Semantic Image Retrieval

Yuanxi LI
Computer Science Department
Hong Kong Baptist University
yxli@comp.hku.edu.hk

Multimedia Sharing from Social Media

Your Searching Experience?

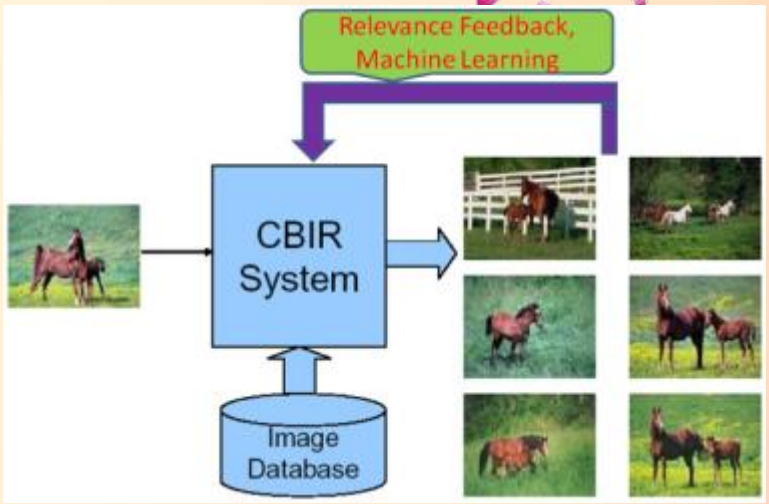
Image Searching Engines:

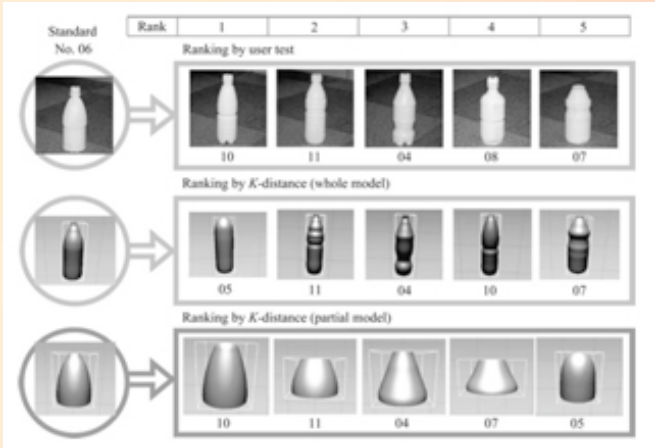
- Google – text annotation, size, file type, coloration, site/domain, color similarity…
- Yahoo – size, coloration, site/domain
- Ask.com – narrowing suggestions
- Exalead – size, file type, coloration, layout
- Being Image
- Flickr Image

Traditional Approaches

Common Approaches for Image Search

- Low-level visual features
  - Spatial relationships
  - Color
  - Texture
  - Shape
  - Edges …



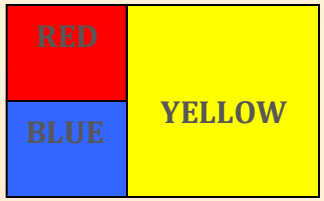


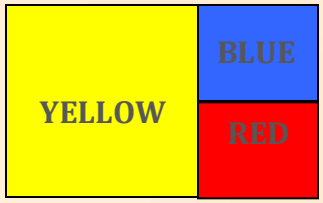
- High-Level Concept

Retrieval by Similarities - Color Similarity

Color Similarity:

Color distribution similarity has been one of the first choices because if one chooses a proper representation and measure it can be partially reliable even in presence of changes in lighting, view angle, and scale.

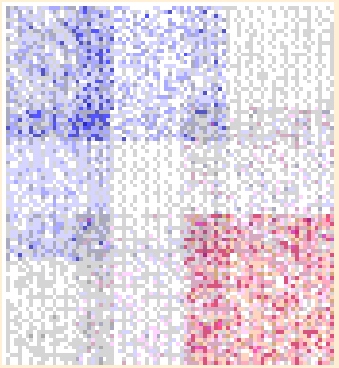


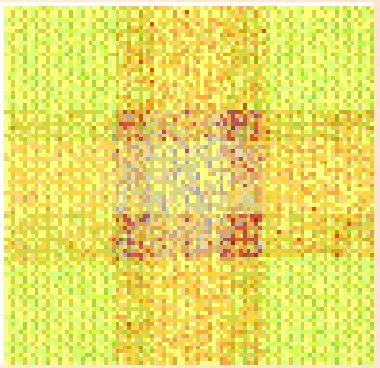


Retrieval by Similarities - Texture Similarity

Texture Similarity:

- Texture reflects the texture of entire image.
- Texture is most useful for full images of textures, such as catalogs of wood grains, marble, sand, or stones.
- Texture images are generally hard to categorize using keywords alone because our vocabulary for textures is limited
- Wold Decomposition
  - Periodic
  - Evanescent
  - Random

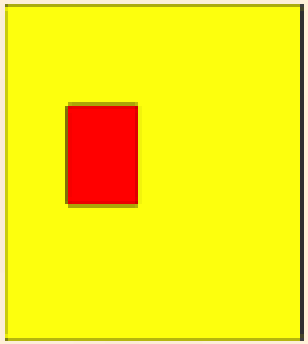


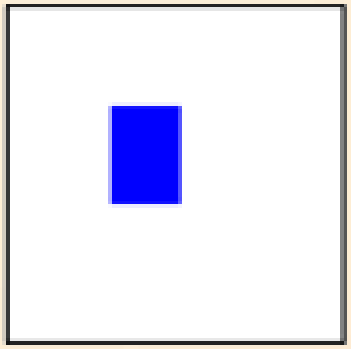


Retrieval by Similarities - Shape Similarity

Shape Similarity:

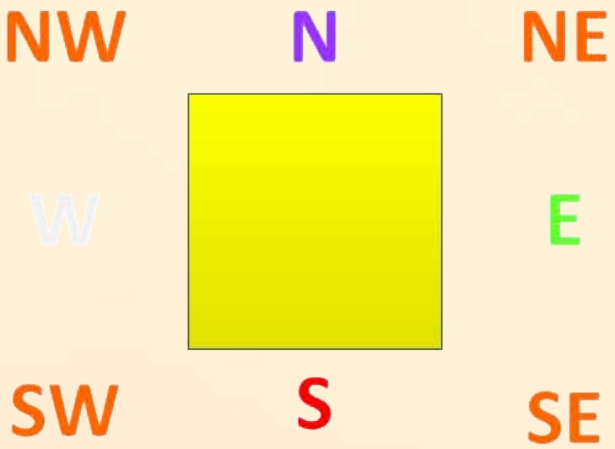
- Shape represents the shapes that appear in the image.
- Shapes are determined by identifying regions of uniform color.
- Shape is useful to capture objects.
- Shape is very useful for querying on simple shapes.



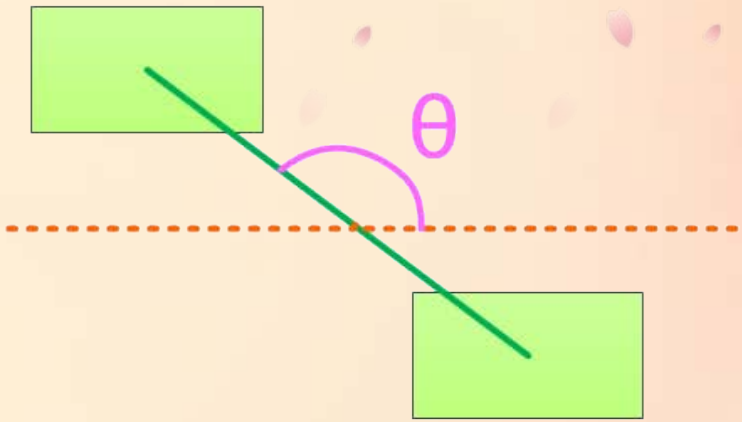


Retrieval by Similarities - Spatial Similarity (1/2)

Directional Relations



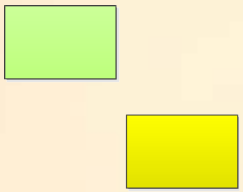
(a) strict and mixed directional relations



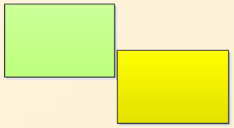
(b) slope directional relations

Retrieval by Similarities - Spatial Similarity (2/2)

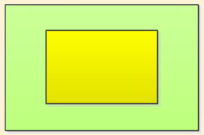
Topological Relationship



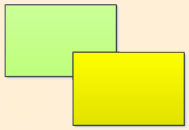
Disjoint



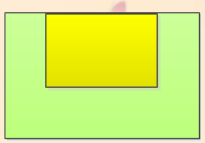
Meets



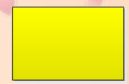
Contains
Inside



Overlap

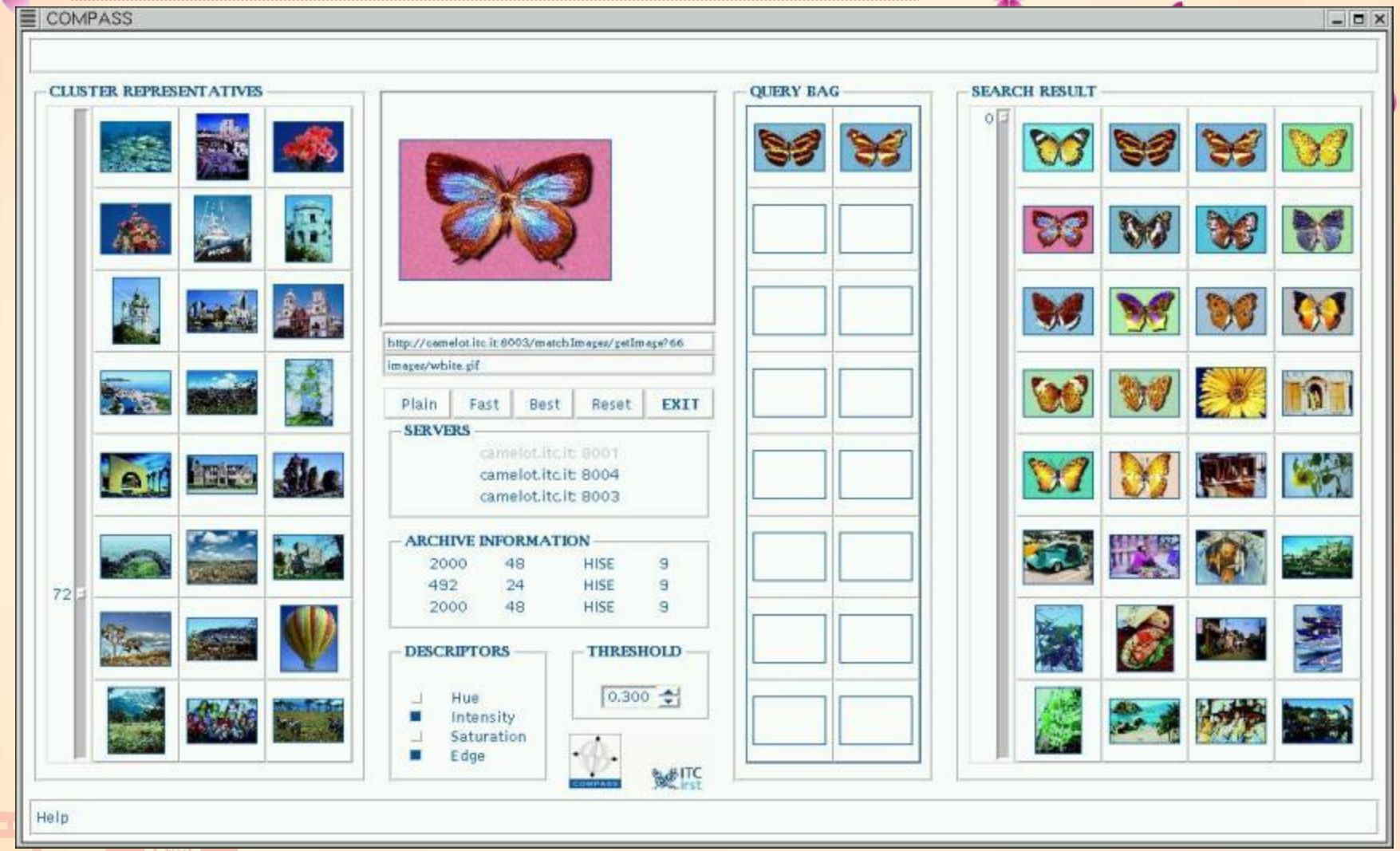


Covers
Covered by



Equals

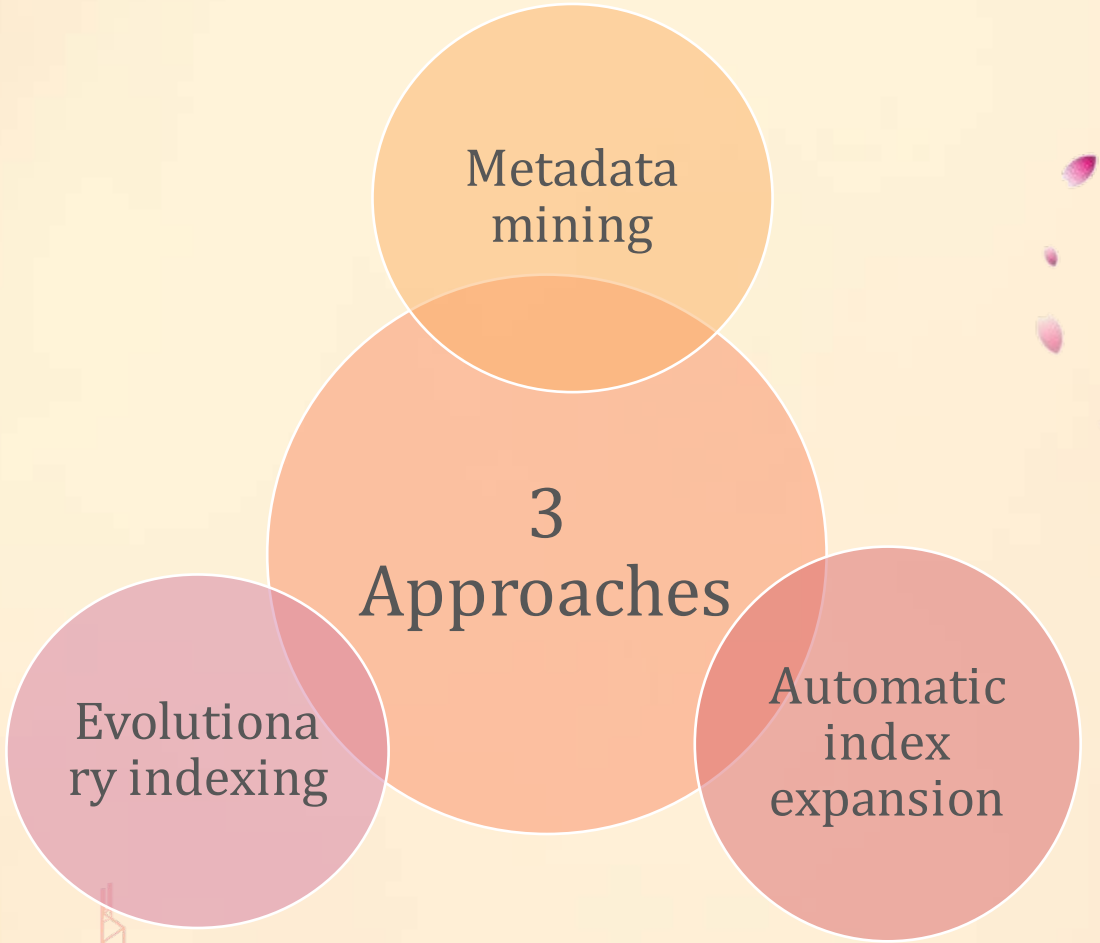
Example: COMPASS



Three Indexing & Annotation Approaches

- Metadata mining
- Automatic index expansion
- Evolutionary indexing

Three Indexing & Annotation Approaches



Automatic Annotation

Metadata Mining

- Metadata
  - Date and time
  - Aperture
  - Exposure time
  - Focal length
  - Flash activation
  - Subject distance
  - GPS information

Classification Used in Conjunction with Feature Extraction

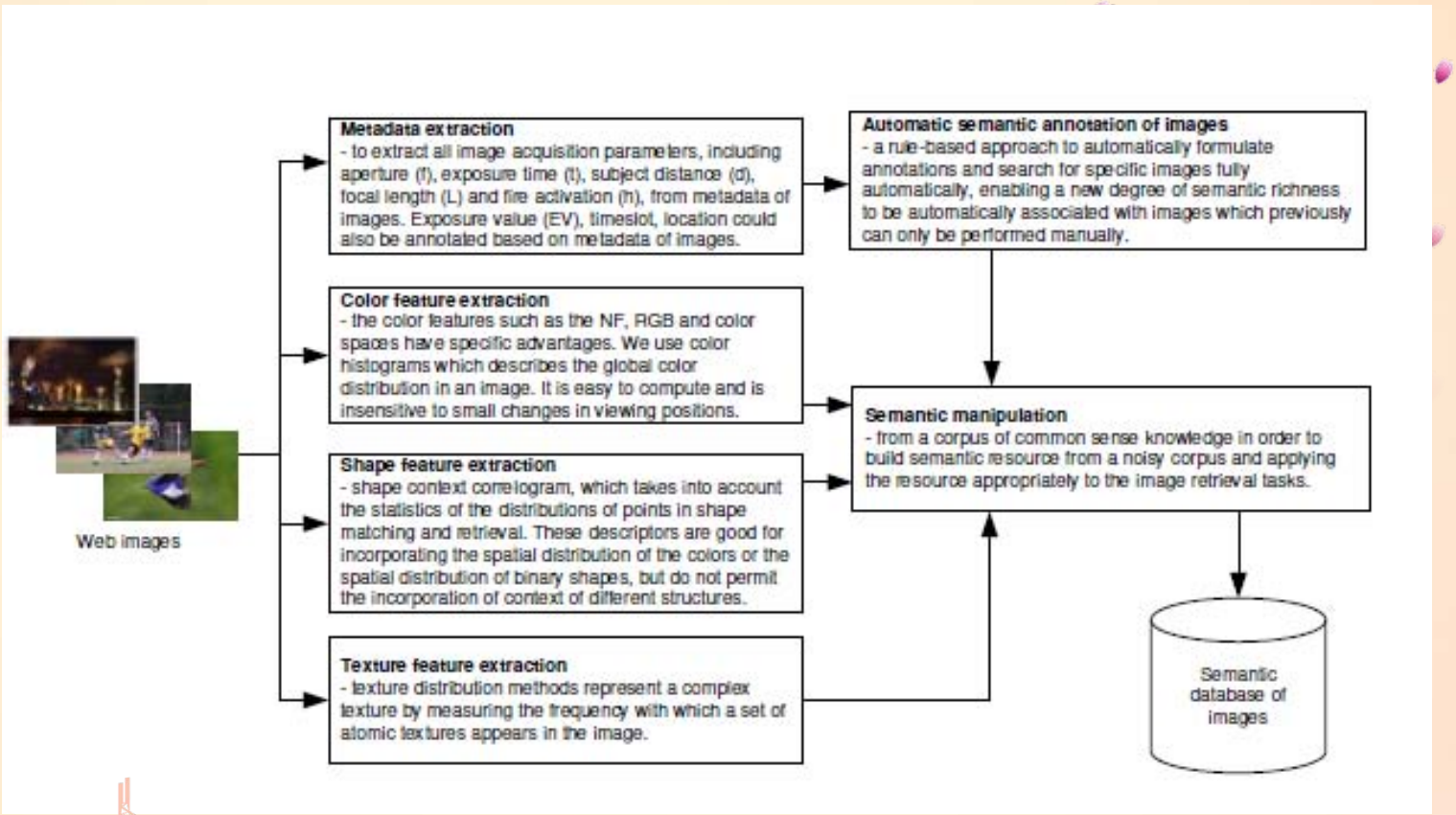
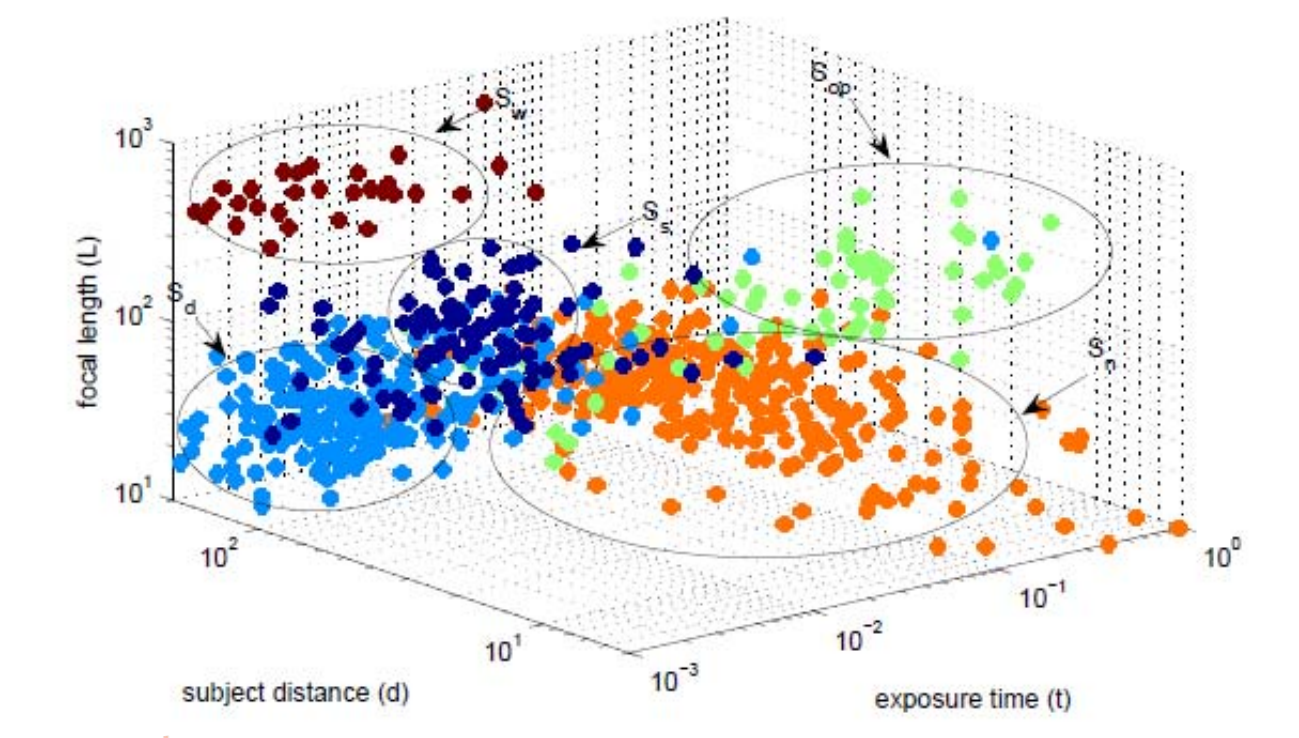


Image distribution in 3-Dimensional space



Rules for ASA

- night scene is having
  - the exposure time exceeding 0.125
  - the subject distance exceeding 30
  - the exposure value not greater than 8

$$\forall i \in I, (t_i > 0.125) \wedge (d_i > 30) \wedge (EV_i \leq 8) \Rightarrow i \in S_n,$$

- day scene is having
  - the subject distance exceeding 30
  - the exposure value greater than 8
  - the exposure time less than 0.125
  - The focal length in between 10 and 100

$$\forall i \in I, (d_i > 30) \wedge (EV_i > 8) \wedge (t_i \leq 0.125) \wedge (10 < L_i \leq 100) \Rightarrow i \in S_d,$$

Rule induction using C4.5 decision trees

Rule 1) $\forall i \in I, (t_i > 0.125) \wedge (d_i > 30) \wedge (EV_i \leq 8) \Rightarrow i \in S_n$

Rule 2) $\forall i \in I, (d_i > 30) \wedge (EV_i > 8) \wedge (t_i \leq 0.125) \Rightarrow i \in S_d$

Rule 3) $\forall i \in I, (f_i > 20) \wedge (d_i > 50) \wedge (EV_i > 11) \Rightarrow i \in S_{ss}$

Rule 4) $\forall i \in I, [(f_i \leq 5.6) \wedge (5 < d_i \leq 8)] \wedge \{[(t_i \leq 0.00625) \wedge (L_i \leq 30)] \vee [(30 < L_i \leq 182) \wedge (ISO_i \leq 250)] \vee (L_i > 182) \vee (t_i \leq 0.003125)\} \Rightarrow i \in S_{op}$

Rule 5) $\forall i \in I, (f_i > 5.6) \wedge (L_i \leq 25) \wedge (5 < d_i \leq 8) \wedge (t_i > 0.003125) \Rightarrow i \in S_{oe}$

Rule 6) $\forall i \in I, (f_i > 5.6) \wedge (0.003125 < t_i \leq 0.011111) \wedge (5 < d_i \leq 8) \wedge (L_i > 25) \Rightarrow i \in S_{ip}$

Rule 7) $\forall i \in I, (5 < d_i \leq 8) \wedge \{(f_i \leq 5.6) \wedge \{[(L_i \leq 30) \wedge (t_i > 0.00625)] \vee [(ISO_i > 250) \wedge (30 < L_i \leq 182)]\}\} \vee [(h_i = 1) \wedge (f_i > 5.6) \wedge (L_i > 25) \wedge (t_i < 0.011111)] \Rightarrow i \in S_{ie}$

Rule 8) $\forall i \in I, (d_i > 10) \wedge (150 < L_i \leq 400) \wedge (t_i \leq 0.005) \Rightarrow i \in S_s$

Rule 9) $\forall i \in I, (d_i \leq 5) \wedge (EV_i > 9) \Rightarrow i \in S_m$

Rule 10) $\forall i \in I, (L_i > 450) \wedge (d_i > 20) \Rightarrow i \in S_w$

Symbols: aperture (f), exposure time (t), subject distance (d), focal length (L) and flash activation (h)

Results of Semantic Queries

"Night scene in Hong Kong in Summer"

Results of Semantic Queries

"Wildlife in Africa in summer"

Comparison with Human Tags

Scenes	Proposed approach	Human tags (including tags, titles and description)
S_m	89.00%	82.29%
S_{ss}	55.00%	92.71%
S_s	82.50%	69.79%
S_n	92.00%	88.54%
S_{op}	91.50%	84.38%
S_w	90.50%	68.75%
S_d	83.50%	82.29%
S_{ie}	51.50%	88.54%
S_{oe}	52.00%	90.63%
S_{ip}	65.00%	92.71%

Sample Annotations

Landscape, night scenes,
Victoria Harbor, Hong Kong,
summer, night,
sea, building

Portrait, indoor events, people,
Cambridge, United Kingdom,
spring, afternoon

Nature, macro, animal,
Tanko National Park, Japan,
summer, morning
leaf

Portrait, outdoor events, people,
Cotton Tree Drive Marriage,
Registry, Hong Kong,
autumn, afternoon

Portrait, sports, people,
Yio Chu Kang Stadium, Singapore,
summer, afternoon,
motion

Nature, wildlife, animal,
Orlando Wetlands Park, Florida,
United States,
autumn, afternoon,
feather, motion

Landscape, day scenes,
Chaopraya, Bangkok, Thailand,
spring, morning,
sea, building, sky

Portrait, indoor events, people,
The Mesa Arts Center,
Mesa, Arizona, United States,
summer, night,
motion

Landscape, sunrise and sunset,
SaiKung, Hong Kong
winter, evening
sea, sky, wood

Nature, wildlife, animal,
Wetland Park, Hong Kong,
autumn, afternoon,
sea, feather

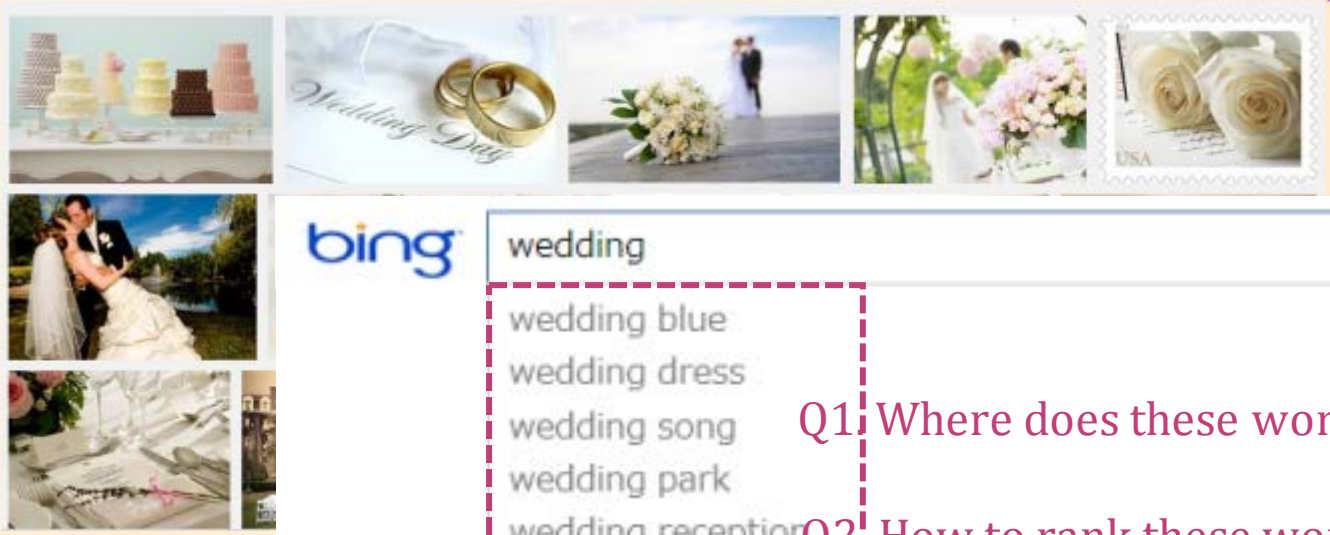
Portrait, outdoor events, people,
Yunlin County Stadium, Taiwan,
winter, afternoon

Portrait, sports, people,
Wuline Stadium, Shenyang, China,
summer, afternoon
motion

Index Expansion

Index Expansion?

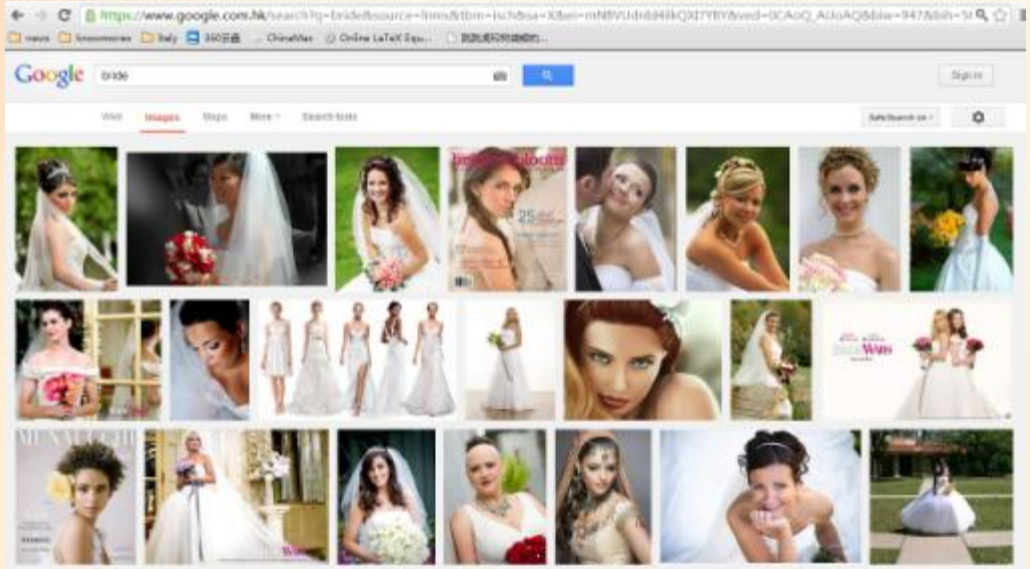
Image an Wedding Picture…



Q1: Where does these words come from?

Q2: How to rank these words?

Index Expansion



Bride

Gown

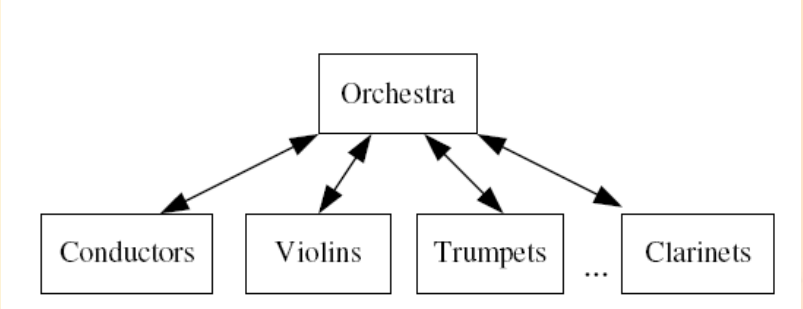
Flower

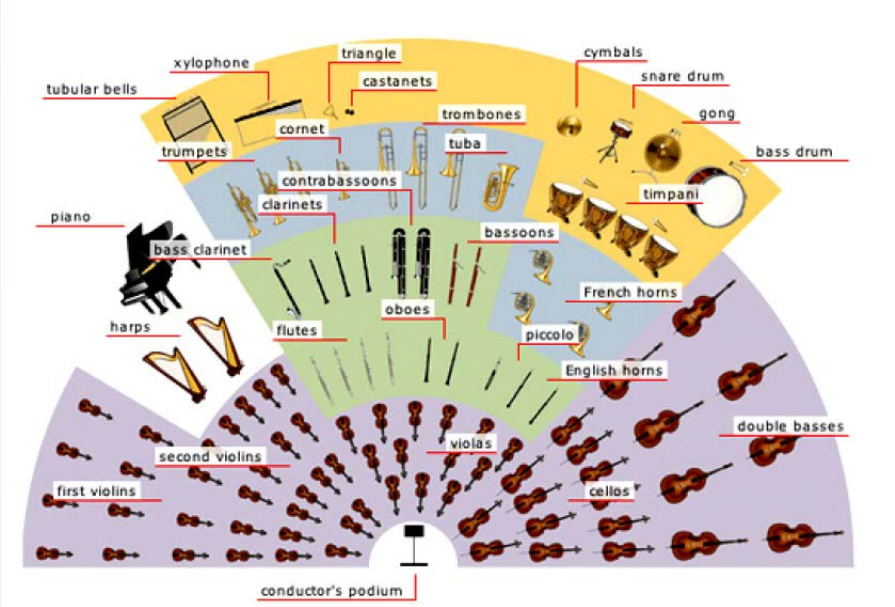
- The presence of particular objects in an image often implies the presence of other objects.
- If term U → V, and if only U is indexed, then searching for V will not return the image in the result, even though V is present in the image.
- The application of such inferences will allow the index elements Ti of an image to be automatically expanded according to some probability which will be related to the underlying ontology of the application.

Where does these words come from? Ontology-based expansion

Aggregation hierarchical expansion

- This relates to the aggregation hierarchy of sub-objects that constitute an object.
- In this example, an orchestra expands to conductors, violins, trumpets, clarinets etc

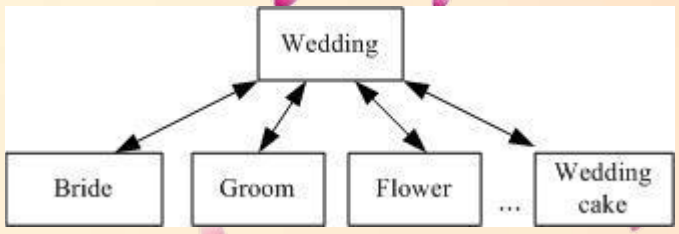




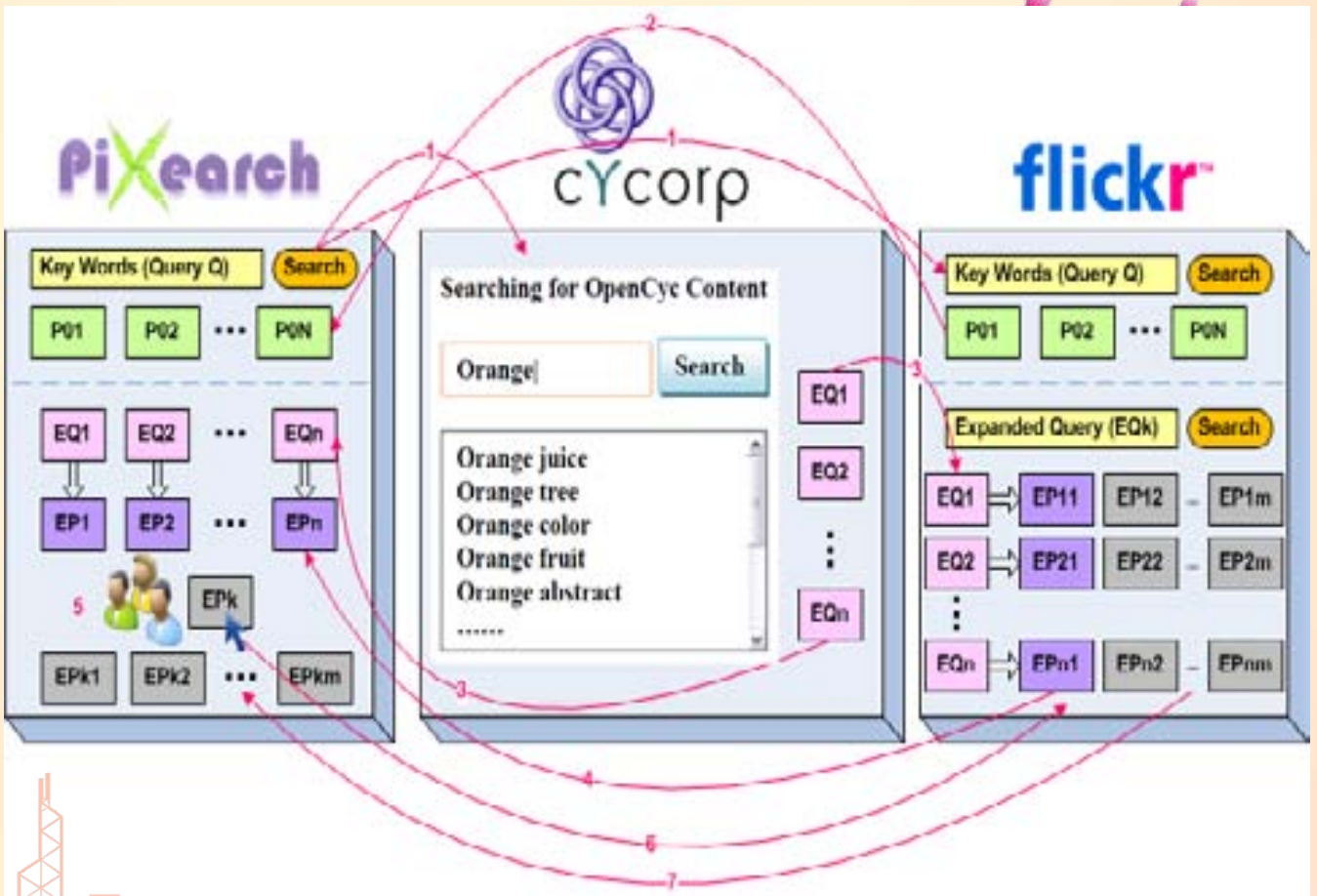
Where does these words come from? Ontology-based expansion

Co-occurrence expansion

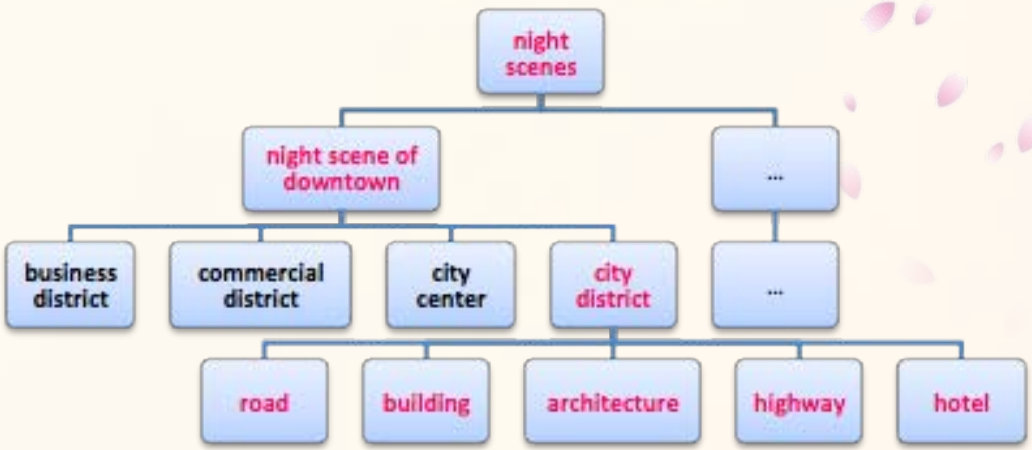
- The relevant weighting may be expressed as a conditional probability given the presence of other objects.
- In this example, it is expected that certain semantic objects (e.g. bride, groom, flower) tend to occur together.



CYC-based Query Expansion



How to rank these words?
Concept Distance



- "downtown" can be expanded to "business district", "commercial district", "city center" and "city district", while "city district" can be expanded to "road", "building", "architecture", "highway" and "hotel".

Ranking

wedding

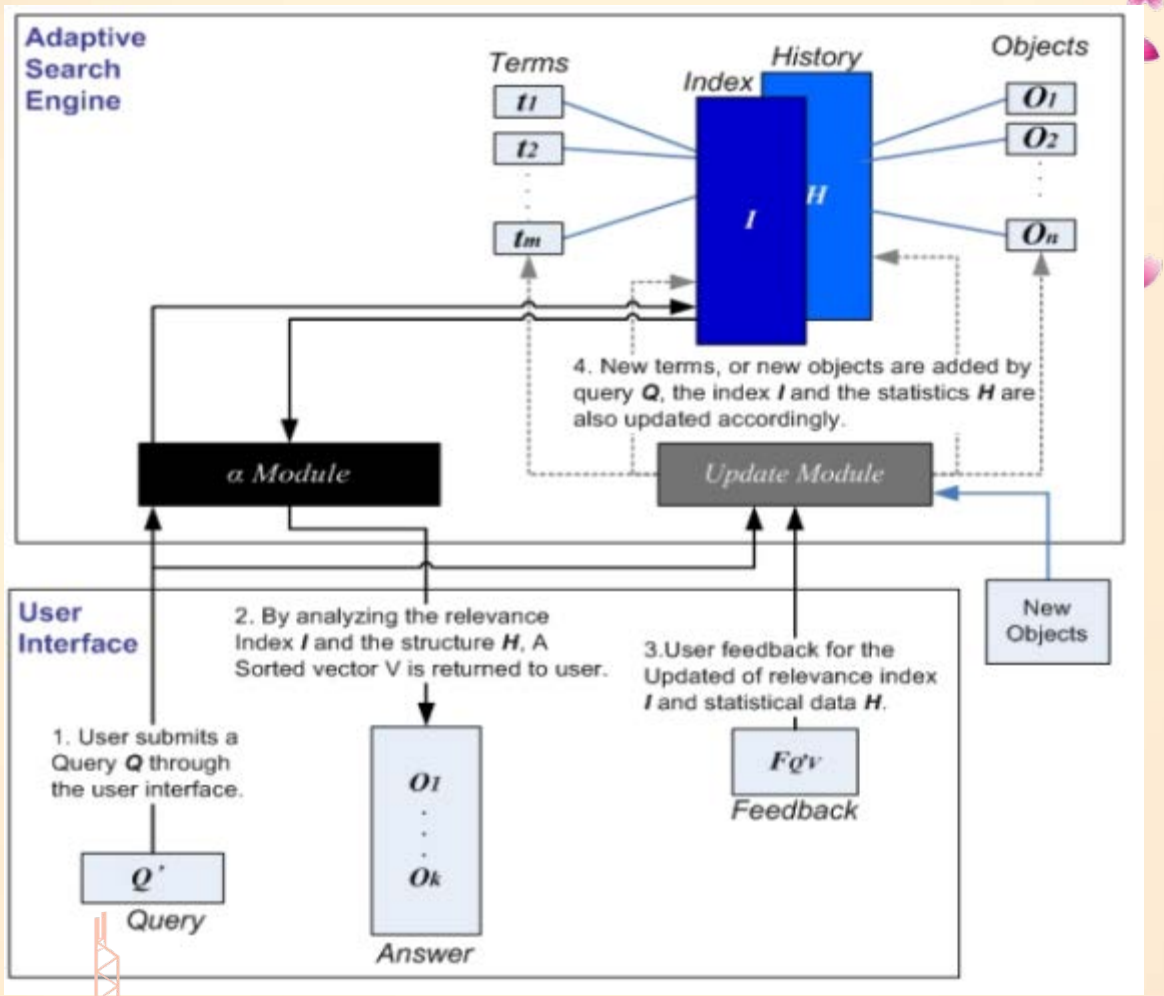
- Wedding Cake
- Wedding Ring
- Wedding Flower
- Wedding Bide
- Wedding bridegroom
- Wedding Party

Evolutionary Indexing

Collaborative Evolutionary Indexing

- Users spend a lot of time on searching and viewing information
  - Exploits visual judgment and perceptive intelligence of web users
- An evolutionary indexing paradigm
  - Capture, analyze, interpret user behavior and response
  - Support semantic visual information search through selection scoring & incremental indexing
  - Allows semantic concepts to be gradually discovered and migrated through an index hierarchy
  - Rich semantics
  - Robust and fault-tolerant

Adaptive Search Engine Architecture



Source: C. H. Leung, W. S. Chan, J. Liu, A. Milani, "Intelligent Social media Indexing and Sharing Using an Adaptive Indexing Search Engine" *ACM Transactions on Intelligent Systems and Technology*, 2012 Vol.3, No.3

Example

In Image Database
Initial Tag:
Clock Tower

- User Search by:

  Kyoto University Clock Tower

- This image returns, user clicks it

  "Kyoto University" mark+1

- User 2, 3, .. N do the same

  "Kyoto University" mark+N

- …

- Tag in Database updated:

  Kyoto University Clock Tower


Thank You
ありがとう