
SNER

Release 0.1.1

Jul 23, 2018

Contents

1	Installation	3
2	Usage	5
3	Options and Arguments	7
4	Contents	9
5	Indices and tables	15

build passing

CHAPTER 1

Installation

```
$ pip3 install sumerian-ner
```

If you are not in a [virtualenv](#) you may need to use `pip3` with `sudo` or `--user`.

If you'd rather clone the repository.

```
$ git clone https://github.com/wwunlp/sner.git
$ cd sner/
$ pip3 install .
```


CHAPTER 2

Usage

If you used `pip3` to install `sner`, you can call `sner` from the command line. If you cloned the repo, you can run `python3 -m snr` from within the repo.

Options and Arguments

- `-r` or `--run`: Run one of the following models: `dec`, `nbc`, `ner`, `rdf`, `sgd`, or `svc`. Or one of the following routines: `analysis`, `export`, `export-atf`, `formatting`, `over-fit`, or `testing`.
- `-cf` or `--config`: Configuration file to use.
- `-p` or `--path`: Path to data directory.
- `-c` or `--corpus`: File name of the corpus.
- `-a` or `--attestations`: File name of the attestations.
- `-sr` or `--seed-rules`: File name of the seed rules.
- `-i` or `--iterations`: Number of iterations.
- `-mr` or `--max-rules`: Max number of rules per iterations.
- `-al` or `--alpha`: Alpha value.
- `-k` or `--k`: K value.
- `-nd` or `--norm-date`: Enable date normalization.
- `-ng` or `--norm-geo`: Enable geographical name normalization.
- `-nn` or `--norm-num`: Enable number normalization.
- `-np` or `--norm-prof`: Enable profession normalization.

`sner` will check `sner.conf` before using default value. The `sner.conf` uses JSON syntax. i.e. `--corpus corpus.csv` would be `"corpus": "corpus.csv"`. If you want to change the hyperparameters used, you can define them in `sner.conf`. If your `sner.conf` is not in the root of the repository, you can set the environment variable `SNER_CONF` to the path of `sner.conf`.



4.1 Changelog

4.1.1 0.1.0 - 2017-02-08

- Initial release on GitHub.

4.1.2 0.1.1 - 2018-07-22

- Initial release on PyPI.

4.2 Contributing to Sumerian Named Entity Recongnition

To start contributing, you can install the project by

```
git clone https://github.com/wwunlp/sner.git
cd sner/
pip3 install .
```

Please follow the [Google Python Style Guide](#).

The project is organized into `classes`, `docs`, `models`, `scripts`, and `tests`.

- Class files should probably go into `classes`.
- The main file for a model should go into `models`.
- General scripts should go into `scripts`, and scripts for models should go into a sub-directory of the model in `scripts`.
- Test files should go into `tests`.
- Compile docs with `sphinx`, and post the html files to `docs/`.

If you're parsing `.atf` files, please look at the [ATF reference](#).

4.3 Contributors to Sumerian Named Entity Recognition

- [Andy Brown](#)
- [Mike Canoy](#)
- [Ian Fisk](#)
- [Matt Adler](#)
- [Luke Terry](#)

4.4 Working with ATF

This is intended to be reference for those working with `sner` and `.atf`, based on material from the [cdli](#) and [oracc](#). Sections are organized by first character in line.

4.4.1 `&`

`&P329447 = CUSAS 03, 0001`: Beginning of a text with CDLI ID and name.

4.4.2 `#`

- `#project: cams`: Project of text. Unique to Oracc.
- `atf: lang akk` or `#atf: lang akk-x-stdbab`: Language of text. CDLI and Oracc, yet Oracc uses `sux` for Sumerian like the CDLI.
- `atf: use unicode`: Unique to Oracc. CDLI uses ASCII.
- `tr.en::`: Transliteration of text.
- Other `#` lines are comment lines.

4.4.3 `@`

Object type

- `@tablet`
- `@envelope`
- `@prism`
- `@bulla`
- `@fragment`
- `@object` - generic tag, i.e. `@object cone` or `@object seal`

Surfaces & locations

- `@observe`

- @reverse
- @left, @right, @top, @bottom
- @face
- @surface - generic tag, i.e. @surface side a
- @edge
- @seal
- @column

Discourse

- @catchline
- @colophon
- @date
- @signatures, @signature
- @summary
- @witnesses

4.4.4 \$

- Breakage
- Blank lines
- Qualification
- Number
- Scope
- State
- Ruling
- Image location

4.4.5 Text

1. 3(disz) gurusz szitim: Number, period, space, then line. Line continuation is space, then line.

4.4.6 >>

>>Q001075 001: Reference to Oracc [Q-catalogue](#) ID and line number.

4.5 sner package

4.5.1 Subpackages

snr.classes package

Submodules

snr.classes.display module

snr.classes.rule module

snr.classes.token module

Module contents

snr.models package

Submodules

snr.models.ner module

snr.models.sklearn_launcher module

Module contents

snr.scripts package

Subpackages

snr.scripts.ner package

Submodules

snr.scripts.ner.contextualfromnames module

snr.scripts.ner.namesfromrule module

snr.scripts.ner.rulefilter module

snr.scripts.ner.rulesperformance module

snr.scripts.ner.spellingfromnames module

snr.scripts.ner.updatetokenstrength module

Module contents

Submodules

4.5. sner package

snr.scripts.analysis module

CHAPTER 5

Indices and tables

- `genindex`
- `modindex`
- `search`