Beyond the Book: Linking Books to Wikipedia

Carlos Martinez-Ortiz - NLeSC
Marijn Koolen - UvA (DH)
Floor Buschenhenke - Huygens ING (Lit. St.)
Karina van Dalen-Oskam - Huygens ING (Lit. St.)
Outline

• Beyond the book
  – What?
  – Why?
  – How?
  – Useful tools
Aim of Beyond the book

• Measure international appeal of books
  – What?

"How international is a work of fiction?"

Will a book written in X be understood by native Y speakers?

How much of the book is X-culture specific?
Aim of Beyond the book

• Measure international appeal of books
  – What?
  – Why?
Aim of Beyond the book

- Measure international appeal of books
  - What?
  - Why?
  - How?
Aim of Beyond the book

Willem-Alexander of the Netherlands

Countries

Expectation
Contribution

Contributions

Countries
Entity extraction

- Semanticizer
- Entities=Wikipedia entries
- Names of:
  - People
  - Places
  - Events
Linking to Wikipedia

- MediaWiki
  - mwclient
- Extract revision information:
  - Contributor IP
  - Contributor country
Revision history

Quetzalcoatl: Revision history

View logs for this page

Browse history

From year (and earlier): 2015  From month (and earlier):  Tag filter:  Go

For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary.
External tools: Revision history statistics • Revision history search • Edits by user • Number of watchers • Page view statistics

(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)

Compare selected revisions

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>User</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>2015-03-27</td>
<td>13:30</td>
<td>201.171.101.220</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-03-14</td>
<td>13:14</td>
<td>185.14.137.41</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-25</td>
<td>00:57</td>
<td>34.348.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-22</td>
<td>01:47</td>
<td>34.348.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-15</td>
<td>01:46</td>
<td>34.371.44.1</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-15</td>
<td>01:45</td>
<td>34.370.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-20</td>
<td>20:47</td>
<td>34.313.44.1</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-18</td>
<td>17:06</td>
<td>34.277.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-18</td>
<td>17:01</td>
<td>34.285.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-18</td>
<td>16:59</td>
<td>34.277.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-18</td>
<td>16:58</td>
<td>34.277.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-02-18</td>
<td>16:55</td>
<td>34.277.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-01-27</td>
<td>07:47</td>
<td>34.256.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-01-31</td>
<td>00:21</td>
<td>34.247.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-01-27</td>
<td>09:27</td>
<td>34.247.86.17</td>
<td>(undo)</td>
</tr>
<tr>
<td>2015-01-26</td>
<td>03:06</td>
<td>34.247.86.17</td>
<td>(undo)</td>
</tr>
</tbody>
</table>
User profile

From Wikipedia, the free encyclopedia

Wikipedian information

This user lives in or hails from Amsterdam, The Netherlands.

This user is a native of México

Wikipedia:Babel

en-3 This user can contribute with an advanced level of English.

es Este usuario tiene el español como lengua materna.

Search user languages

Categories: User en-3 | User es-N | Wikipedians in North Holland
Tools for publishers

• Identify culture-specific words
  – Assess translatability
  – Deal with difficult concepts appropriately
Tools for publishers

From *Boven is het stil*
Conclusion

• Innovative approach
• Enhancing literary studies
• Applicable to other areas

– https://github.com/semanticize/st
– c.martinez@esciencecenter.nl