

Reclassify the
PyConJP proposal
categories.

Motoki Hirao

Analysis method

- Getting the talk session details in PyCon JP Site by my hand...
- Parsing and extracting only nouns by Janome
- Creating WordCloud.
- Retag....?

Sorry, This Poster is failed, so...

- PyCon JP talk sessions have a difficult for analysis.
 - In 2012 and 2013, too short abstract and many tags added, I can't retag by my hand.
 - Unfortunately, In 2015 datas could not be accessed temporarily.
 - In 2018, having many tags...
 - and PyCon JP have both language session(JP and EN), those are different nature, I can't retag...

Annual analysis ?

- In a PyCon JP Session, it have many categories sessions.
 - I looked for words with high frequency, but I found only common words in sessions...
 - Because sessions use different libraries and analysis methods according to properties.
 - But common words are used by any sessions.
 - The result did not improve even though I deleted some common words.

Example (PyCon 2011)



Example (PyCon 2017)



What do I try next ?

- Normalizing the data
- Doing more pre-processing
- Inputting more data
 - If I want to see year's trend, some overseas PyCon.
 - Translating Japanese session content by Google Translate, Getting more English data!

Future prospects

- I want to be able to deliver session data information in JSON
 - It's too hard for scraping
 - PyCon JP WebSite are recreated every year
 - It's too hard for hands
 - About 50 sessions in a year
- It maybe be wanted by PyCon JP Contents Staffs !