

(1) 準備

専門演習B：Pythonによるデータ分析入門

環境の整備

- ATOM（エディタ）のインストール
- Python
 - パッケージ管理ツールのインストール
 - パッケージのインストール
- Yahoo!デベロッパーネットワークの登録（グループで1個でいい）

F501における全般的な注意

- 管理者権限がない → ユーザ領域にインストール
 - ユーザ領域：自分のホームディレクトリ以下
- 特にPythonのパッケージインストール時に問題になる場合がある

ATOMのインストール

ATOM

- ウェブサイト <https://atom.io>
 - ダウンロードしてパスの通ったところにコピーすればインストール完了
 - パスの通ったところ = ホーム直下の Applications あたり？
- その他必要なパッケージはATOMの Settings からインストール
(以下はオススメパッケージ)
 - autocomplete-python : Pythonの入力の補完をしてくれる。
 - atom-beautify : フォーマット整形ツール

ところで…

Q. エディタって何？

A. 文書を入力・編集するソフト

Q. ワープロとは違うのか？

A. エディタは文書の整形・装飾機能を省き，高速に入力することに特化したソフト。ワープロは文書の整形・装飾機能も重要。

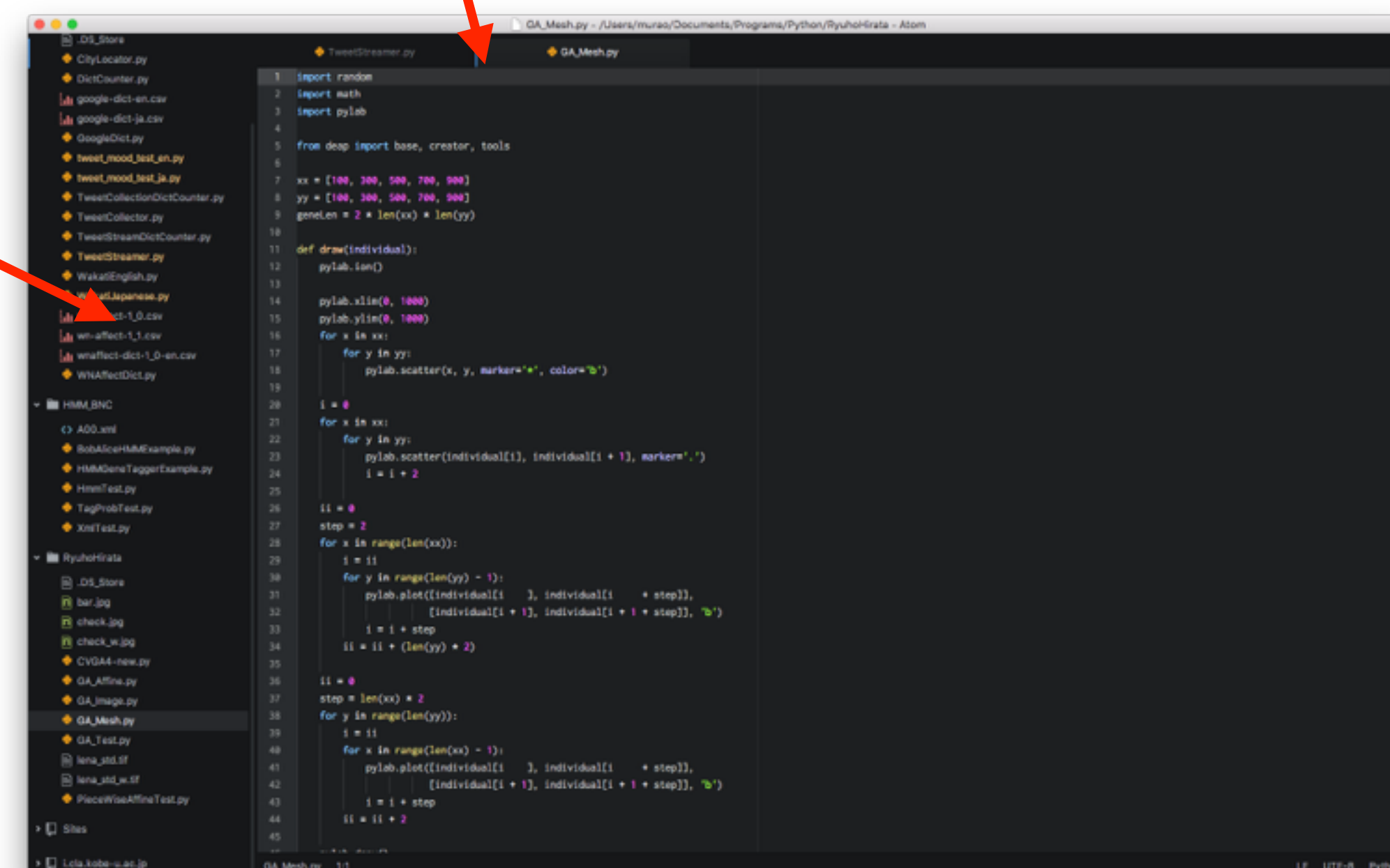
ATOMの画面

ブラウザ領域：

「ファイル」メニューの

「Add Project Folder...」で登録

タブ：複数のファイルを切り替え



*ブラウザ領域に演習用のフォルダを登録しておくると便利!!

Pythonの環境設定

Python

- pip (パッケージ管理ツール) のインストール (青字が入力)

```
$ curl -kL https://raw.githubusercontent.com/pypa/pip/master/contrib/get-pip.py > get-pip.py  
$ python get-pip.py --user
```

pipを実行できるようにする（パスを通す）

1. ATOMで新規ファイルを作成

2. 下記の内容を記述

```
export PATH=$HOME/Library/Python2.7/bin:$PATH
```

4. ホーム直下に `.bash_profile` という名前で保存

*すでにファイルがある場合は上書きせず、既存のファイルを書き換える

5. その後ターミナルを再起動し、下記のように動作確認（青字が入力）

```
$ pip --version
pip 7.1.2 from /home1/hjmr/Library/Python/2.7/lib/python/site-packages (python 2.7)
$
```

pipの簡単な使い方

```
pip list
```

インストールされているパッケージの一覧を表示（全体）

```
pip list --user
```

ユーザ領域にインストールされているパッケージの一覧表示

```
pip search <名前>
```

名前を含むパッケージの検索

```
pip install <パッケージ名>
```

パッケージのインストール

```
pip install <パッケージ名> --user
```

ユーザ領域へのパッケージのインストール

```
pip uninstall <パッケージ名>
```

パッケージのアンインストール

必要なパッケージのインストール (pip)

- nltk
 - 自然言語処理ツールキット (Natural Language Toolkit)
- scikit-learn
 - 機械学習関係
- requests-oauthlib
 - OAuth認証対応のHTTPリクエストライブラリ
- beautifulsoup4
 - XMLのパパーサ

nltk用データのインストール（青色が入力）

- コーパスなどnltkの実行に必要なデータをインストールする

```
$ python  
>>> import nltk  
>>> nltk.download()
```

- データダウンロード用のウィンドウが出るので全て（all）インストール（場所はホーム直下の nltk_data になるはず）
- 終わったらウィンドウを閉じ下記を入力して終了

```
>>> exit()  
$
```

nltkの確認(1)

- Pythonを対話モードで起動

```
$ python  
>>>
```

- nltkを読み込み

```
>>> import nltk
```

- 解析するテキストをセット

```
>>> text = "Arthur didn't feel very good at eight o'clock on  
Thursday morning."
```

nlTKの確認(2)

- トークン分割と結果の確認

```
>>> tokens = nltk.word_tokenize(text)
>>> tokens
```

- 品詞タグ付けと結果の確認

```
>>> tagged = nltk.pos_tag(tokens)
>>> tagged
```

- 構文木の解析と結果の確認

```
>>> entities = nltk.chunk.ne_chunk(tagged)
>>> entities
```

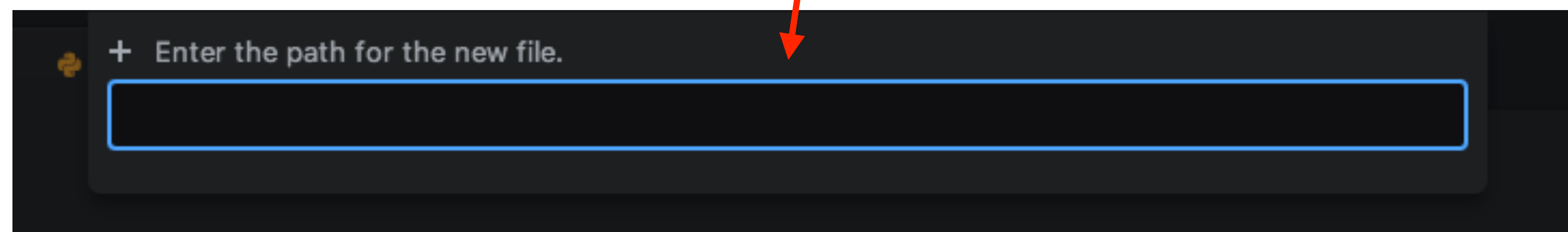
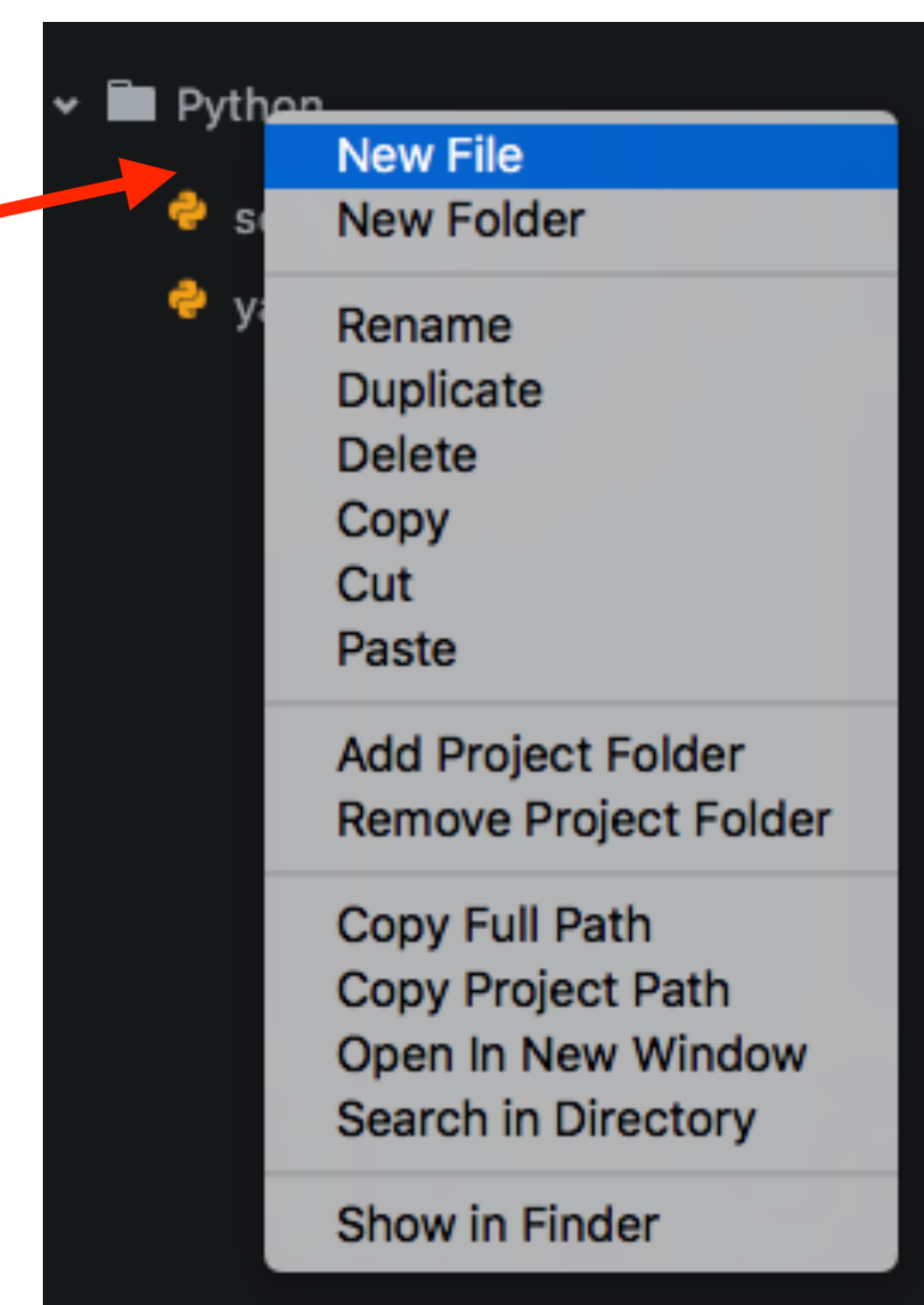
Yahoo! デベロッパーズネットワーク

Yahoo!デベロッパネットワーク

- Yahoo! IDに連携してYahoo! Japanの提供するWebサービスを利用
 - 今回は形態素解析のサービスを利用する予定
- ID発行の手順
 1. Yahoo! IDを取得
 2. アプリケーションIDを取得
 - クライアントサイドにする
 - アプリケーション名は自由
 3. 画面にアプリケーションIDが表示されるのでメモっておくこと

Yahoo! 形態素解析のテスト（入力開始!!）

1. ATOMを起動
2. ブラウザ領域に登録済みの作業用フォルダで右クリック
3. ポップアップメニューから「New File」を選択
4. 画面上部に出る入力パネルにファイル名を入力



ファイル名は yahoo_test.py としよう

Yahoo! 形態素解析のテスト (絶賛入力中)

```
#!/usr/bin/env LC_ALL=ja_JP.utf-8 python
# -*- coding: utf-8 -*-

import urllib
import sys
reload(sys)
sys.setdefaultencoding('utf-8')

from bs4 import BeautifulSoup

appid = 'XXXXXXXXXX' # Application ID
url = 'http://jlp.yahooapis.jp/MAService/V1/parse'
```

(右上に続く)

```
def morph(sentence, appid=appid, results="ma"):
    postdata = {
        'appid'      : appid,
        'results'    : results,
        'sentence'   : sentence.encode('utf-8')
    }

    params = urllib.urlencode(postdata).encode('ascii')
    result = urllib.urlopen(url, params)

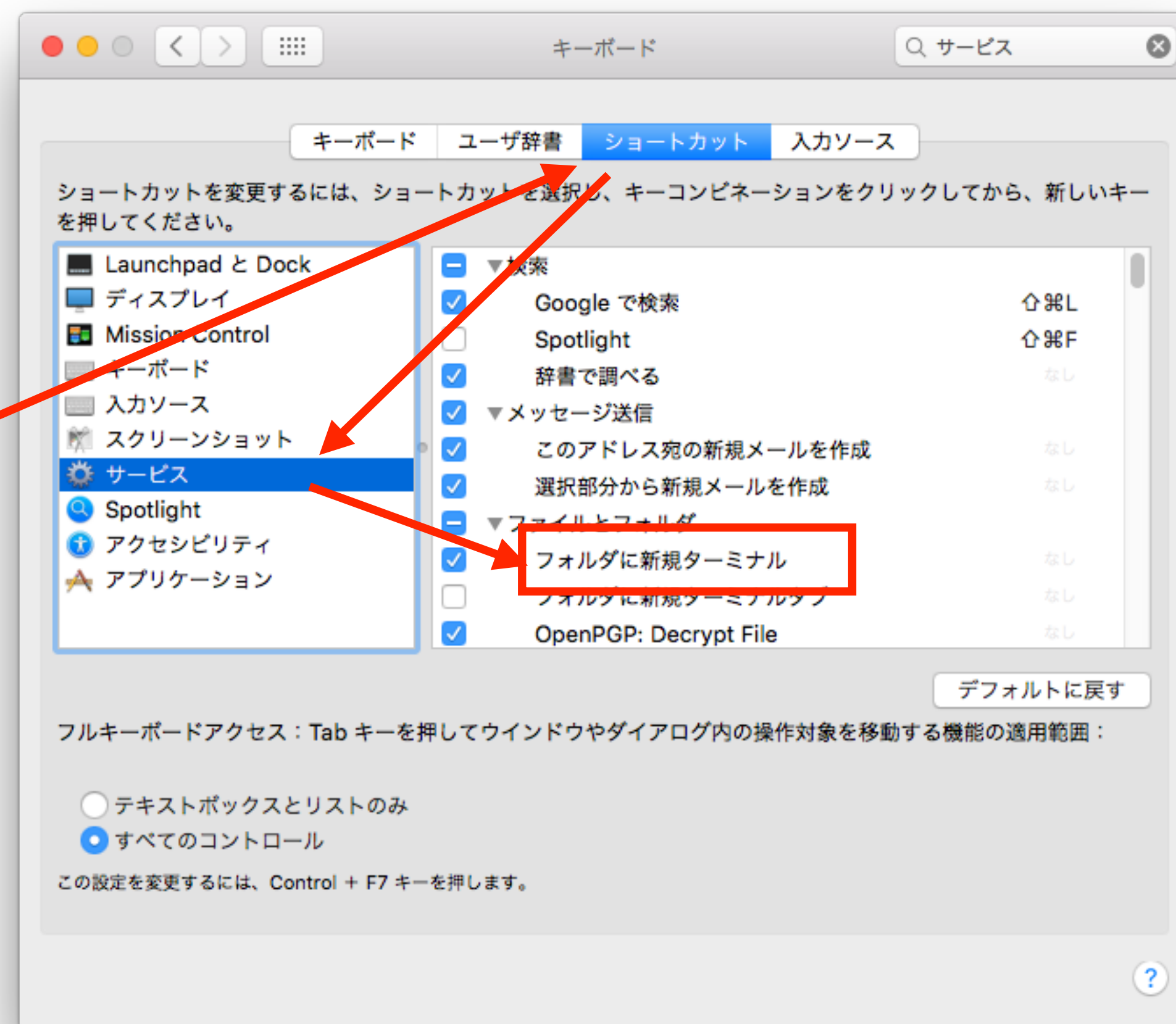
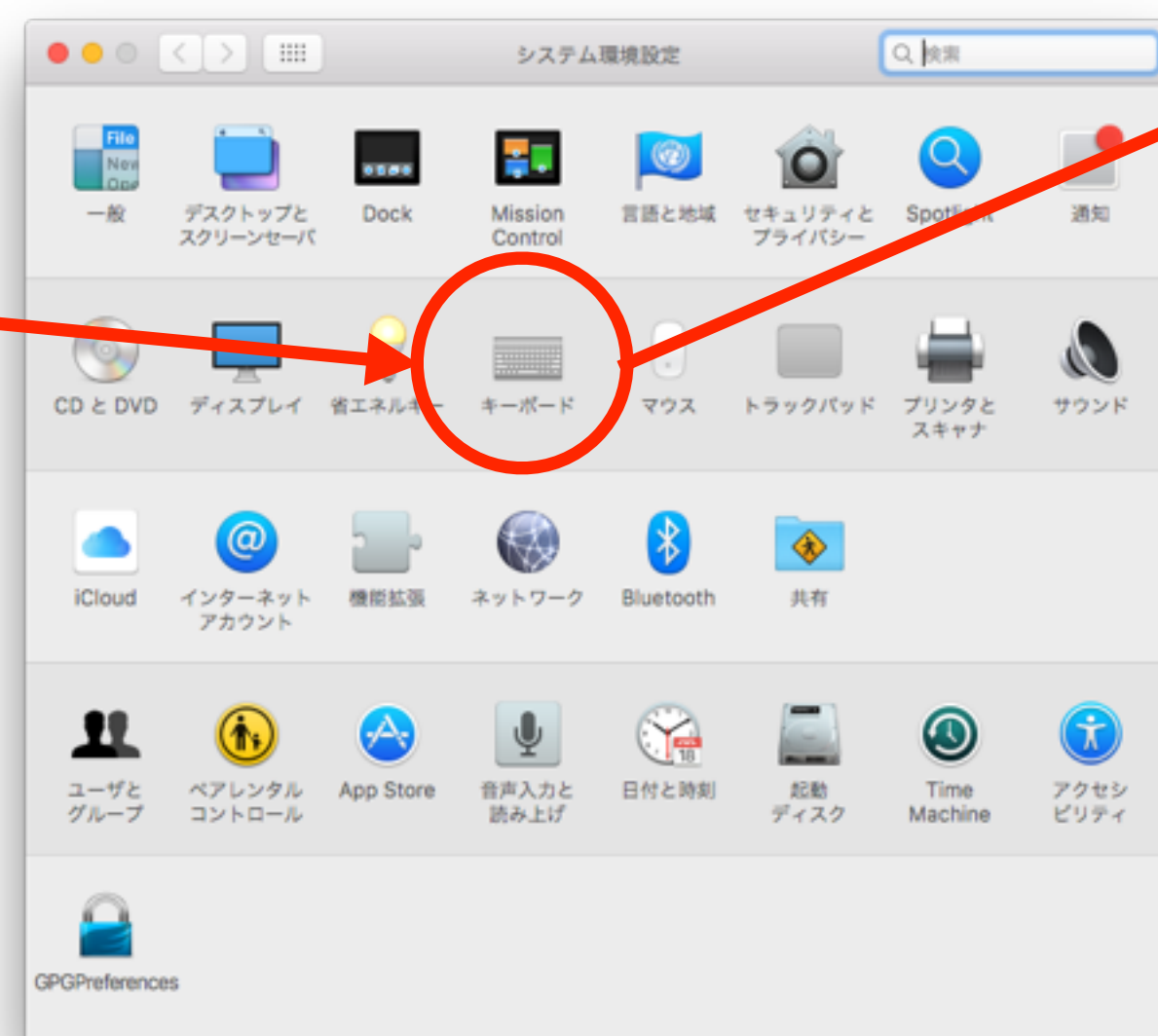
    soup = BeautifulSoup(result.read(), "html.parser")
    return [(w.surface.string, w.reading.string, w.pos.string)
            for w in soup.ma_result.word_list]

sentence = '庭には二羽ニワトリがいる。'
result = morph(sentence, appid=appid)
for word, reading, pos in result:
    print("%s(%s)/%s"%(word, reading, pos))
```

*インデント (字下げ) もこの通りに入力

Yahoo! 形態素解析のテスト（実行準備）

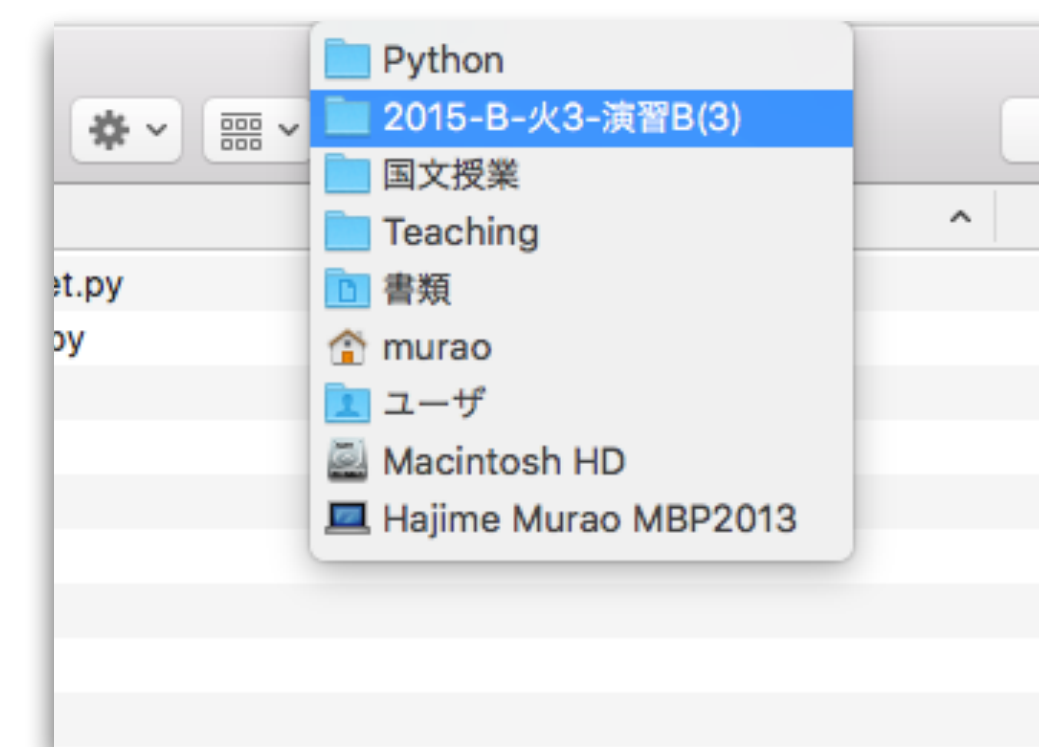
1. Macのシステム環境設定を開く
2. キーボードを開く
3. 「ショートカット」 → 「サービス」 を選択
4. 「フォルダに新規ターミナル」 にチェック



Yahoo! 形態素解析のテスト（実行手順1）

1. ファイルのあるフォルダを開く

- ✓ 方法1：ATOMのブラウザ領域でファイルを右クリック
→ ポップアップメニューから「Show in Finder」を選択
- ✓ 方法2：Finderでファイルを探す



2. フォルダを一つ上げる

1. Finderのタイトルを右クリック
2. ポップアップメニューから一つ上のフォルダを選ぶ（上から2つ目）

3. ファイルのあるフォルダでターミナルを開く

1. フォルダを右クリック
2. ポップアップメニューの「サービス」→「フォルダに新規ターミナル」

Yahoo! 形態素解析のテスト (実行手順2)

