MUFIN Image Annotation vs. Google Label Detection:

Who Wins?

Petra Budíková, Michal Batko

Motivation

- There are many images out there…

- To enable text search, we need images with keywords

Flower, yellow, dandelion, detail, close-up, nature, plant, beautiful

- Manual annotation
- Automatic annotation
  - MUFIN Image Annotation
  - Google Label Detection

P. Budikova, M. Batko, P. Zezula. *Semantic Image Annotation by ConceptRank.* Submitted to Multimedia Tools And Applications, October 2016.

Presentation Outline

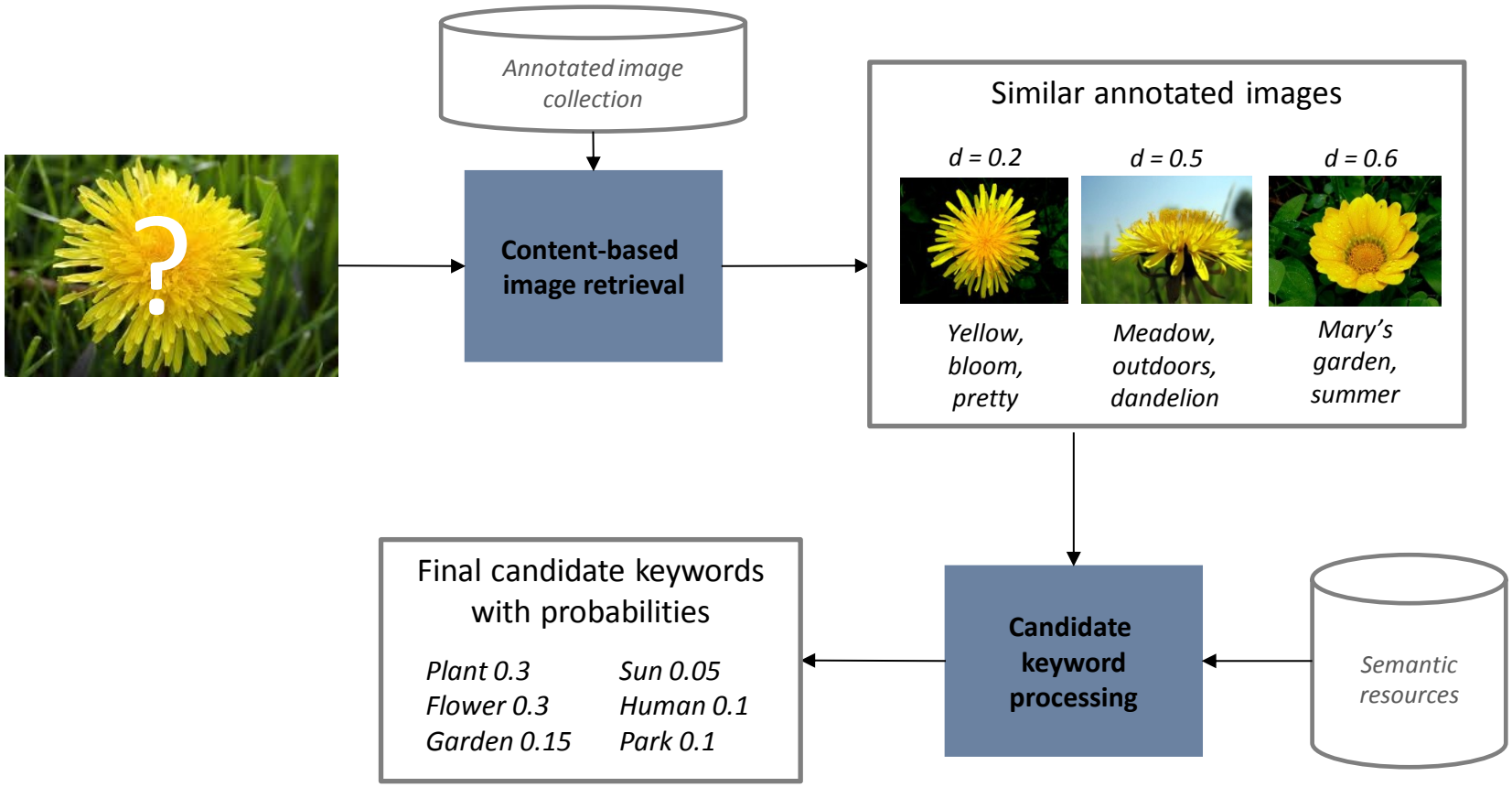
- MUFIN Image Annotation
  - Basic idea of the search-based approach
  - ConceptRank algorithm outline
  - Implementation details
- Google Label Detection
  - Basic idea of the model-based approach
  - Known and unknown details
- Comparison
  - Data and metrics
  - Results
  - Examples
- Conclusion

Part I

MUFIN Image Annotation

Solution Overview

- Search-based annotation



Phase 1: Content-based retrieval for annotations

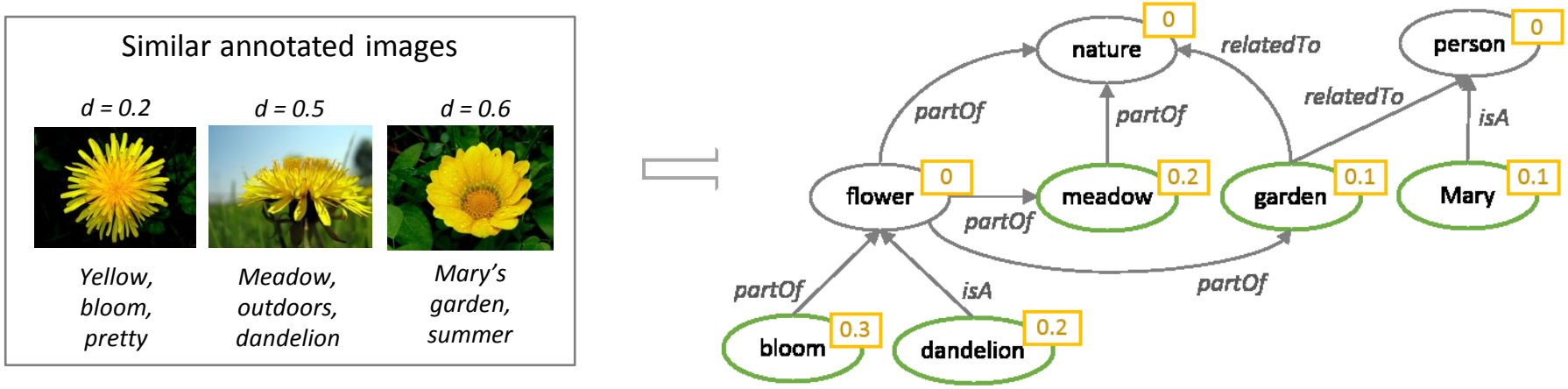
- What we need:
  - Large collection of reliably annotated images: Profiset
    - 20 million general-purpose photos from the Profimedia photostock company
    - Descriptive keywords for each photo provided by authors who want to sell the pictures → rich and reliable annotations

Profiset keywords: botany, close, closeup, color, daytime, detail, exterior, flower, germany, hepatica, horticulture, laughingstock, liverwort, lobed, mecklenburg, nature, nobilis, outdoor, outside, plant, pomerania, purple, round, western

  - Efficient and effective search: DeCAF descriptors and PPP-codes
    - DeCAF: 4096-dimensional vector obtained from the last layer of a neural network image classifier
    - PPP-codes: effective permutation-based metric space indexing method

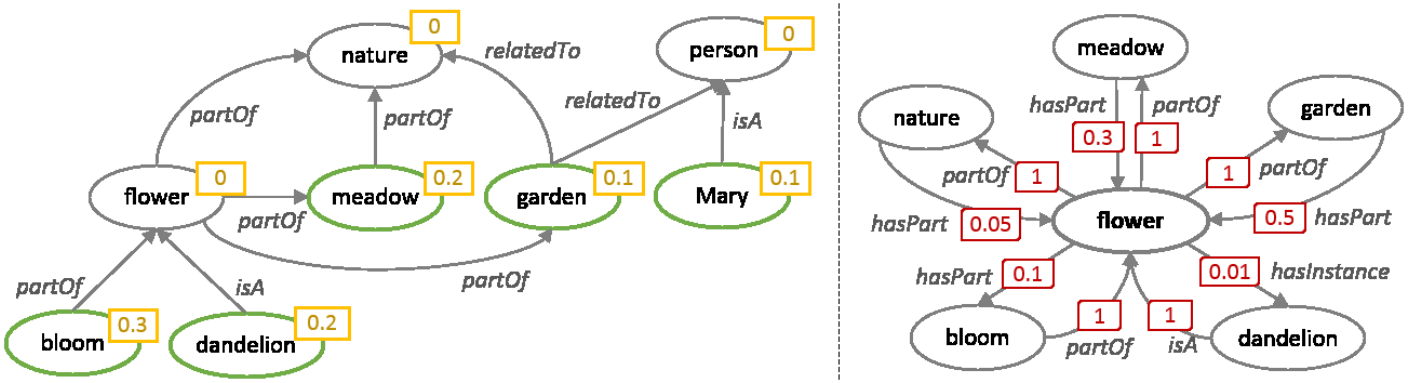
Phase 2: ConceptRank

- Candidate keyword analysis inspired by Google PageRank
- Uses semantic connections between candidate keywords to determine the probability of individual candidates
- Main steps:
  - Construct a graph of candidate keywords related by WordNet semantic links
  - Apply biased random walk with restarts to compute the score of each keyword



ConceptRank – semantic network

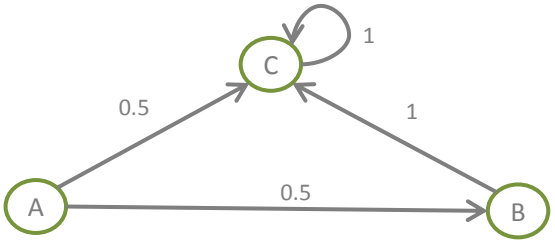
- Semantic network: graph representation of semantic relationships
  - Nodes: candidate objects
    - Node probability: current probability of the respective candidate concept
  - Edges: relationships between candidate objects
    - Edge weight: conditional probability of the target node concept, given that the source node concept is relevant



- Semantic network construction
  - Initial nodes and their probabilities taken from CBIR result
  - For each node, relationships are found in the WordNet lexical database -> new edges and nodes

ConceptRank – node probability computation

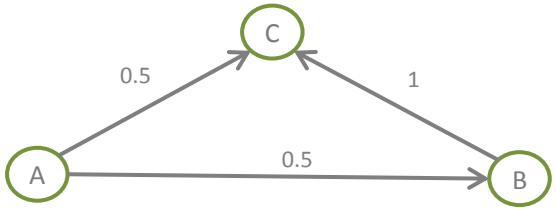
- Random walk idea (general directed graph)
  - User starts in random node and walks randomly in the graph. The importance of each node is equal to the probability that the random walker ends up in the given node.
  - Let M be a matrix describing the edges in the graph. Mathematically, we are looking for a vector r of node weights that satisfies the equation r = M.r
    - This can be computed by repeatedly multiplying a random initial vector r_0 by M until the steady state is found



Transition matrix:

Node probabilities:

  - Problem: in many real-world graphs, the matrix M is such that the equation does not work as expected

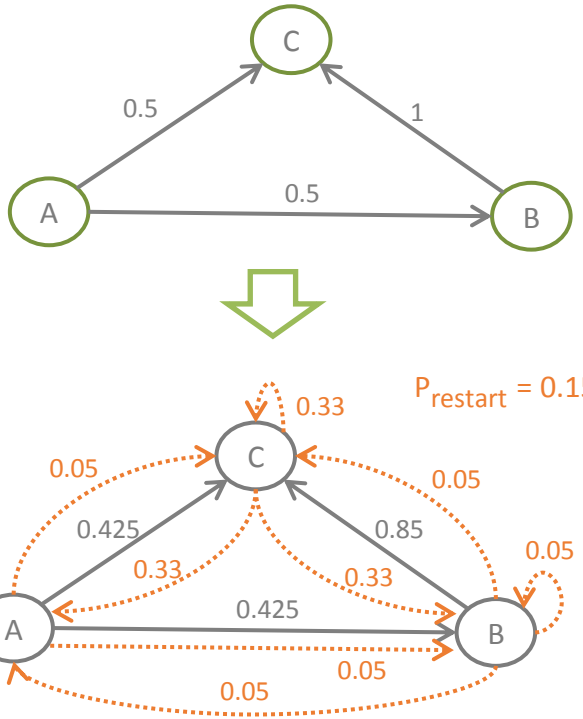


Transition matrix:

Node probabilities:

ConceptRank – node probability computation II

- Random walk with restart
  - Proposed by Google to eliminate the problems of basic random walk + model more realistically real web users
  - With a given probability $P_{restart}$, the random walker can decide in each step whether to follow the links, or to randomly restart in any node



Transition matrix: ( )

Node probabilities: ( ) ( ) ( ) ( )

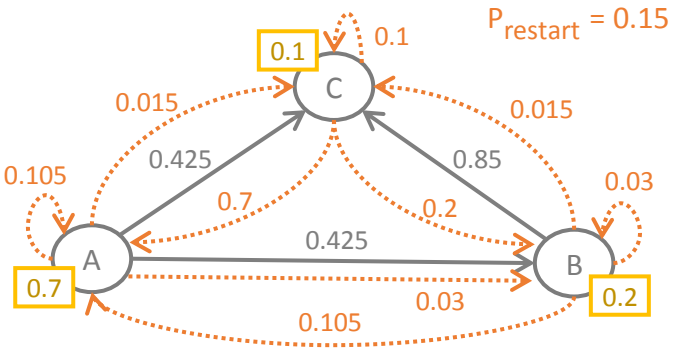
Stochastic transition matrix: ( )

Node probabilities: ( ) ( ) ( ) … ( )

ConceptRank – node probability computation III

- Random walk with biased restart
  - In standard RWR, all nodes are equal – the restart is equally probable in any node
  - Biased restart prefers some nodes over others for the restart
    - e.g. selected reliable web nodes

- ConceptRank
  - Biased RWR on the semantic network model of candidate concepts
  - The probability of restart reflects the initial probability of nodes
  - The non-restart edges represent semantic relationships



Stochastic transition matrix: $\begin{pmatrix} & \end{pmatrix}$

Node probabilities: $\begin{pmatrix} & \end{pmatrix} \begin{pmatrix} & \end{pmatrix} \begin{pmatrix} & \end{pmatrix} \ldots \begin{pmatrix} & \end{pmatrix}$

MUFIN Image Annotation – recapitulation

- CBIR:
  - 20M Profiset images
  - DeCAF descriptors, PPP-codes
  - 100-NN query
- Semantic analysis:
  - Initial candidate keywords provided by CBIR results
  - Semantic network built from initial candidates, using selected semantic relationships from WordNet
  - ConceptRank algorithm computes the final probability of all nodes
- Output selection:
  - Postprocessing: remove instances and auxiliary semantic nodes
  - Return the most probable keywords

MUFIN Image Annotation – example

1. Retrieve 100 similar images from Profiset
2. Merge their keywords, compute frequencies
3. Build the semantic network using WordNet
4. Compute the ConceptRank
5. Apply postprocessing & return 20 most probable keywords

Candidate keywords after CBIR
church, architecture, travel, europe, building, religion, germany, buildings, north, churches, christianity, america, religious, exterior, st, historic, world, tourism, united, usa, …

Semantic network
4 relationships: hypernym *(dog → animal)*, hyponym *(animal → dog)*, meronym *(leaf → tree)*, holonym *(tree → leaf)*
270 network nodes, 471 edges

ConceptRank scores
building (2.53), structure (2.41), LANDSCAPE (2.10), BUILDINGS (1.87), OBJECT (1.84), NATURE (1.78), place_of_worship (1.75), church (1.74), Europe (1.68), religion (1.64), continent (1.51), …

Final keywords
building, structure, church, religion, continent, group, travel, island, sky, architecture, tower, person, belief, locations, chapel, christianity, tourism, regions, country, district

Part II

Google Label Detection

Google Cloud Platform

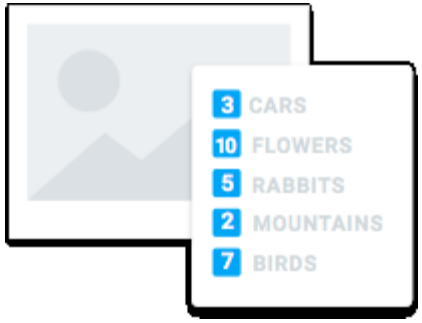
https://cloud.google.com/

- Commercial service
- Offers various tools for developers
  - Computation power (Batch processing, Web Application Engines, Containers, …)
  - Storage and Databases (Cloud key-value store, Bigtable NoSQL, …)
  - Networking (Virtual networks, Load-balancing, …)
  - Big Data support (Warehousing, Data exploration, …)
  - Machine learning (Model training, Deep neural networks, …)
    - Tools for: speech, **vision**, language translation, natural language processing
  - Management tools (Monitoring, Logging, Debugging, …)
  - Developer Tools (Cloud SDK, Application deployment, …)
  - Identity and Security (Access control, Authentication, …)
- Support for mobile applications
- Trial period for any services
  - Free credit for using any commercial service
  - Small amounts of data can be processed for free

Google Cloud Vision API

https://cloud.google.com/vision/

- Image analysis tools
  - Derive insight from images based on content
  - Exploits machine learning models
- Classification of images
  - From flowers, animals, or transportation to thousands of categories
    - e.g., "sailboat", "lion", "Eiffel Tower"
  - Improves over time as new concepts are introduced and accuracy is improved
- Detection of faces
  - Sentiment analysis
- Text recognition within images
- Offensive content filtering
- Product logo detection



- Available via REST API
  - Works either on images in Google storage or uploaded in the request

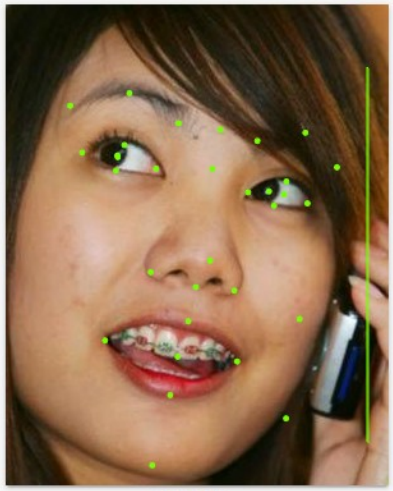
Vision REST API Example

https://cloud.google.com/vision/

dog5.jpg

Dog	99%
Mammal	94%
Vertebrate	93%
Dog Breed	91%
Labrador Retriever	89%
Puppy	86%
Dog Like Mammal	83%
Nose	81%


```
"labelAnnotations": [
  {
    "mid": "/m/0bt9lr",
    "description": "Dog",
    "score": 99
  },
  {
    "mid": "/m/04rky",
    "description": "Mammal",
    "score": 94
  },
  {
    "mid": "/m/09686",
    "description": "Vertebrate",
    "score": 93
  },
  {
    "mid": "/m/0kpmf",
    "description": "Dog Breed",
    "score": 91
  },
  {
    "mid": "/m/0km3f",
    "description": "Labrador Retriever",
    "score": 89
  },
```

oblicej.jpg

Joy	Very Likely
Sorrow	Very Unlikely
Anger	Very Unlikely
Surprise	Unlikely
Exposed	Very Unlikely
Blurred	Very Unlikely
Headwear	Very Unlikely

Roll: 10° Tilt: 8° Pan: 18°

Confidence 99%


```
"landmarks": [
  {
    "type": "LEFT_EYE",
    "position": {
      "x": 87.51825,
      "y": 118.21665,
      "z": -0.00086109631
    }
  },
  {
    "type": "RIGHT_EYE",
    "position": {
      "x": 206.26216,
      "y": 146.21912,
      "z": 38.988995
    }
  },
  {
    "type": "LEFT_OF_LEFT_EYEBROW",
    "position": {
      "x": 50.562824,
      "y": 79.266365,
      "z": 2.5175068
    }
  },
```

Vision REST API

https://cloud.google.com/vision/

- One request method

  POST https://vision.googleapis.com/v1/images:annotate

- Request specifies the image and the features to extract as JSON

```
{
  "requests":[
    {
      "image":{
        "content":"/9j/7QBEUGhvdG9...image contents...eYxxxzj/Coa6Bax//Z"
      },
      "features":[
        {
          "type":"LABEL_DETECTION",
          "maxResults":1
        }
      ]
    }
  ]
}
```

- Types of features to extract
  - LABEL_DETECTION, TEXT_DETECTION, FACE_DETECTION, IMAGE_PROPERTIES, LANDMARK_DETECTION, LOGO_DETECTION, SAFE_SEARCH_DETECTION

Pricing

Feature	1 - 1000 units/month	1001-1,000,000 units/month	1,000,001 to 5,000,000 units/month	5,000,001 - 20,000,000 units/month
Label Detection	Free	$5.00	$4.00	$2.00
OCR	Free	$2.50	$1.50	$0.60
Explicit Content Detection	Free	$2.50	$1.50	$0.60
Facial Detection	Free	$2.50	$1.50	$0.60
Landmark Detection	Free	$2.50	$1.50	$0.60
Logo Detection	Free	$2.50	$1.50	$0.60
Image Properties	Free	$2.50	$1.50	$0.60

Vision API – Under the Hood?

- Only vague phrases mentioned by Google
  - Uses deep neural network
  - Classification into several thousands of labels
  - Specifics are not disclosed

- Our guess
  - Probably some improved deep convolutional “Inception” model
    - Currently v3 (https://github.com/tensorflow/models/tree/master/inception)
    - Based on ImageNet training data
    - TensorFlow implementation (https://www.tensorflow.org)

  - We have seen quite specific detection of animals and cars
  - Not so good detection of person-related labels
    - But face-detection seems to work well, so it can be potentially combined
      - Google does not include the results of face detection and sentiment in labels?
  - Presents labels only if their scores are greater than 50%

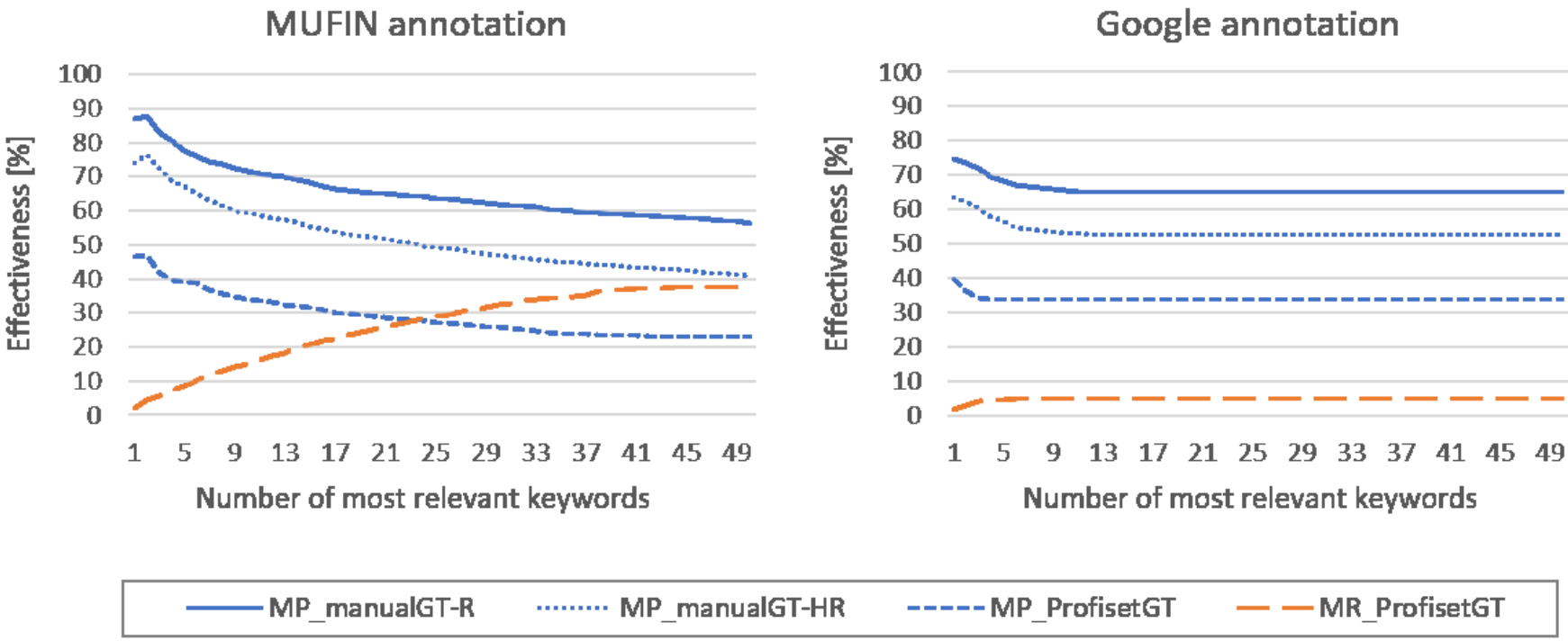
Part III

Comparison

Data & Evaluation Metrics

- Queries
  - 166 images from Promedia: 86 photos selected manually, 80 chosen randomly
  - The query images were removed from the Profiset collection
    - so there is no overlap between the test queries and the annotated image collection used as knowledge base for the MUFIN Image Annotation processing
- Ground Truth
  - Manual GT – created by manual evaluation of keywords provided by MUFIN and Google
    - Two types: GT-R contains all keywords assessed as “relevant”, GT-HR contains only “highly relevant”
  - Profiset GT – the original image descriptions
- Quality measures
  - Precision: can be computed w.r.t. all types of GT
  - Recall w.r.t. Profiset GT
    - The manual GT is not complete

Results



- On the first positions, MUFIN about 8 % better!
- For the same precision, MUFIN gives significantly better recall

Results (cont.)

- Failed annotations:
  - Google returned no keywords for 11 images out of 166
  - MUFIN failed to return anything relevant among the top 5 keywords only for 3 images
- Average result size:
  - MUFIN: 50 keywords
  - Google: 5.7 keywords (maximum 16)
- Average overlap:
  - 1.9 keywords appear in both results
  - out of these, 1.75 keywords is relevant

Where MUFIN wins

Google keywords
product

MUFIN keywords
person, **adult**, animals, activity, **scientist**, woman, knowledge, **people**, health, **research**, work, wellbeing, **science**, **indoors**, **one**, **laboratory**, **head**, mid, **years**, **man**, **clothing**, female, medical, doctor, coat, care, prosperity, hospital, **men**, worker, male, think, **equipment**, personnel, technician, **researcher**, working, young, professionals, color, **occupations**, technology, bioscience, organization, two, **photography**, healthcare, holding, african

Google results
t_shirt

MUFIN results
person, **school**, juvenile, blackboard, adult, **classroom**, **mathematics**, knowledge, science, **student**, subject, woman, female, objects, **room**, communication, child, **one**, activity, **education**, **people**, young, **years**, educator, **teenager**, teacher, **indoors**, professionals, **youth**, males, girl, hair, boy, teen, learning, mid, **length**, arithmetic, **writing**, building, head, color, man, teaching, part, board, ethnicity, high, schools, location

Google results
line

MUFIN results
fingerprint, finger, **print**, individual, group, identification, hand, crime, evidence, thumbprint, identity, ideas, **white**, finding, concept, black, information, digit, change, smudge, discovery, biometrics, thumb, unique, id, security, recognition, closeup, representation, people, vector, close, criminal, privacy, police, science, safety, photo, tech, **background**, touch, heritage, theft, curves, verify, investigation, offender, ink, state, symbol

Where Google wins

Google keywords
bumper, automotive_design, automotive_exterior, vehicle, car, wheel, land_vehicle, sports_car, mercedes_benz, supercar, automobile_make, mercedes_benz_slr_mclaren, model_car

MUFIN keywords
car, show, vehicle, travel, transport, sports, motor, automobile, speed, person, **luxury**, coupe, **new**, museum, road, indoors, concept, color, view, manufacturers, front, three, automotive, horizontal, **expensive**, nobody, convertible, business, photography, roadster, industry, european, study, transportation, **fast**, photo, silver, **modern**, salon, make, street, white, showpiece, cars, black, republic, city, studio, district, state

Google results
volcanic_landform, lava, phenomenon, geological_phenomenon, landform

MUFIN results
sky, evening, water, ocean, change, island, set, cloud, sunrise, formations, dusk, clouds, light, sunset, morning, mountain, group, sundown, national, lava, nature, outdoors, daylight, volcanoes, color, sea, travel, weather, big, **natural**, geyser, scenery, sun, red, park, **night**, horizontal, scenic, coast, gap, vacation, region, rock, people, environment, **eruption**, power, shore, landscape, countries

Google results
insect, pollen, pattern, membrane_winged_insect, honey_bee, flower

MUFIN results
tree, autumn, plant, season, travel, change, **yellow**, fall, leaves, flower, color, aspen, **animal**, quality, nature, poplar, sunflower, insect, discolored, person, arthropod, water, **nobody**, colors, horizontal, **close**, background, flora, **invertebrate**, summer, forest, image, detail, colour, creek, group, natural, outdoors, river, bee, mountains, grunge, deciduous, national, sierra, new, beautiful, supply, locations, treetop

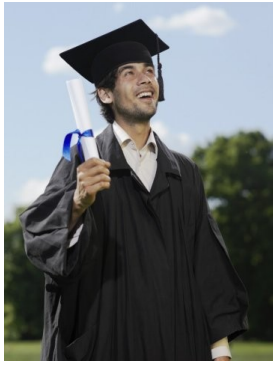
Where both MUFIN and Google are successful

Google results
penguin, flightless_bird, vertebrate, **bird**,

MUFIN results
penguin, **animal**, **group**, **bird**, seabird, aptenodytes, wildlife, chicks, **snow**, **continent**, **baby**, **emperor**, children, hill, offspring, **young**, island, **outdoors**, **sea**, ice, **water**, **nobody**, birds, daytime, cold, weather, nature, color, **flightless**, wild, laughingstock, colony, adult, day, glacier, fauna, outdoor, body, polar, travel, **photography**, marine, antarctic, horizontal, region, natural, peninsula, cute, outside, regions,

Google results
goal, soccer_kick, soccer_player, player, football_player, sports, **soccer**, kick,

MUFIN results
person, activity, **football**, recreation, **sport**, **soccer**, golf, **adult**, **young**, **years**, woman, game, **men**, ball, man, features, player, **people**, **length**, **group**, male, **outdoors**, **color**, athlete, **view**, **playing**, two, play, **green**, **grass**, child, equipment, one, **team**, compete, action, lifestyle, examining, juvenile, **competition**, female, rugby, baseball, ballgame, full, attitude, locations, lawn, outside, **field**,



Google results
graduation, academic_dress, **mortarboard**,

MUFIN results
person, **graduation**, group, completion, **student**, body, **clothing**, **adult**, college, communication, woman, **diploma**, **gown**, **young**, juvenile, **people**, school, university, get, **dress**, **cap**, **certificate**, **achievement**, activity, **graduate**, **education**, **headgear**, years, **document**, female, **smiling**, man, teenager, male, glasses, portrait, eye, youth, asian, **length**, kids, happy, communicate, **academic**, ethnicity, **holding**, mid, **men**, caucasian, studio,

Efficiency

- MUFIN: approximately 700 ms needed for single image annotation
  - 54 ms for DeCAF descriptor extraction (GPU implementation)
  - 390 ms for content-based search in 20M images (PPP-codes + PCA)
  - 40 ms for semantic network construction
  - 200 ms for ConceptRank computation (approximate RWR)

- Google: approximately 200 ms needed for single image annotation
  - Including uploading the image and the REST service overhead
  - Network overhead (RTT) about 8ms

Part IV

Conclusions

Conclusions

- MUFIN Image Annotation works!
  - In our experiments, even better than Google
  - Very good results also in the ImageCLEF competition
- MUFIN Image Annotation is effective, efficient, and scalable
- The MUFIN and Google solutions are in several aspects complementary
  - different basic approach (search-based vs. model-based)
  - provides different types of annotations
  - what is problematic to MUFIN is often easy for Google and vice versa
- Promising direction for future: combining the two approaches
  - Ideally in a generalized ConceptRank model