27. Multimedia Search and Retrieval

INFO 202 - 1 December 2008

Bob Glushko

Plan for Today's Lecture

"Describing Things" in Search and Retrieval

"Describing (Text and Non-Text) Things" with Text

"Describing Non-Text Things" with Non-Text Descriptions

Demos

The Demo-ers

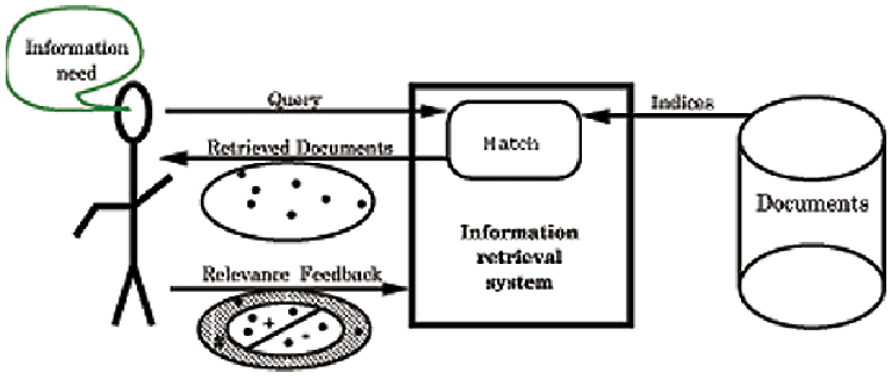
Heather Dolan

Sarah Van Wart

Ljuba Miljkovic

your name here

Reminder: Schematic View of Classical Search



Classical Search: Processing [1]

The user translates an information need or question(s) into a QUERY

The query expresses the information need in a format or as a set of DESCRIPTIVE FEATURES that the system can handle

The processable representation of these features make up the INDEX or INDICES

Classical Search: Processing [2]

The system matches the descriptive features in the query against the features that describe the "documents" or "information objects" (or pointers to them) stored by the system

Items are retrieved when the degree of the match exceeds some measure of similarity (which might be "exact match" for some queries or systems)

The system presents the retrieved items according to the measure of similarity

What Does it Mean to Describe Something?

Identify / scope the thing to be described

Study it to identify its important properties or features

Compare it with other things like and unlike it

Select or develop a system / vocabulary for description using "good" categories and terms that enables particular things to be identified and different things to be distinguished

Create the descriptions, measurements, and other statistics about the object, either "by hand" or by some automated / computational process

Why "Describing Text" Is Relatively Straightforward

Most of the concepts and techniques that authors or other people might use for "describing things" were designed for text information

The text content suffers from the vocabulary problem and the text can vary in formats, fonts, etc. -- but at least the alphabet defines "equivalence classes" for these different representations

... so that many techniques for extracting text descriptions from the information being described can automated

The internal structure of text information and collections is explicit, which enables descriptions to be assigned at objectively consistent granularities

Describing Non-Text With Text - By People

Professional cataloguers of "museum objects," images/paintings and other "cultural works" often use the Getty "Categories for the Description of Works of Art"

A "CDWA-Lite" is being developed a la Dublin Core for use by non-specialists

- http://www.getty.edu/research/conducting_research/standards/cdwa/index.html

ID3 tags on MP3 audio files contain a very limited amount of song metadata

MPEG-7 is the newest, most standard, and most complicated specification for "semantic" image and video metadata

Annotating Audio

Tom Coates created a prototype system for BBC Radio and Music Interactive in 2005 with several others

The idea was to mark up/make semantically useful radio broadcasts from the BBCs 80 year history

Never deployed, but some excellent demos:

- Creating audio annotations (http://www.plasticbag.org/files/misc/audio_annotation_playback.mov)
- Editing (http://www.plasticbag.org/files/misc/audio_annotation_edit.mov)

Metadata-Assisted Image Retrieval

Title	Roomy Fridge
Date	circa 1952
Description	An English Electric 76A Refrigerator with an internal storage capacity of 7.6 cubic feet, a substantial increase on the standard model.
Subject	Domestic Life
Keywords	black & white, format landscape, Europe, Britain, England, appliance, kitchen appliance, food, drink, single, female, bending

Table 1. Metadata used for resolving the request of the query 'A photo of a 1950s fridge'.

CDWA Lite

1. Object/Work Type
2. Title
3. Display Creator
4. Indexing Creator
5. Display Measurements
6. Indexing Measurements
7. Display Materials/Techniques
8. Indexing Materials/Technique
9. Display State/Edition
10. Style
11. Culture
12. Display Creation Date
13. Indexing Dates
14. Location / Repository
15. Indexing Subject
16. Classification
17. Description / Descriptive Note
18. Inscriptions
19. Related Works
20. Rights for Work
21. Record
22. Resources

The ESP Game - Labeling Images

von Ahn & Dabbish got pairs of people to label images in the "ESP Game"; recently introduced by Google as "Image Labeler"

Player 1 guesses: purse
Player 1 guesses: bag
Player 1 guesses: brown

Success! Agreement on "purse"

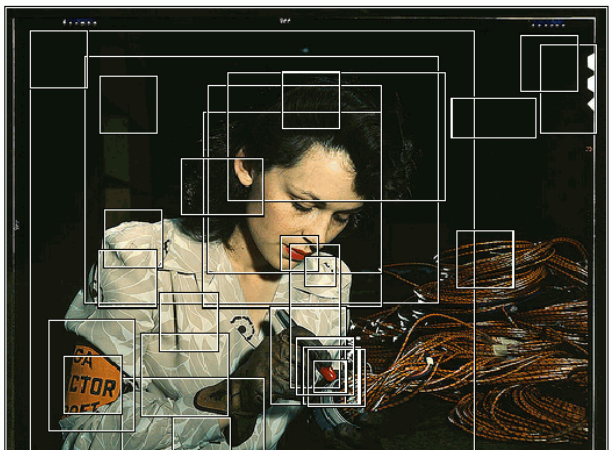
Player 2 guesses: handbag

Player 2 guesses: purse
Success! Agreement on "purse"

Getting the Masses to Tag Photos: LOC on Flickr

http://blogs.ischool.berkeley.edu/i202f08/2008/11/23/lib-ocongress-on-flickr/ (Becky Hurwitz)

Woman aircraft worker, Vega Aircraft Corporation, Burbank, Calif. Shown checking electrical assemblies (LOC)



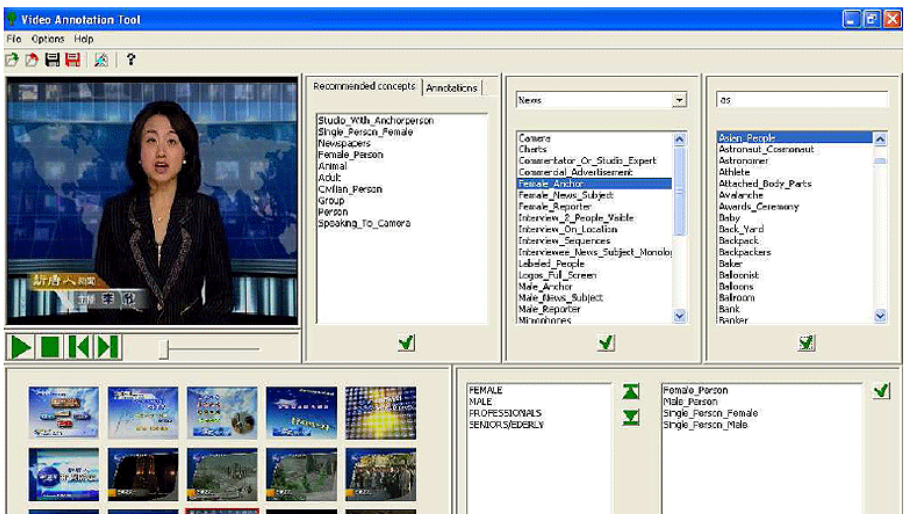
Uses of Video Metadata

Video search (asset level)	Mashups/Remixes
Video search (scene level)	Advertising (in-stream, overlay, banner)
Seek and skip functions	Personalization and targeting
Video packaging and presentation	Sharing and social networking
Playlisting	Reporting and analytics
Dynamic program updates	Recommendations
Multiple navigation paths within or across videos	

"Vertical" and "Ontological" Annotation

Many metadata data elements are used in many contexts and applications

But many applications require specialized controlled vocabularies in video or photo annotation and archiving



NFL Fantasy Video

http://www.gotuit.com/customers/sprint_ff.html

Sprint uses MPEG 7 metadata to create a new application for fantasy football with the National Football League.

Each week during the season, every play of every game in the NFL is indexed using metadata. Sprint customers can then set up their fantasy team and see the video highlights of just their players, and even jump to a specific play

Using "Channel Correlations" to Annotate & Search Multimedia

Text overlays (captions) can be used to identify people or places in videos

Location information (e.g, GPS) attached to images or video can be used to infer content descriptions

"Embedded" or "Scene text" can sometimes be extracted to identify photo or video objects or settings (highway signs, restaurant, shopping center, numbers on sport player uniforms)

Narration and dialog in video can be used a scene keywords

Radio stations publish metadata about what's playing now *http://api.yes.com/)

Describing Non-Text With Text - Automatic

Other textual metadata can be assigned by the devices or mechanisms that created the non-text objects

EXIF (Exchangeable Image File Format) is used in digital cameras

Most professional digital audio applications (DTV, DVD, etc) use metadata about the Dolby encoding to enable adjustment and optimization of audio output

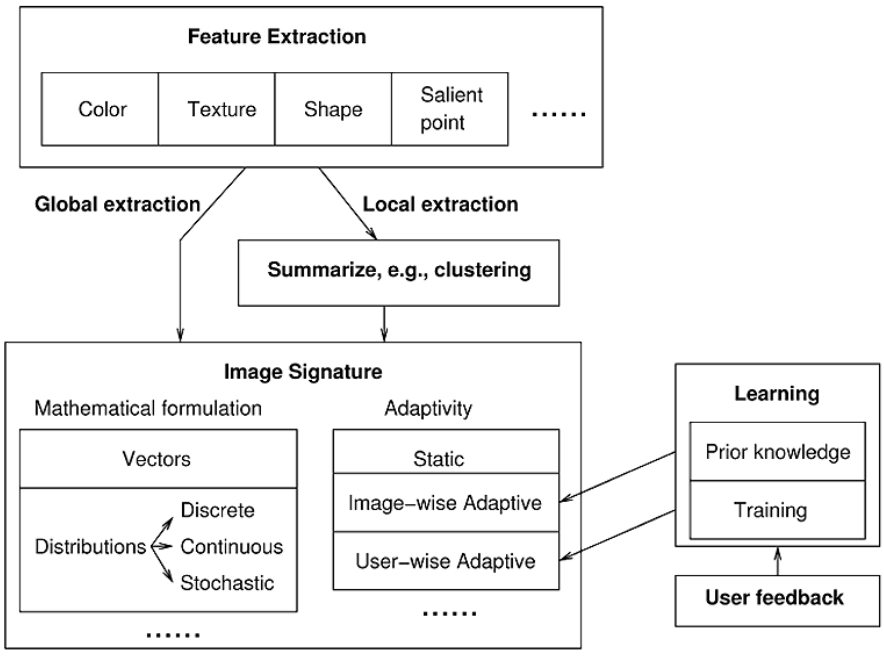
Reminder: The Semantic Gap

Instruments, devices, sensors and so on encode data in formats that are optimized for efficient capture, storage, decoding, or other criteria

As a result, the content / representation / encoding / material of the object is semantically opaque, and can't be (easily) processed to understand what the object "means"

So there is a gap between the semantic descriptions that people assign to objects and the descriptions that can be assigned by computers or other automated mechanisms

Creating the "Image Signature"



Typical Features Extracted from Images

Face detection	
Layout	
Shape	Fourier descriptors
	Elementary description
	Angles between edges, and cross ratios of them
Texture	Wavelet, Fourier transform
	Edge-orientation histogram
	Local binary patterns
	Automatic texture features
Color	Eigen image
	Dominant colors
	Region histogram
	Fixed subimage histogram
	Farthest neighbor histogram
	Global histogram
	Laplacian

The Semantic Gap

Descriptors *feature-vectors*	Segmented blobs, Salient regions, Pixel-level histograms, Fourier descriptors, etc...
Raw Media *images*	

Crossing the Semantic Gap Through Computation

A consequence of the semantic gap for mm is that there are a very large number of low-level features that can be reliably identified

So any description using these features will be "sparse" - lots of missing values

Dimensionality reduction techniques can exploit correlations between low-level features

Machine learning techniques can use extract the features that distinguish mm objects given the same text labels

Christel's "Killer Functionality" for Multimedia

In the "why isn't anyone using it" article, Mike Christel argues that we must "transform our capability to produce and store massive amounts of multimedia materials into a benefit"

We should use social tagging and tagged multimedia to train systems to classify objects and extract descriptions from them

There are over 20 million unique tags and over 3 billion images on Flickr as of November 2008

"Supervised Learning" of Object Categories in Images

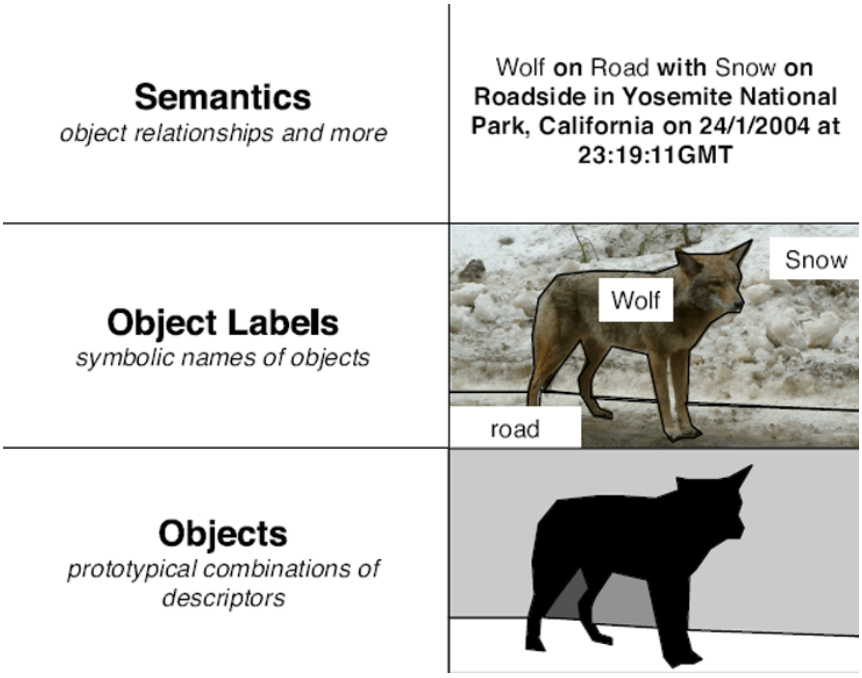
http://blogs.ischool.berkeley.edu/i202f08/2008/11/10/a-new-era-for-image-annotat (Janani Vasudev)

The image labeling system is trained to identify classes of objects, such as “tigers,” “mountains” and “blossoms,” by exposing the system to many different pictures of tigers, mountains and blossoms

The supervised approach allows the system to differentiate between similar visual concepts – such as polar bears and grizzly bears

http://www.jacobsschool.ucsd.edu/news/news_releases/release.sfe?id=650

The Semantic Gap - Crossed



Ignoring the Semantic Gap in Search

But maybe we don't need to cross the semantic gap to have effective multimedia IR

We can use the low-level features that can be extracted automatically to index the multimedia collection and then extract the same ones from a multimedia "query by example"

- Shazam - use audio "fingerprinting" http://www.shazam.com/music/web/home.html
- Query by humming
- Music recommendation by genre
- Face recognition and classification

"Fingerprinting"

Audio Fingerprinting is the attempt to song based on its “signature”

It differs from Query-By-Humming in that it’s often looking for specific recordings rather that any version of a song

It can also be used for song ID (recording a song off the radio for later query) or labels to track sales of their music (“intellectual property”), but often as a precursor to legal action by the labels (“network scanning”) to detect copyright infringement

Systems either perform a match for the whole recording on a Table of Content (TOC) look up (Gracenote/CDDB for example) or perform computation to find a specific song (Audible Magic, being used in Myspace)

Similar work is being done with Video Fingerprinting for YouTube

Readings for 12/3

Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, Introduction to Information Retrieval, Chapter 13, through Section 13.1

Adam Kilgarriff and Gregory Grefenstette, "Introduction to the Special Issue on the Web as Corpus," Computational Linguistics 29(3) (2003)

"From Babel to Knowledge Data Mining Large Digital Collections" Daniel Cohen, -Lib Magazine (March 2006)

Weiguo Fan, Linda Wallace, Stephanie Rich, and Zhongju Zhang, "Tapping the Power of Text Mining," Communications of the ACM, September 2006

Paul Graham, "A Plan for Spam" http://www.paulgraham.com/spam.html