From Humanities Data to Digital Humanities Data: a new opportunity for technical services

Haiqing Lin

C.V. Starr East Asian Library
University of California, Berkeley

CEAL, Denver
2019
Humanities Data vs. Digital Humanities
Data: a Case Study

Self-Portrait with Straw Hat, 1887
Metropolitan Museum of Art
(F365v)
Colour Analysing: a Digital Humanities Approach

- 35 self-portraits are clustered into 6 groups by periods
# Colour Objects

Cluster colors, sized by number of pixels:

<table>
<thead>
<tr>
<th>cluster</th>
<th>pixels</th>
<th>name</th>
<th>HEX</th>
<th>RGB</th>
<th>HSV</th>
<th>LCH</th>
<th>Lab</th>
<th>tags</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>39.05%</td>
<td>#9A9C01</td>
<td>154</td>
<td>156</td>
<td>97</td>
<td>62</td>
<td>38</td>
<td>61</td>
</tr>
<tr>
<td></td>
<td>21.05%</td>
<td>#A7B7AE</td>
<td>167</td>
<td>183</td>
<td>174</td>
<td>145</td>
<td>9</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>20.29%</td>
<td>#67B432</td>
<td>103</td>
<td>180</td>
<td>130</td>
<td>141</td>
<td>43</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>10.19%</td>
<td>#BEA74B</td>
<td>190</td>
<td>167</td>
<td>75</td>
<td>48</td>
<td>60</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>9.43%</td>
<td>#43412D</td>
<td>67</td>
<td>65</td>
<td>45</td>
<td>53</td>
<td>32</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Colour Analysing

<table>
<thead>
<tr>
<th></th>
<th>avg</th>
<th>med</th>
<th>min</th>
<th>max</th>
<th>Avg Color</th>
<th>Med Color</th>
<th>Min Color</th>
<th>Max Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1886</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>49</td>
<td>44</td>
<td>15</td>
<td>161</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>39</td>
<td>38</td>
<td>11</td>
<td>139</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>30</td>
<td>29</td>
<td>10</td>
<td>114</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1886-1887</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>100</td>
<td>104</td>
<td>23</td>
<td>235</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>98</td>
<td>102</td>
<td>29</td>
<td>208</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>90</td>
<td>85</td>
<td>29</td>
<td>184</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1887</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>111</td>
<td>102.5</td>
<td>22</td>
<td>238</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>102</td>
<td>96</td>
<td>28</td>
<td>212</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>87</td>
<td>87</td>
<td>15</td>
<td>176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1887-1888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>114</td>
<td>115.5</td>
<td>36</td>
<td>213</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>106</td>
<td>101</td>
<td>26</td>
<td>203</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>91</td>
<td>88</td>
<td>38</td>
<td>172</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1888</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>141</td>
<td>144</td>
<td>36</td>
<td>228</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>159</td>
<td>167</td>
<td>35</td>
<td>229</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>113</td>
<td>104</td>
<td>26</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1889</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:R</td>
<td>124</td>
<td>132</td>
<td>18</td>
<td>227</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:G</td>
<td>123</td>
<td>119.5</td>
<td>33</td>
<td>206</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RGB:B</td>
<td>107</td>
<td>105</td>
<td>30</td>
<td>173</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Result: A Colour Curve of Portraits of Vincent van Gogh
Other Examples: Text

Object.person: 陸德明

Object.offical: 刺史

Object.word: 文言
Other Examples: Map

Object.place: 句容縣
Object.place: 溧水縣
Object.place: 溧陽縣
Humanities Data

• Definition: “Humanities data might be about the record of events – what we commonly call facts – or about texts created about them, whether these texts are written, oral, visual, aural or spatial.” (Gardiner, E., Musto, R. G., The digital humanities: a primer for students and scholars, New York: Cambridge University Press, 2015, page 31)

• Humanities data include text, image, audio, video, object, space and so on.
Digital Humanities Data

- **Definition of Digital Humanities Data**
  
  Digital humanities data is a set of objects which are extracted from the digitalized humanities data. This set of objects consist of people, organizations, places, events and all other real-world objects. Each object has its own attributes and properties which are usually described by metadata to enable human understanding, as well as to be explicit to machine understanding. With machine understanding, digital humanities data is able to be processed by algorithms.

- **From Humanities Data to Digital Humanities Data**
Linking Digital Humanities Data to the Library Data: Person

From Library of Congress Name Authority File

Lu, Deming, 1956-

URI(s)
- http://id.loc.gov/authorities/names/n50022511

Instance Of
- MARC/RDF Person's Name
- MARC/RDF Authority
- SKOS Concept rdf

Scheme Membership(s)
- Library of Congress Name Authority File

Collection Membership(s)
- Names Collection - Authorized Headings
- LC Names Collection - General Collection

Variants
- 陸德明, 1956-
- 陸德明, 1956-

Additional Information
- http://id.loc.gov/authorities/names/n50022511

Exact Matching Concepts from Other Schemes
- http://viaf.org/viaf/authority/LC%3C强调%3E00022511#skos:concept

Earlier Established Forms

Object.person: 陸德明
Linking Digital Humanities Data to the Library Data: Place

Object.place: 句容縣

From Library of Congress Name Authority File

Jurong Xian (China)

URI(s)
- http://id.loc.gov/authorities/names/m82046009

Instance Of
- MAP/RI/DE/Varvadik
- MAP/RI/DE/Authority
- SKOS Concept

Scheme Membership(s)
- Library of Congress Name Authority File

Collection Membership(s)
- Names Collection - Authorized Headings
- LC Names Collection - General Collection

Variants
- Chuir-yung (China)
- Juyong (China)
- Kusagi-ken (China)
- Kusugi-ken (China)
- Kiyoy (China)

Additional Information
- http://id.loc.gov/works/term/m82046009

Exact Matching Concepts from Other Schemes
- http://viaf.org/viafsource/LOC/6701/52046008#scope:Concept
Why Library Data

- Reliable
- Evidence based
- Rich contents
- Consistency
- Systematic
- Linking to library collection
Library data as a digital humanities infrastructure
New Role of Technical Services

- Modeling
  FRBR, BIBFRAME, Library Reference Model (LRM)
- Describing
  AACR II, RDF
- Encoding
  MARC21, RDF/XML, MARCXML
- Evaluating
Towards Data-oriented technical services: next generation

• By the data
• Of the data
• For the data
Thank you!