Content-Based Annotation and Classification Framework: A General Multi-Purpose Approach

Michal Batko, Jan Botorek, Petra Budikova, Pavel Zezula

Laboratory of Data Intensive Systems and Applications
Faculty of Informatics, Masaryk University, Brno

IDEAS 2013

Outline

- Motivation
  - Why annotations?
  - State-of-the-art in multimedia annotation
- General annotation model
  - Global architecture
  - Application to selected tasks
  - Specification of components
- Web image annotation
  - Current implementation
  - Experimental evaluation
- Conclusions and future research directions

Motivation

Yellow flower

nature

dandelion

Taraxacum officinale

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

The first dandelion that bloomed this year in front of the White House.

„Image is worth a thousand words.“

Why do we need the thousand words?

- Keyword-based image retrieval
  - Popular and intuitive
  - Needs pictures with text metadata, we do not want to create them manually
- Information seeking: “What is in the photo I just took?”
  - Tourist information / Plant identification / …
- Impaired users
- Classification tasks
  - Scientific data (medicine, astronomy, chemistry, …)
  - Improper content identification
  - Personal image gallery
- Data summarization: “What images are on this computer?”

- Not only images!
  - Sound, video, ….

Several dimensions of the annotation problem

- Input
  - Image / Image and seed keyword / Image and text / Text
- Type of information needed
  - Identification / Detection / Categorization
- Vocabulary
  - Unlimited vocabulary / Controlled vocabulary
- Form of annotation required
  - Sentence / Set of keywords / All relevant categories / A single category / Localization in a taxonomy
- Interactivity
  - Online / offline annotation

- **Easy tasks**: identify a single relevant category from a short list
- **Difficult tasks**: wide (unlimited) vocabulary, “all relevant needed”, online processing, very little or no input text

State-of-the-art text-extraction techniques

- Pure text-based
  - Analyze the text on a surrounding web page
- Content-based / Content- and text-based
  - Mainly exploit visual properties (+ text when available)

- Content-based annotation scenario:
  - Basic annotation
    - Model-based: train a model for each concept in vocabulary
    - Search-based: kNN search in annotated collection
  - Annotation refinement
    - Statistical
    - Ontology-based
    - Secondary kNN search
    - …

Existing approaches – summary

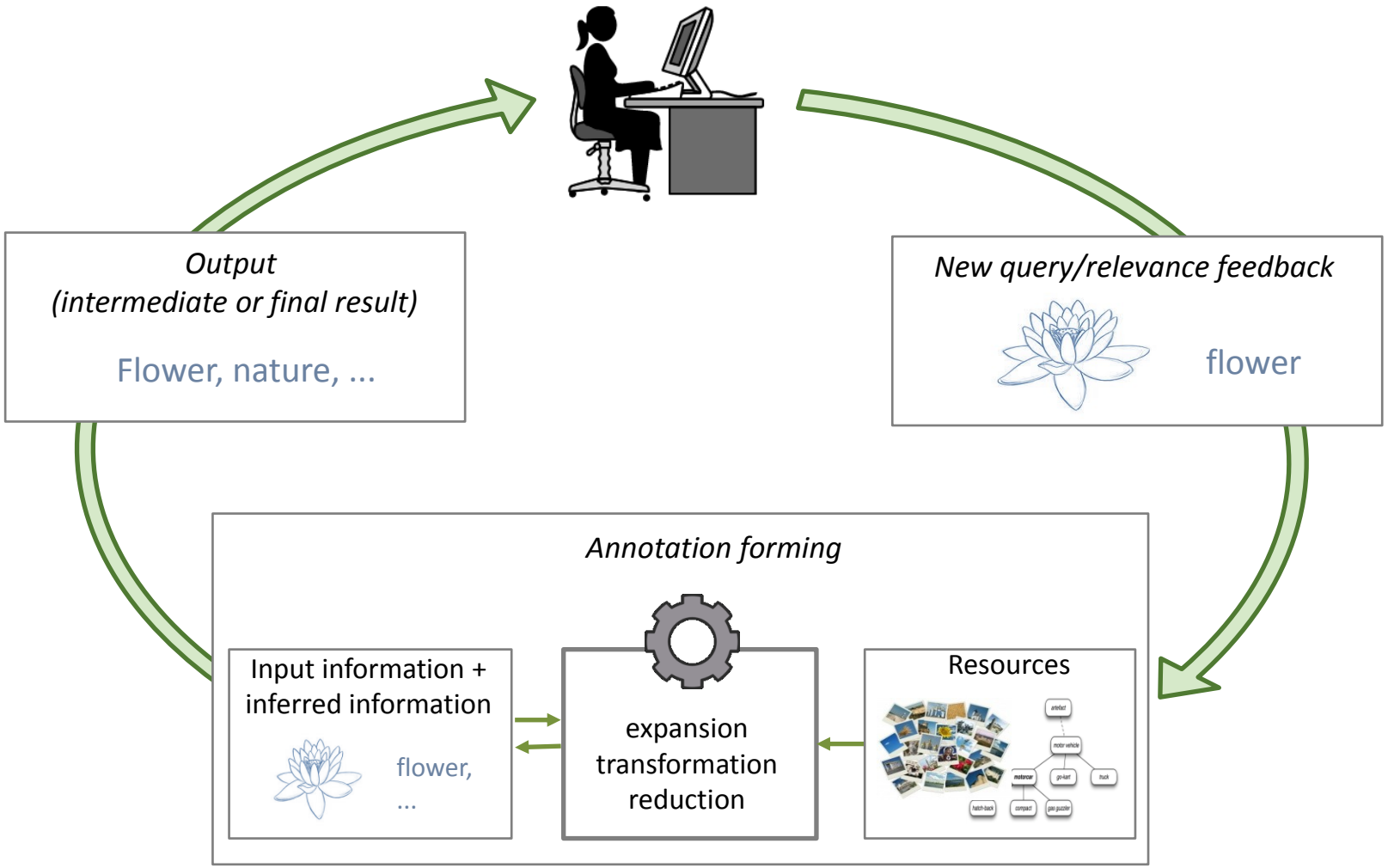
- Model-based techniques:
  - \+ Specialized classifiers can achieve high precision
  - \+ Fast processing
  - \- Training feasible only for a limited number of concepts feasible, high-quality training data needed
- Search-based techniques:
  - \+ Can exploit vast amounts of annotated data available online
  - \+ No training needed, no limitation of vocabulary
  - \- Costly processing when large datasets need to be searched
  - \- Content-based similarity measures often not precise enough

- Summary of state-of-the-art:
  - Mostly specialized solutions for a specific type of application
  - Reasonable results only for simple tasks

Our approach

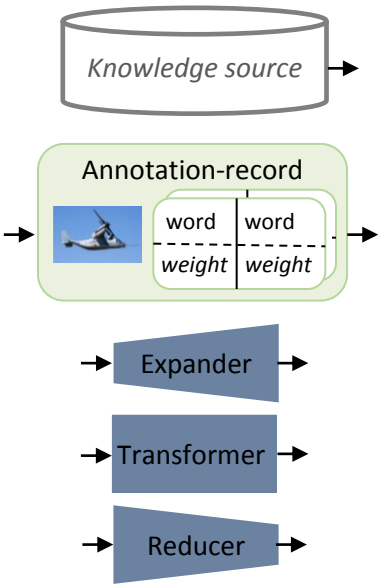
- Facts
  - Experiments show that state-of-the-art solutions are not very successful for complex problems
  - Psychologic research suggests hierarchical annotation
- Our vision:
  - Broad-domain annotation is a complex process, needs to be modeled as such
    - Multiple processing phases
    - Modular design
    - Hierarchic annotation
    - Combine multiple knowledge sources
    - User in the loop
  - The same infrastructure can be used for different applications (annotation, classification, …)
    - The principal components are the same
    - Easy evaluation, comparisons

General annotation model



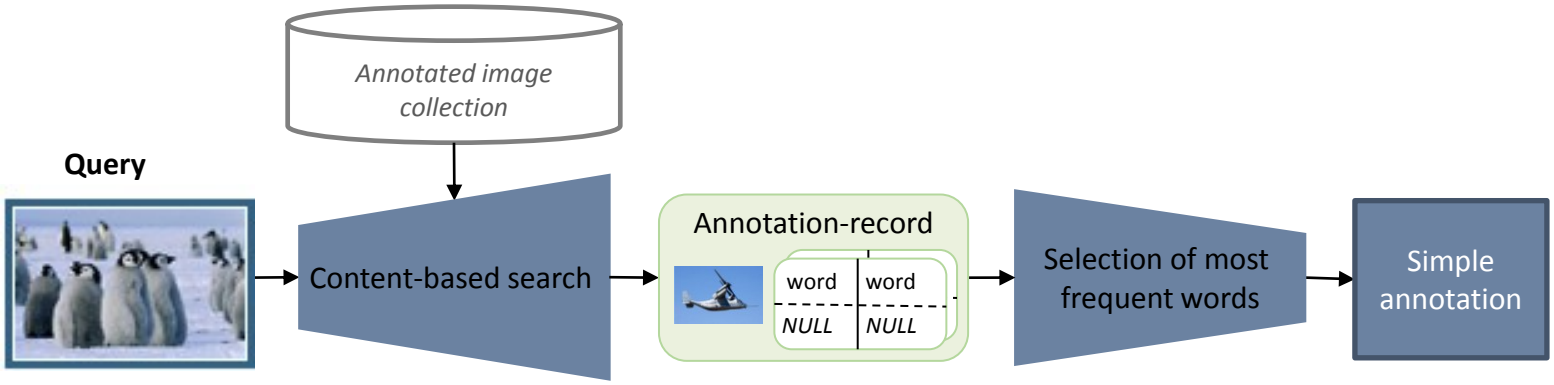
General annotation model (cont.)

- Framework components
  - Query
    - Image / image + text / (text)
  - Knowledge sources
    - Annotated image collection, WordNet, ontologies, internet, …, user
  - Annotation-record
    - Query + candidate keywords, weights, any other knowledge
  - Processor modules
    - Expander, transformer, reducer
  - Evaluation scenarios
- Properties
  - Clear structure, modularity
  - Can be adapted to various annotation/classification tasks
  - Supports extensive experiments, comparison of techniques

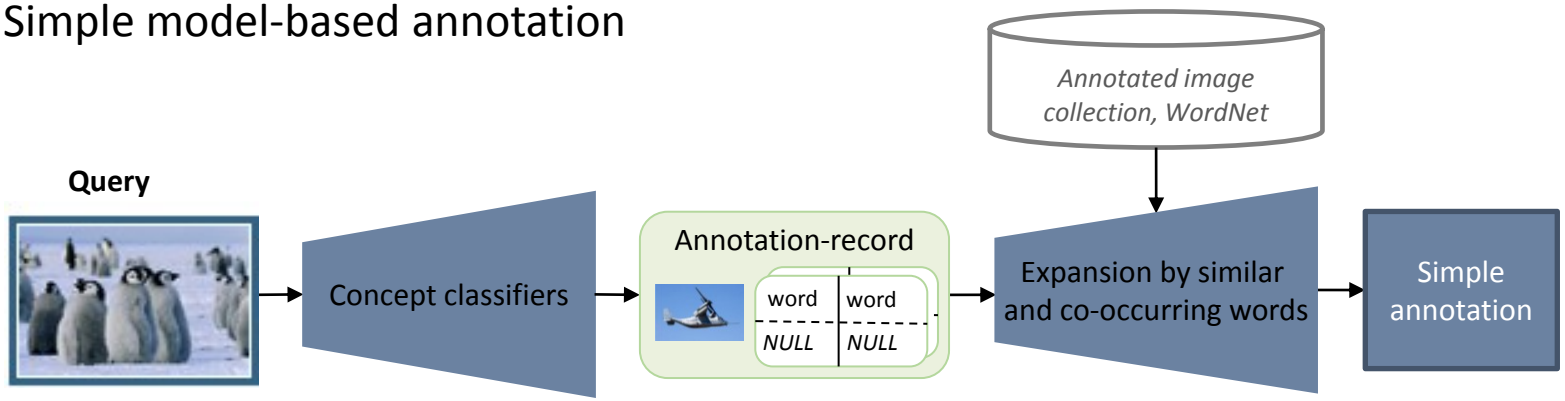


Simple examples

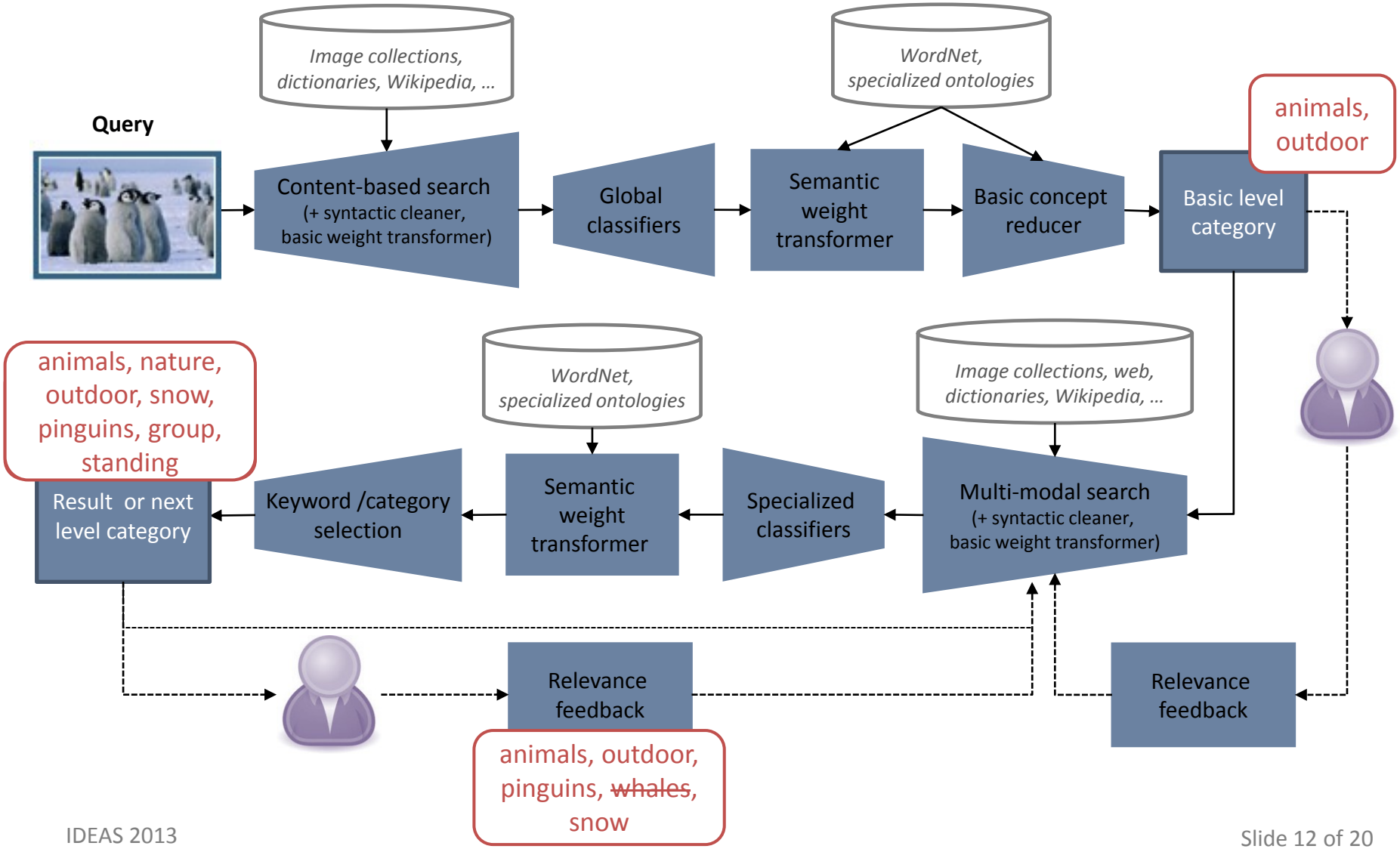
- Basic search-based annotation



- Simple model-based annotation



Advanced example: Hierarchic image annotation



Processing modules

- "The brain of the annotation process"

- Expanders
  - Provide candidate keywords
  - Visual-based nearest-neighbor search
    - Similarity measured by MPEG-7 global descriptors
    - Metric search provided by efficient M-index structure
    - Knowledge source: annotated image collection
  - Face detection software
    - Luxand FaceSDK
      - commercial library for detection and recognition of faces
    - Depending on number of faces detected, people-related concepts are added to annotation-record

Processing modules (cont.)

- Transformers
  - Adjust weights of candidate keywords
  - Basic weight transformer
    - Frequency of a keyword in the descriptions of similar images
    - Similarity score of each image with the particular keyword
    - Knowledge source: descriptions of similar images
  - Semantic transformer
    - Uses WordNet hierarchies to cluster related words
    - Keyword weight increased proportionately to the size of containing cluster
    - Knowledge source: WordNet
- Reducers
  - Remove unsuitable candidates
  - Syntactic cleaner
    - Stopword removal, translation, spell-correction
    - Knowledge sources: WordNet, dictionaries, Wikipedia

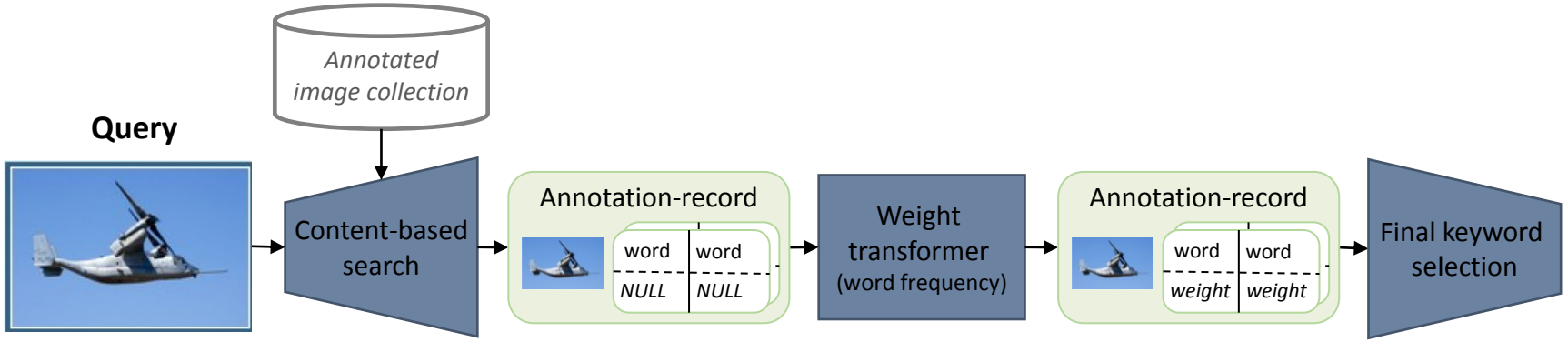
Web image annotation problem

- Task specification

  "Given an image, provide the *K* most relevant keywords that describe the content of this image."

- Use case
  - A professional photographer uploading images to a photo-selling site needs to provide accompanying keywords to enable text search

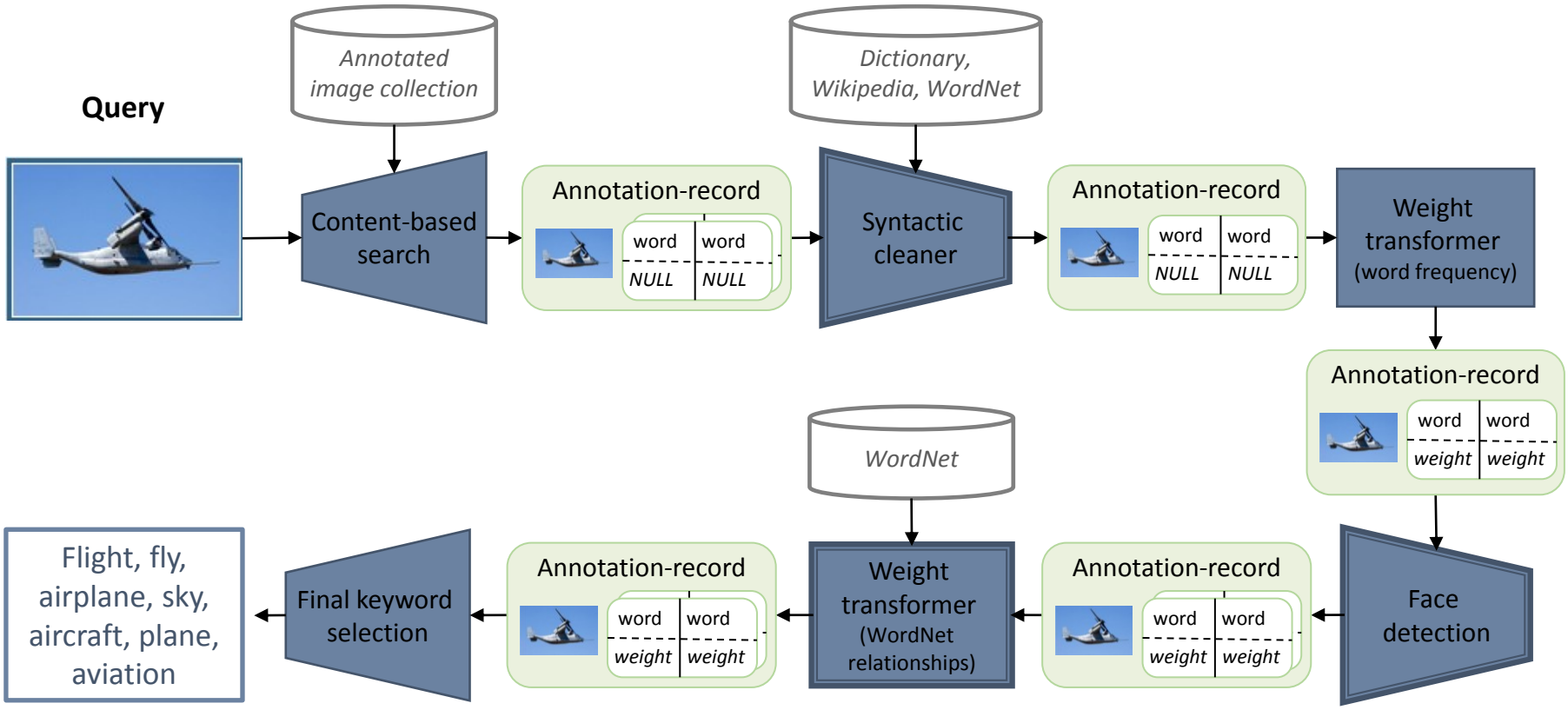
- Basic solution



Budikova, Batko, Zezula: *Online image annotation*. SISAP 2011.

Web image annotation problem (cont.)

- A more complex solution



Web image annotation – evaluation

- Methods under comparison
  - Original search-based annotation
  - Cleaned keywords
  - Boosting by distance
  - Clustering by WordNet meaning
  - Face detector boosting
  - Face detector enrichment
- Evaluation methodology
  - 160 test queries
    - Categories easy/medium/difficult
    - 20 best keywords requested
  - Result relevance evaluation:
    - User-provided (result relevance assessments)
    - Automatic (comparison to image description provided by author)

Web image annotation – evaluation (cont.)

Easy query

entertainment, art, **sparkling**, **event**, **enjoyment**, show, display, air, **celebration**, **festival**, flash, level, **fireworks**, cracker, **explosion**, **fire**, excitement, **firecracker**, **light**, bang

Medium query

blossom, location, **plant**, bird, food, **trees**, **natural**, **citrus**, flowers, generic, antique, destinations, **nature**, recreation, **tree**, **foliage**, botany, **fruit**, determination, flower

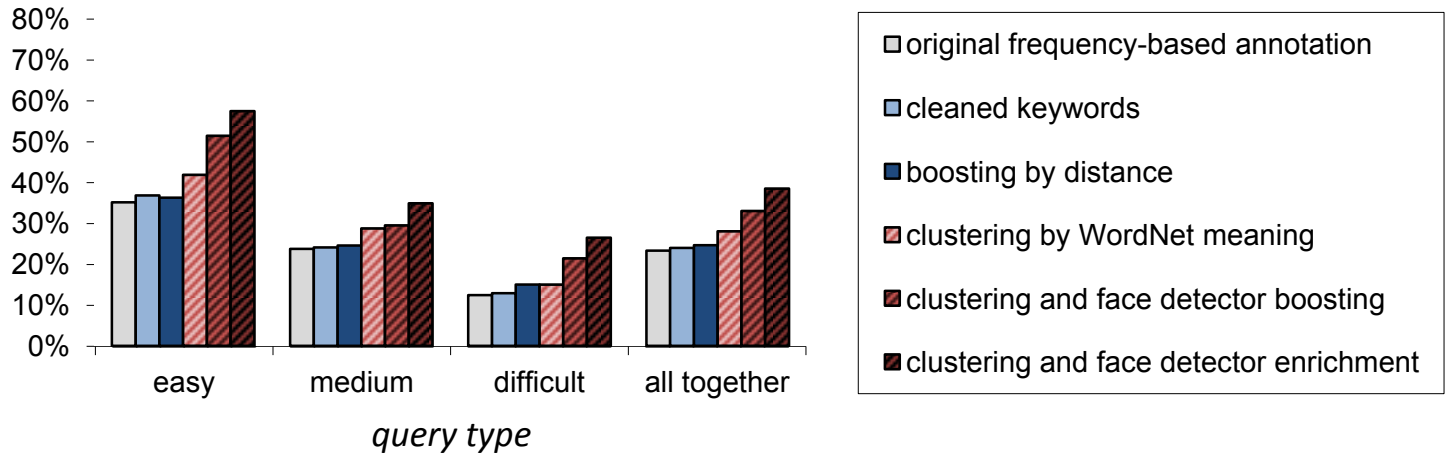
Difficult query

form, station, antique, **interior**, frame, **bookcase**, indoors, group, animal, antiques, snack, **person**, construction, food, chinese, **study**, wood, architecture, dynasty, building

Web image annotation – evaluation (cont.)

Precision – user evaluation

highly relevant keywords (dark) + relevant keywords (light)



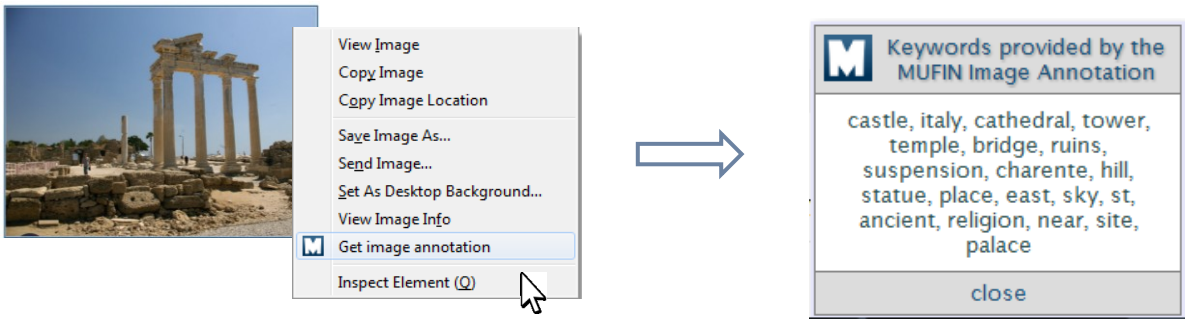
Processing costs

Pipeline type	$k = 5$	$k = 10$	$k = 15$	$k = 30$
original frequency-based annotation	0.94 s	1.34 s	4.35 s	7.44 s
cleaned keywords[a]	0.94 s	1.34 s	4.35 s	7.44 s
boosting by distance	1.04 s	1.44 s	4.53 s	8.15 s
clustering by WordNet meaning	1.98 s	2.94 s	4.52 s	13.89 s
clustering & face detector boosting	2.24 s	3.27 s	4.81 s	14.47 s
clustering & face detector enrichment	2.13 s	3.18 s	4.74 s	14.33 s

Conclusions

- Image annotation remains a challenging task
  - Broad domains, interactive applications, lack of training data, …
- Our contributions
  - General annotation model & implementation framework
  - Implementation & evaluation of several processing components
  - Improved annotation tool

    http://disa.fi.muni.cz/prototype-applications/image-annotation/



- Future work
  - Refinement of semantical analysis
  - Development of new components, hierarchic annotation processing
  - Relevance feedback strategies for image annotation

More experimental results

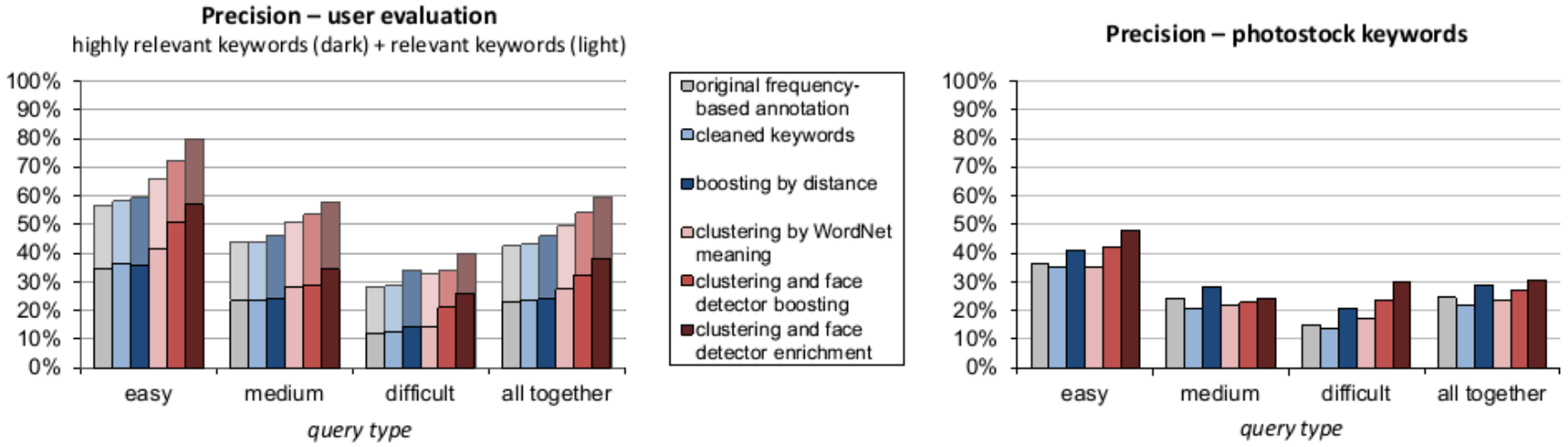


Figure 7: Profimedia test-set results.

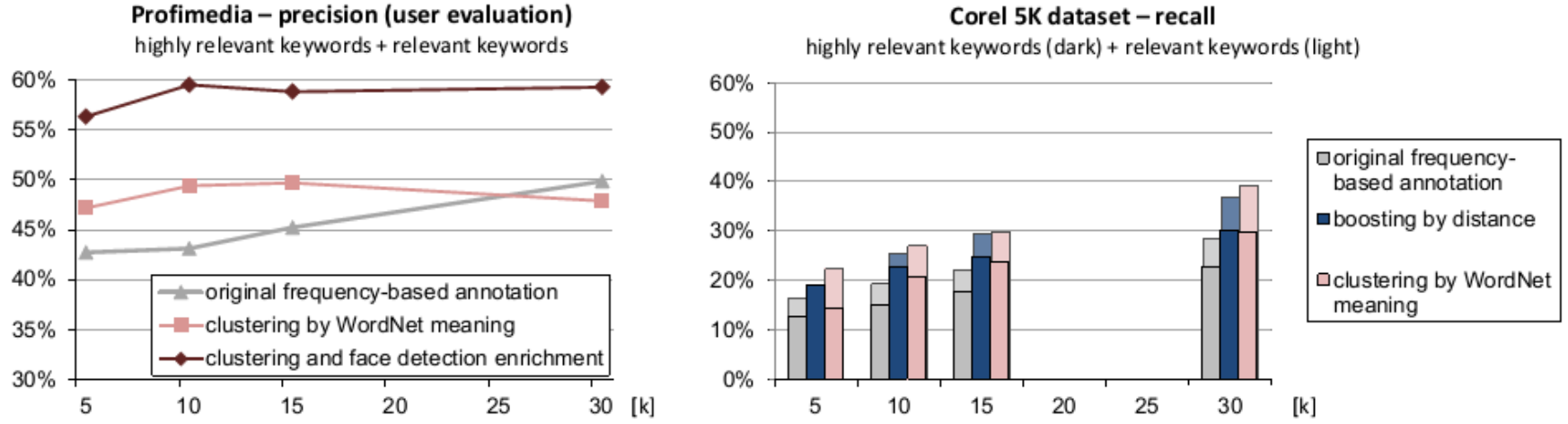


Figure 8: Influence of k on annotation precision (left), Corel dataset evaluation results (right).