


研究ファイルと使い方

1. Pythonの環境構築をする。
2. SudachiのGithub
<https://github.com/WorksApplications/Sudachi#sudachi-日本語readme>
3. 以下を理解するためには、まずロードマップをやることをおすすめ
 [学習ロードマップ](#)

すべてのファイルの配置を記述する

https://s3-us-west-2.amazonaws.com/secure.notion-static.com/07f76b0f-28ed-41ce-b897-14830b6f5d61/sudachipy_env.zip

```
sudachipy_env
- oitama_system (メインのシステム)
- outputFig (作図データの保存先)
- share (よくわからない)
- outputCSV (出力CSVの保存先)
- lib -> python3.9 -> -> site-packages ->ここの中に、いろいろなパッケージが入っている。
この中の、sudachipy -> resources -> この中に、バイナリ化した辞書を配置することになる。
- user_dict ( この中に、バイナリ化したいSCV辞書を格納しておく )
- requirements.txt
- auto_evaluation.csv ( ここに正解データを入れたCVSを置いておく。名前は、auto_evaluation.csv )
```

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/75e36a8c-444e-43b3-9c1b-c976b280ea06/requirements.txt>

requirements.txtを実行すれば使用しているすべてのパッケージがinstallされるはず。あとは、sudachi_coreなどの辞書はその時の最新を入れる。

CSVで作成した辞書を、バイナリ化 (01表記) する。

<https://s3-us-west-2.amazonaws.com/secure.notion-static.com/d462f8ab-43a6-41bd-81a8-72480734465a/binaryCode.txt>

↑をコマンドライン（ターミナル）で、実行してバイナリ化する。（難所、要注意）

【重要】sudachipyでユーザー辞書をバイナリ化

→ バイナリ化、実行成功したら、sudachipy_envの中にuser.dicが生成されるので、sudachipy

lib -> python3.9 -> -> site-pachages -> sudachipy → resources → の中に格納する。

そこにある sudachi.jsonの中に、そのファイル名を適切な場所に記述する。

→ バイナリ化失敗したら、CSVの中にミスがあるはず。

空欄に、asteriskがないとか、情報が埋まってないとか、接続IDが8000以上だとか、エラーを読めばCSVの何行目のどこがミスっているかわかる。

実行

pythonを実行する。

sudachipy_env に移動して、oitama_system に移動して、、、

そこで、python3 main.pyを実行する。

“なにをしますか？”

'0:分割と評価','1:翻訳と評価','2:作図'

その後、言われた通りに選択する。