FIN-CLARIAH 2024/25

Authors:
Matti Nelimarkka
University of Helsinki; Aalto University
Anton Berg
University of Helsinki

D:3.3.4 TOOLS FOR VISUAL ANALYSIS

NEXT GENERATION OF VISUAL ANALYTICS

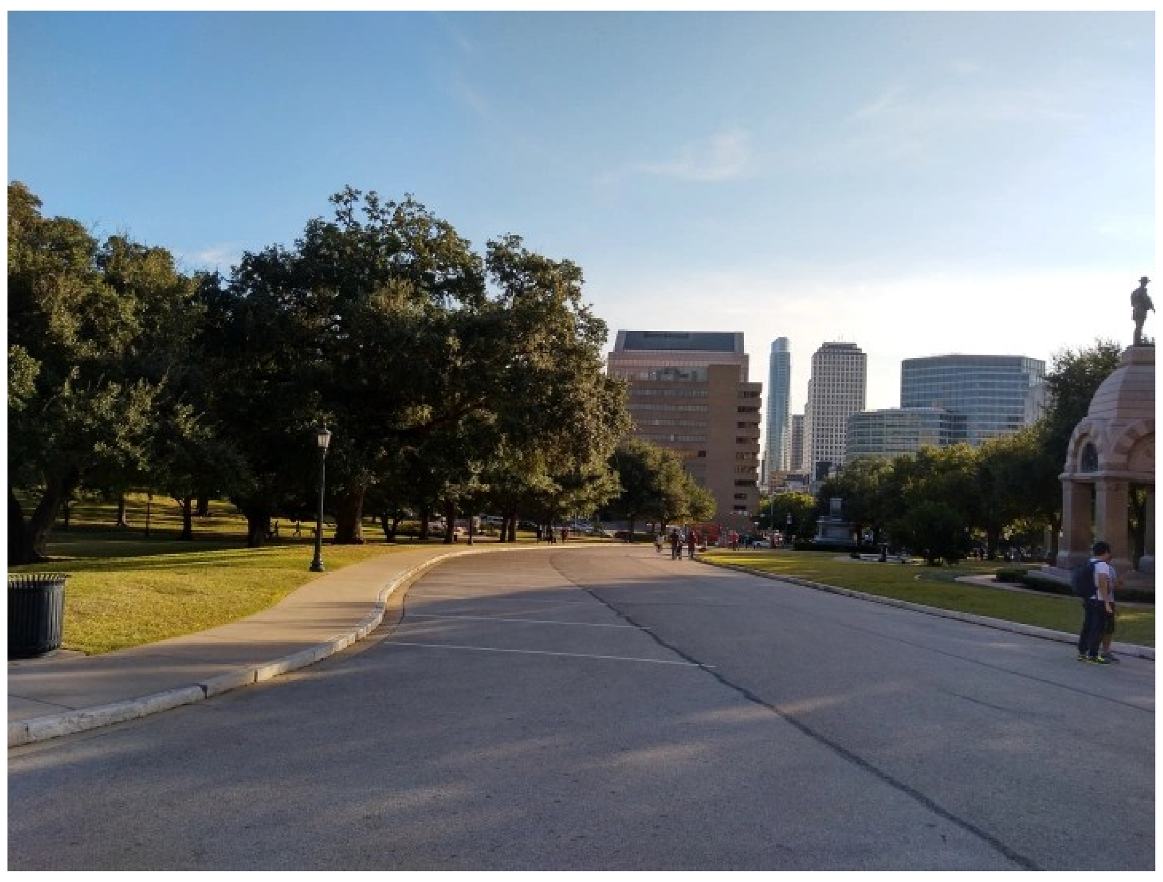
In an increasingly visual world, scholars need methods to analyse visual material at scale. Current best practice would be to develop a classifier for the content analysis, but classifier development is a resource intensive effort.

To address this gap, we examine alternative directions to work with visual materials such as:

- using labelling system to for content analysis
- unsupervised methods for content analysis
- zero-shot methods and visual question query approaches for content analysis

CROSS-SERVICE LABEL AGREEMENT SCORE

Both commercial and open-source models exist for image recognition and labelling, where a system tags image with a set of provided labels. However, the labels provided for a single image differ both in quantity and content.



Google Vision describes this image as daytime, city, road, tree, lane and various other labels.

Microsoft Azure agrees on city, road and street, does not recognise lane or daytime and sees a bus and a train.

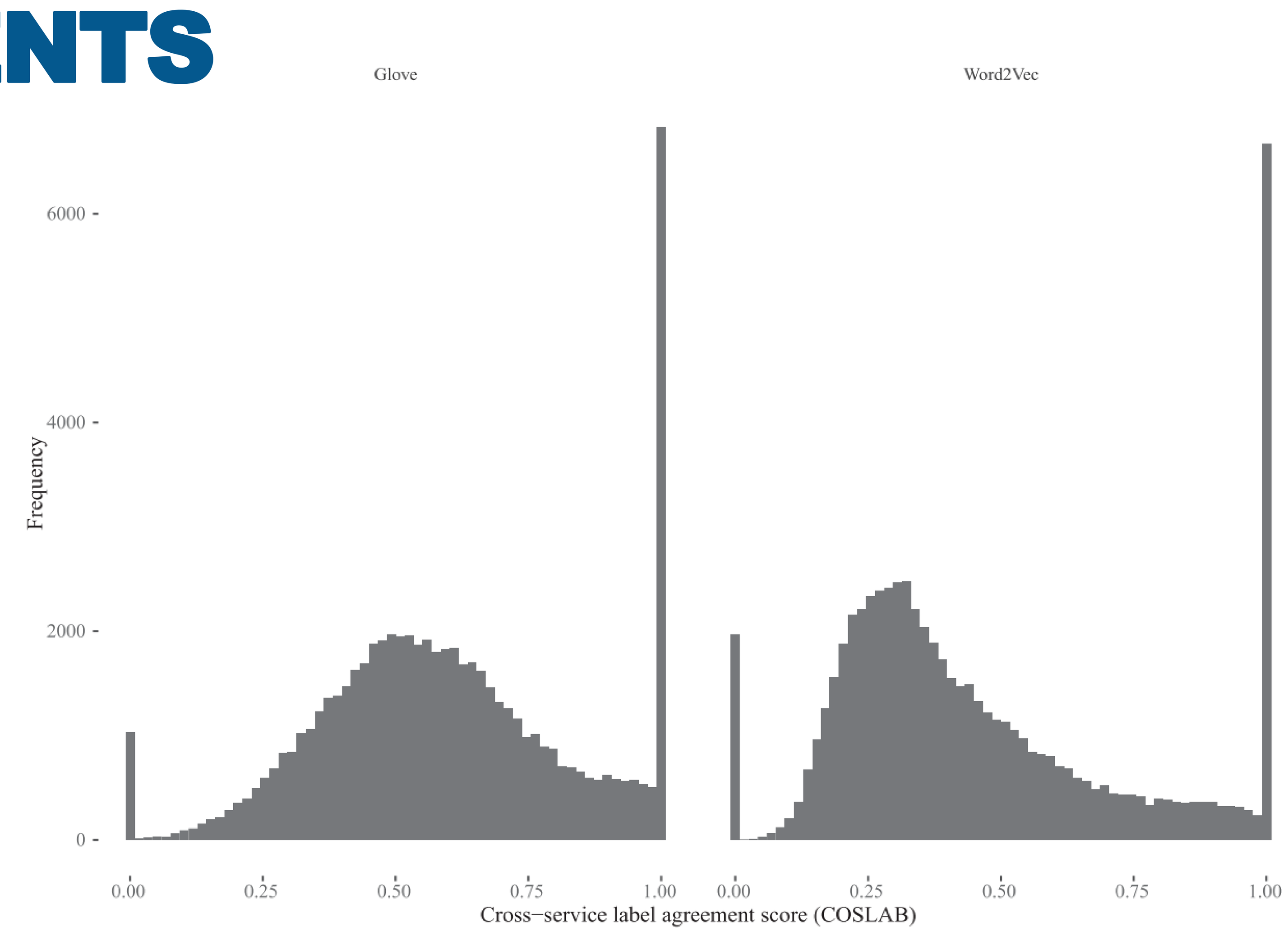
TIMELINE

- The command line interface for cross-service label agreement score is available for use.
- A graphical user interface for the cross-service label agreement score is developed by end of Q2/2025.
- During 2025, we work to explore and validate methods for clustering of visual content.

DIFFERENCES BETWEEN SERVICES

We tested Google Vision, Microsoft Azure and Amazon Rekognition services and observed that services differ on their provided labels, not only on the content of the labels but also on the quantity: even when requesting as much images as possible, Microsoft Azure gave around ten labels for each image, Google Vision around 25 and Amazon Rekognition more than 100. This illustrates the differences between these services.

MEASURING CROSS-SERVICE AGREEMENTS



To further examine these differences, we developed an approach to match each label from service 1 to its best counter-label from service 2 using word embedding models. This approach addressees the synonym differences – such as lane, road and street highlighted above – and allows thus using several services to build more comprehensive interpretation on the content of the images. We found out that while for some labels a perfect or nearly perfect fit could be found, for about ¼ of the labels, no match could be found. This did not seem to differ across services nor on the image content.

Our approach addresses a core weakness of using image recognition and labelling services for scholarly work. DARIAH helps to make this more accessible for social scientists.

SEE MORE

https://github.com/uh-soco/coslab-core

Berg, A., & Nelimarkka, M. (2023). Do you see what I see? Measuring the semantic differences in image-recognition services' outputs. *Journal of the Association for Information Science and Technology*, (74)11, 1307-1324.

Nelimarkka, M., & Berg, A. (2023). Is the World Different Depending on Whose AI Is Looking at It? Comparing Image Recognition Services for Social Science Research. *Information Matters*, 3(8).

UNIVERSITY OF HELSINKI
FACULTY OF SOCIAL SCIENCES
Aalto University
C. V. ÅKERLUNDIN MEDIASÄÄTIÖ
DARIAH-FI
KIELIPANKKI The Language Bank of Finland
FIN-CLARIN