

mi-8. 自然言語処理

人工知能を演習で学ぶシリーズ（8）

<https://www.kkaneko.jp/dblab/intro/m/index.html>

金子邦彦



アウトライン



8-1 自然言語処理

8-2 機械翻訳

8-3 合成, パターン認識



8-1. 自然言語処理

自然言語処理



- 人間のことば（日本語や英語など）を，コンピュータが処理することを，**自然言語処理**という

自然言語処理の技術



- 文を単語に分割（**単語の切り出し**）
- 単語の種類（**品詞**）の判定
- 係り受けの関係などの分析（**構文解析**）
- **意味解析**

単語の切り出しと品詞の判定の例



白い雲と青い空が美しい

```
>>> print(m.parse ("白い雲と青い空が美しい"))
白い      シロイ   白い      形容詞-自立      形容詞・アウオ段      基本形
雲        クモ     雲        名詞-一般
と        ト       と        助詞-並立助詞
青い      アオイ   青い      形容詞-自立      形容詞・アウオ段      基本形
空        ソラ     空        名詞-一般
が        ガ       が        助詞-格助詞-一般
美しい    ウツクシイ 美しい    形容詞-自立      形容詞・イ段      基本形
EOS
```

単語ごとの読み仮名，品詞などが，
コンピュータにより得られる



⑥ 次のコマンドを入れる

python

```
-bash-4.4$ python
Python 2.7.15rc1 (default, Nov 12 2018, 14:31:15)
[GCC 7.3.0] on linux2
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

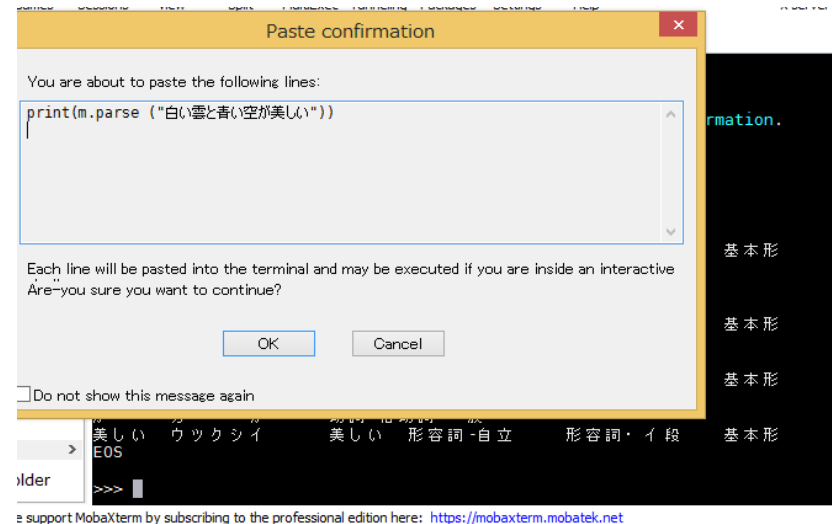
⑦ 次のプログラムを入れる。結果を確認

```
import sys
import MeCab
m = MeCab.Tagger ("-Ochasen")
print(m.parse ("海は広い"))
```

```
>>> import sys
>>> import MeCab
>>> m = MeCab.Tagger ("-Ochasen")
>>> print(m.parse ("海は広い"))
海      ウミ      海      名詞 -一般
は      ハ      は      助詞 -係助詞
広い    ヒロイ    広い    形容詞 -自立
EOS
```

			形容詞・アウオ段	基本形
--	--	--	----------	-----

プログラムを入れるとき



Webブラウザの
右クリックメニューで
コピー

MobaXTermの
右クリックメニューで
貼り付ける (**Paste**)
確認表示には OK



⑧ 引き続き，次のプログラムを入れる．結果を確認

```
print(m.parse ("白い雲と青い空が美しい"))
```

```
>>> print(m.parse ("白い雲と青い空が美しい"))
白い      シロイ   白い      形容詞-自立      形容詞・アウオ段      基本形
雲        クモ      雲        名詞-一般
と        ト        と        助詞-並立助詞
青い      アオイ   青い      形容詞-自立      形容詞・アウオ段      基本形
空        ソラ      空        名詞-一般
が        ガ        が        助詞-格助詞-一般
美しい    ウツクシイ 美しい    形容詞-自立      形容詞・イ段      基本形
EOS
```



⑨ ネットの記事や文書など, 1文, コピーして
試すことを数回行いなさい

print(m.parse ("あなたの文章"))

```
>>> print(m.parse ("日本国民は、正当に選挙された国会における代表者を通じて行動し"))
日本      ニッポン      日本      名詞-固有名詞-地域-国
国民      コクミン      国民      名詞-一般
は        ハ          は        助詞-係助詞
、        、          、        記号-読点
正当      セイトウ      正当      名詞-形容動詞語幹
に        ニ          に        助詞-副詞化
選挙      センキョ      選挙      名詞-サ変接続
され      サ          する      動詞-自立          サ変・スル      未然レル接続
れた      レ          れる      動詞-接尾          一段      連用形
た        タ          た        助動詞      特殊・タ      基本形
国会      コウカイ      国会      名詞-一般
における      ニオケル      における      助詞-格助詞-連語
代表      ダイヒョ      代表      名詞-サ変接続
者        シャ      者        名詞-接尾-一般
を        ヲ          を        助詞-格助詞-連語
通じて      ヲツウジテ      通じて
行動      コウドウ      行動      名詞-サ変接続
し        シ          する      動詞-自立          サ変・スル      連用形
EOS
```

構文解析の例



・ 構文解析 KNP のデモサイト

<http://lotus.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/knp.cgi>

← → ↺ ① lotus.kuee.kyoto-u.ac.jp/nl-resource/cgi-bin/knp.cgi

KNPを試してみる

入力文:

☐ 解析結果の詳細を表示

空は青い

S-ID:1 KNP:4.2-407aed4a DATE:2019/06/07 SCORE:-5.37833

空は——— <体言>
青い<用言:形><格解析結果:ガ/空;ニ/->

EOS

入力文:

☐ 解析結果の詳細を表示

白い雲と青い空が美しい

S-ID:1 KNP:4.2-407aed4a DATE:2019/06/07 SCORE:-25.24505

白い——— <用言:形><格解析結果:ガ/雲>
雲と(P)——— <体言>
青い——— <用言:形><格解析結果:ガ/空;ニ/->
空が(P)——— <体言>
美しい<用言:形><格解析結果:ガ/雲;ガ/空;ニ/-;デ/-;カラ/-;時間/-;ガ 2/->

EOS

意味解析の例



- それぞれの言葉の意味を，コンピュータが理解しながら，他の言葉の意味を判断するなど

駅前のコンビニは，**油を売っていた**

あの人は，仕事中に，**油を売っていた**

- 単語の切り出し, 品詞の判定
 - 辞書との照合により, ある程度可能
- 構文解析, 意味解析
 - コンピュータは, 前もって, 大量の文書を用いて**事前学習**を行う.

自然言語処理の用途



- スマホやパソコンでの**かな漢字変換**，**予測変換**
- **機械翻訳** 「Google 翻訳」など
- **対話システム** 「Siri」，「Amazon Echo」など

まとめ



- 単語 単一の意味のまとまり
- 品詞 単語の種別
- 係り受け 主語, 述語, 修飾語などの, 単語間の関係の分析
- 意味解析 語句の意味の特定などを行うこと



8-2. 機械翻訳

機械翻訳システム



ある自然言語



言語
理解



言語
生成



別の自然言語

機械翻訳の例



📄 テキスト 📄 ドキュメント

日本語 - 自動検出 英語 日本語 韓国語 ⇄ 日本語 英語 韓国語

白い雲と青い空が美しい ✕ Beautiful white clouds and blue sky ☆

Shiroi kumo to aoi sora ga utsukushi 11/5000 🗣️ 🗣️ 📄 ⋮

フィードバックを送信

📄 テキスト 📄 ドキュメント

言語を検出する 英語 日本語 韓国語 ⇄ 日本語 英語 韓国語

Once upon a midnight dreary, while I pondered,
weak and weary,
Over many a quaint and curious volume of
forgotten lore,
| ✕ 深夜の惨めな時に、私が熟考しながら、弱くて疲
れきった、
忘れられていた伝説の古風で不思議なボリューム
の上に、 ☆

Shin'ya no mijimena toki ni, watashi ga jukukō shinagara,
yowakute tsukare kitta, wasure rarete ita densetsu no kofūde
fushigina boryūmu no ue ni, 🗣️ 🗣️ 📄 ⋮

120/5000 🗣️ 🗣️ 📄 ⋮

機械翻訳の例



- ワード、パワーポイントファイルの中から、文字を読み取ることもできるように

ue-2. C++ プロジェクトの新規作成 (Unreal Engine 演習シリーズ)

<https://www.kkaneko.jp/cc/ue/index.html>

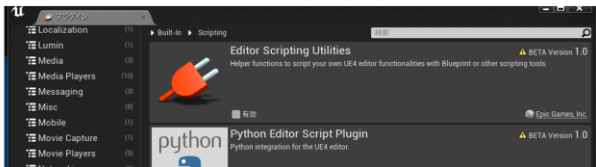
金子邦彦



- ① Unreal エディタのメインメニューで、「編集」、「プラグイン」と操作



- ② 「Scripting」を選び、「Python Editor Script Plugin」の「有効」をチェック



ue-2. Create a new C ++ project (Unreal Engine exercise series)

Kunihiko Kaneko

<https://www.kkaneko.jp/cc/ue/index.html>

- 1 In the main menu of the Unreal Editor, operate "Edit" and "Plug-in"

- 2 Select "Scripting" and check "Enable" in "Python Editor Script Plugin"

- 3 Confirm the message and "Yes"

- Create Project Visual Studio code is generated
- Visual Studio project is open