

Crowdsourcing Knowledge-Intensive Tasks In Cultural Heritage

Jasper Oosterman¹, Alessandro Bozzon, Geert-Jan Houben
Archana Nottamkandath, Chris Dijkshoorn, Lora Aroyo²

¹ j.e.g.oosterman@tudelft.nl
² l.m.aroyo@vu.nl

Delft University of Technology
VU University Amsterdam

Challenge

Cultural Heritage Collections

Enrich data collections by tapping into the **interest** and **expertise** of **crowds** to create knowledge;
Crowd Generated Knowledge.

- ✓ Data-intensive
 - Rijksmuseum has 1M art pieces requiring annotation
- ✓ Knowledge-intensive
 - Diverse and specific knowledge needed
- ✓ Goals
 - Coverage: Enrich complete (sub)collection
 - Quality: High quality annotations

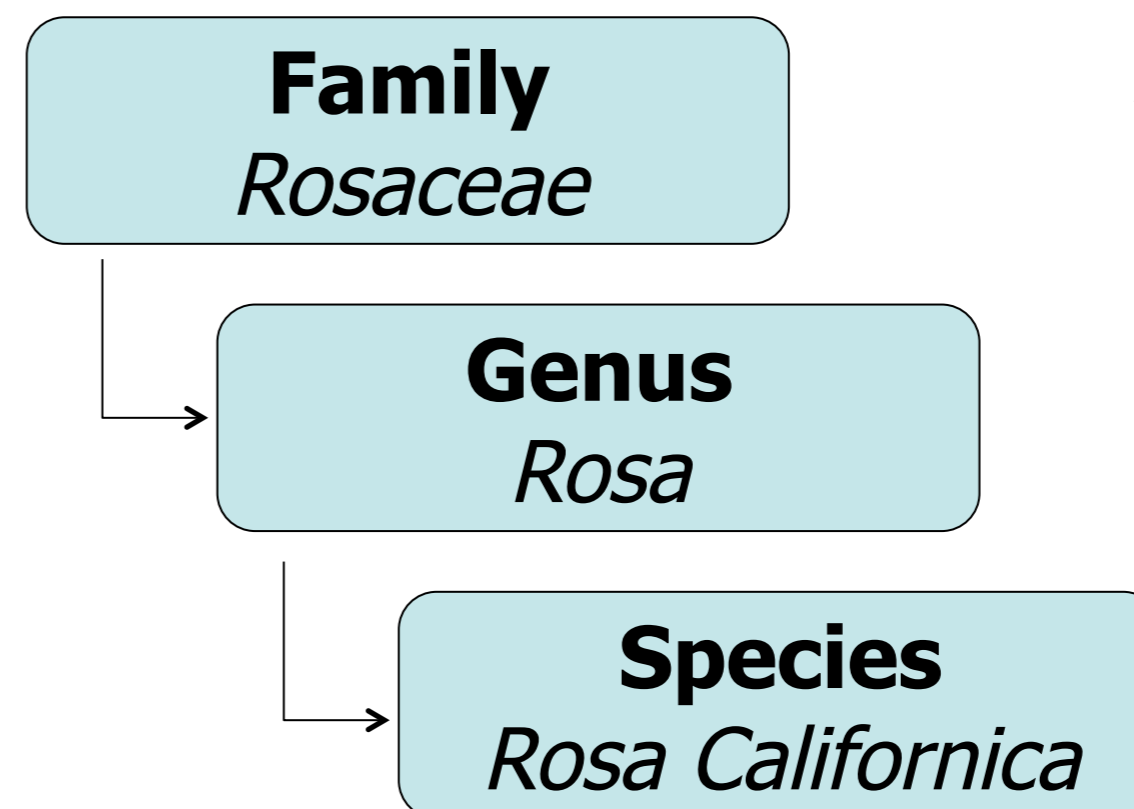


1

Aspects of Knowledge Intensive Tasks

Identification

- Identify relevant entities
- Prominence and amount of entities
- Artistic interpretation, lack of detail, fantasy



Annotation

- Tag identified entities
- Specificity of tags
- Domain and culture specific knowledge

What is the relation between entity identification difficulty and crowd annotation behavior?

Links

Demo



Video



Experiment performed within the SEALINCMedia project. Scan for a demo of Accurator or a video explaining our research together with the Rijksmuseum.

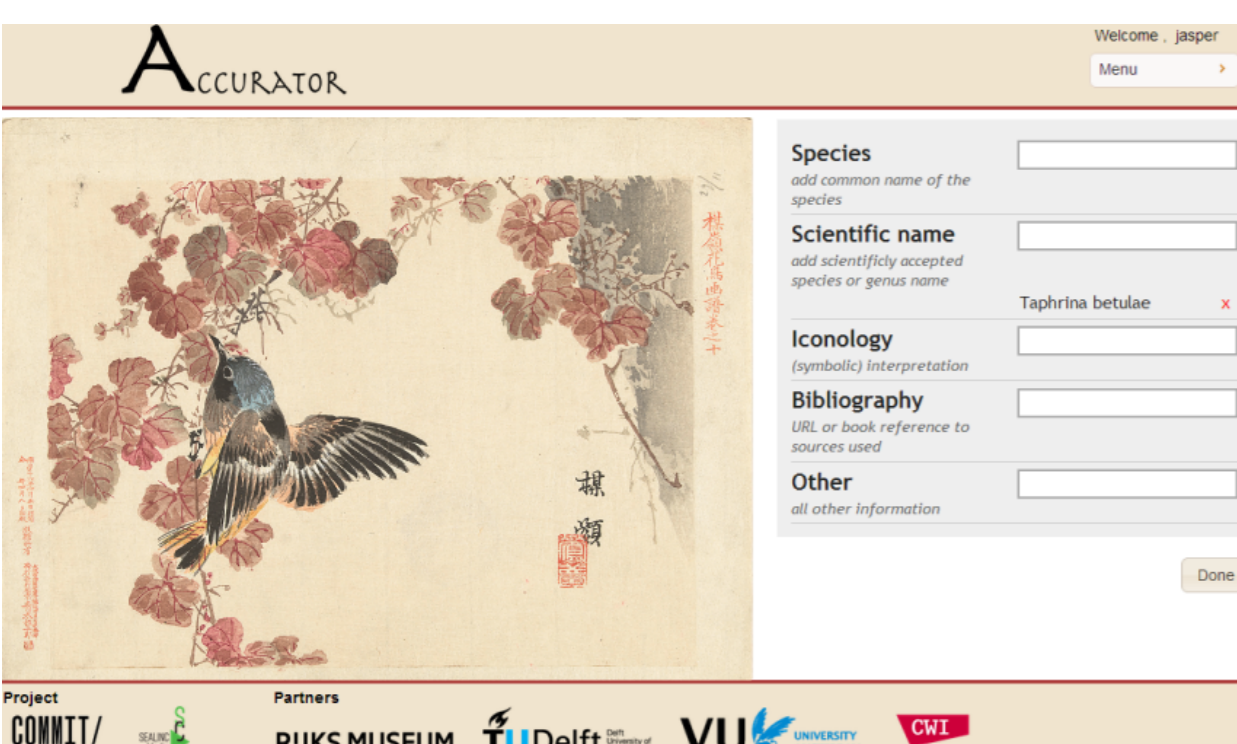
2

Experiment

Setup

- 82 prints from the Rijksmuseum containing flowers
- Tasks: annotate prints with specific flower names
- Executed by experts and crowd workers via crowdsourcing platforms

Experimental platform: Accurator

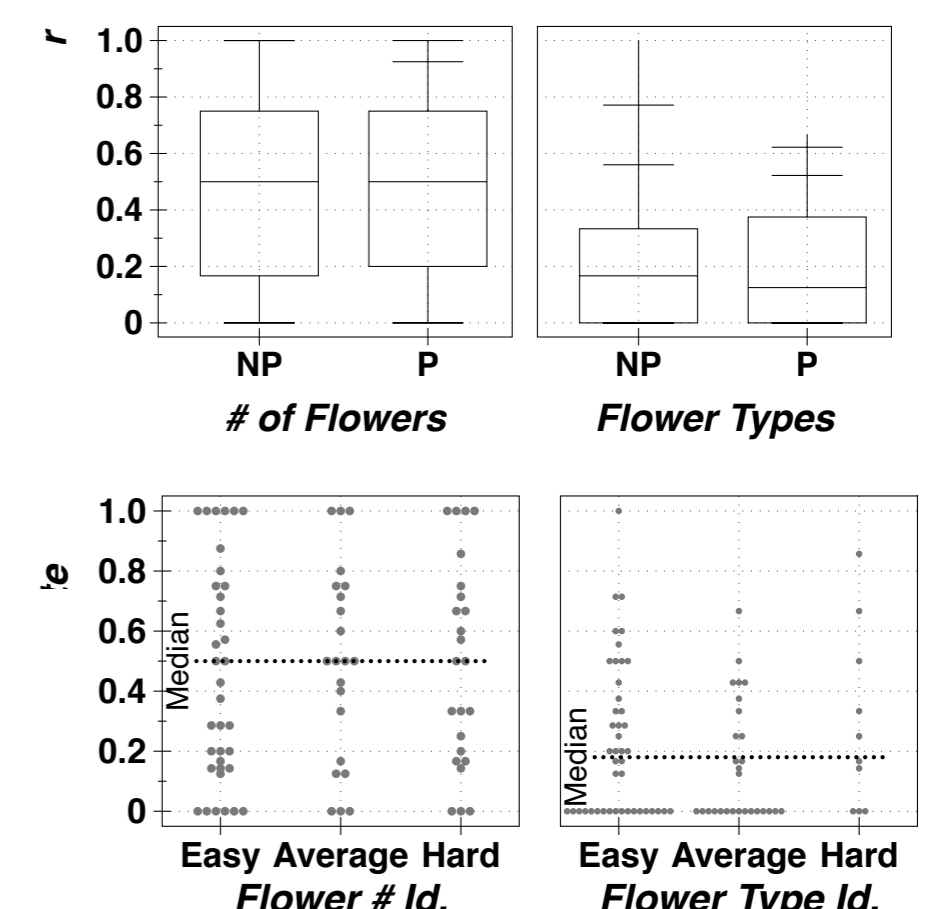
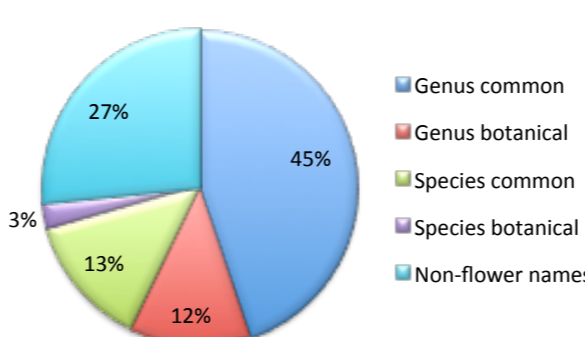


A platform to support crowd-enabled, collaborative annotation processes.

3

Conclusions

# Workers opened task	732
# Workers passed test questions	84
# Selected workers	44
# Annotation tasks performed	488
# "Fantasy" task	58
# "Unable" task	70
# "Flowers" task	360
# Flower labels	465



Insights

- Domain specific tasks on CS platforms not popular, but knowledge is present in some workers.
- Flower prominence does not affect identification
- Print difficulty only affects flower types identification
- Low crowd annotator agreement → worker selection and task orchestration are required