

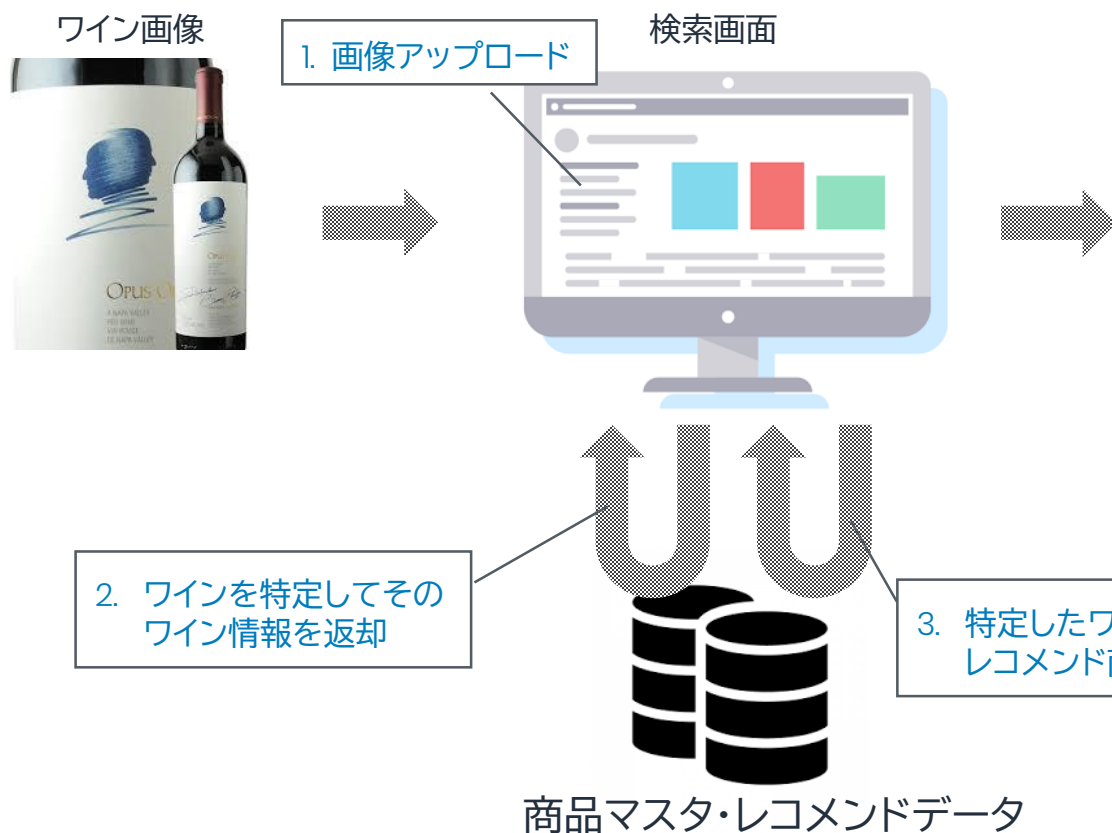
補足資料

2020年10月31日

全体像

- ワイン検索&レコメンデーションシステム

1. アップロードしたワイン画像の種類を特定
2. ワインを特定してそのワイン情報をDBから返却
3. 特定したワインに対するレコメンド商品を返却
4. 返却された情報をブラウザへ表示
5. そのワインと関連性の高いワインをレコメンド



※画面イメージ

4. ワイン情報を表示



5. 関連性の高いワインをレコメンド

実際の画面

■検索画面(画像を選択して「検索する」ボタンを押下)

ワイン検索&レコメンデーションシステム

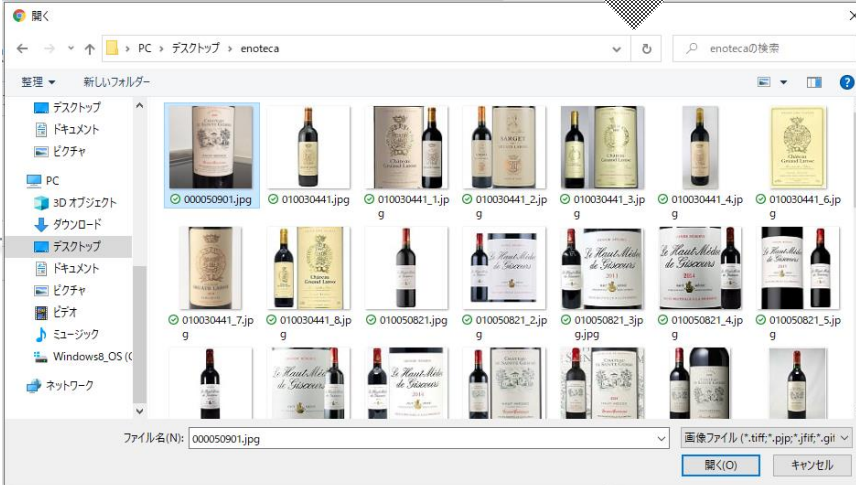
画像ファイルを選択して検索ボタンをクリックしてください。

Browse

検索する

検索したいキーワードを入力してください。


検索する



■結果画面(詳細は次ページ)

ワイン検索&レコメンデーションシステム







アップロード画像



商品番号	010050901
推定確率	98%
名称	CH.DE SAINTE GEMME (シャトー・ド・サン・ジェム)
種類	ワイン / 赤ワイン
価格	2000円
生産国	フランス
品種	
詳細	シャトー・ラネッサンで知られるブーティエ家が所有するシャトー。複雑味に満ちた仕上がり、付けシャトーに匹敵する傑出したワイン。著名なワイン評論家ロバート・パーカー氏が絶賛する、格付けシャトーに近い実力を誇るシャトー。熟した果実味とほのかなスパイスの風味が重なり合い、熟成を経たことで複雑味に溢れるスタイルに変化しています。まさに今飲み頃を迎えており、親しみやすい柔らかい味わいをお楽しみいただけます。

ワイン検索画面へ戻る

この商品を検索した人へおすすめのワイン

 4,500円 (税別)	 3,500円 (税別)	 13,500円 (税別)	 3,000円 (税別)	 11,000円 (税別)	 12,000円 (税別)
---	---	--	---	--	--

多くの著名なシャトーに隣接する、シャトー・パデスクローのセカンドラベル。メルロ由来の柔らかな優しい味わいのお値打ちボトル。

あのロバート・パーカー氏のお気に入りシャトー。格付けシャトーに匹敵するクオリティを備えた、お値打ちボトルワイン。

サン・ジュリアンを代表するシャトーのひとつ、レオフィル・パレトンの兄弟シャトー。伝統を守り続けるクラシックなスタイルが魅力。

ホルダー右隣のブティックワイナリーが仕立てる、有名シャトーに劣らない上質な赤ワイン。滋味深い果実味と複雑味溢れる味わい。

ロバート・パーカー氏が最高のジュヴレ・シャンベルタンの選り手と評する生産者。知る人ぞ知る銘醸店。価格で気品溢れる仕上がり。

変化した品質とクラシックなスタイルが魅力の有力シャトー。熟した果実の味わいとしっとりとした骨格を備える正統派ボトル。

結果詳細

類似の説明文の商品が表示されているかの確認

■検索したワインの情報

ワイン検索&レコメンデーションシステム

アップロード画像

商品番号	010050901
推定確率	98%
名称	CH.DE SAINTE GEMME (シャトー・ド・サン・ジェム)
種類	ワイン / 赤ワイン
価格	2000円
生産国	フランス
品種	
詳細	シャトー・ラネッサンで知られるブーティエ家が所有するシャトー。複雑味に満ちた仕上がりの、付けシャトーに匹敵する傑出したワイン。著名なワイン評論家ロバート・パーカー氏が絶賛する、格付けシャトーに近い実力を誇るシャトー。熟した果実味とほのかなスパイスの風味が重なり合い、熟成を経たことで複雑味に溢れるスタイルに変化しています。まさに今飲み頃を迎えており、親しみやすい柔らかな味わいをお楽しみいただけます。

ワイン検索画面へ戻る

■レコメンド商品の表示

この商品を検索した人へおすすめのワイン

赤ワイン

赤ワイン

赤ワイン

赤ワイン

赤ワイン

赤ワイン

多くの著名なシャトーに隣接する、シャトー・ペデスクローのセカンドラベル。メルロ由来の柔らかく優しい味わいのお値打ちボルドー。

あのロバート・パーカー氏のお気に入りシャトー。格付けシャトーに匹敵するクオリティを備えた、お値打ちボルドーワイン。

サン・ジュリアンを代表するシャトーのひとつ、レオヴィール・ノリトンの兄弟シャトー。伝統を守り続けるクラシックなスタイルが魅力。

ボルドー右岸のプティックワイナリーが仕立てる、有名シャトーに劣らない上質な赤ワイン。滋味深い果実味と複雑味溢れる味わい。

ロバート・パーカー氏が最高のジュヴレ・シャンベルタンの造り手と評する生産者。知る人ぞ知る銘醸畑。繊細で気品溢れる仕上がり。

安定した品質とクラシックなスタイルが魅力の実力派シャトー。熟した果実の味わいとしっかりとした骨格を備える正統派ボルドー。

4,500円 (税抜)

3,500円 (税抜)

13,500円 (税抜)

3,000円 (税抜)

11,000円 (税抜)

12,000円 (税抜)

※別途、検索・閲覧した商品と同価格帯の商品に絞って表示するなどの対応を検討する必要あり。

※説明文は上記に表示されている文以外にも学習に使用しているため、上記の赤枠で囲った箇所以外にも関連のあるキーワードは存在しています。

学習できているかの確認①

任意のキーワードと距離の近いワインを検索し、類似したワインが表示されるか(学習できているか)を確認

検索したいキーワードを入力してください。

オーパスワン

検索

検索したキーワードと関連したワイン



シャトー・ムートン・ロスチャイルド、オーパス・ワンと同じコンセプトで造られる、チリ最高のプレミアムワイン。

144,000円 (税抜)



オーパス・ワンを超える評価を獲得した、モンテス社渾身のプレミアム・ワイン。継ぎ目のないシルキーなタンニンとエレガントな風味。

12,000円 (税抜)



フィリップ・ド・ロスチャイルド男爵とロバート・モンダヴィ氏のコラボレーションで生まれた、カリフォルニアワインのトップ。

360,000円 (税抜)



オーパス・ワンの元醸造家と「獺祭」の旭酒造による、日本酒酵母で仕立てた世界初の白ワイン。ほのかな吟醸香とふくよかな旨味が魅力。

2,980円 (税抜)



オーパス・ワンの元醸造リーダーが手掛けるワイナリー。海に近く冷涼な気候から生み出される突き抜けるようなパワーとゴージャスな味わいが魅力。

57,000円 (税抜)



シャトー・ムートン・ロスチャイルド、オーパス・ワンと同じコンセプトで造られるチリ最高峰のプレミアムワイン。力強くフィネスに溢れるスタイル。

22,000円 (税抜)

学習できているかの確認②

任意のキーワードと距離の近いワインを検索し、類似したワインが表示されるか(学習できているか)を確認

検索したいキーワードを入力してください。

爽やか

検索

検索したキーワードと関連したワイン



スペインワインを牽引する名生産者が手掛ける、世界中で親しまれている人気シリーズ。爽やかでフルーティな、何度でも飲みたくなる白ワイン。

1,600円 (税抜)



スペインワインを牽引する名生産者が手掛ける、世界中で親しまれている人気シリーズ。爽やかでフルーティな、何度でも飲みたくなる白ワイン。

1,600円 (税抜)



ポルトガルベストワインメーカー賞受賞の実力派。3種の伝統品種をブレンドして造る白。ブドウ本来の味わいを活かした爽やかでエレガントなスタイル。

1,850円 (税抜)



イングルヌックを再建した、フランシス・コッポラがプライドをかけて磨いた、品種の個性を堪能するシリーズ。魅惑的な丸みある輪郭を備えたワイン。

3,980円 (税抜)



世界遺産チンクエ・テッレの断崖の畑の最高のブドウから生み出される逸品。エレガントな香りとミネラル感あふれるフレッシュな味わいが魅力。

3,640円 (税抜)



チリ最高峰のワイナリーが手掛ける、世界で愛される天使のワイン待望の新作。フレッシュで爽やかな果実味を最大限に表現したソーヴィニヨン・ブラン。

2,000円 (税抜)

自然言語処理(前処理)

■説明文(味わい)

グラスに注ぐと紫がかった濃いルビーの色合い。熟したプラムやブラックベリー、カシスなどの赤系、黒系果実にイチジクのアロマ。ほのかにブラック・ペッパーやフレンチオーク由来のタバコやバタースコッチのニュアンスも重なり、チリワイン独特の複雑かつ官能的な香りと見事に調和しています。口を含むと芳醇な果実味と滑らかなタンニンとのバランスが良く、エレガントな酸味を備えた見事な仕上がりは余韻まで心地よく続きます。このコクのあるカベルネ・ソーヴィニヨンには、スパイスをきかせて焼き上げた赤身肉のステーキや、濃厚な味わいのスパゲティ・ボロネーズやラムチョップなどが好相性。熟成のポテンシャルもありますが、今からでも十分お楽しみいただける1本です。

■前処理

```
def __preprocessing(self, text):
    """テキスト前処理"""
    # urlを除去
    text = re.sub(r'https?://[w/:%#\$\&?\(\)\~\.\=\+\-]+', '', text)

    # 改行、スペースを削除
    text = re.sub('\n', '', text)
    text = re.sub(' ', '', text)
    text = re.sub(' ', '', text)

    # 記号削除
    del_symbols = re.compile(
        '[!"#%&'\\"\\\\()*+,-./:;<=>?@[\\]^_`{|}~「」□□""◇『』□&*・○$#<>@.、?!`´+¥%~\']')
    text = del_symbols.sub('', text)

    # 数文字削除
    text = re.sub(r'[0-9 0-9]', '', text)

    return text

def __split_into_words(self, doc, name=''):
    """形態素解析"""
    words = self.__del_vocab(self.tokenizer(doc))
    return TaggedDocument(words=words, tags=[name])

def __del_vocab(self, words):
    """特定の文字列を削除"""
    del_list = ['1本', 'ワイン', '味', '味わい', '果実', '年', 'こちら', 'さ', ]
    words = [w for w in words if w not in del_list]
    return words
```

■形態素解析(Janome) ※「名詞」と「形容詞」のみを学習に使用

【名詞】

['グラス', '紫', 'ルビー', '色合い', '熟', 'プラム', 'ブラックベリー', 'カシス', '赤', '系', '黒', '系', '果実', 'イチジク', 'アロマ', 'ほのか', 'ブラックペッパー', 'フレンチ', 'オーク', '由来', 'タバコ', 'バター', 'スコッチ', 'ニュアンス', 'チリ', 'ワイン', '独特', '複雑', '官能', '的', '香り', '見事', '調和', '口', '芳醇', '果実', '味', '滑らか', 'タンニン', 'バランス', 'エレガント', '酸味', '見事', '仕上がり', '余韻', 'コク', 'カベルネソーヴィニヨン', 'スパイス', '赤身', '肉', 'ステーキ', '濃厚', '味わい', 'スパゲティ', 'ボロネーズ', 'ラム', 'チョップ', '相性', '熟成', 'ポテンシャル', '今', '十分', 'み']

【形容詞】

['濃い', '良く', '心地よく', '愉し']

【動詞】

['注ぐ', 'がかつ', 'し', '重なり', 'し', 'い', '含む', '備え', '続き', 'ある', 'きかせ', '焼き上げ', 'あり', 'いただける']

【助詞】

['に', 'と', 'の', 'や', 'など', 'の', 'に', 'の', 'に', 'や', 'の', 'や', 'の', 'も', 'の', 'と', 'に', 'て', 'に', 'と', 'と', 'と', 'の', 'が', 'を', 'は', 'まで', 'の', 'に', 'は', 'を', 'て', 'の', 'や', 'の', 'や', 'など', 'が', 'の', 'も', 'が', 'から', 'でも']

【助動詞】

['た', 'た', 'な', 'ます', 'な', 'な', 'な', 'た', 'な', 'ます', 'た', 'な', 'ます', 'です']

自然言語処理(学習)

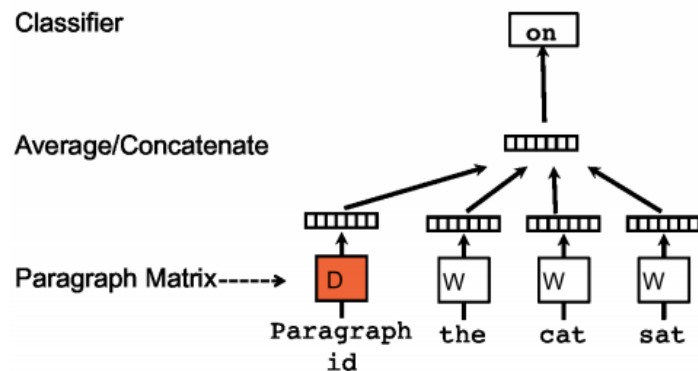
■学習時のポイント

- windowサイズは小さめ

※今回のケースでは5以上にすると意図した結果が得られなかった

- Epoch数は1000と多めに設定

- dm=1 (PV-DM)を使用



```
def __train(self, epochs):
    """学習用"""
    logger.info('Reading files and text preprocessing start...')
    try:
        corpus = self.__read_files(character_limit=1000)
        sentences = list(self.__corpus_to_sentences(corpus))
    except Exception as e:
        logger.error('Reading files and text preprocessing error. {}'.format(e))
        sys.exit(1)
    else:
        logger.info('Reading files and text preprocessing ended normally.')

# 学習
logger.info('Doc2Vec model training start...')
try:
    self.__model = models.Doc2Vec(sentences, dm=1, window=5, alpha=0.03, epochs=1000,
                                   min_alpha=0.05, min_count=10, sample=1e-5, negative=5, workers=1)
    for epoch in range(epochs):
        logger.info('-----')
        logger.info('Epoch: {}/{} start...'.format(epoch+1, epochs))
        self.__model.train(sentences,
                            total_examples=sum([len(s) for s in sentences]),
                            epochs=self.__model.iter)

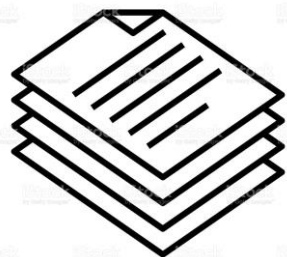
        # テストスコア表示
        self.__test_proc()
except Exception as e:
    logger.error('Doc2Vec model training error. {}'.format(e))
    sys.exit(1)
else:
    logger.info('Doc2Vec model training ended normally.')

# 学習結果保存
logger.info('Model save process start...')
try:
    self.__model.save(self.model_path)
except Exception as e:
    logger.error('Model save process error. {}'.format(e))
else:
    logger.info('Model save process ended normally.')
```


自然言語処理(学習度合の確認)

コサイン類似度で学習の進み具合を評価

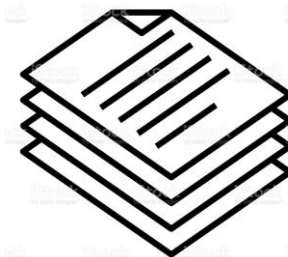
N エポック目



学習



N+1 エポック目



```
array([ 1.3845743e-04,  2.2527881e-03,  2.1567733e-03,  1.6176930e-03,
        -3.9818049e-03, -3.3417728e-03,  4.6719741e-03, -5.5631623e-04,
        -3.6124536e-03,  1.6725663e-03,  1.6517657e-03,  1.5800855e-03,
        -1.4602476e-04,  2.9591618e-03, -2.1307389e-03, -1.0844590e-03,
        3.4925586e-03,  2.2113184e-03, -3.0602498e-03, -9.8337812e-05,
        -2.6016932e-03, -1.0951535e-04,  4.7823223e-03, -1.7470532e-03,
        -3.0310219e-04, -3.3226029e-03, -3.0137205e-03,  4.8629693e-03,
        -3.7409852e-03,  3.4510186e-03, -4.4884216e-03, -4.0413721e-05,
        -1.1230287e-03, -1.5799218e-03, -3.6428488e-04, -1.6990561e-03,
        1.2628553e-03, -1.2385391e-03,  3.1174365e-03,  4.1266554e-03,
        4.8873252e-03,  4.1540726e-03, -4.7755805e-03, -2.8908814e-03,
        -3.3284728e-03, -3.0358674e-03, -1.1272541e-04,  2.8342942e-03,
        2.2518486e-03, -1.5560046e-03, -3.3887075e-03, -2.0355131e-03,
        4.3546520e-03, -3.0600666e-03,  1.4264594e-03, -6.7142624e-04,
        -4.8649642e-03,  2.2846726e-03,  7.6802634e-04, -1.7428732e-03,
        4.5045703e-03,  4.5922468e-03, -4.9618906e-03,  3.4423330e-04,
        -2.5883720e-03, -3.9950218e-03,  2.1540236e-03,  1.4958148e-03,
        -2.6098392e-03, -4.7335816e-03,  1.7470537e-03,  2.0343617e-03,
        -2.8249219e-03, -4.6004552e-05,  1.2904147e-03, -1.7202720e-03,
        1.5653541e-03,  8.9911051e-04, -1.9822754e-03, -1.8130733e-03,
        2.6344298e-03,  2.3862957e-03,  4.9604508e-03,  9.5812022e-04,
        -4.4024349e-03,  2.6171419e-03,  4.6850555e-03,  3.7352871e-03,
        4.9331336e-04, -7.1522308e-04,  2.2784085e-03, -2.0288767e-03,
        -4.9580941e-03,  2.3133697e-03,  4.8894892e-03, -1.4455436e-03,
        4.2582802e-03, -3.7040468e-03,  1.3155816e-03, -1.5798656e-04],
        dtype=float32)
```

$$\text{Similarity} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

コサイン類似度

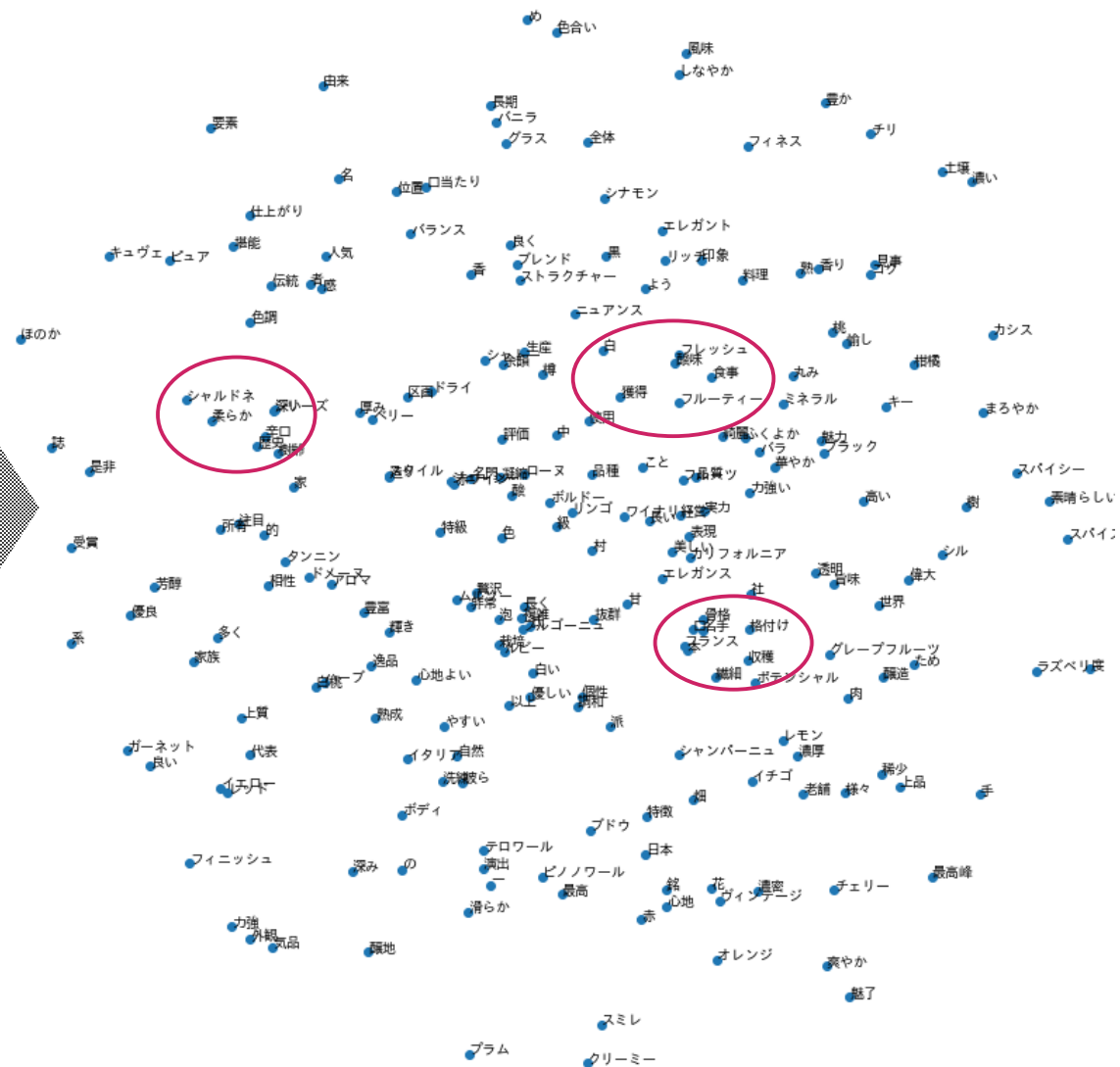
1に近づくほど収束していると解釈できる

```
array([ 1.3845743e-04,  2.2527881e-03,  2.1567733e-03,  1.6176930e-03,
        -3.9818049e-03, -3.3417728e-03,  4.6719741e-03, -5.5631623e-04,
        -3.6124536e-03,  1.6725663e-03,  1.6517657e-03,  1.5800855e-03,
        -1.4602476e-04,  2.9591618e-03, -2.1307389e-03, -1.0844590e-03,
        3.4925586e-03,  2.2113184e-03, -3.0602498e-03, -9.8337812e-05,
        -2.6016932e-03, -1.0951535e-04,  4.7823223e-03, -1.7470532e-03,
        -3.0310219e-04, -3.3226029e-03, -3.0137205e-03,  4.8629693e-03,
        -3.7409852e-03,  3.4510186e-03, -4.4884216e-03, -4.0413721e-05,
        -1.1230287e-03, -1.5799218e-03, -3.6428488e-04, -1.6990561e-03,
        1.2628553e-03, -1.2385391e-03,  3.1174365e-03,  4.1266554e-03,
        4.8873252e-03,  4.1540726e-03, -4.7755805e-03, -2.8908814e-03,
        -3.3284728e-03, -3.0358674e-03, -1.1272541e-04,  2.8342942e-03,
        2.2518486e-03, -1.5560046e-03, -3.3887075e-03, -2.0355131e-03,
        4.3546520e-03, -3.0600666e-03,  1.4264594e-03, -6.7142624e-04,
        -4.8649642e-03,  2.2846726e-03,  7.6802634e-04, -1.7428732e-03,
        4.5045703e-03,  4.5922468e-03, -4.9618906e-03,  3.4423330e-04,
        -2.5883720e-03, -3.9950218e-03,  2.1540236e-03,  1.4958148e-03,
        -2.6098392e-03, -4.7335816e-03,  1.7470537e-03,  2.0343617e-03,
        -2.8249219e-03, -4.6004552e-05,  1.2904147e-03, -1.7202720e-03,
        1.5653541e-03,  8.9911051e-04, -1.9822754e-03, -1.8130733e-03,
        2.6344298e-03,  2.3862957e-03,  4.9604508e-03,  9.5812022e-04,
        -4.4024349e-03,  2.6171419e-03,  4.6850555e-03,  3.7352871e-03,
        4.9331336e-04, -7.1522308e-04,  2.2784085e-03, -2.0288767e-03,
        -4.9580941e-03,  2.3133697e-03,  4.8894892e-03, -1.4455436e-03,
        4.2582802e-03, -3.7040468e-03,  1.3155816e-03, -1.5798656e-04],
        dtype=float32)
```

自然言語処理(学習結果を一部可視化)

T-SNEで可視化

※単語を削減して拡大表示(頻度>200)



今後の課題

■単純な類似だけでは**ユーザの嗜好と異なる可能性がある**ため下記を検討する必要あり

- ・協調フィルタリングと組み合わせて精度を担保する
- ・協調フィルタリングと今回のようなディープラーニング系のレコメンドエリアを目的別に分けて(複数枠)表示する

■下記のような**評価指標を検討**する必要あり

1. Serendipity: 未知の推薦対象に対するポジティブな驚きの感情の指標
2. Novelty: ユーザーにおける推薦対象の未知割合の指標
3. Diversity: 推薦対象の多様性の指標
4. Coverage: 推薦対象/全アイテムの指標

【参考論文】

PURS: Personalized Unexpected Recommender System for Improving User Satisfaction
<https://dl.acm.org/doi/abs/10.1145/3383313.3412238>



End