The Primary Sources Tool
Curating citations for Wikidata statements

Marco Fossati - User:Hjfocs
The story so far

- The globe
- Freecode (2005)
- Google (2010)
- WIKIDATA (2016)

Leonardo da Vinci

Mathematician

Leonardo di ser Piero da Vinci, more commonly Leonardo da Vinci or simply Leonardo, was an Italian polymath whose areas of interest included invention, painting, sculpting, architecture, science, music, ... Wikipedia

Born: April 15, 1452, Anchiano
Died: May 2, 1519, Clos Lucé, Amboise, France
Primary sources tool
Wikidata reliability

- Need for trust on the content addition process
- Lack of validation against third-party resources
- Few references to external authoritative sources
- A reference for each statement is not ensured
Data-driven solution

- Harvest factual information from a corpus of reliable Web sources
- Generate Wikidata-compliant statements
- Validate existing ones via references to such sources
- Suggest new (referenced) statements
Outcome: StrepHit

a project funded via
Wikimedia Individual Engagement Grants
StrepHit progress

- 2016: beyond Freebase & Knowledge Vault
- 2017: next level of the primary sources tool
Time for an uplift!
IEG renewal
kick-off

Primary sources tool
Frustrating workflow

1. No filtering capabilities
2. No reference previews
3. No extracted fact highlighting
4. Too many page reloads
5. Duplicated data
1. Filtering, take 1
2. Reference preview
### 3. Extracted fact preview

<table>
<thead>
<tr>
<th>Name in Kana</th>
<th>Occupation</th>
<th>Reference URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>シャオ・シェンヤン</td>
<td>Actor, Singer</td>
<td><a href="http://video.xbox.com/en-us/movie/a-woman-a-gun-and-a-noodle-soup/7fbd3866e9134921-6">link</a></td>
</tr>
</tbody>
</table>

[Link to Wikidata page: Xiaoshenyang](https://www.wikidata.org/wiki/Q1018752)
Primary sources filter

Filtering, take 2
What do you think?

Marco Fossati - User:Hjfocs