Enriching Cultural Heritage Collections with Crowd Generated Knowledge

Jasper Oosterman, Alessandro Bozzon, Geert-Jan Houben
Archana Nottamkandath, Chris Dijkshoorn, Lora Aroyo
Mieke H. R. Leyssen, Myriam C. Traub

Delft University of Technology
VU University Amsterdam
Centrum Wiskunde & Informatica

1 j.o.g.oosterman@tudelft.nl

**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

**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

**Family**
- Rosaceae

**Genus**
- Rosa

**Species**
- Rosa Californica

**How can experts from the crowd support the enrichment of Cultural Heritage collections?**

**Experiment**

**Setup**
- 86 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.

**Links**

**Demo**

Scan for a demo of Accurator or a video explaining our research together with the Rijksmuseum.

**Video**

**Conclusions**

<table>
<thead>
<tr>
<th>Number of</th>
<th>Experts</th>
<th>Crowd workers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annotators</td>
<td>4</td>
<td>75</td>
</tr>
<tr>
<td>Annotation tasks performed</td>
<td>128</td>
<td>982</td>
</tr>
<tr>
<td>Tags (total)</td>
<td>161</td>
<td>1077</td>
</tr>
<tr>
<td>Tags (flower names)</td>
<td>119</td>
<td>831</td>
</tr>
<tr>
<td>Distinct flower names</td>
<td>47</td>
<td>123 (17 overlapping names)</td>
</tr>
</tbody>
</table>

**Type of experts annotations**
- 25% Genera common
- 20% Scientific names
- 15% Scientific names
- 10% Botanical names
- 7% Common names
- 5% Other flower names

**Type of crowd workers annotations**
- 35% Genera common
- 22% Scientific names
- 6% Scientific names
- 9% Botanical names
- 28% Common names
- 2% Other flower names

**Insights**
- Botanical flower knowledge is present in the crowd...
- ... but image difficulty affects crowd annotation quality
- Low crowd annotator agreement → worker selection and task orchestration are required