 | 26.6 | 26.8 | 23.8 | $\boxed{\mathbf{29.6}}$ | 29.5 | 25.7 | 26.6 |

➢ Observations:

- ✓ Fusion (VF) outperforms unimodal (BU)
- ✓ Style Link: concept order, threshold, and combination matter
- ✓ Influence of Style Link is evident: R4 14% better than CC
- ✓ Learning improves upon rule based ontology with 25%
- ✓ MAP fair measure for comparison?
- ✓ SVC improves P@100 concept detection performance

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion

# Feature Extraction Task

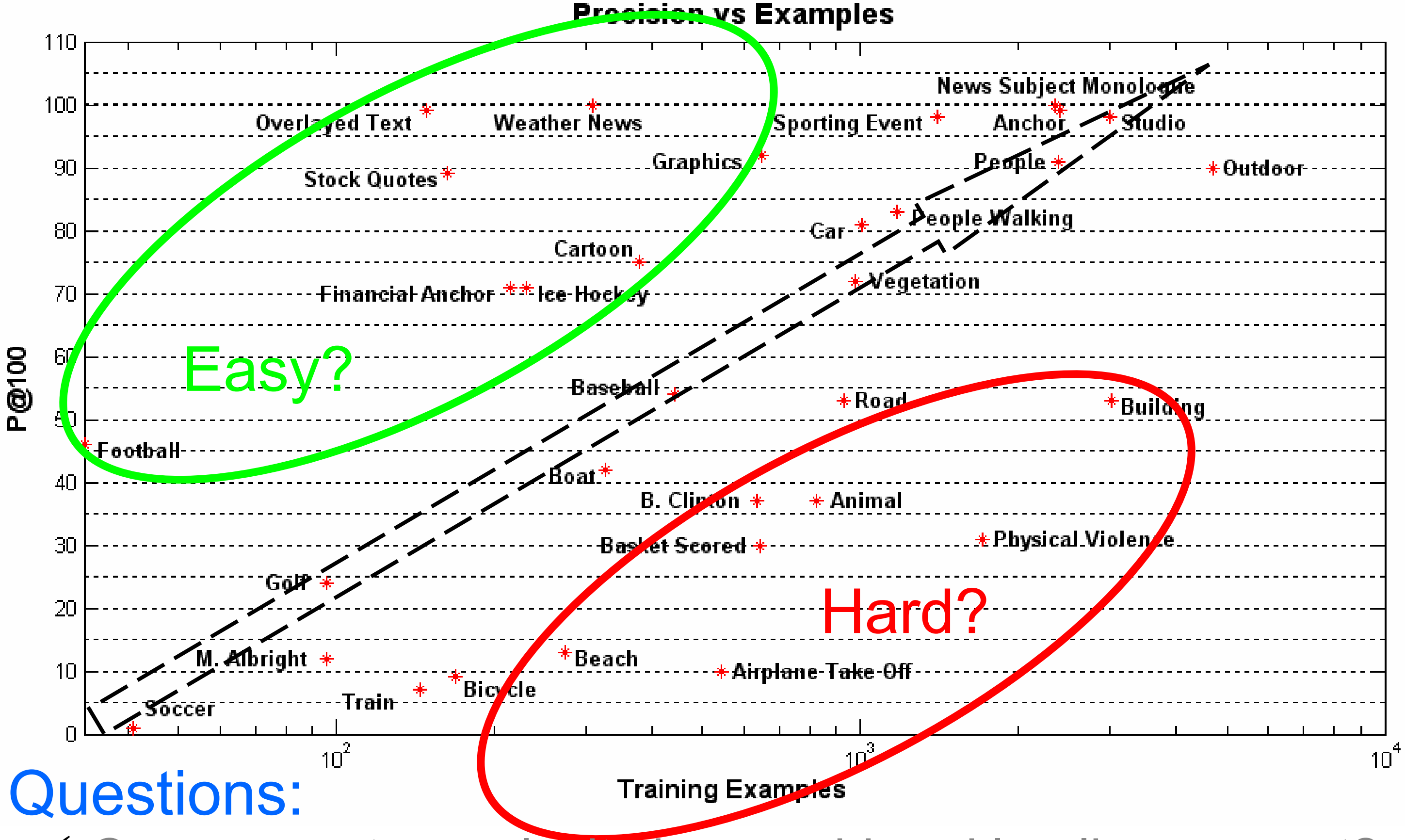



➢ Observations:
- ✓ Best performer for *People Walking* and *Physical Violence*
- ✓ Top 3 system ranking in 8 concepts
- ✓ Good performance for setting, objects, people, events
- ✓ Semantic Value Chain allows for generic semantic concept detection

References:
Naphade & Smith, ACM MM'04

# Concept Complexity

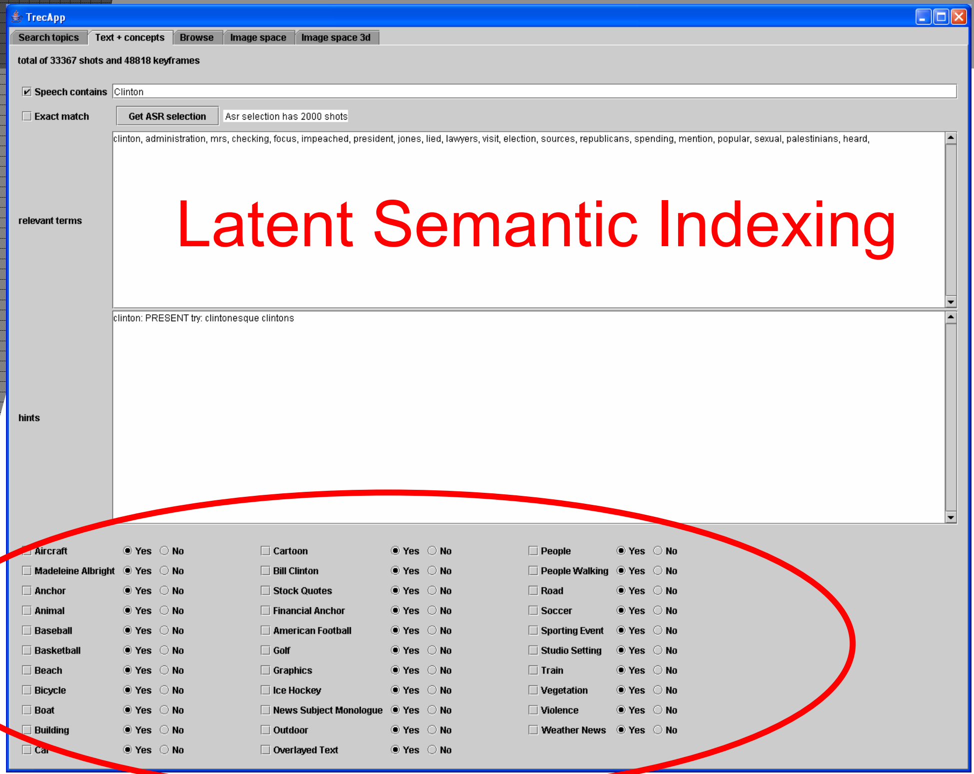





- **Questions:**
  - ✓ Can concept complexity be considered log-linear or not?
  - ✓ How about easy and hard semantic concepts?
  - ✓ Can we define a technique taxonomy?
  - ✓ 1000 examples a safe upper bound?

- Introduction
- Semantic Value Chain
- Results
- **Semantic Search Engine**
- Results
- Conclusion

# Semantic Search Engine

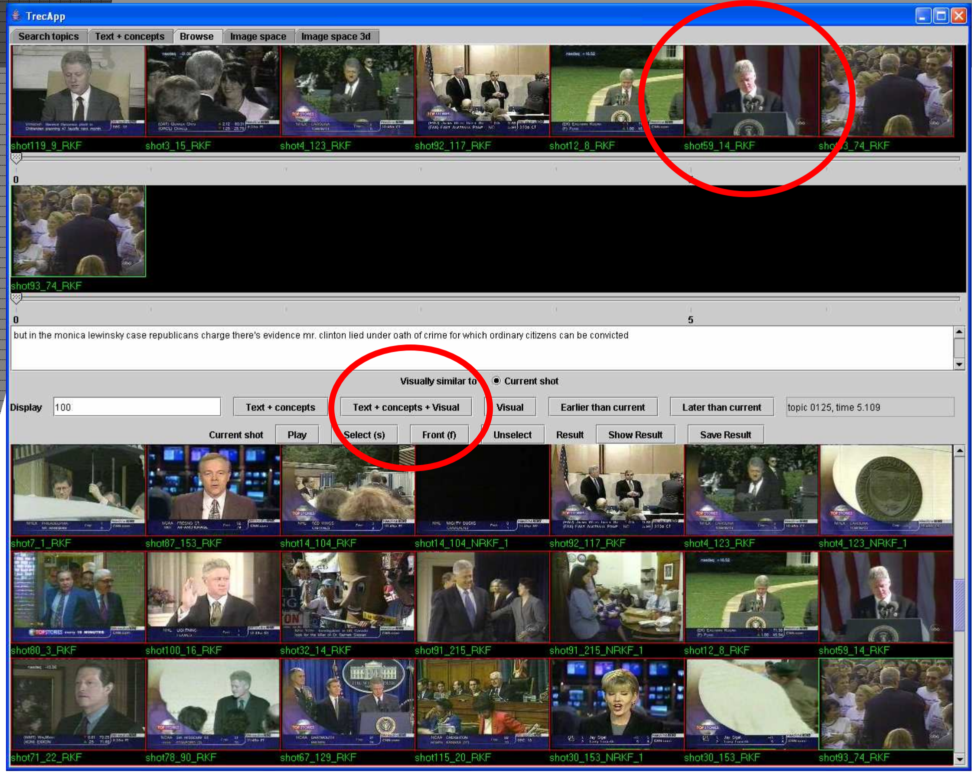



➢ Features:

- ✓ Query by concept (using lexicon)
- ✓ Query by keyword (using LSI space)

- Introduction
- Semantic Value Chain
- Results
- **Semantic Search Engine**
- Results
- Conclusion

# Semantic Search Engine

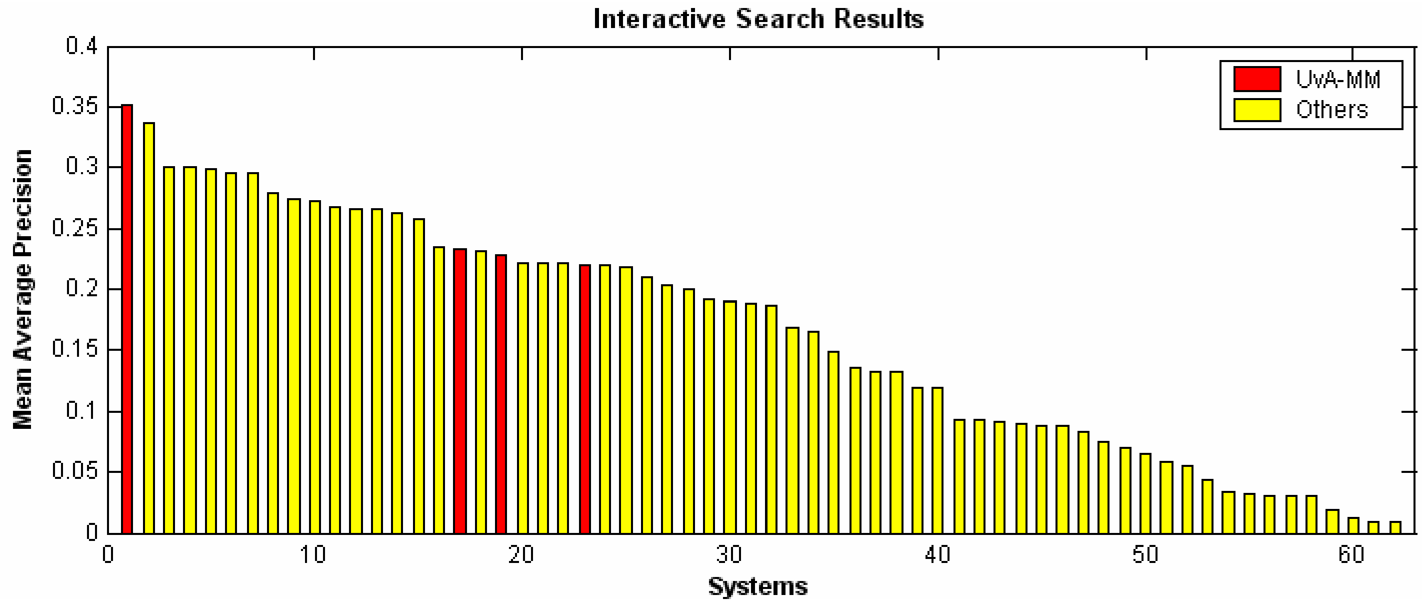



User selected results

Storyboard query results

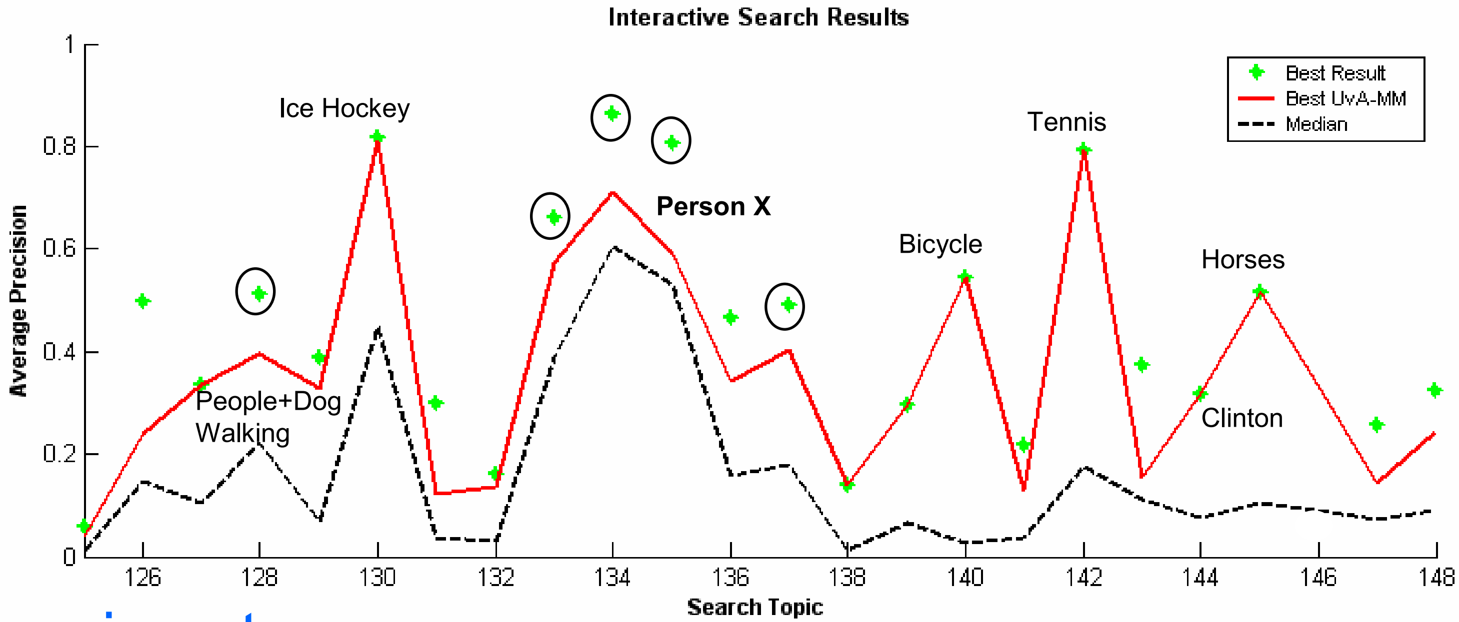
- Features:
  - ✓ Query by concept (using lexicon)
  - ✓ Query by keyword (using LSI space)
  - ✓ Query by similarity (using *Lab* space)

# Interactive Search Task




- Experiment:
  - 4 expert users: 1 with knowledge of semantic concepts
- Observations:
  - Best MAP of all interactive systems

# Interactive Search Task

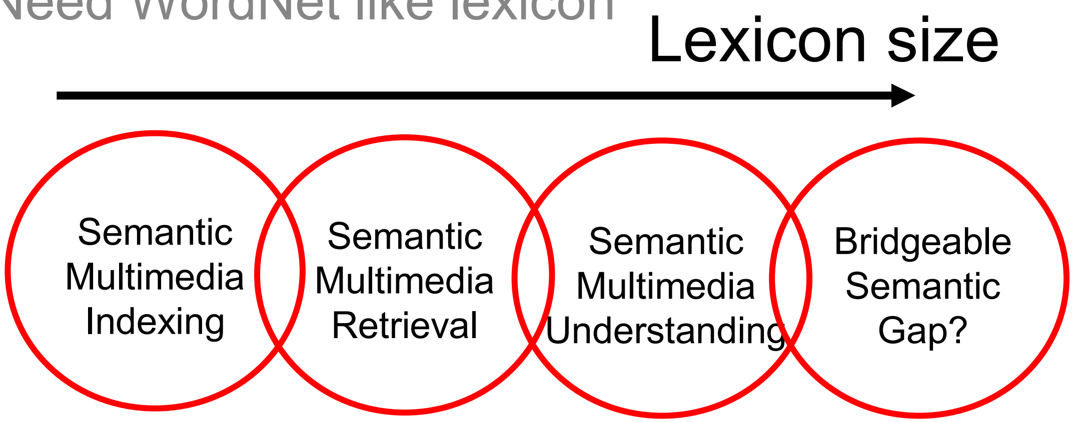



- **Experiment:**
  - ✓ 4 expert users: 1 with knowledge of semantic concepts
- **Observations:**
  - ✓ Best MAP of all interactive systems
  - ✓ Lexicon aids in retrieval when overlap in topics exists
  - ✓ Query by keyword/similarity aids in refining results
  - ✓ More concepts needed, e.g. person x related

- Introduction
- Semantic Value Chain
- Results
- Semantic Search Engine
- Results
- Conclusion

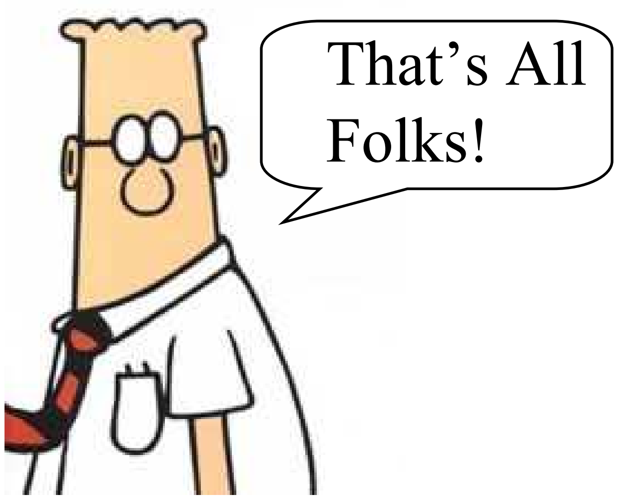
# Conclusions & Future Work

- Weak retrieval allows for semantic access
  - ✓ Video analysis is inverse authoring process
  - ✓ Semantic Value Chain facilitates generic concept detection
  - ✓ Limited lexicon boosts interactive retrieval results
  - ✓ Combine query by concept, similarity, & keyword
- Future work
  - ✓ bigger, better, more…
  - ✓ Need WordNet like lexicon

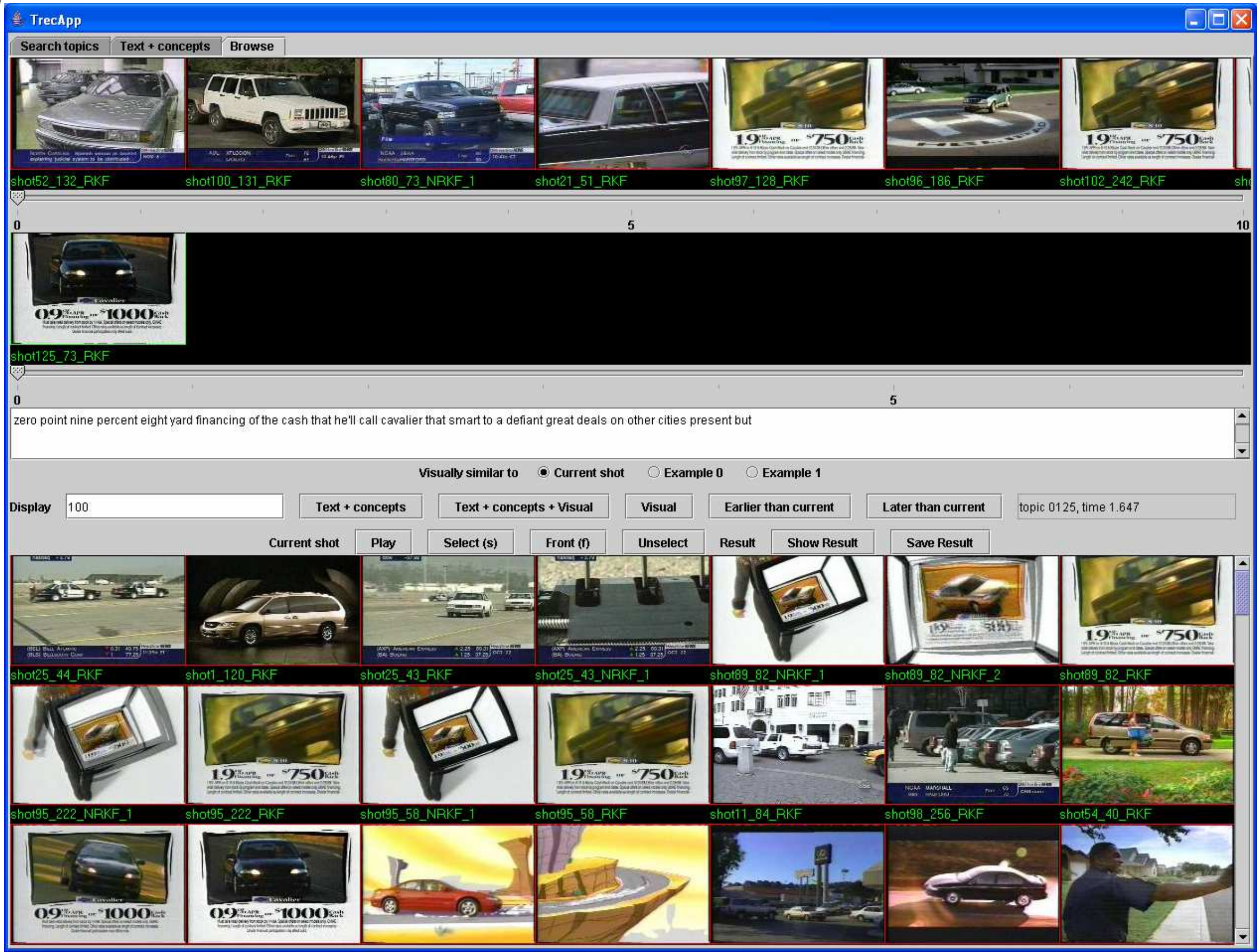


# Acknowledgement

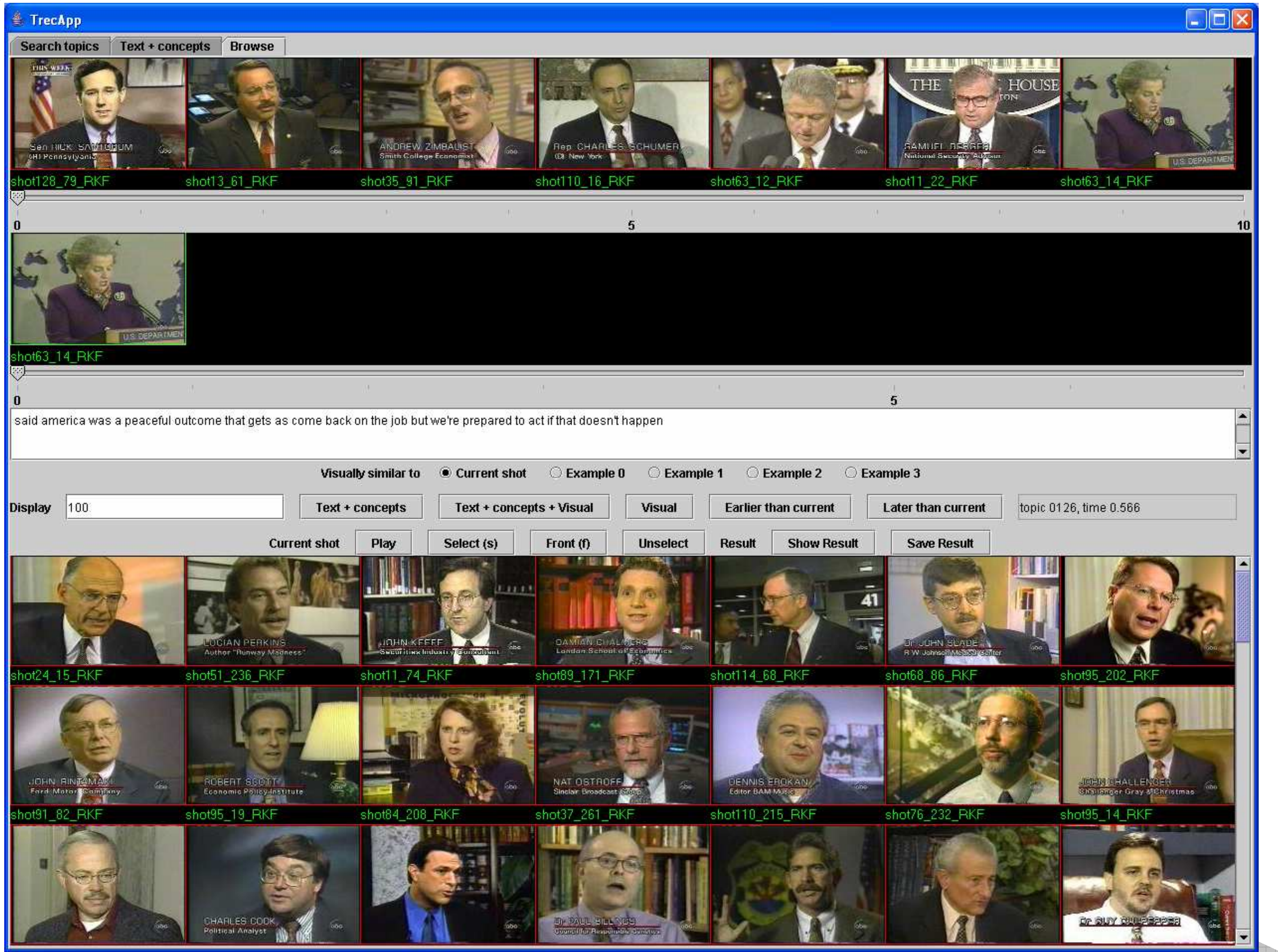

- NIST and coordinators for organization and evaluation effort
- Kees Verstoep, Vrije Universiteit Amsterdam, for DAS-2 support
- Giang Nguyen and Jan van Gemert, University of Amsterdam, for search experiment participation

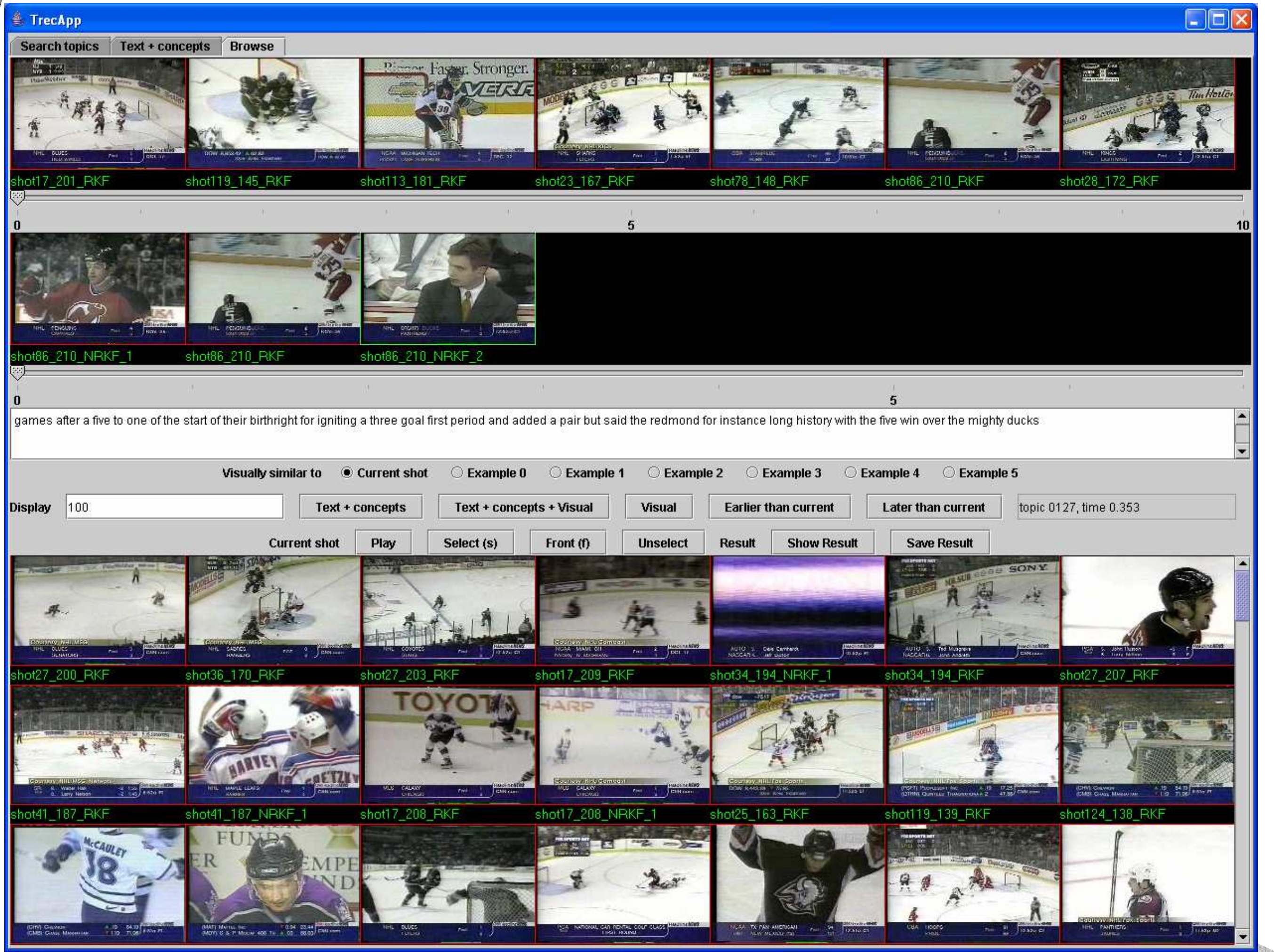
# Demo - Car


TrecApp
Search topics
Text + concepts
Browse
shot52_132_RKF
shot100_131_RKF
shot80_73_NRKF_1
shot21_51_RKF
shot97_128_RKF
shot96_186_RKF
shot102_242_RKF
shot125_73_RKF
zero point nine percent eight yard financing of the cash that he'll call cavalier that smart to a defiant great deals on other cities present but
Visually similar to
Current shot
Example 0
Example 1
Display
100
Text + concepts
Text + concepts + Visual
Visual
Earlier than current
Later than current
topic 0125, time 1.647
Current shot
Play
Select (s)
Front (f)
Unselect
Result
Show Result
Save Result
shot25_44_RKF
shot1_120_RKF
shot25_43_RKF
shot25_43_NRKF_1
shot89_82_NRKF_1
shot89_82_NRKF_2
shot89_82_RKF
shot95_222_NRKF_1
shot95_222_RKF
shot95_58_NRKF_1
shot95_58_RKF
shot11_84_RKF
shot98_256_RKF
shot54_40_RKF

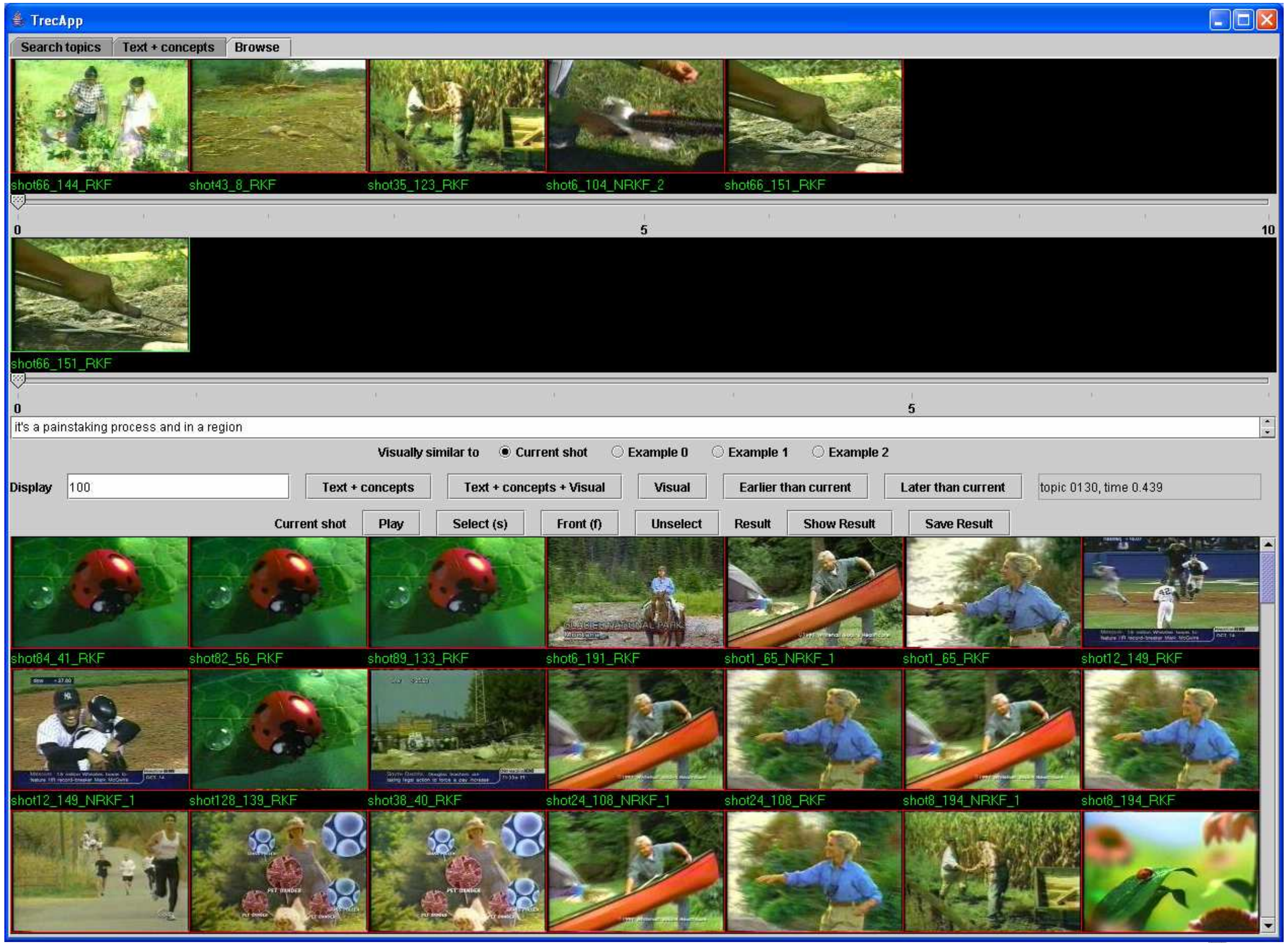
# Demo - TV News Monologue



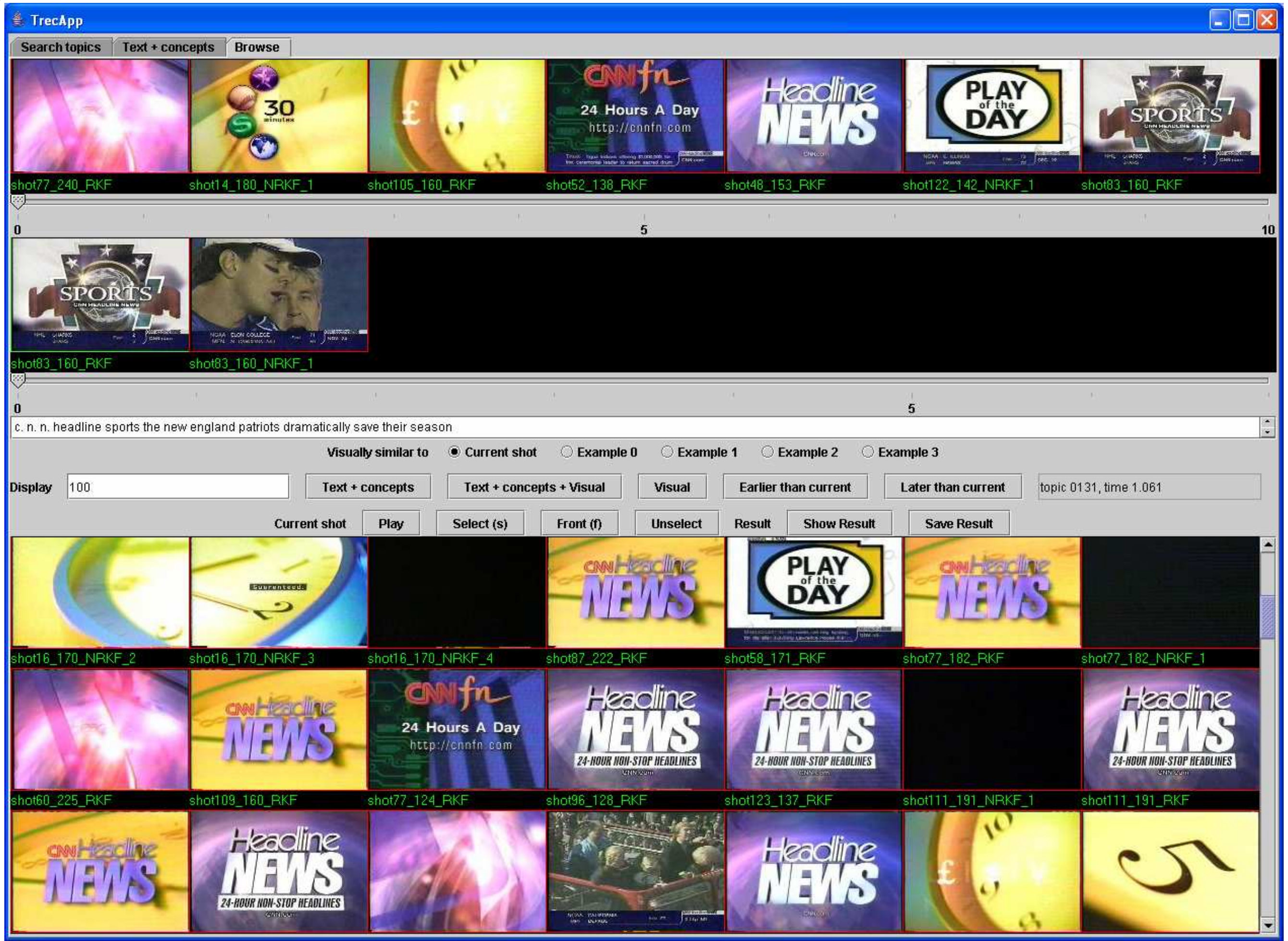
# Demo - Ice Hockey



# Demo - Vegetation



# Demo - Graphics



# Contact Info

- Cees G.M. Snoek, M.Sc.

  Informatics Institute
  University of Amsterdam
  Kruislaan 403
  NL-1098 SJ Amsterdam
  The Netherlands

  E: cgmsnoek@science.uva.nl
  T: +31-(0)20-525-7528
  F: +31-(0)20-525-7490
  W: http://www.science.uva.nl/~cgmsnoek/