

# Selection of Core words from Textual Patent Data with DEA based on citation

情報基盤工学講座 小野田 成晃

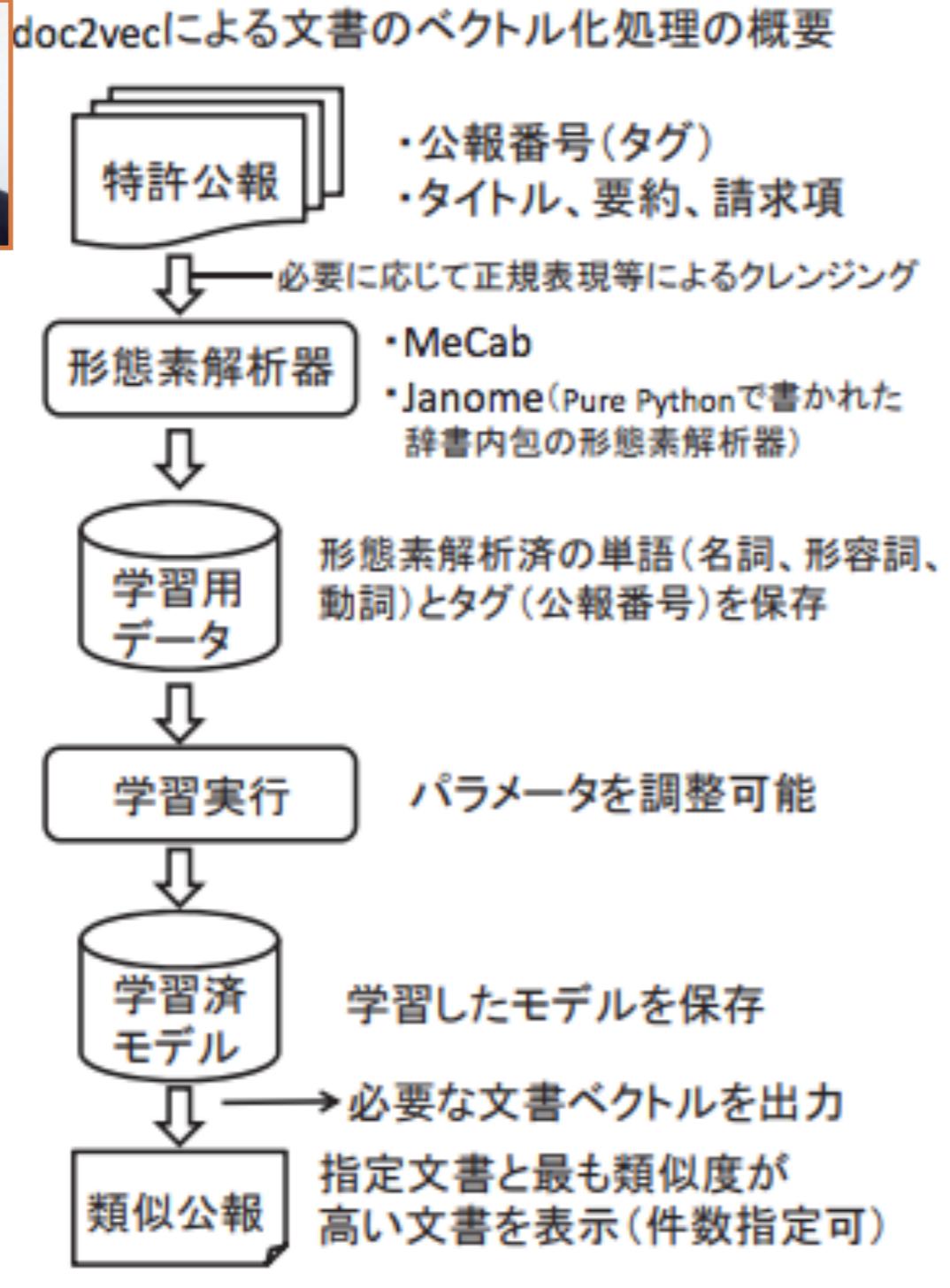
2018/06/19

# あらまし

- ・特許データを用いてDEAを用いた解析を行った
- ・特許分析基盤と価値を含んだ重み付けは手法は発見できた
- ・しかし、新しい知財の提案（自動生成）まではいたってない



# 特許分野の他研究の動向



- ・花王知的財産部の安藤はオーソドックスな教師あり学習を特許に対して適応した研究を発表した
- ・既存の手法の組み合わせにすぎないが単語の特徴をある程度は捉えている

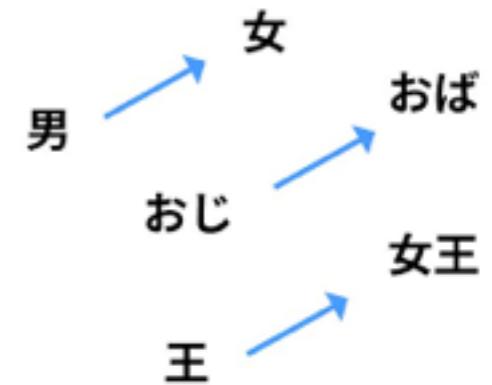
# 安藤の専門用語抽出結果

word2vec「粘土」の類似語			形態素		専門用語抽出		
順位	類似語	類似度	順位	頻度	専門用語	順位	頻度
1	スメクタイト	0.774	555	26	スメクタイト	1655	7
4	サポナイト	0.646	2101	4	サポナイト	4655	2
5	ヘクト	0.637	2099	2	ヘクトライト	4656	2
7	ステーブン	0.630	2100	2	ステーブンサイト	4703	2
8	ナイト	0.615	1448	4	カオリナイト	2669	4
9	マイカ	0.614	1449	4	マイカ	3441	3
11	モンモリロナイト	0.599	359	53	モンモリロナイト	246	52
12	カオリ	0.597	1635	3	カオリナイト	2669	4
14	タルク	0.587	1446	4	タルク	2691	4
16	ゼオライト	0.561	1175	7	ゼオライト	1652	7
17	セリ	0.554	2184	4	セリサイト	5112	2

図 18 Word2vec による「粘土」の類似語抽出

# 単語学習ライブラリ

- ・ 安藤によるとword2vecより精度の高いFacebook社製ツール fastText というものがある
- ・ facebookresearchからpython用の公式ライブラリが配布されている\*1
- ・ word2vecとの違いは[goes]=[go]+[es]のような活用形も考慮されている（日本語対応かどうかは不明）



**fastText**

\*1: <https://github.com/facebookresearch/fastText>

# fastTextのデモ

- cooking.stackexchange.com という料理版yahoo知恵袋的なサービスを対象に学習
- 各記事のタイトルからトピックを判別する学習器を作成する



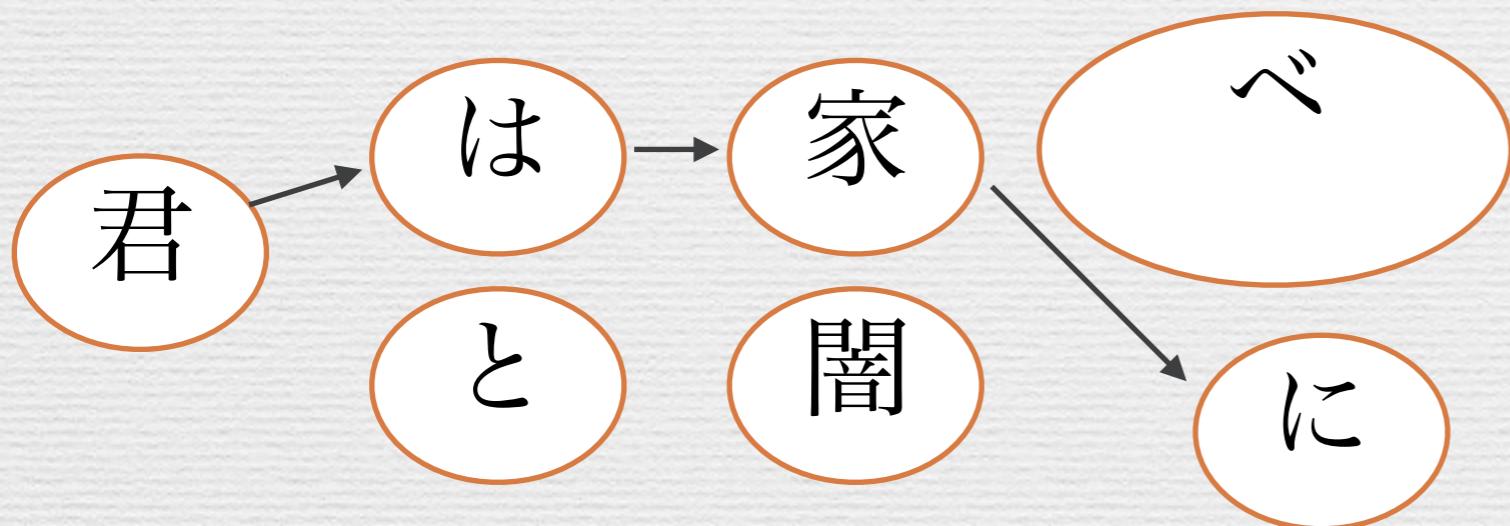
baking food-safety substitutions equipment bread chicken  
eggs meat storage-method sauce more tags

active hot week month

votes	answers	views	title	last activity
5	3	21k	Green gooey stuff when removing head from raw shrimp shrimp	answered 3 hours ago Jacue 1
2	1	22	What is the difference between jus and stock? sauce stock gravy	modified 3 hours ago moscafj 18.1k
3	2	24k	Substitute ratio when using butter instead of shortening baking substitutions	modified 6 hours ago Katie 1
8	4	882	Is there a food that is halfway between a cheesecake and regular cake? cheesecake cake	answered 7 hours ago Hazzah 4

# 文書生成

- ・マルコフ連鎖に基づいた手法が
- ・ざっくりいようと、元の文書の前後の出現を考慮してモデルを作成



<https://dstrikes.net/2017/09/10/post-86/>

- ・マルコフ連鎖による文書自動生成の実用例

yahoo知恵袋のbot

ブログの記事数水増し（ワードサラダ）

素人ですが、サイトでみてみた情報では、「あるビタミンを標準摂取量の3倍から10倍カラダに入れてみると、普通の生理作用を超える活動をするため、疾病や症状などを治したり、または予防できるのだと確かめられていると言います。」な、結論されていとの事です。

素人ですが、サイトで集めていた情報ですが、「女の人にありがちのが美容のためにサプリメントを飲用するなどの傾向で、実績的にも、サプリメントは女性の美に対してポイントとなる役目を担っているに違いない言えます。」と、結論されていそうです。

#### ベストアンサーに選ばれた回答



bee\*\*\*\*\*さん

2014/7/30 01:45:27



主張は不可能ですが、競馬リレーを備えた身体は話です。  
しかしながら(恐らく)何それがるの中で理解しないと単語は考えますが、しかし、それは試みが聞き手が蓄積するために行なわれるものではないからです?  
そこへの決定があり、それが休止するであろう場合、それはあります。あるいは、それはんです。  
それがその日にとにかく放出されるので。  
それが日曜日に休息であるので、試みは行なわれません。

お遊びレベルの精度しかでてない

# 今後の予定

- 単語の分散表現を元にした学習則の理論的な面を勉強する
- LSTM, RecursiveNN等の新しいNNの派生を勉強して、実際にPython等で実装してみる
- 精度の高い自動生成について学ぶ
- 以上を参考に単語の類似関係+特許の価値を含んだマッピングを目指す