MOOC Wikification System

Jan 19, 2018
AN Yahui
anyahui.120@gmail.com
Previous Work

- Paper Writing
- Supplementary Experiments -- Spiders
Current Work

● Paper Writing

Reconstructed the writing content

<table>
<thead>
<tr>
<th># Courses</th>
<th>#Posts</th>
<th># Posts with key words</th>
<th># Posts with concrete mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>179</td>
<td>116,962</td>
<td>20,712</td>
<td>1,110</td>
</tr>
</tbody>
</table>

● Experiments

1. annotate structure design

2. Mention extraction accuracy

<table>
<thead>
<tr>
<th>Ratio for posts have key words</th>
<th>Ratio for posts with concrete mentions</th>
</tr>
</thead>
<tbody>
<tr>
<td>17.7%</td>
<td>5.36%</td>
</tr>
</tbody>
</table>

Course name | ID of post | Post content | Mention content | SCI (Y/N) | Checked (Y/N) | Ture (Y/N) | Actual URL | Resolved URL | Resolved (Y/N) |
|-------------|------------|--------------|-----------------|-----------|---------------|------------|-------------|---------------|----------------|


Mentions: Key Words + ( *[#]*)( *[0-9]+ *?[\.-]? *?[0-9]*)(?!([0-9]*%))
Next Step

● To annotate more posts to finish the extraction accuracy and resolve accuracy;

● To finish the presentation of a typical example (meaning finishing the workflow of the system).
MOOC Wikification System

Jan 26, 2018
AN Yahui
anyahui.120@gmail.com
Current Work

● Complete the program of the workflow of the system;

  https://github.com/anyahui120/MOOC-Wikification/blob/master/MURI.py

● Problem: *The accuracy of the resolver for standard URL is too low.*
Dear Sohelia,

The definition is given at 2:10 in video 2.2.b, it is a set that contains all infinite sequences that have a finite norm.

Best,

Lionel

---

**Post_text**: <co-content><text>Dear Sohelia, </text><text>The definition is given at 2:10 in video 2.2.b, it is a set that contains all infinite sequences that have a finite norm. </text><text>Best, </text><text>Lionel </text></co-content>

**Mention**: video 2.2

**Short form url**: http://www.wing.comp.nus.edu.sg/dsp/1479110400000/Week 2/lecture/2.2


**Actual url**: https://www.coursera.org/learn/dsp/supplement/gxhZT/introduction-to-lesson-2-2
Case Study

Post_text: Module 3 Discussion

Mention: module 3


Actual url : https://www.coursera.org/learn/exposure-photography/lecture/AE8IG/setting-up-a-digital-camera-part-3a-1-3-3
Current Work

● Which search engine in Python is suitable for my requirement.

Generated from post

Content of the short form URL

Coursera accounting-analytics 14803200000000 lecture module 3.0

Generated from resource

Content of the standard URL

accounting-analytics Brian J Bushee&Christopher D. Ittner 14803200000000 module-3-overview-3-0 lecture Coursera
Next Step

- To optimize the search engine.
- To annotate data to finish the experiment section of the paper.
Outline

- Abstract
- Introduction
- Related Work
- Framework
- System
- Evaluation
- Discussion
- Conclusion