

情報実験第 4 ビッグデータ処理

全体スケジュール

- 第1回 ガイダンス・環境構築
- 第2回 講師紹介,Python基礎
 - 文法、データ型
 - テーブルデータ操作(pandas),グラフ
- 第3回 テキストマイニング基礎
 - 形態素解析、係り受け解析、文書類似度
- 第4回 ～ 第5回 テキストマイニング（個人課題）
- 第6回 チーム作り / チーム課題設定
- 第7回～第9回 チーム課題
- 第10回 チーム課題発表
- Extra
- 第11回 まとめ(課題フィードバック等)

講師紹介



たなか こうじ
田中 公司

(株) 日立公共システム
事業戦略統括本部 研究・開発部
主任技師

2010 筑波大学 ビジネス科学研究科 経営システム科学専攻

2012 同修了 経営学修士(MBA)

2012 筑波大学 ビジネス科学研究科 企業科学専攻システムズ・マネジメントコース

2015 同修了 博士 (システムズ・マネジメント)

専門： **自然言語処理**

- テキストマイニング系論文多数、
その他、ソフトウェア欠陥件数予測技術に関する論文など

Python基礎

- とりあえず実行 編 -



Python

- 汎用高水準言語
- クロスプラットフォーム
- 文法がシンプル
(学習しやすい/読みやすい)
- データ分析/AIの分野で広く使われている
(ライブラリが充実)
- 動的な型付け
- ガベージコレクション
- マルチパラダイム
(手続き型、関数型、オブジェクト指向)

前準備

- プログラム保存用のフォルダを作る

書類/exp4bd/1st/

#Finderで 書類 を表示して、右クリックメニュー
(orメニュー [ファイル]) で[新規フォルダ]

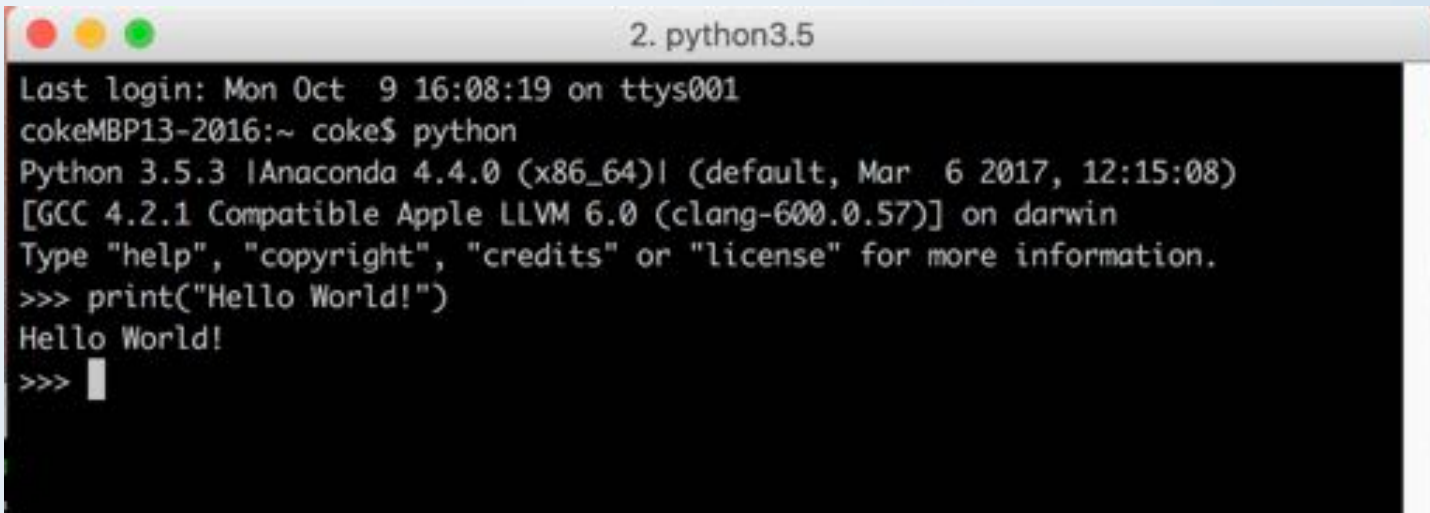
- 1stの下にdataフォルダを作る
- <http://www.itpro.titech.ac.jp/exp4/>からダウンロードした
weather.csv,weather2.csvをdataフォルダに格納

#フォルダのパス・名前は任意だが、以降の説明では上記パスを利用するので、適宜読み替えること

Pythonの実行 :

①インタラクティブシェル

- ターミナル で
python
と入力
- 「>>>」の後に
print("Hello World!")



```
2. python3.5
Last login: Mon Oct  9 16:08:19 on ttys001
cokeMBP13-2016:~ coke$ python
Python 3.5.3 |Anaconda 4.4.0 (x86_64)| (default, Mar  6 2017, 12:15:08)
[GCC 4.2.1 Compatible Apple LLVM 6.0 (clang-600.0.57)] on darwin
Type "help", "copyright", "credits" or "license" for more information.
>>> print("Hello World!")
Hello World!
>>> 
```

- exit()
と入力してインタラクティブシェルを終了

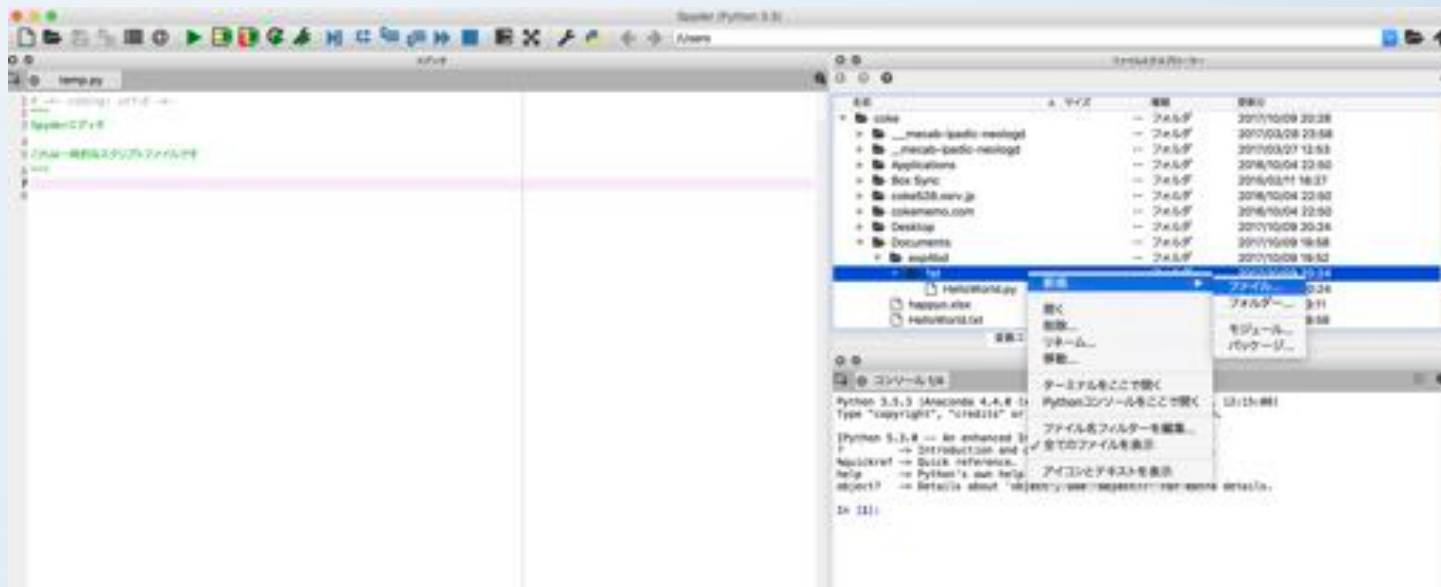
Pythonの実行 : ②ファイル

- テキストエディタでフォルダ : 書類/exp4bd/1st/
に新規ファイルを作成
#macの場合 テキストエディットを開き
書類/exp4bd/1st/
を表示して、[新規書類]
→メニューの[フォーマット | 標準テキストにする]
[編集 | 自動置換 | スマート引用符]のチェックを外す
- `print("Hello World!")`
と入力して、
ファイル名: HelloWorld.py
で保存
- ターミナル で
`python ~/Documents/exp4bd/1st/HelloWorld.py`
と入力

Pythonの実行 : ③IDE

<<起動とファイル作成>>

- Anaconda NavigatorからSpyderを起動
- ファイルエクスプローラで (ユーザ名)/Documents/exp4bd/1st/を選択して右クリックメニュー [新規 | ファイル]



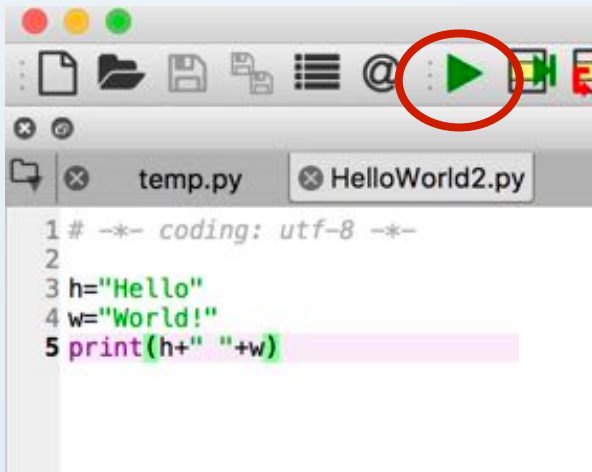
- ファイル名 : HelloWorld2.py

Pythonの実行 : ③IDE

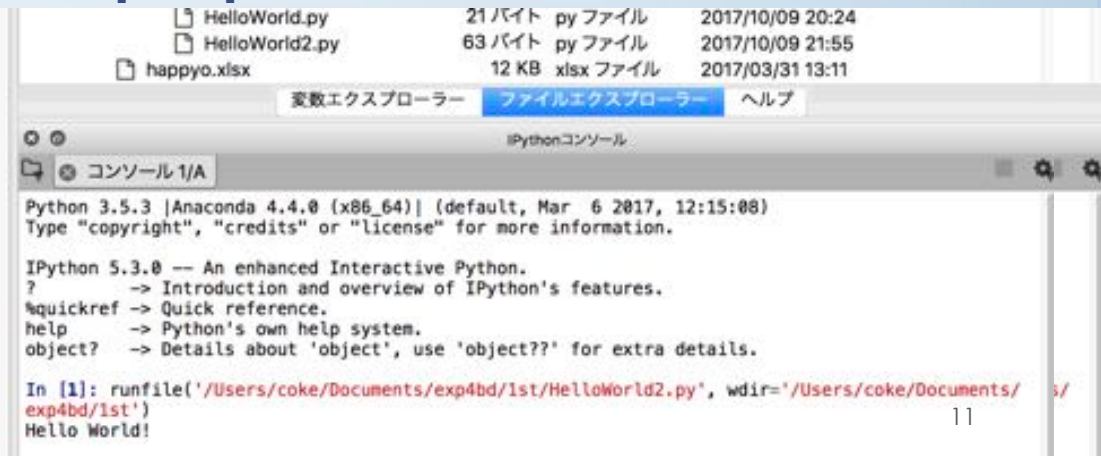
<<編集と実行>>

- 画面左のエディタに作成したファイルが開くので、以下を入力

```
h="Hello"
w="World!"
print(h+" "+w)
```



- 実行ボタン(緑三角)を押すと、画面右下のIpythonコンソールに実行結果が表示される
#実行設定のダイアログが出た場合はそのまま[実行]



Pythonの実行 : ③IDE

<<デバッグ>>

- エディタのw="hello"の行の先頭ダブルクリックしてブレークポイントを設定
→行の頭に●が表示される

```

1 # -*- coding: utf-8
2
3 h="Hello"
4 w="World!"
5 print(h+" "+w)
6
7
8

```



- デバッグ開始ボタンを押す



- 次のブレークポイントまで実行ボタンを押す

Pythonの実行 : ③IDE

<<デバッグ>>

①エディタのw="hello"の行の先頭ダブルクリックしてブレークポイントを設定

→行の頭に●が表示される

```
1 # -*- coding: utf-8
2
3 h="Hello"
4 w="World!"
5 print(h+" "+w)
6
7
8
```



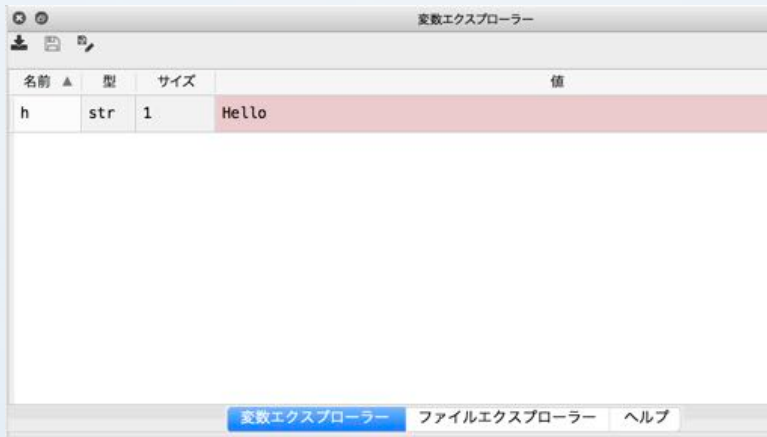
②デバッグ開始ボタンを押す



③次のブレークポイントまで実行ボタンを押す

Pythonの実行 : ③IDE

④変数エクスプローラ（ファイルエクスプローラから切り替え）
に変数hの内容が表示される



⑤ステップ実行ボタンを押す

⑥変数エクスプローラに変数w
の内容も表示される



⑦次のブレークポイントまで実行ボタンを押して最後まで実行させる
(Ipythonコンソールに"Hello World!"が表示される)

Pythonの実行 : ④jupyter notebook

Let's
try it

- Anaconda Navigatorからjupyter notebookを起動
- Documents/exp4bd/1st/を開く
- [New | Python3]を選択
- [File | Rename]で適当な名前で作成する
- Cell(In []:と表示されている枠) に HelloWorld2.pyと同じ内容を入力

```
In [ ]: h="Hello"  
        w="World!"  
        print(h+" "+w)
```

Pythonの実行 : ④jupyter notebook

- cell実行ボタンをクリック
→結果が表示され、新たなcellが表示される



- 先ほどのcellの内容を編集 ("World!"を"World!!!")に変更) して再度実行
→結果が更新される



Pythonの実行 : ④jupyter notebook

<<Markdownでメモ>>

- 新しいcellを選択し、メニューバーのドロップダウンを[code]から[Markdown]に変更
- CellにMarkdown記法で何か記載する

#Markdown記法については
メニューの[Help | Markdown]を
選択して表示されたページの
[
[Basic writing and formatting syntax](#)]
のリンクを参照

```
## Markdown練習
### サブタイトル
#### テキスト
テキストテキスト
#### コード
```
some code
some code
```

#### リンク
テキストテキスト \[リンク\](http://localhost:8888)テキスト
#### リスト
- list1
- list2
- List3

#### 番号付きリスト
1. Num1
2. Num2
3. Num3
```

Pythonの実行 : ④jupyter notebook

<<Markdownでメモ>>

- cell実行ボタンをクリック
→表示がHTMLに変わることを確認
#ダブルクリックすると編集画面に戻る

Markdown練習

サブタイトル

テキスト

テキストテキスト

コード

```
some code  
some code
```

リンク

テキストテキスト [リンク](#) テキスト

リスト

- list1
- list2
- List3

番号付きリスト

1. Num1
2. Num2
3. Num3

Python基礎

- 最低限の文法 編 -

→ jupyter notebook
"Python文法基礎"



Python基礎 - Pandasによる データ操作 編-

→ jupyter notebook
"pandas基礎"
"グラフ基礎"

