

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

# Courses	#Posts	# Posts with key words	# Posts with concrete mentions
179	116, 962	20, 712	1, 110

- Experiments

1. annotate structure design

2. Mention extraction accuracy

Ratio for posts have key words	Ratio for posts with concrete mentions
17.7%	5.36%

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)
-------------	------------	--------------	-----------------	-----------	---------------	------------	------------	--------------	----------------

Key Words: ([Pp]roblem|[Qq]uiz*|[Qq]uestion|[Mm]odule|[Ll]ecture|[Ss]lide|[Vv]ideo|[Hh]omework)

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.*

Case Study

Post_text: <co-content><text>Dear Sohelia,</text><text>The definition is given 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](http://www.wing.comp.nus.edu.sg/dsp/1479110400000/Week%202/lecture/2.2)

Standard form url : *[http://www.wing.comp.nus.edu.sg/Coursera/dsp/1479110400000/Prandoni&Martin Vetterli&Lionel Coulot/supplement/introduction-to-lesson-2-2](http://www.wing.comp.nus.edu.sg/Coursera/dsp/1479110400000/Prandoni&Martin%20Vetterli&Lionel%20Coulot/supplement/introduction-to-lesson-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*

Short form url : *[http://www.wing.comp.nus.edu.sg/digital-analytics/1480320000000/Week 3/lecture/3](http://www.wing.comp.nus.edu.sg/digital-analytics/1480320000000/Week%203/lecture/3)*

Standard form url : [http://www.wing.comp.nus.edu.sg/Coursera/exposure-photography/1480320000000/Peter Glendinning&Mark Valentine Sullivan/lecture/setting-up-a-digital-camera-part-3a-1-3-3](http://www.wing.comp.nus.edu.sg/Coursera/exposure-photography/1480320000000/Peter%20Glendinning&Mark%20Valentine%20Sullivan/lecture/setting-up-a-digital-camera-part-3a-1-3-3)

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 1480320000000
lecture module 3.0

Content of the
standard URL

accounting-analytics Brian J
Bushee&Christopher D. Ittner 1480320000000
module-3-overview-3-0 lecture Coursera

Generated from resource

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