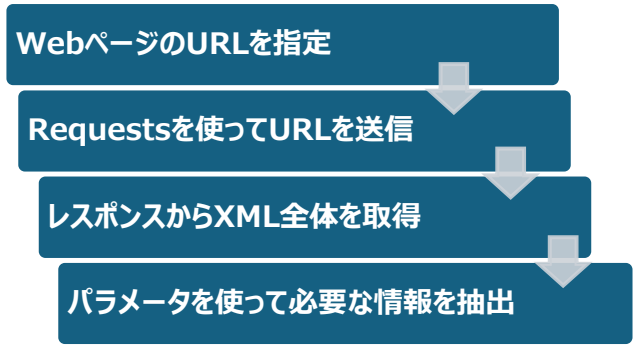
◆Pythonを使った機械学習・ディープラーニングの腐食への適用(5)

概要

- Pythonを用いたWebスクレイピングで、J-Stage内の腐食に関する論文収集を行った。
- 形態素解析によるAI学習モデルを用いることで、論文が分類できることを確認した。

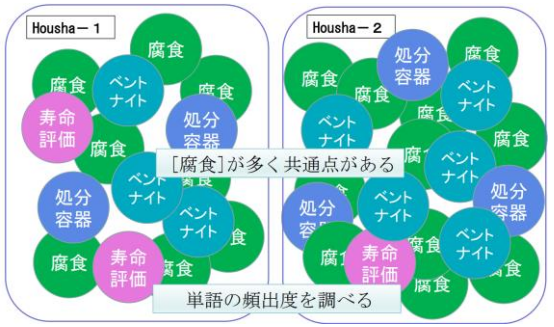


Webスクレイピングのしくみ

テスト用WebページのHTMLの内容

今後の展開

- 腐食研究の情報収集が効率的に行える。



AI（人工知能）を使った解析

```
import MeCab    # 形態素解析エンジン

tagger = MeCab.Tagger()                  # 形態素解析をするための辞書を指定
tagger_text = tagger.parse(text)         # 指定した辞書を使用して形態素解析
tagger_texts = tagger_text.split('\n')[:-2] #リストの後ろ2つを消す
#リスト形式に成形
results = []
for txt in tagger_texts:
    parts = txt.split(',')
    print(parts)
    surface_pos, pos1, base = parts[0], parts[1], parts[-3]
    surface, pos = surface_pos.split('\t')
    results.append(dict(表層形 = surface, 基本形 = base, 品詞 = pos, 品詞1 = pos1))
word_freq = defaultdict(int)    # 単語の出現頻度を調べる
for result in results:
    if result['品詞'] != '記号': # 記号以外を取得
        word_freq[result['基本形']] += 1
print(dict(word_freq))
```

MeCabを使った形態素解析