
Draft Topic

Release 0.4.1

Scoring Platform Team

Dec 21, 2022

CONTENTS

1	Changelog	3
2	Indices and tables	7
	Python Module Index	9
	Index	11

This package contains a set of utilities and assets for predicting topics to new drafts based on Wikiprojects on English Wikipedia.

In this package, you'll find the `feature lists` used to train models for the supported wiki. There's a set of `command-line utilities` that are used to perform data pipeline operations specific to training and testing draft topic models.

See the API Reference for low level details.

CHANGELOG

All notable changes to this project will be documented in this file.

1.1 [0.4.1]

1.1.1 Added

- Add Github Action that build & pushes to Pypi index

1.2 [0.4.0]

1.2.1 Changed

- Rebuild to add compatibility for revscoring 2.11

1.3 [0.3.0]

1.3.1 Added

- Adds feature lists for arwiki, cswiki, enwiki, kowiki, and viwiki
- Adds `fetch_draft_text`, `fetch_article_text`, and `taxo_label` utilities

1.3.2 Changed

- Use fasttext 100 cell vectors for enwiki

1.4 [0.2.0]

1.4.1 Added

- Added `extract_from_text` utility.
- Added `fetch_text` script to text for a list of page titles.
- Feature extraction rule using `revscoring` `extract`.
- Gradient boosting config.
- Added Word Vectors feature lists.
- Added `mid_level_wp` to arguments.
- Added `fetch_page_wikiprojects` script to label pages with all wikiprojects.
- Parser code for generating mapping of mid-level topics to wikiprojects.
- Exception handling for request failures.
- Added requirement for `revscoring` v2.5.1
- Added requirement for `mediawiki-utilities` v0.4.18.
- Parser Tests.
- Release Criteria.

1.4.2 Fixed

- Headings regex bug.

1.4.3 Changed

- Escape angular brackets.
- Refactored logic for request processing.
- Use Pytest for testing.
- Dataset output file contains a date now.

1.4.4 Removed

- `mediawiki-utilities==0.4.18`

1.5 [0.1.1] - 2017-09-05

1.5.1 Added

- Bootstrap code.
- WikiProjects Parsing Script.

INDICES AND TABLES

- genindex
- modindex
- search

PYTHON MODULE INDEX

d

drafttopic, ??

D

drafttopic (*module*), 1