



# 特許情報に関する言語生成モデルを 活用した知的財産創造手法の開発

Development of Intellectual Property Creation Method  
Using Language Generation Model on Patent Information

Shigeaki Onoda

Graduate School of Information Engineering, Toyama Prefectural University  
t855005@st.pu-toyama.ac.jp

Wednesday, 7 24, 2019,  
Toyama Prefectural Univ.



## 行ったこと

### 入力用データの作成

- キーワード抽出付きクロール
- モデルの改良



# キーワード付きクロールの例

キーワード抽出付きの特許文書クロールを automation 分野に対して行った

以下は図 1 はその 1 サンプルをあらわしたものであるこのキーワードを部分をモデルの入力として利用する

```
In [26]: 1 df
Out[26]:
```

	0
<b>_id</b>	5d2851a807e7717271733b77
<b>PID</b>	US20100138015A1
<b>inventors</b>	[Armando Walter Colombo, Axel Bepperling, Rona...
<b>title</b>	Collaborative automations system and method for the co...
<b>grant_date</b>	[]
<b>CNUM</b>	2
<b>CBNUM</b>	3
<b>description</b>	The invention relates to a collaborative autom...
<b>keywords</b>	[service, agent, device, device service, agent...

Figure: 1: 収集データの一例



# 前回までの問題点

いままでモデル作成において以下問題点があった

- 1 CUDA のメモリ不足
- 2 先行研究モデルへのベクトルサイズが合わない問題

これらは以下の本のコードを参考にすることで解決の見込みが見えた

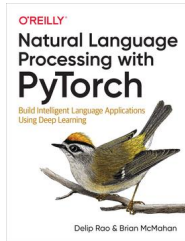


Figure: 2: 参考図書



# 解決策

## メモリ不足

隠れ層や単語埋め込みベクトルの次元数をへらすという手もあるが  
CPU モードで動かすことでとりあえずメモリ溢れは起こらなくなった  
=> 今後ビッグデータ処理するときに GPU でできるように検討する  
必要あり

## ベクトルサイズ不一致問題

こちらは小野田は単語を数値化する際に word2vec というものを使っていたが web 上に公開されている類似ニューラルネットで適用したものはほとんどなかったそのため小野田が下地にした facebook 社のコードでは上手くマッチしなかった  
先週購入した本に実践的な適応方が載っていたのでそちらを土台にニューラルネットを作ることにした