Usage based Tag enhancement of Images

PROJECT TAKEN AT

BIG DATA EXPERIENCE LAB

ADOBE SYSTEMS

by

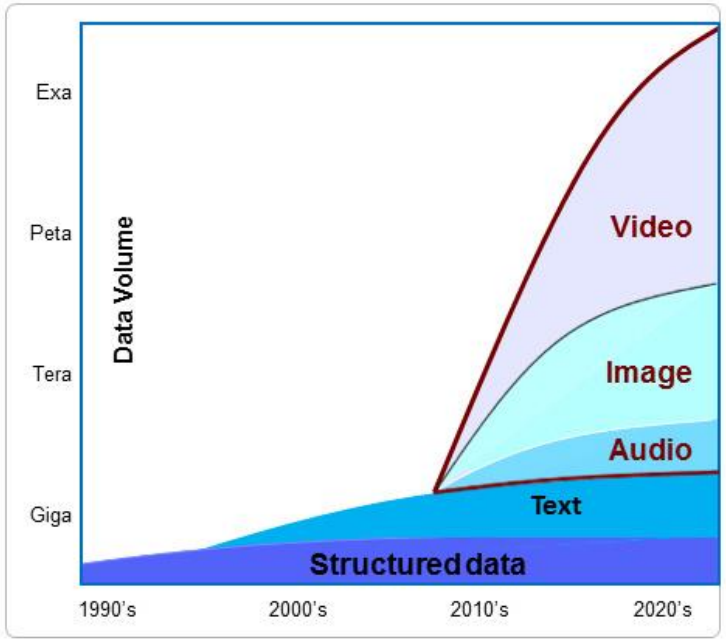
Saurabh Verma

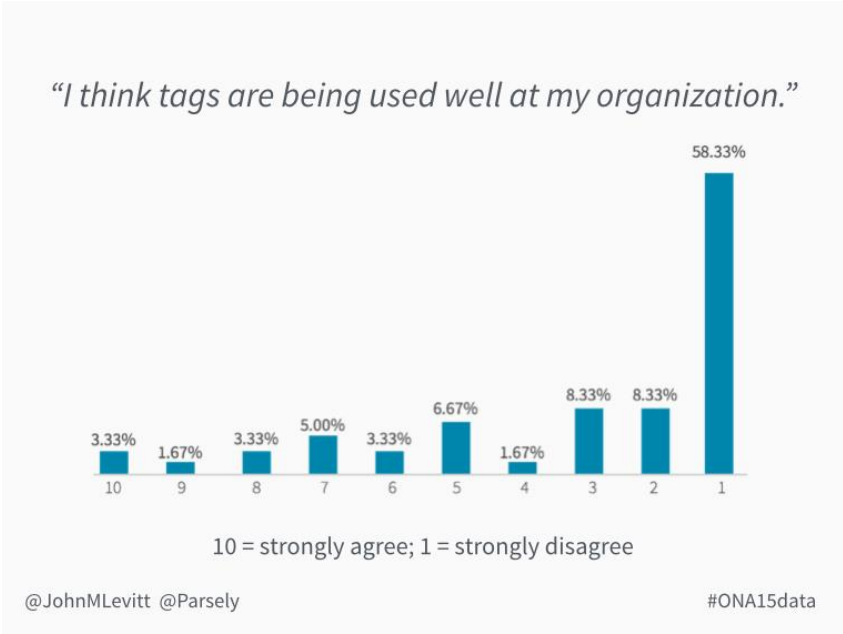
IIT Roorkee

Acknowledgements

- I would like to take this opportunity to sincerely thank my team members, Balaji Vasan(BEL Adobe), Noman Sheikh(IIT Delhi) and Roshan Kumar(IIT Kanpur) for their collaboration in the project.
- I am also thankful to Lab Members for their invaluable ideas and suggestions.
- I am highly indebted to Prof. Niloy Ganguly(IIT Kharagpur) for making the project reach these heights.

What is tagging? Why do we need it?





https://avindh2014.wordpress.com/background/

http://blog.parsely.com/post/2565/ona15data-a-year-in-data-from-parse-ly/parse-ly/

What if there was a system that could bridge the gap between what images contain and how they are perceived?

What if there was a system that can give a set of tags that maximizes their significance in retrieval?

“Problem statement: Given an image and set of associated content, enhance the set of tags based on it's usage”

- Develop a semantic understanding of the image content
- Enhance a given set of tags using background knowledge
- Improve retrieval and recommendation by using the given tags

Running Example

http://www.npr.org/sections/parallels/2016/03/22/471401729/in-vulnerable-europe-a-third-major-terrorist-attack-in-a-year

Tags from Clarifai API

clarifai ABOUT PRICING DEVELOPER BLOG CONTACT US SIGN UP NOW

Predicted Tags

bridge people travel city architecture building vehicle
man tourism landscape outdoors tourist road adult
sky group transportation system

https://www.clarifai.com/#demo

Tags from Microsoft Vision API

Features:

Feature Name	Value
Description	{ "type": 0, "captions": [{ "text": "a man standing on top of a bridge", "confidence": 0.13269974921728017 }] }
Tags	[{ "name": "sky", "confidence": 0.9977843165397644 }, { "name": "outdoor", "confidence": 0.9974787831306458 }, { "name": "building", "confidence": 0.996910035610199 }, { "name": "arch", "confidence": 0.9431114792823792 }, { "name": "person", "confidence": 0.9322513937950134 }, { "name": "bridge", "confidence": 0.9236949682235718 }]
Image Format	jpg
Image Dimensions	900 x 599
Clip Art Type	0 Non-clipart
Line Drawing Type	0 Non-LineDrawing
Black & White Image	False

https://www.microsoft.com/cognitive-services/en-us/computer-vision-api

Tags derived by our pipeline

Prior Art (Academic)

Text mining for automatic image tagging

- Leong et al., *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. Association for Computational Linguistics, 2010.

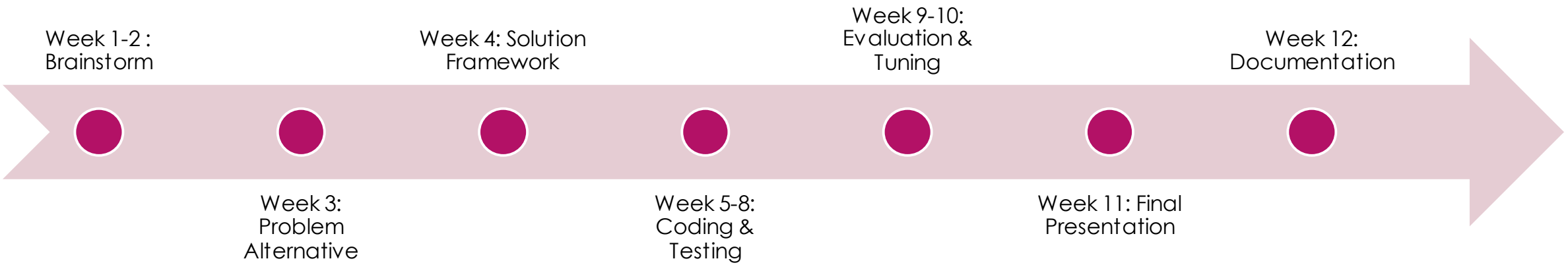
A review on automatic image annotation techniques

- Zhang, Dengsheng, Md Monirul Islam, and Guojun Lu. *Pattern Recognition* 45.1 (2012): 346-362.

Prior Art (Industry)

- https://cloud.google.com/vision/
- https://www.microsoft.com/cognitive-services/en-us/computer-vision-api
- https://www.ibm.com/watson/developercloud/visual-recognition.html
- https://imagga.com/solutions/auto-tagging.html
- https://www.clarifai.com/api

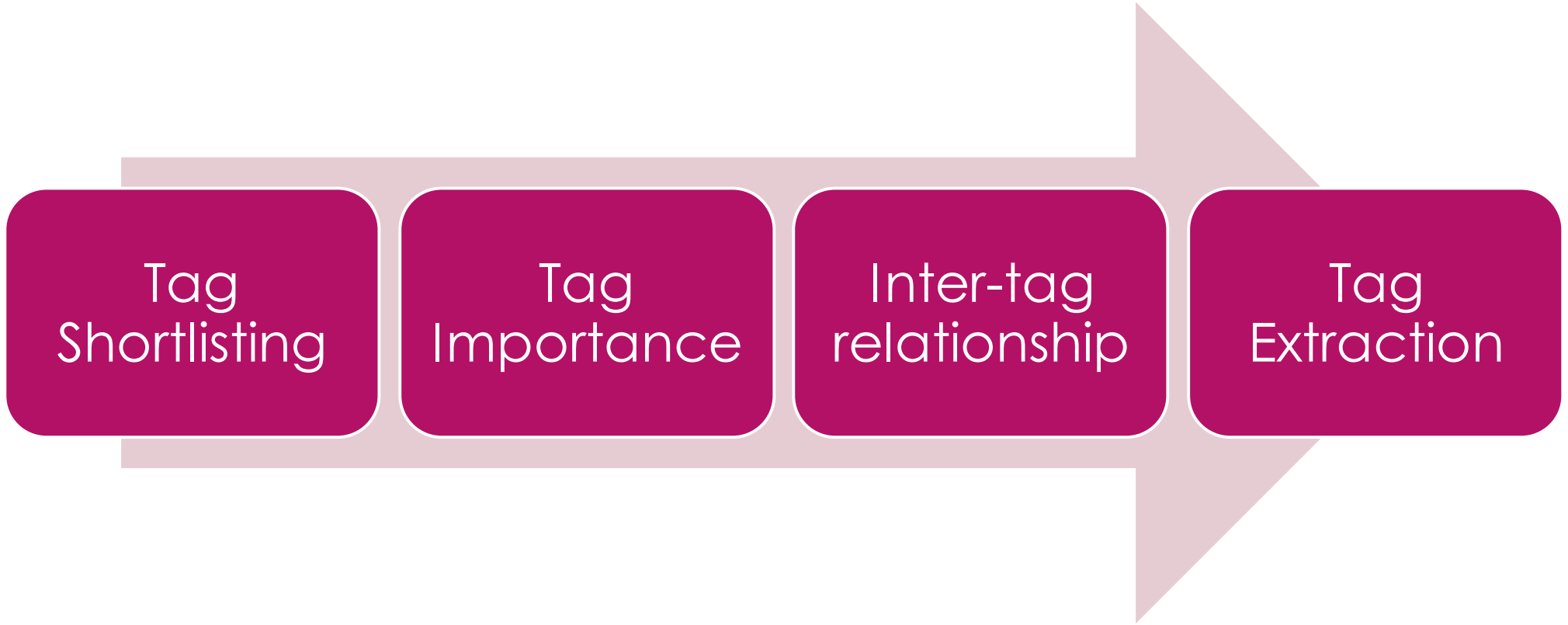
Workplan


Week 1-2 : Brainstorm
Week 3: Problem Alternative
Week 4: Solution Framework
Week 5-8: Coding & Testing
Week 9-10: Evaluation & Tuning
Week 11: Final Presentation
Week 12: Documentation

Objectives

- Combine visual and textual features for tagging of images
- Enhance the given set of tags using background knowledge from large knowledge bases
- Use the final tags to improve retrieval and recommendation engines
- Develop a cross-platform and re-entrant pipeline

Solution Framework



Tag Shortlisting

Input Image

Clarifai API

Candidate Tags with weights

Input associated Text

AIDA - YAGO

Stanford NLP Parser

Named Entities and Noun Phrases

Tag Importance

Frequency	Distance in dependency tree	Final Score
• Total frequency of the candidate tag occurrence in the associated text accounting for the co-reference or indirect mentions	• Inverse of the average distance of the entity from the root of the corresponding dependency tree obtained through a dependency parser.	• The average of the two measures above.

Inter-tag relationship

Word2Vec

- Pre-trained model trained on a corpus of Google News dataset with 100 billion words resulting in a final corpus of 3 million word representations.
- Yields a 300-dimensional vector for candidate tag representing the word in the space of trained deep neural network.
- Cosine-similarity to capture semantic closeness between the tags for interchangeable use.

Co-occurrence

- Point-wise Mutual Information between two entities based on their co-occurrences in the Wikipedia articles.
- Similarity based on how coherent the two tags are. It measures topical closeness.
- $pmi(x; y) = log \frac{p(x,y)}{p(x)p(y)}$

Inter-tag relationship

Algorithm 1 Tag Unifier

```
1: procedure UNIFY(tagsFromImage, TagGraph)
2:     for tag ∈ tagsFromImage do
3:         tag ← normalize(tag)
4:         for node ∈ TagGraph do
5:             val ← similarity(tag,node) (from Sec.3.3)
6:             if val > σ1 then
7:                 MergeNodes(tag,node)
8:                 node.weight ← MergedWeight()
9:                 PropagateWeight(node)
10:            else if val > σ2 then
11:                edge ← createNewEdge(tag,node)
12:                edge.weight ← val
13:            else
14:                continue
15:            end if
16:        end for
17:    end for
18: end procedure
```

Tag Extraction

Problem : Identify top nodes in the tag graph that are closely connected to the visual tags.

$$P(i \rightarrow j) = e_{ij} * nj$$

$$P(i \rightarrow i) = n_i$$

Chitrapura, Krishna P., and Srinivas R. Kashyap. "Node ranking in labeled directed graphs." *Proceedings of the thirteenth ACM international conference on Information and knowledge management*. ACM, 2004.

$$y[i](t+1) = y[i](t) + \sum_{j:(j,i)\in E} P(j \rightarrow i) * y[j](t)$$

$$y[i](0) = \begin{cases} 1 & for\ imageTag \\ 0 & otherwise \end{cases}$$

Evaluation & Results

Dataset

Leong et al. "Text mining for automatic image tagging." Proceedings of the 23rd International Conference on Computational Linguistics: Posters. ACL, 2010.

300 image-text pairs collected by issuing a query to Google Image API and processing one of the query results that has a significant amount of text around the images. Also created a gold standard tag set based on manual annotations from 5 annotators via Amazon Mechanical Turk accepting annotations from annotators with approval rating > 98%. We used the Clarifai API [18] to generate the visual tags for all our experiments.

http://lit.csci.unt.edu/index.php/Downloads

Metrics – Term Significance

The intuition here is to compute how important a tag is to the given context(usage) and normalize it with its “commonness” across a bigger corpus

Significance(tags) = NDCG(TF-IDF(tag) for tag in tags)

Lu, Yu-Ta, et al. "A Content-Based Method to Enhance Tag Recommendation." *IJCAI*. Vol. 9. 2009.

Metrics – Tag Relevance

To capture the relevance of the tags to the image in hand using gold-standard human annotations.

$$sim = \frac{1}{N}\sum_{i}\frac{\sum_{g_j \in TopK(G,Ti)} \mathrm{sim}(g_j, T_i)\gamma^j}{\sum_j \gamma^j}$$

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th international joint conference on Artificial intelligence - Volume 1* (IJCAI'95), Chris S. Mellish (Ed.), Vol. 1. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 448-453.

Metrics – Cophenet Correlation Coeffiecient

How faithfully a dendogram preserves the pairwise distances between the original un-modeled data points?

Higher value of the cophenet correlation coefficient indicates presence of more significant clusters and hence, more tag diversity.

$$c = \frac{\sum_{i<j}(x(i,j) - \overline{x})(t(i,j) - \overline{t})}{\sqrt{[\sum_{i<j}(x(i,j) - \overline{x})^2][-\sum_{i<j}(t(i,j) - \overline{t})^2]}}$$

Sokal, et al. "The comparison of dendrograms by objective methods." Taxon (1962): 33-40.

Baselines

Imagga

- https://imagga.com/auto-tagging-demo

Microsoft Vision

- https://www.microsoft.com/cognitive-services/en-us/computer-vision-api

Text mining for automatic image tagging

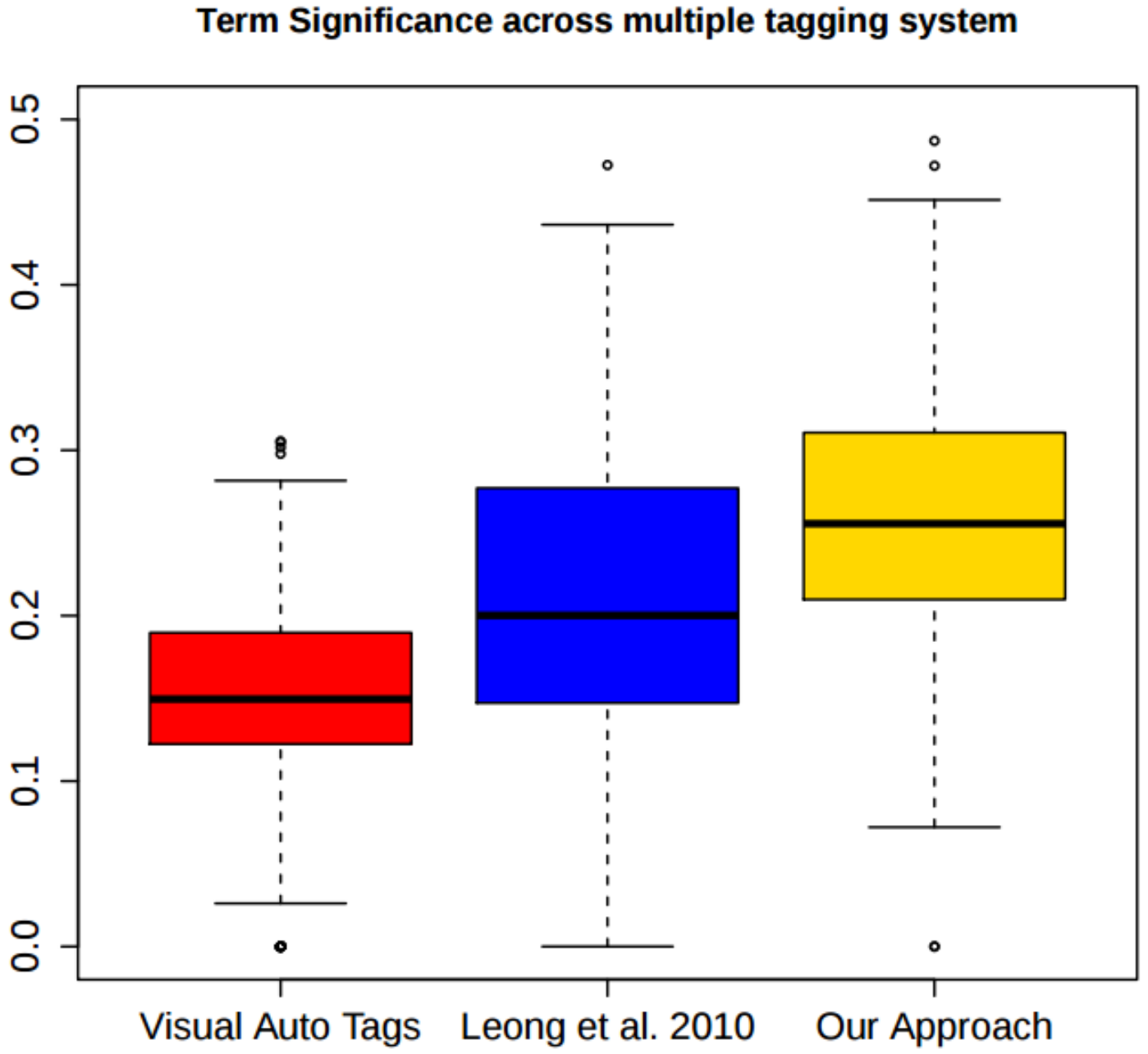
- Leong et al. "Text mining for automatic image tagging." *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*. ACL, 2010.

Gold-standard Human Annotations and Visual autotagging engine

- http://lit.csci.unt.edu/index.php/Downloads

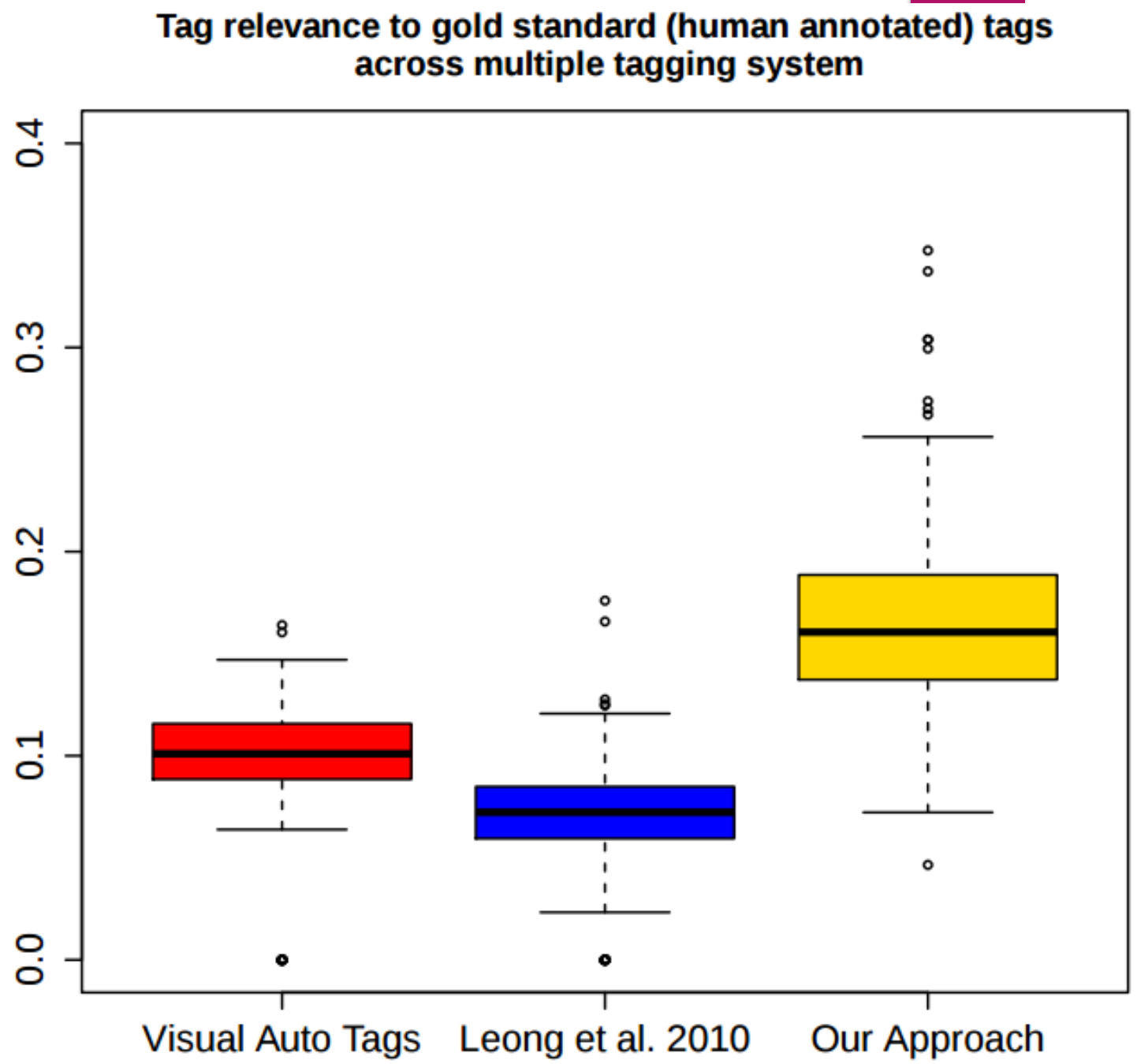
Tagging Performance

Term Significance



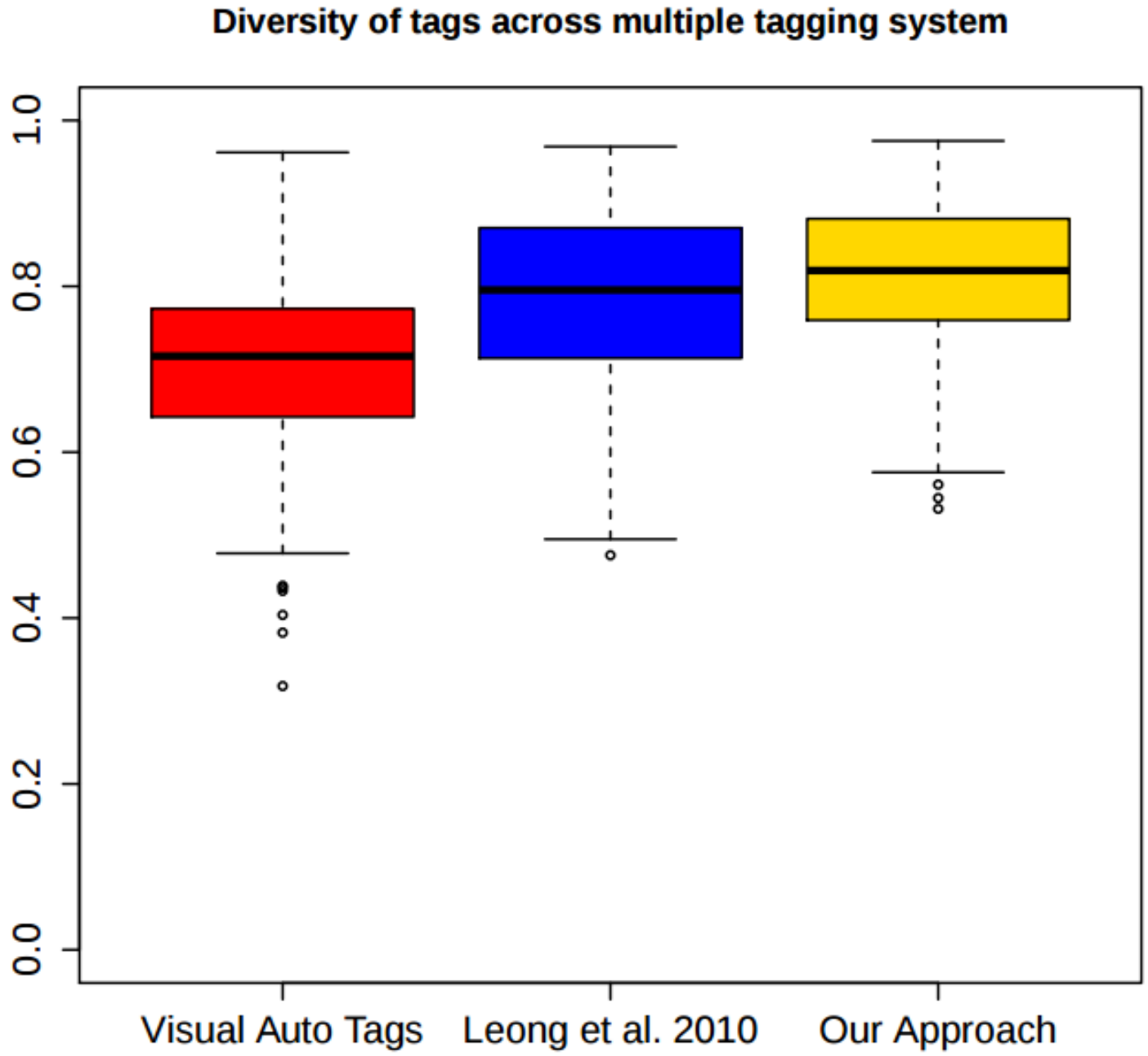
Tagging Performance

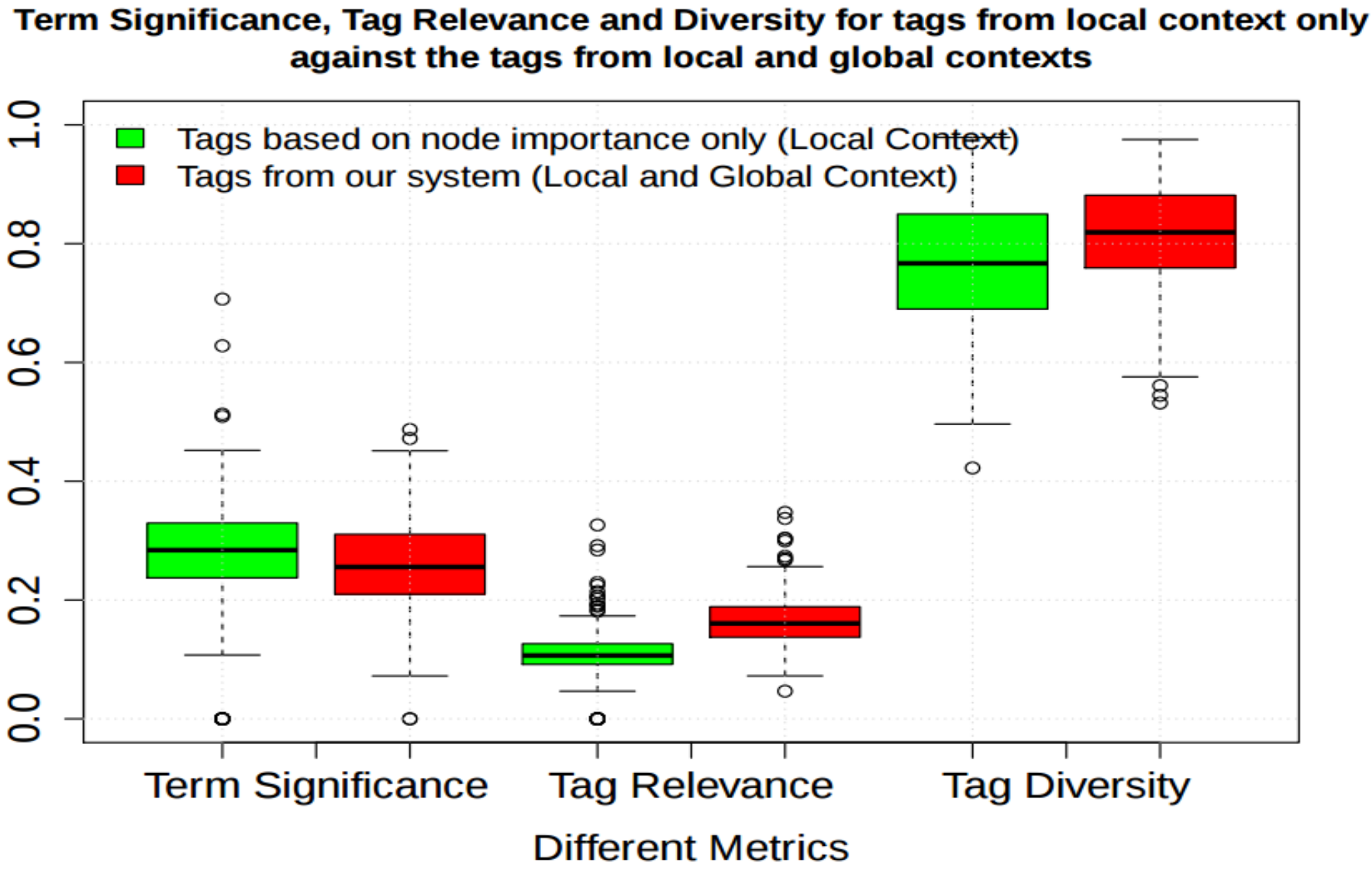
Tag Relevance

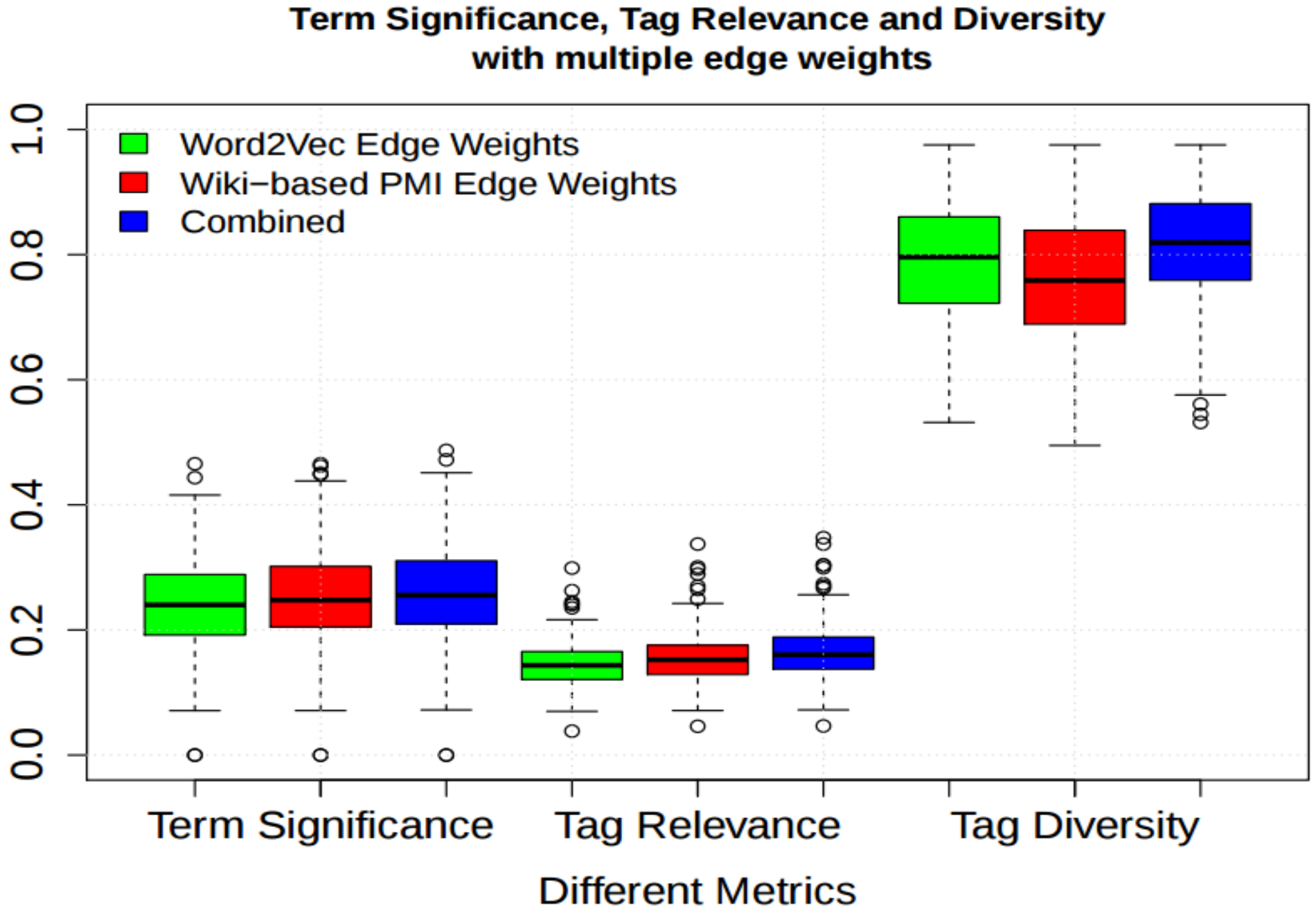


Tagging Performance

Div ersity




Term Significance, Tag Relevance and Diversity for tags from local context only against the tags from local and global contexts
Tags based on node importance only (Local Context)
Tags from our system (Local and Global Context)
0.0
0.2
0.4
0.6
0.8
1.0
Term Significance
Tag Relevance
Tag Diversity
Different Metrics


Term Significance, Tag Relevance and Diversity
with multiple edge weights
Word2Vec Edge Weights
Wiki−based PMI Edge Weights
Combined
0.0
0.2
0.4
0.6
0.8
1.0
Term Significance
Tag Relevance
Tag Diversity
Different Metrics

Future Work

- Extend this work for multimedia content using the same principal pipeline
- Capture multiple word senses in associated content.
- Optimize the application for low latency solution
- Offer solution for rarely used images

Publication

In the Proceedings of Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD2017) Conference as a long presentation paper with the title "Usage based Tag enhancement of images".

Out of 458 submissions, 45 papers are selected as long presentation paper.

Thank you