



# CuZero: Interactive Video Search for Informed User

---

***Shih-Fu Chang***  
***(joint work with Eric Zavesky)***

**Digital Video and Multimedia Lab**  
**Columbia University**

**[www.ee.columbia.edu/dvmm](http://www.ee.columbia.edu/dvmm)**

**Oct. 2008**

# The Goggle Era Search Paradigm

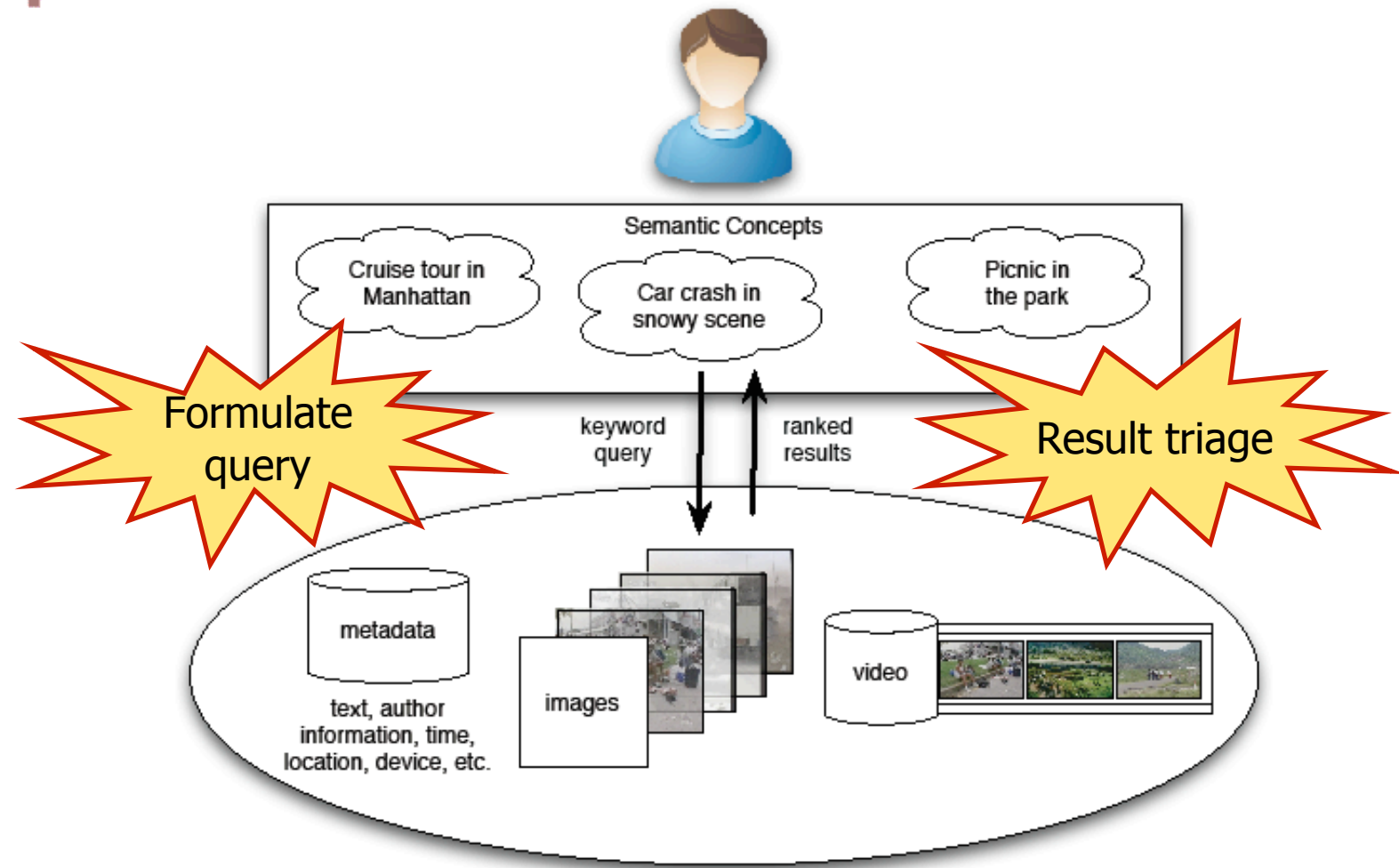
“...type in a few words at most, then expect the engine to bring back the perfect results. More than 95 percent of us never use the advanced search features most engines include, ...”

– *The Search*, J. Battelle, 2003

- Simple keyword search is the primary search method
- Extension to visual search seems to be a natural desirable step
- Examples: text search on Google, Flickr, Youtube

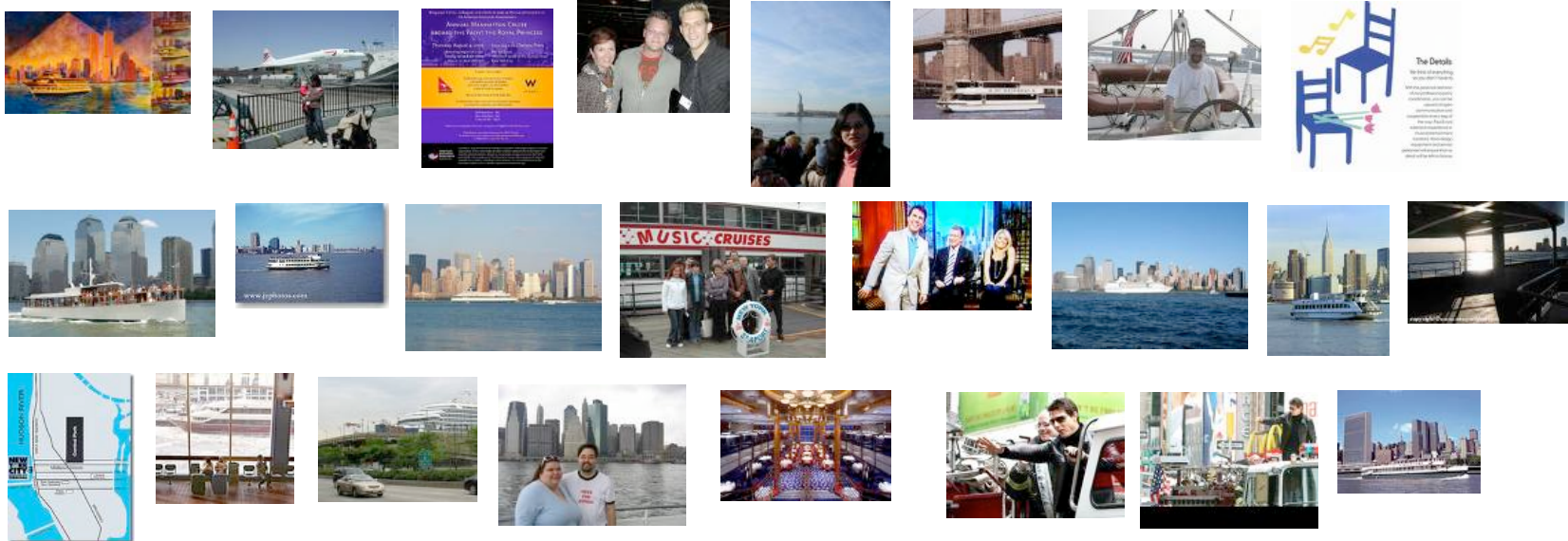


# The Role of User



# An Example of Web Image Search

- Text Query : “*Manhattan Cruise*” over Goggle Image



- How to visualize large image/video sets?
- Why are these images returned? No word-document matching
- If this does not work, how do we choose better search terms?



# Minor Changes in Keywords → Big Difference

- Text Query : ***“Cruise around Manhattan”***

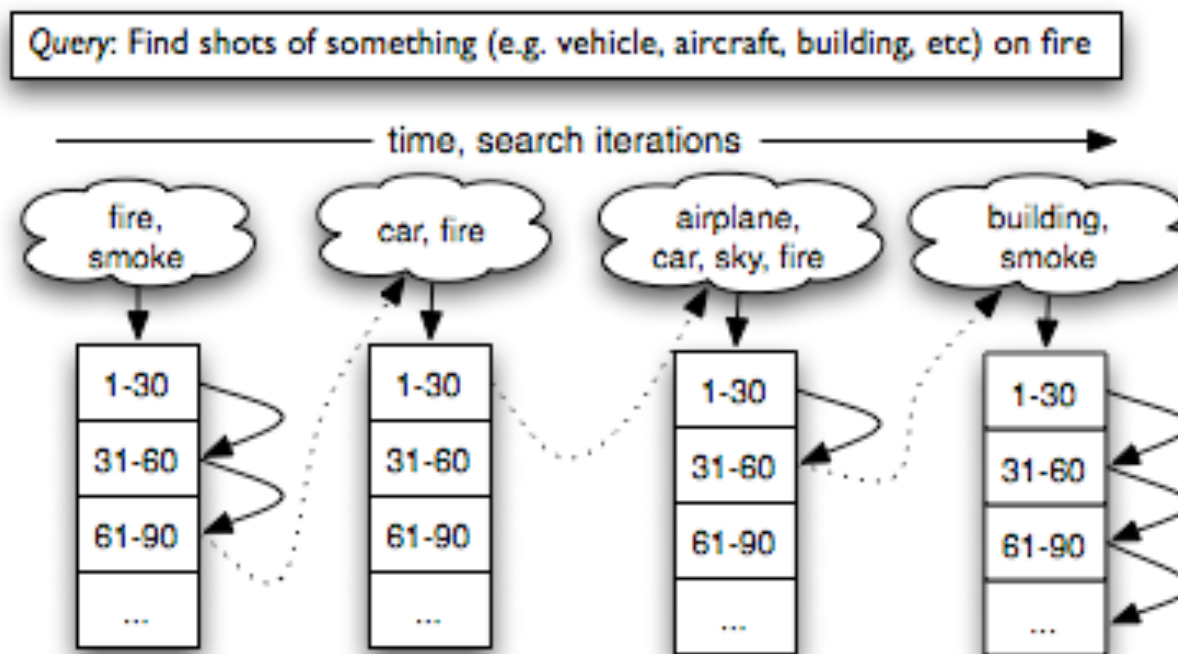


»digital video | multimedia lab»

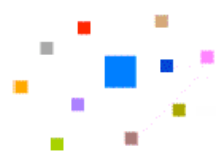


# Pains of Frustrated Users

- Forced to take “one shot” searches, iterating queries with a trial and error approach...

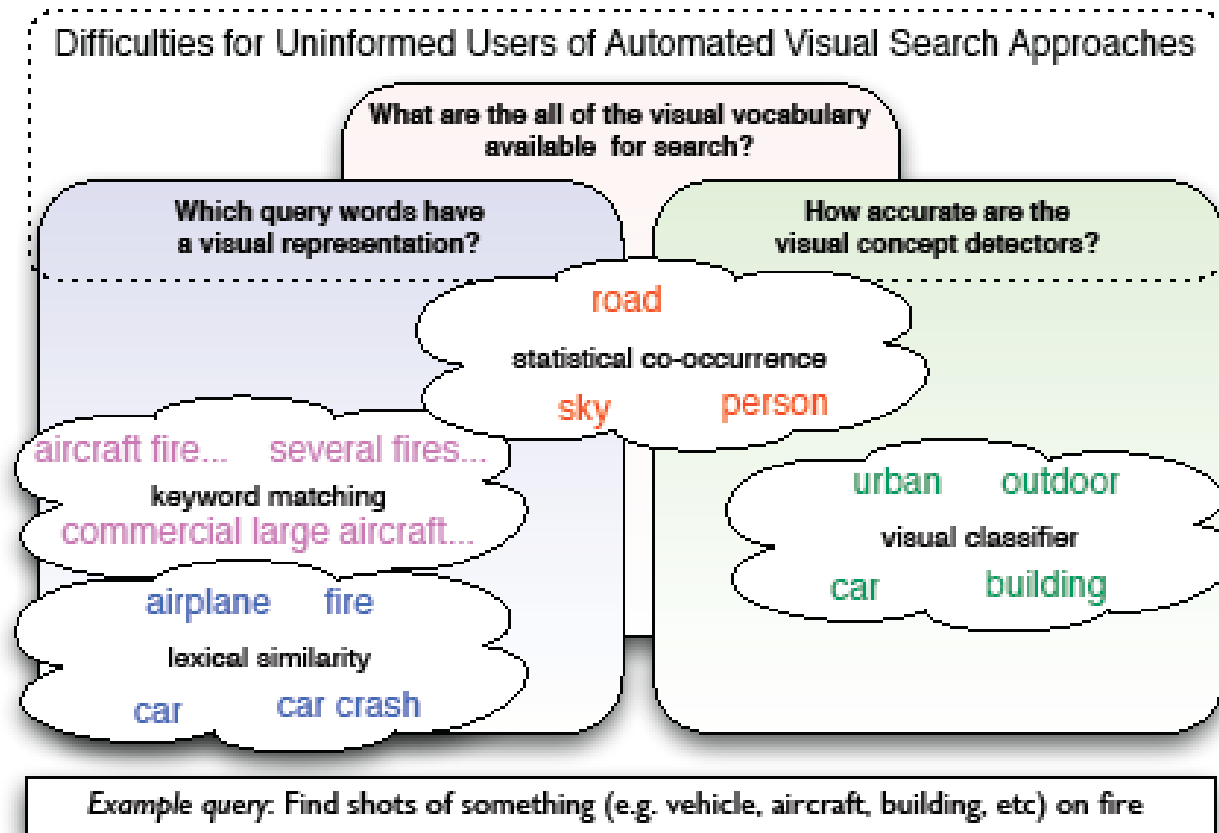


# Problem:



## User's Inability in Forming Queries

- Difficult to choose words/concepts without in-depth knowledge of data and vocabulary

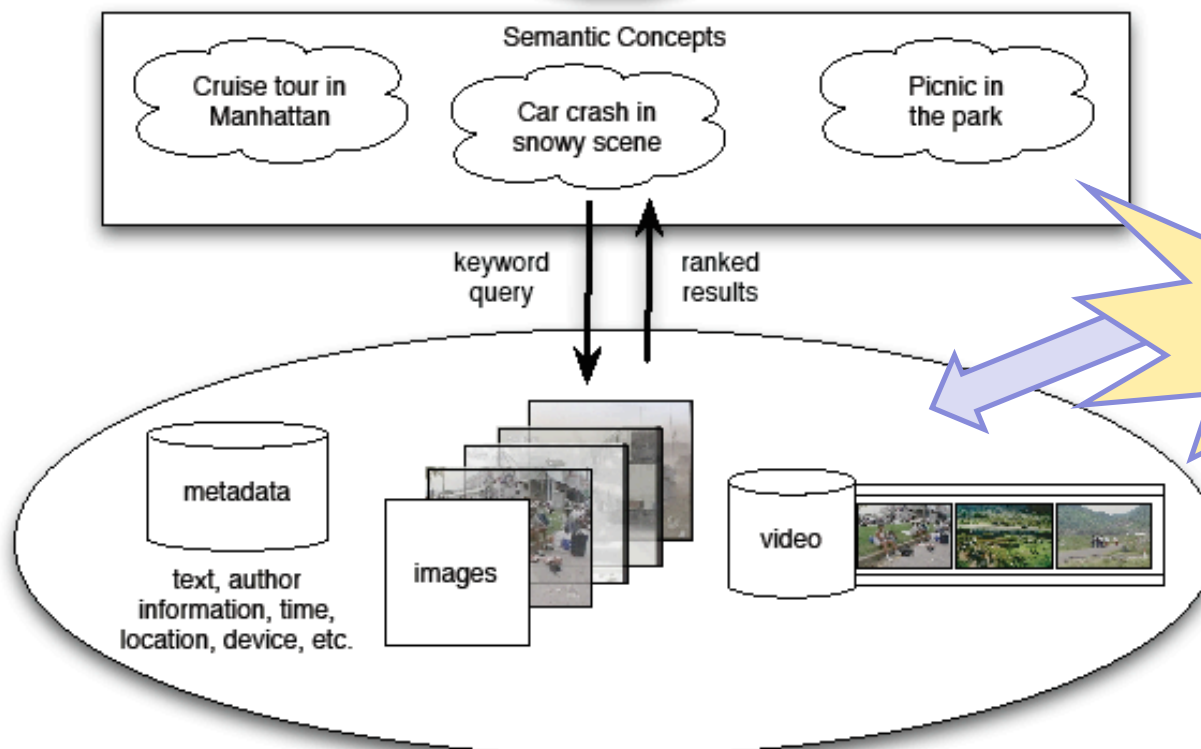




# A Call for Research Attention



Research needed  
to address  
user frustration



A lot of work  
on content  
analytics

Solution--



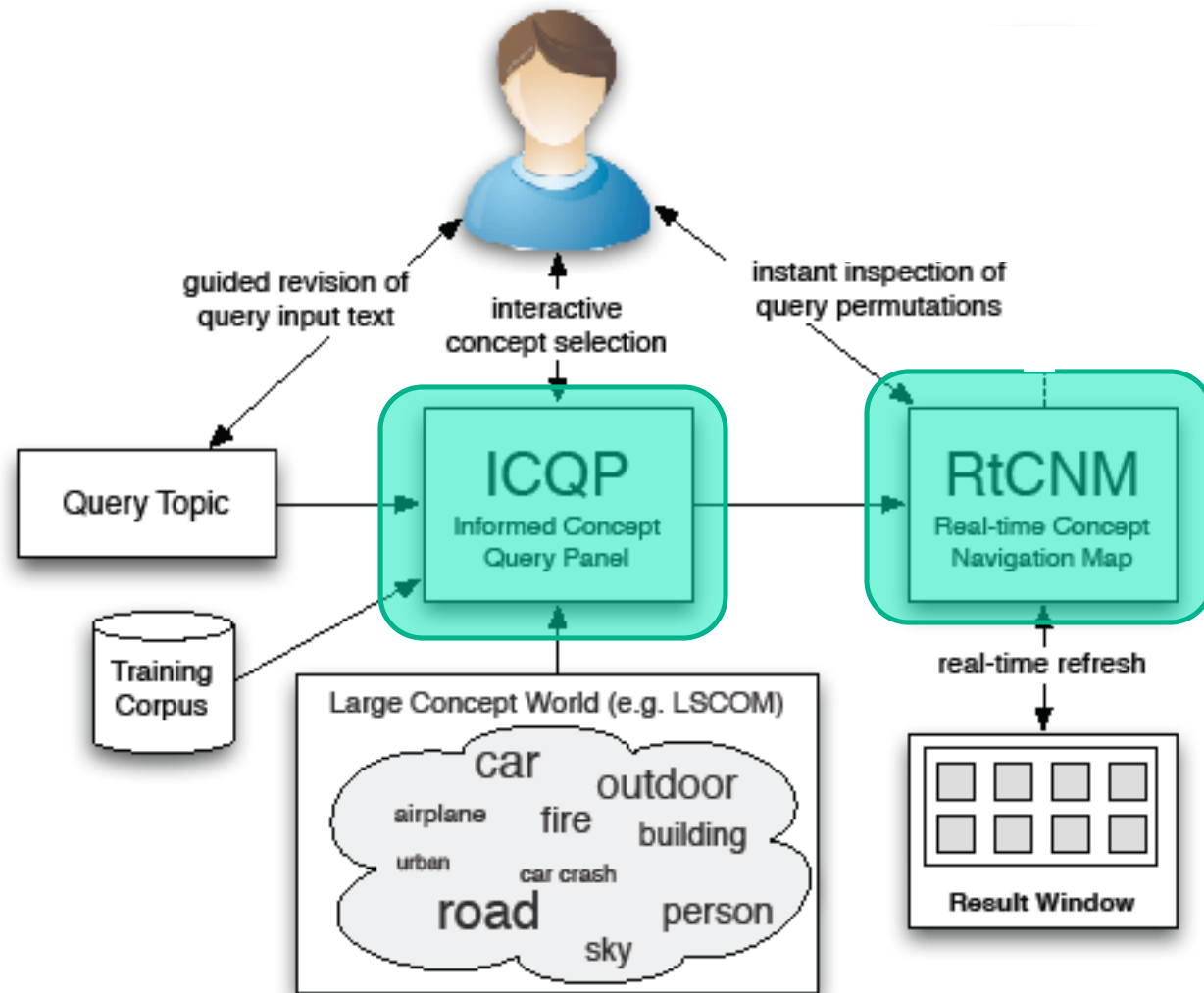
## Keep User Informed and Engaged

---

- Help users stay “informed”
  - in formulating queries
  - in understanding search results
  - in changing search rapidly and flexibly
- Keep users “engaged”
  - Instant feedback with minimal disruption as opposed to slow “trial-and-error”

An attempt: *CuZero*

## Zero-Latency Informed Search & Navigation



# A lot of efforts in visual tagging

- Multi-source tagging
  - Social tagging
  - Tagging by game
  - Geo-context tagging
  - Auto-tagging by content analysis
- Useful for keyword search

flickr GAMMA

1 MILLION LABELS COLLECTED

## The ESP Game

As seen on CNN and newspapers around the world!

beta

64 Players LOGGED IN

TOP SCORES

HOW TO PLAY

New to the ESP Game? Sign up for FREE!

Already have an account?

Screen Name:

Password:

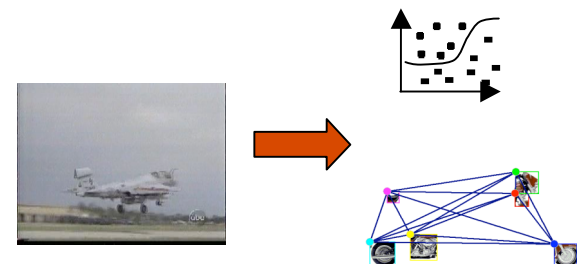
Sign In

Did you know? The ESP Game is helping to label all images on the Web! learn more...

Play our new game: [NEW Pictch](#) [NEW](#)

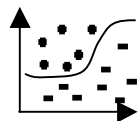
[Terms of Service](#) | [FAQ](#) | [ESP Image Search](#) | [Contact Us](#) | [Credits](#)

Endowed in part by the National Science Foundation (NSF)  
© 2006 Carnegie Mellon University, all rights reserved. Patent Pending.

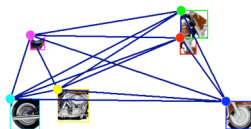


# Tagging by Content Analytics

- Audio-visual features
- Machine learning models
- Web metadata
- Context (geo, user etc)
- Large pools of semantic tags, concept labels

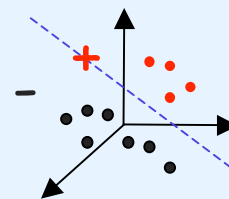


■  
■  
■

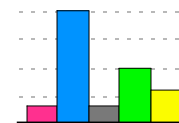
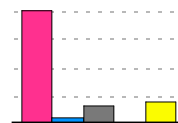


Statistical models

## Semantic Indexes



- Anchor
- Snow
- Soccer
- Building
- Outdoor





# Example: LSCOM concept pool/models



**374 concept detection models:**  
objects, people, location, scenes,  
events, etc

airplane airplane\_takeoff airport\_or\_airfield armed\_person building car  
cityscape crowd desert dirt\_gravel\_road entertainment explosion\_fire forest  
highway hospital insurgents landscape maps military military\_base  
military\_personnel mountain nighttime people-marching person  
powerplants riot river road\_rpg shooting smoke tanks urban vegetation  
vehicle waterscape\_waterfront weapons weather

# A User Community Approach to Vocabulary Development

## ■ Large Scale Concept Ontology for Multimedia

*(IBM-Columbia-CMU joint effort in 2005-6)*

- Ontology for tagging multi-source news video for analyst tasks
- Joint effort by government analysts, librarians, researchers
- Criteria: Useful, Observable, and Machine Detectable
- 834 concepts defined, extended to 2000+ by Cyc, 449 concepts annotated
- Labeled over TRECVID 2005 development set, 61,000 shots
  - 30+ annotators at Columbia and CMU
  - 33 million judgments
- Free for download (350+ downloads so far)
  - <http://www.ee.columbia.edu/dvmm/lscm/>

## Results of sample concept detectors

# waterfront

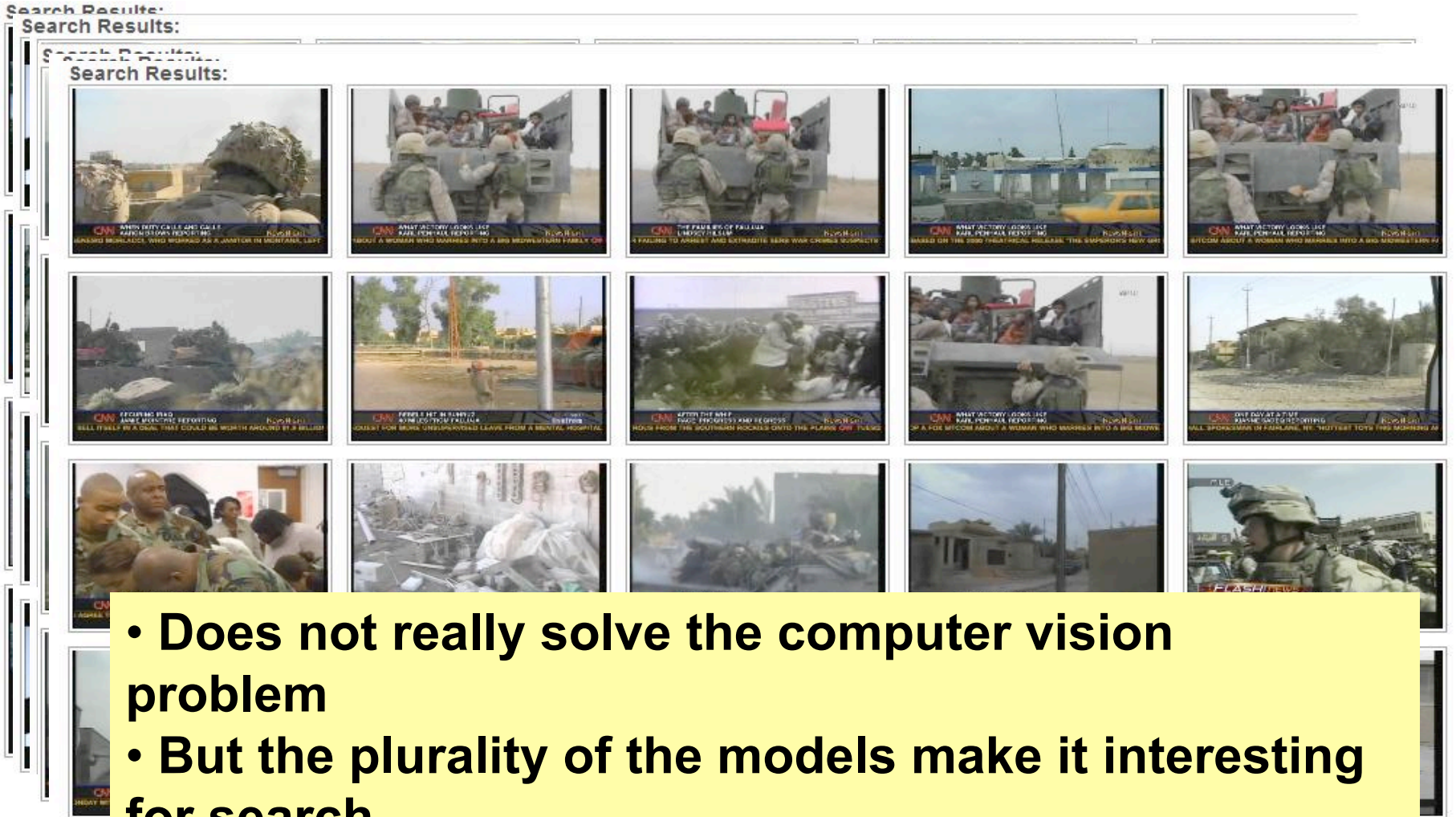
bridge

crowd

explosion fire

# US flag

## Military personnel



- Does not really solve the computer vision problem
- But the plurality of the models make it interesting for search
- Similar to speech retrieval from imperfect ASR



# Sample detection results (TRECVID20)

**Airplane flying**



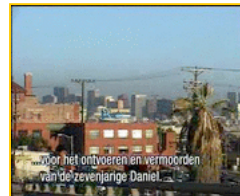
**Classroom**



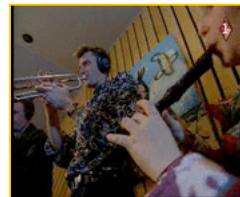
**Demonstration Or Protest**



**Cityscape**



**Singing**

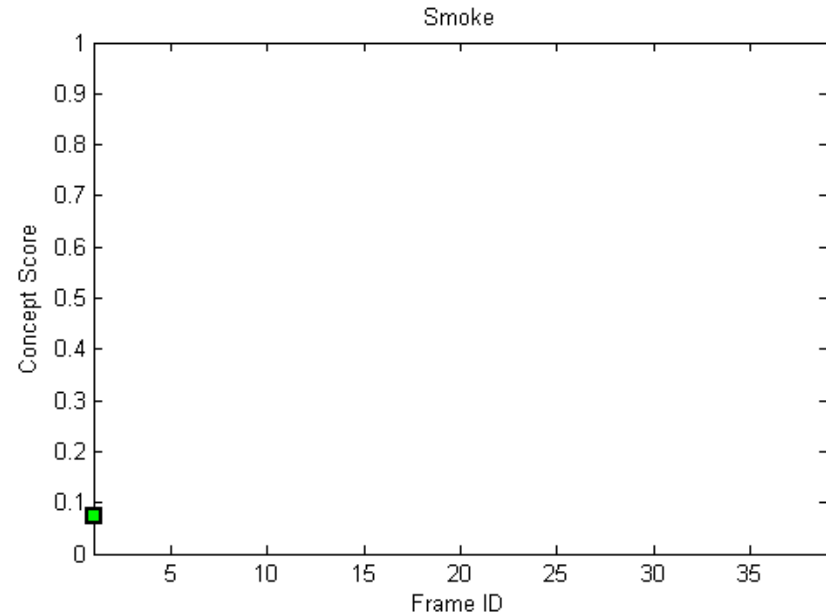


# Example: “Smoke” Detector



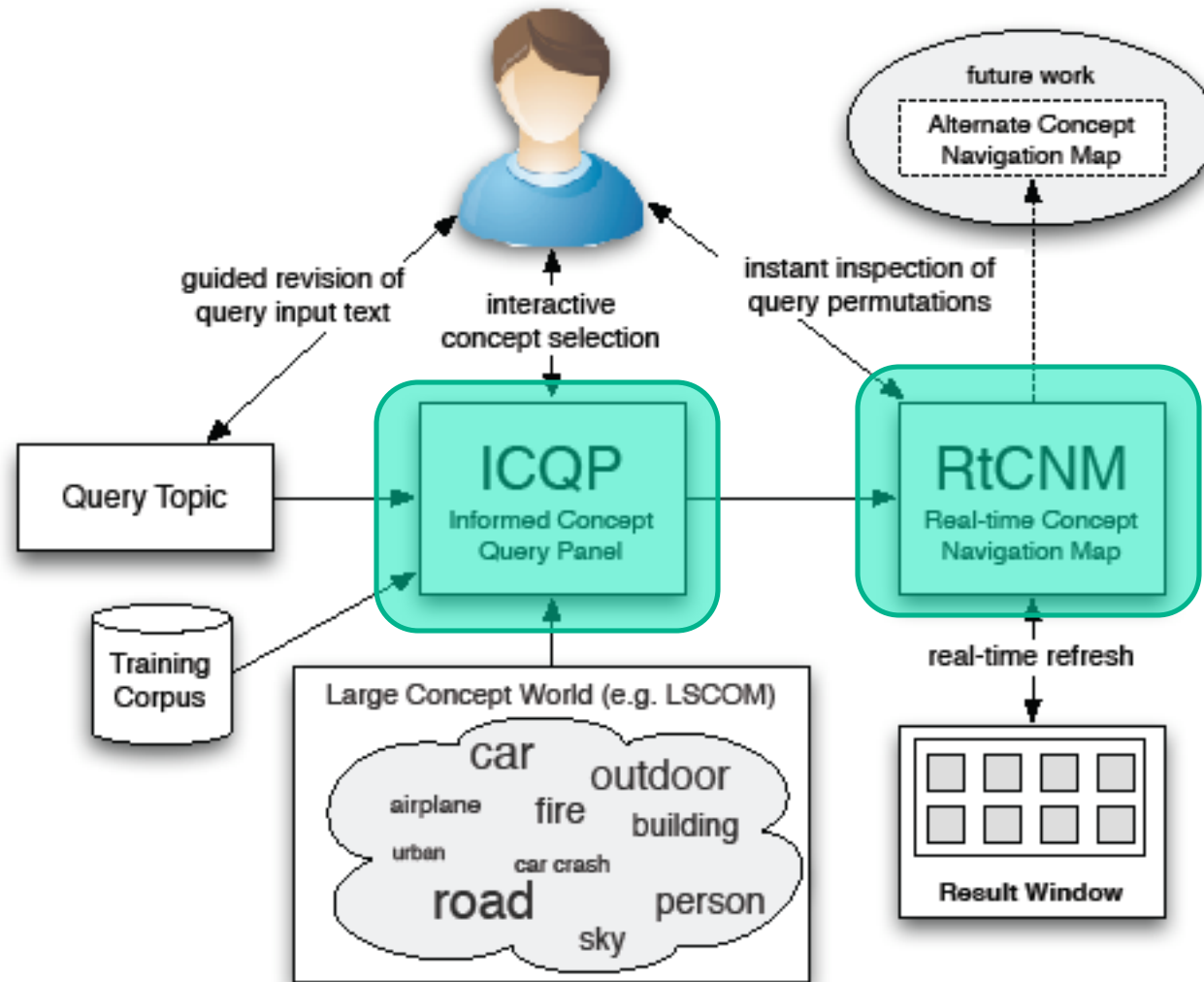
Frame 1

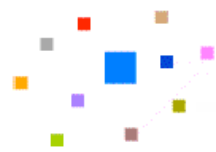
(Video Source: Youtube)



- Compare to other tags, content-based analysis provide precision for deep tagging
- Ping point exact segment of concept

# A prototype *CuZero*: Zero-Latency Informed Search & Navigation





# Informed User:

## Instant Informed Query Formulation

The image illustrates the concept of 'Instant Informed Query Formulation' by showing how a user's search on YouTube can be refined based on suggestions and how a browser search can be refined based on results.

**YouTube Interface:**

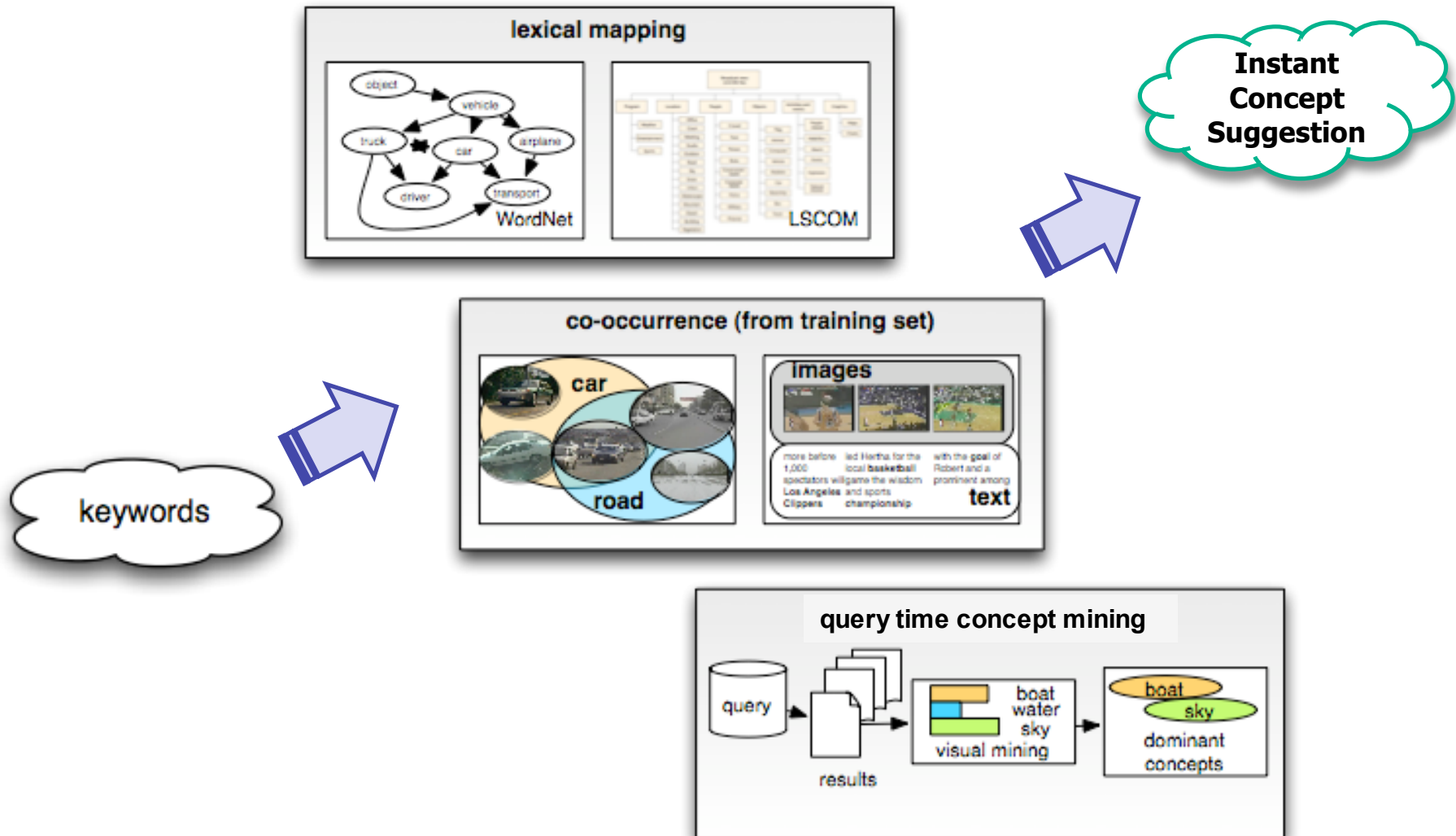
- Logo: **You Tube** Broadcast Yourself™
- Navigation: Home, Videos, Channels
- Search Bar: car explode
- Suggestions:
  - car explodes
  - car exploded
  - car explodes autobahn
  - car explodes on autobahn
  - car explodes on highway
- Videos being watched right now:
  - Video 1: 02:52
  - Video 2: 01:34
  - Video 3: 01:31

**Mozilla Firefox Browser:**

- Address Bar: news
- Search Results:
  - Google News: <http://news.google.com/>
  - CNN.com - Breaking News, U.S., World, Weather, Entertainment & Video News: <http://www.cnn.com/>
  - NPR : National Public Radio : News & Analysis, World, US, Music & Arts: <http://www.npr.org/>
  - Columbia News :: The Home Page of the Office of Communications and Public Affairs: <http://www.columbia.edu/cu/news/>
  - columbia news - Google Search: <http://www.google.com/search?q=columbia+news&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&...>
  - political news - Google Search: <http://www.google.com/search?q=political+news&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-US:official&cli...>

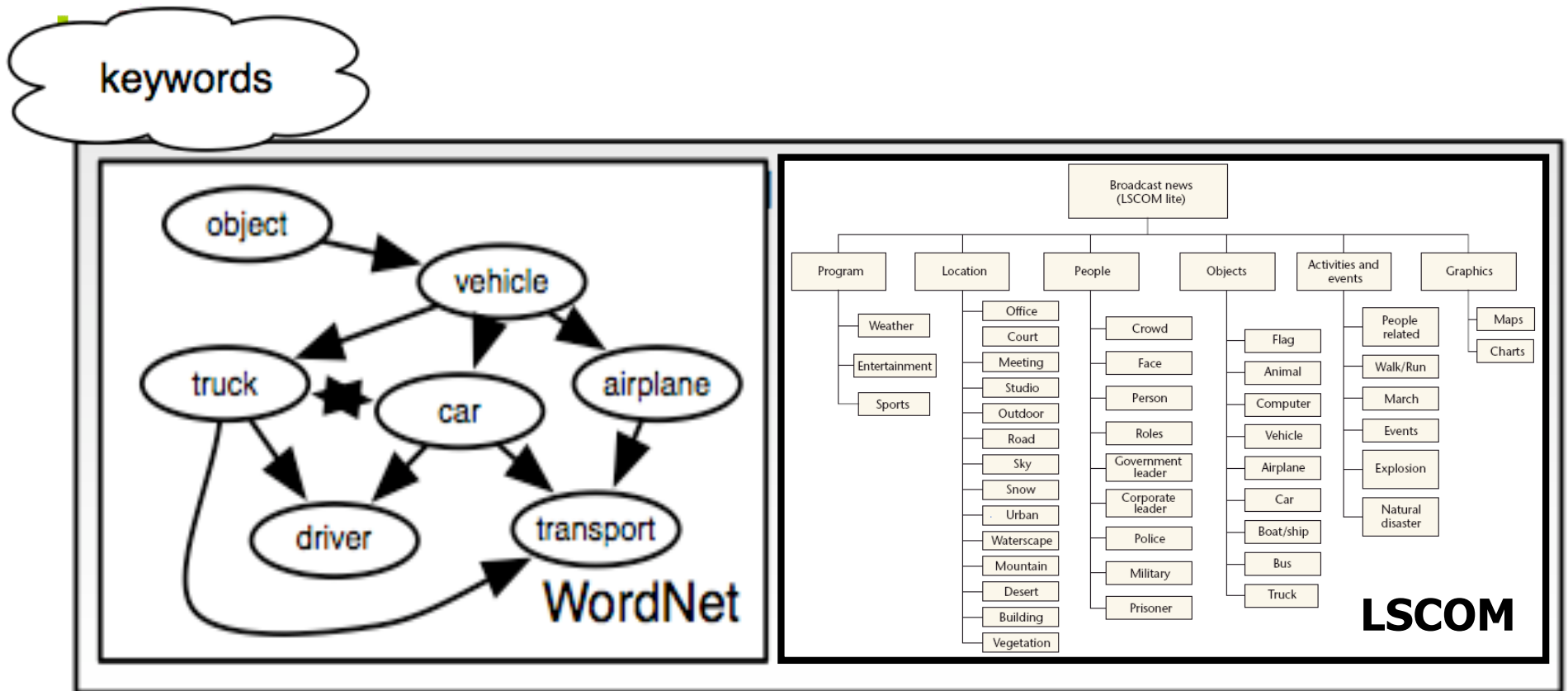
# Informed User for Visual Search:

## Instant visual concept suggestion



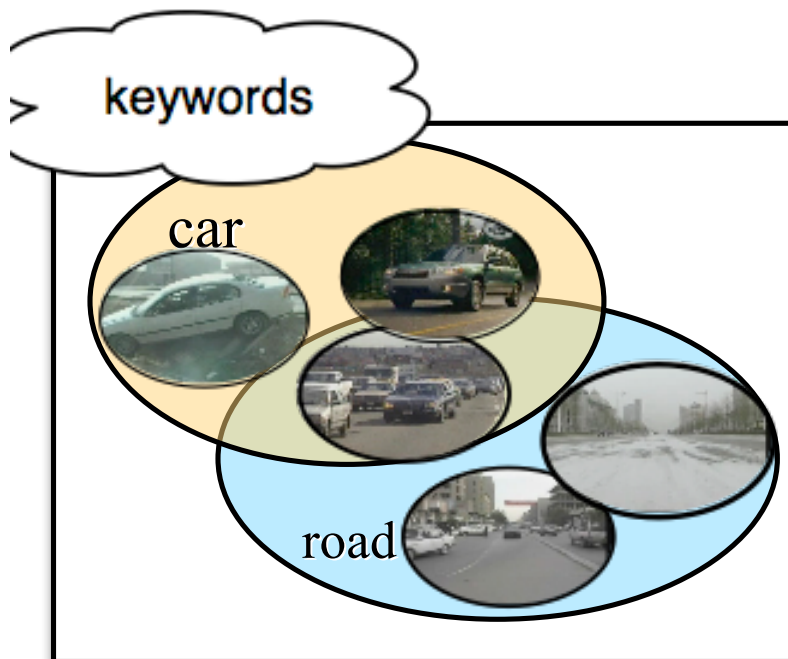


# Lexical mapping



Mapping keywords to concept definition, synonyms, sense context, etc

# Co-occurrent concepts



images



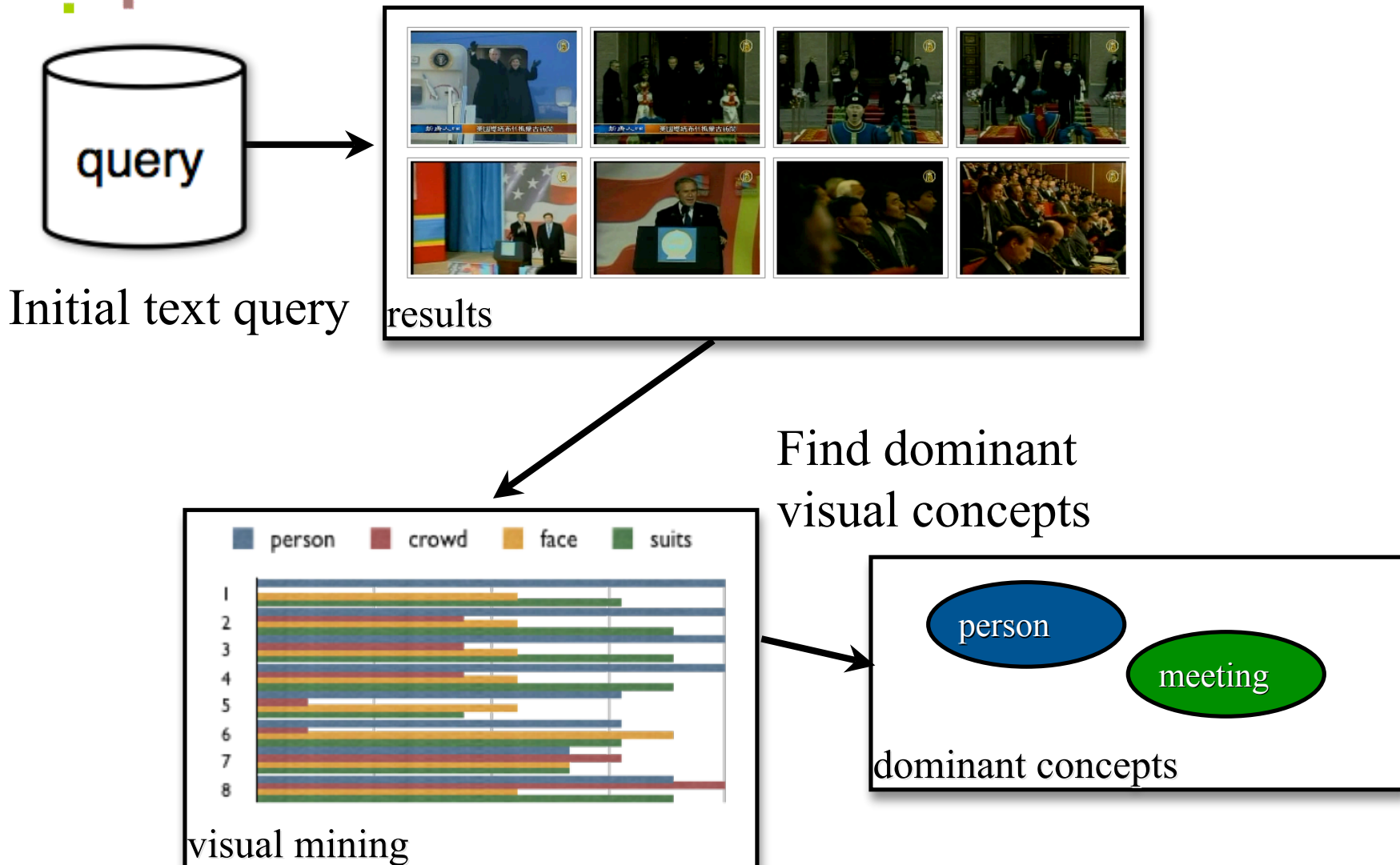
**Basketball courts** and the American won the Saint Denis on the Phoenix Suns because of the 50 **point** for 19 in their role within the National Association of **Basketball**

George Rizq **led** Hertha for the local **basketball** game the wisdom and sports **championship** of the president

Baghdad to attend the game I see more **goals** and the **players** did not offer great that Beijing Games as the beginning of his brilliance Nayyounf 10 this atmosphere the culture of sports **championship**

text

# Query-Time Concept Mining



# CuZero Real-Time Query Interface

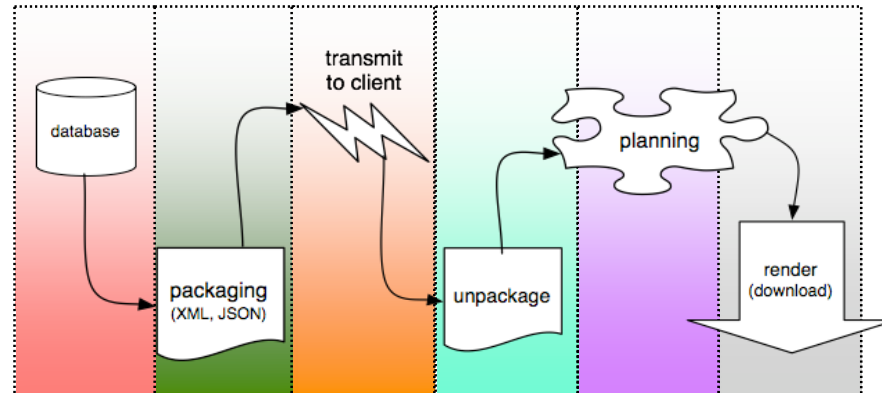
(demo)

The image displays the CuZero Real-Time Query Interface. On the left, a search bar shows the query "police snow accident car road" with a dropdown menu of suggestions: "road", "policemen", "road work", "future scores road", and "announced road work". A callout box labeled "Query auto completion & expansion" points to this dropdown. The main interface on the right shows the same query in a search bar, with buttons for "Search Speech" and "Search Concepts". Below the search bar are tabs for "query formulation", "results", "settings", and "CuZero". The "results" tab is active, displaying four sections of concept suggestions:

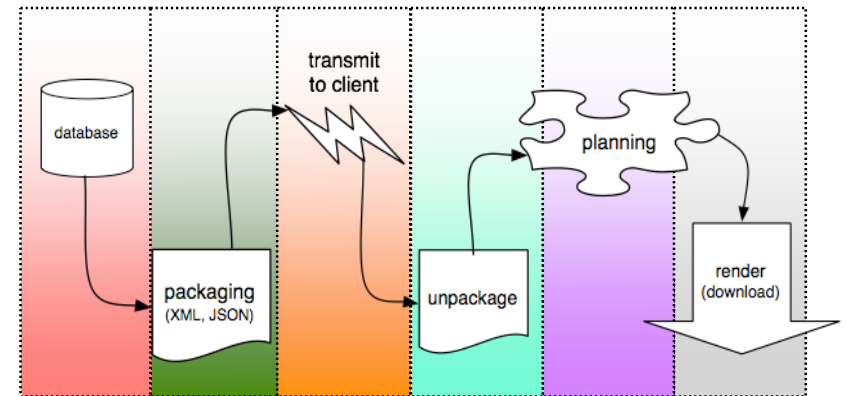
- Text to Concept Ontological Mapping**  
Car Car\_Crash Car\_Racing Dirt\_Gravel\_Road Emergency\_Vehicles Exiting\_Car Highway Officers  
Police Police\_Private\_Security\_Personnel Police\_Security Road Road\_Block Road\_Overpass Snow
- Text to Concept Mapping (by LSCOM definition)**  
Antenna Car Car\_Crash Car\_Racing Civilian\_Person Construction\_Vehicles  
Cul\_de\_sac Dirt\_Gravel\_Road Emergency\_Vehicles Exiting\_Car Ground\_Vehicles  
Highway Parking\_Lot Rainy Road Road\_Block Road\_Overpass Sidewalks Snow Street  
Street\_Battle Traffic Vehicle Weapons
- Co-Occurring Concepts (with above)**  
Commercial\_Advertisement Construction\_Site Landscape  
Male\_Person Motorcycle Scene\_Text Single\_Person  
Single\_Person\_Male Sky
- Dominant Visual Concepts fom Text Results**  
Bridges Explosion\_Fire Face Military\_Buildings Observation\_Tower Ship Snow

A callout box labeled "Instant Concept Suggestion" points to the "Text to Concept Ontological Mapping" section.

# Achieving Zero Latency Exploration



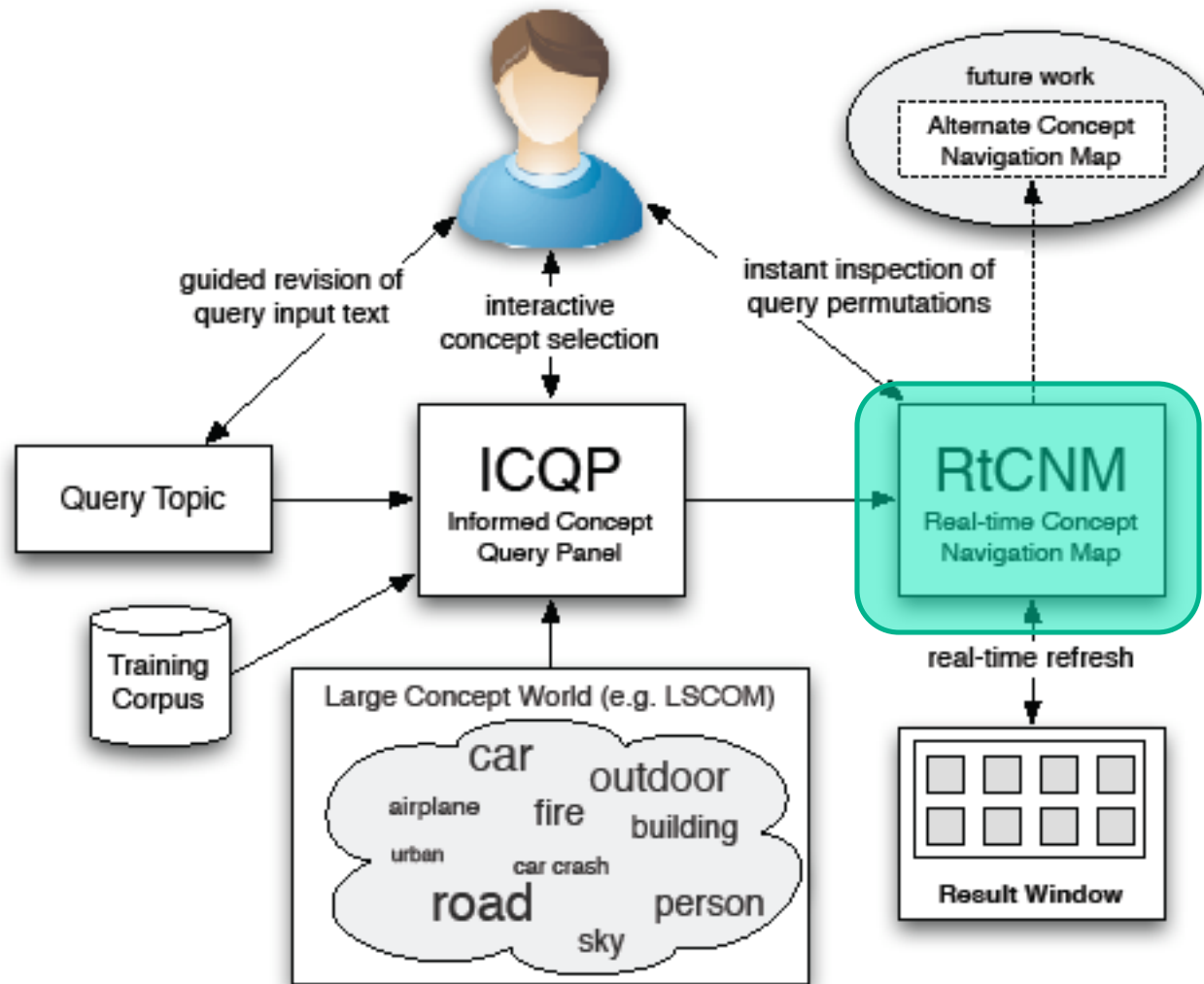
Concept formulation (“car”)



concept formulation (“snow”)

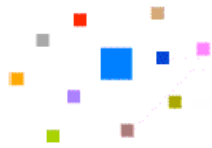
- Overlap (concept formulation) with (map rendering)
- Hide rendering latency during user interaction
- Course-to-fine concept map planning/rendering
- Speed optimization on-going ...

# A prototype *CuZero*: Zero-Latency Informed Search & Navigation

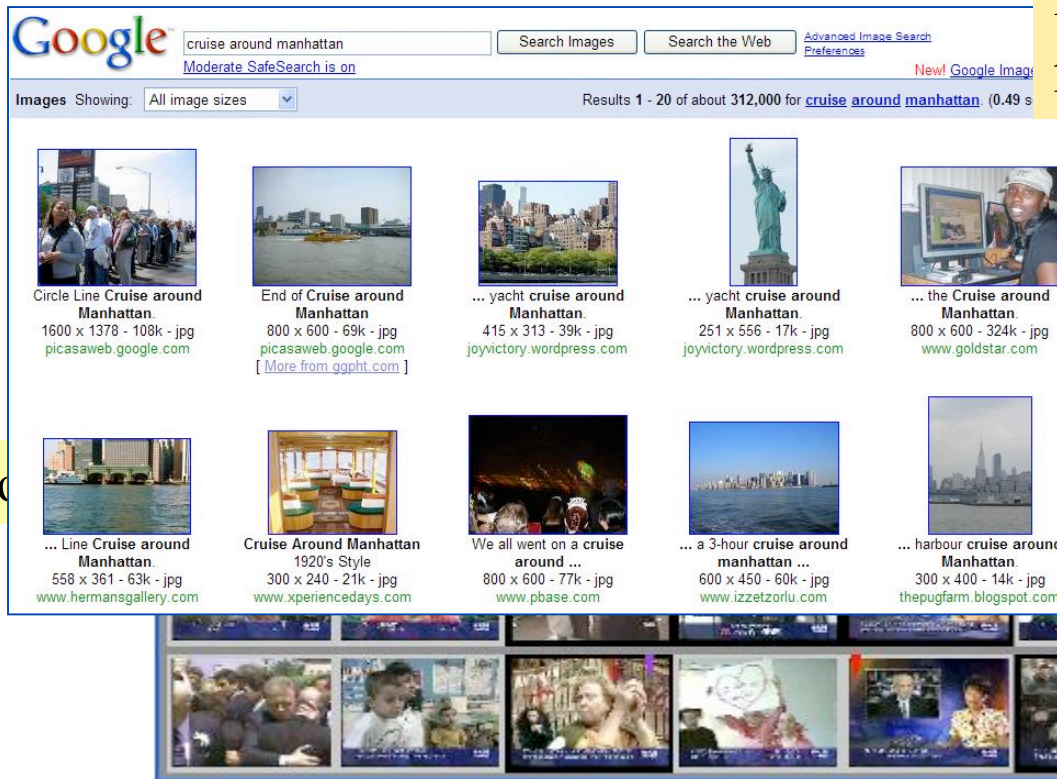




# Address Information Overflow: Effective Visualization



linear browser restricts  
inspection flexibility



Informedia Concept Filter

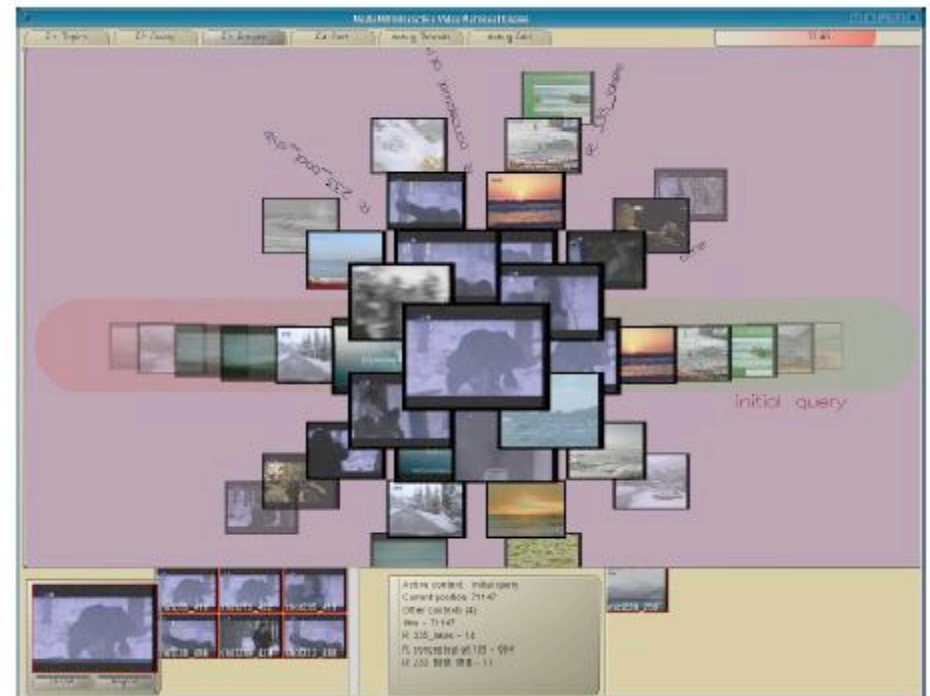
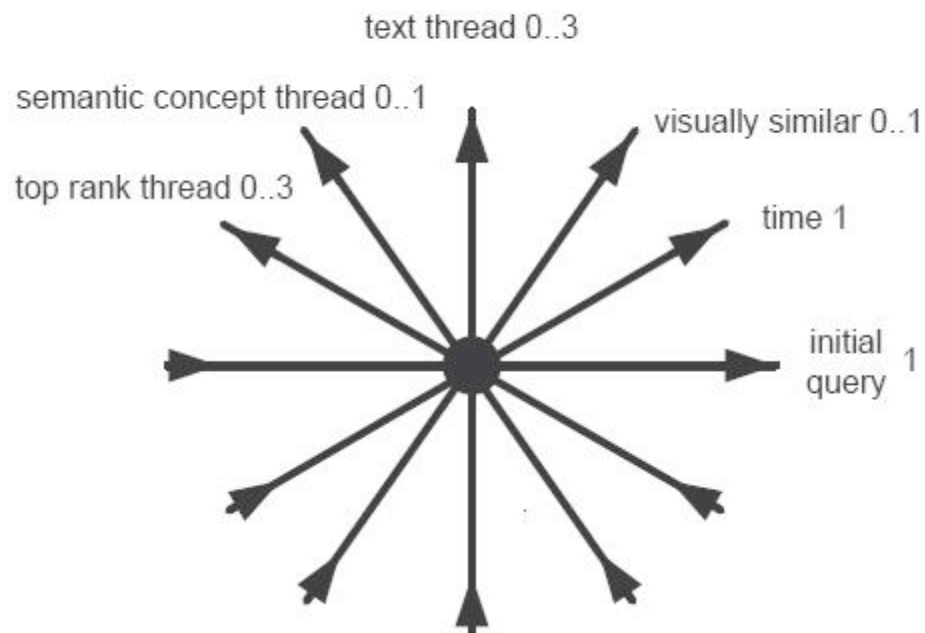


only outdoor

# Informed User: Rapid Exploration of Results



## Media Mill Rotor Browser



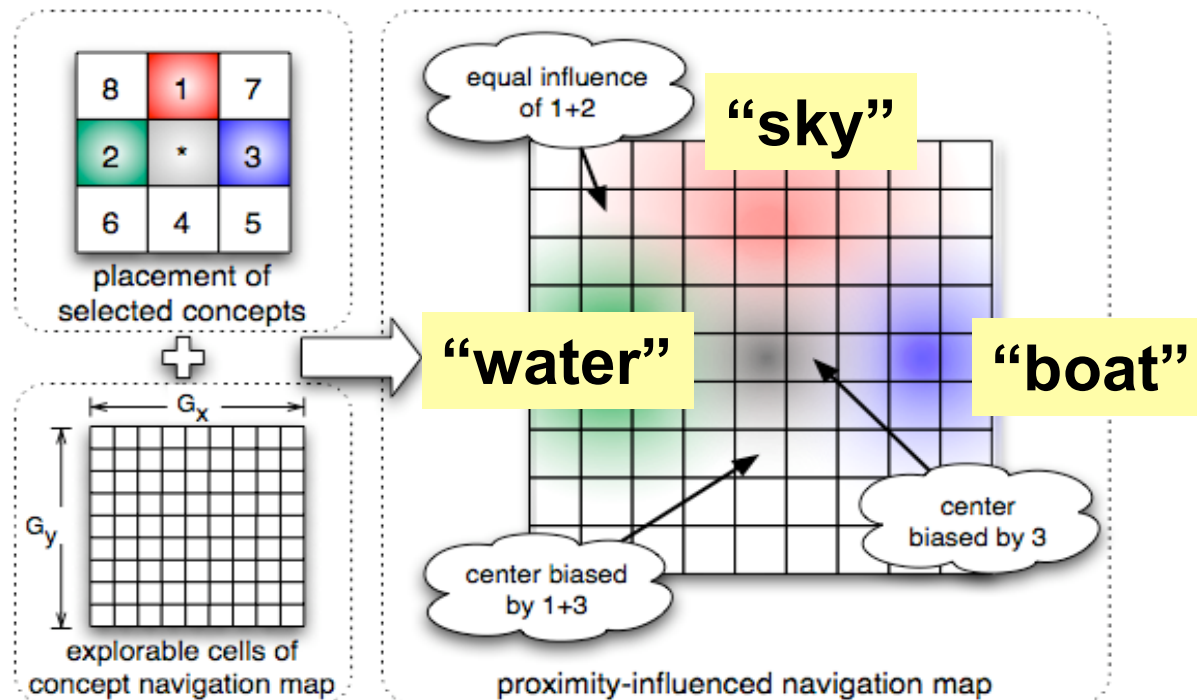


# CuZero visualization:



## Help users stay informed and engaged

- Create a concept map as search anchors
- Direct user control by sliding through the map
- Instant display for each location, without new query



# Achieve Breadth-Depth Flexibility by instant concept map navigation (demo)

- Breadth: Quick sliding in the concept map
- Depth: Deep exploration of each result list

Query: police snow accident car road

Search Speech Search Concepts

query formulation results settings CuZero

**Concept map**

**Deep exploration of single permutation**

0.36419 Full Text + Concept  
0.08004 Emergency Vehicles  
0.08004 Road  
0.01759 Snow  
0.00019 Police  
0.00019 Car

# Columbia CuZero Video Search Engine

- addressing the search frustration

## Query Formulation

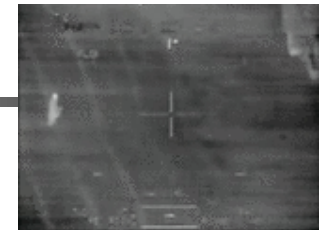
Support **intuitive** formation in arbitrary **combination** of semantic concepts and visual examples

## Navigation and exploration

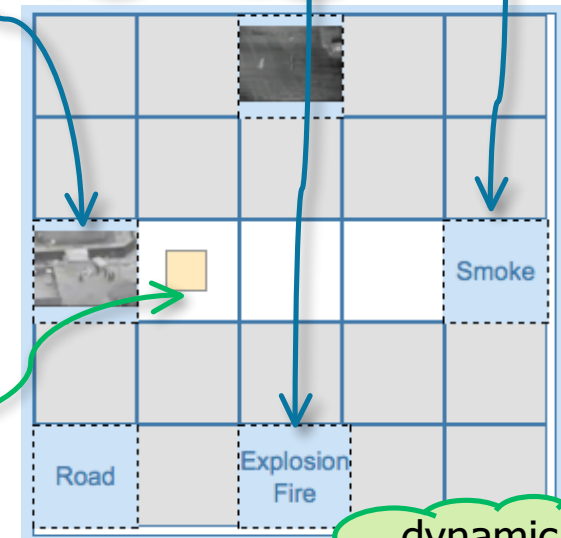
**Real-time** refinement and update of query.

## Fast, lightweight environment

Web-based client-server approach for instant deployment with large data archive.

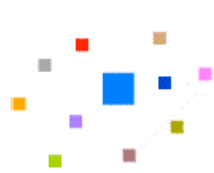


(Video Source: Youtube)



Smoke  
Road  
Explosion  
Running 1  
Running 2

dynamic refinement



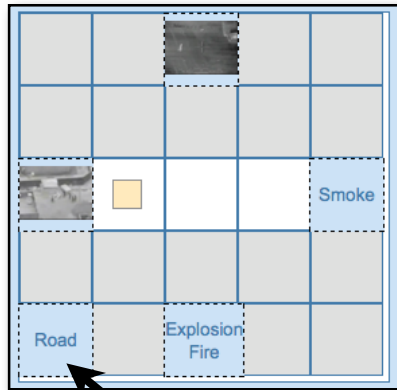
# Simple Extensions

---

- More flexibility in modifying concept map
- Construct composite “super” concepts from simple ones
- Construct personal search toolbox
- Apply concept template to streaming mode

# Concept Map Refinement

(current method requires multi-panel interaction)



1 User wants to replace 'road' with a related concept

2 Co-occurrence can suggest related concepts

Co-Occurring Concepts (with above)  
Daytime\_Outdoor Ground\_Vehicles Outdoor

Highly Correlated Concepts (with above)  
Building Car Daytime\_Outdoor Ground\_Vehicles  
Outdoor Sidewalks Urban Urban\_Scenes Vehicle

3 Changing navigation map requires additional query reformulation and search execution

Query: road | ☐ Include Keyword Search over Speech

query formulation ( 3:00 )

**Text to Concept Ontological Mapping**

Dirt\_Gravel\_Road Highway Road Road\_Block Road\_Overpass

**Text to Concept Mapping (by LSCOM definition)**

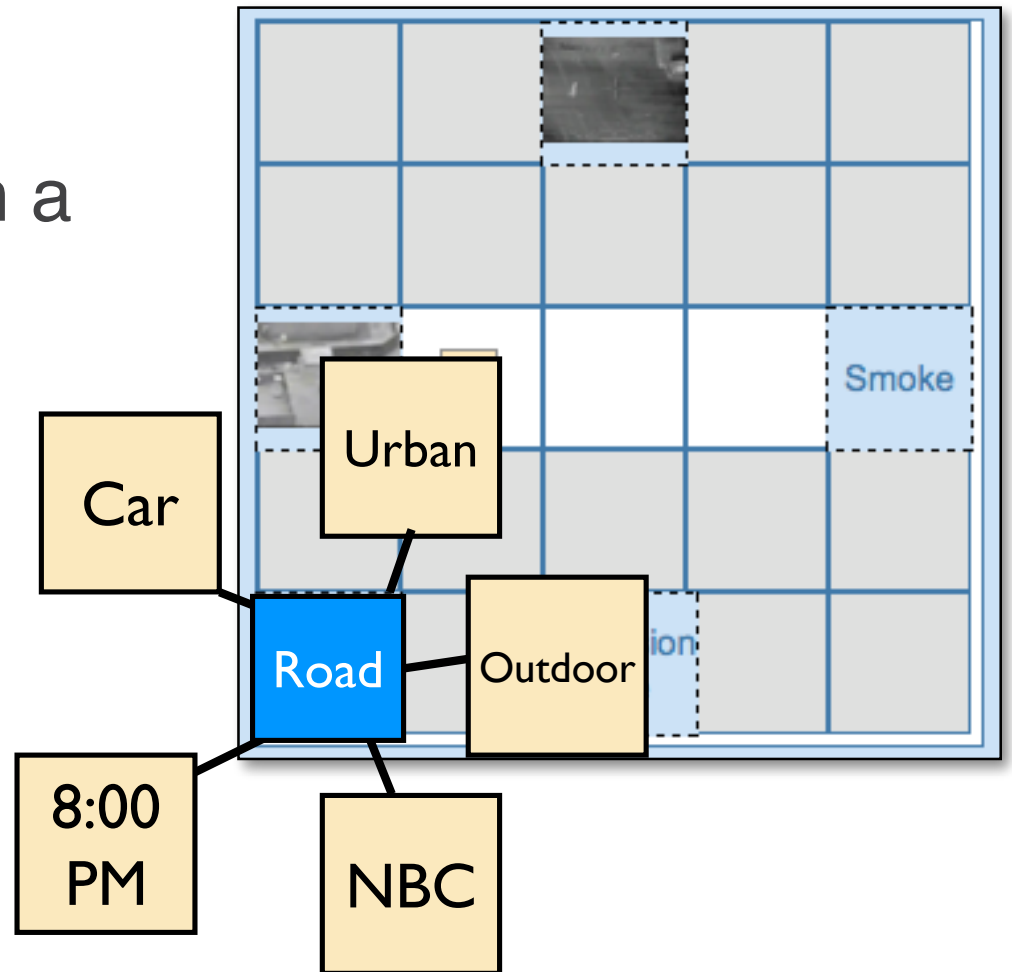
Construction\_Vehicles Cul\_de\_sac Dirt\_Gravel\_Road Highway Rainy Road Road\_Block

Road\_Overpass Traffic

4 Will interrupt the browsing task and requires cumbersome text-entry and manual concept selection

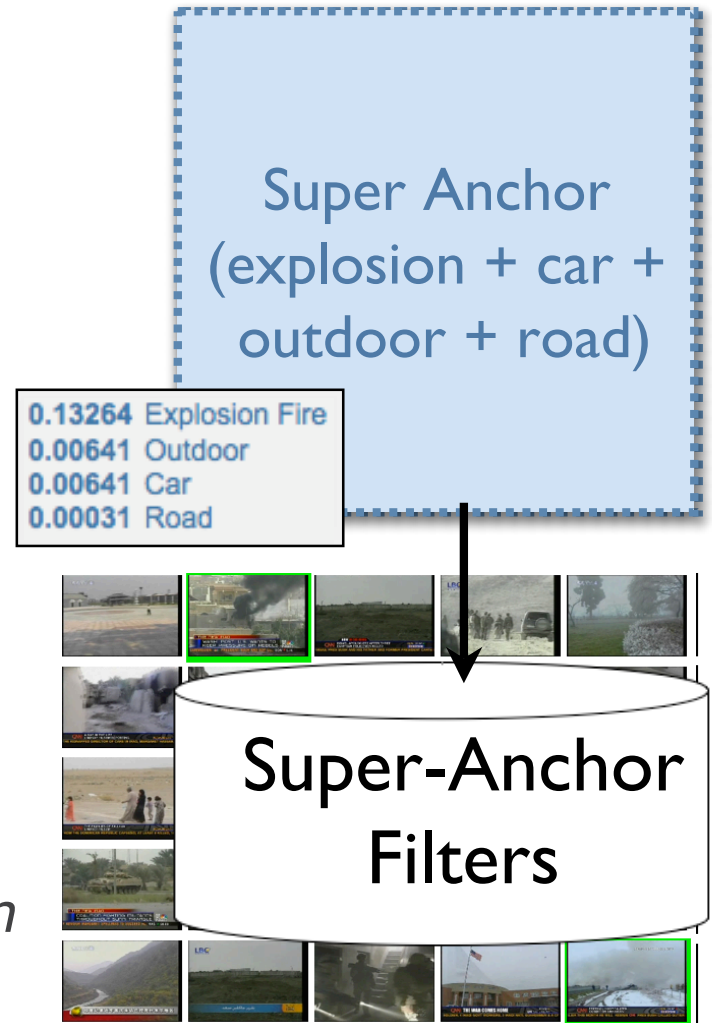
# In-place Anchor Editing

- Allow fast lookup of related anchors or allow user to swap in a 'back-up' anchor
- May also suggest strong metadata relationships
  - Capture time
  - Author/channel
  - Geo location



# Formulating a Super-Anchor

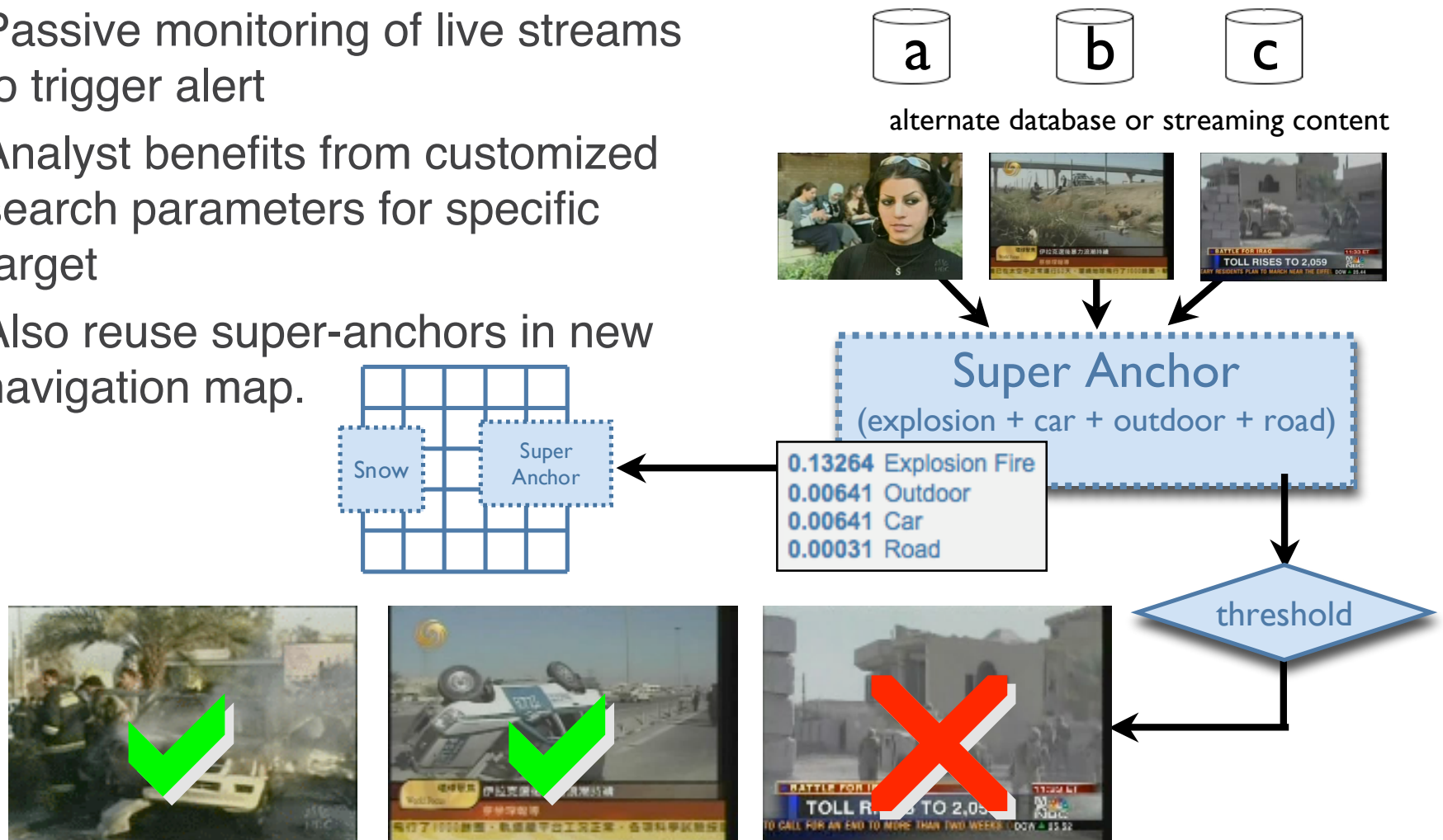
- Navigation map allows precise specification of anchor weights
- For search topics of high-interest, user may have invested significant time in formulation/browsing
- With ideal location in navigation map, want to save this anchor configuration
  - *Apply on new search at later time*
  - *Share with other users for collaborative filtering*
  - *Eventually build a personalized search tool box*





# Super-Anchors: Automated Alerts

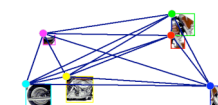
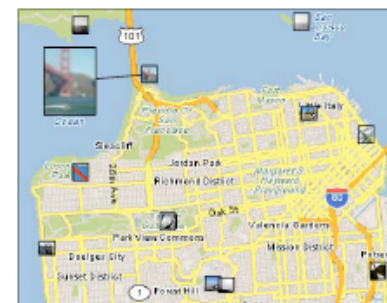
- Passive monitoring of live streams to trigger alert
- Analyst benefits from customized search parameters for specific target
- Also reuse super-anchors in new navigation map.





# Deeper issues

- Is keyword-concept search the right paradigm?
- Perhaps for certain vertical domains?
- How far can automatic content analytics go?
  - Scale up the number of concepts?
  - Find the right visual lexicon?
  - Are they accurate enough for search?





# Summary

---

- Effective visual search requires attention to help improve user experience
- More challenges than tagging and retrieval
- CuZero Design Principles
  - Keep user informed and engaged in
    - Query formulation and result visualization
  - Instant feedback
  - Support fusion of search modalities
    - Visual concept
    - Example
    - Keyword /speech transcripts



# More Information

---

- LSCOM lexicon and annotation
  - Include 449 concepts (object, scene, event et al)
  - <http://www.ee.columbia.edu/lscocom>
- Columbia DVMM Lab
  - <http://www.ee.columbia.edu/dvmm>
  - 374 SVM-based concept detectors
  - Video search demos