

14. 自然言語処理と AI

(コンピューターサイエンス)

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

金子邦彦



謝辞：この資料では「いらすとや」のイラストを使用している



アウトライン

1. イントロダクション
2. 自然言語処理と人工知能
3. 単語ベクトル
4. チャットボットの概要
5. チャットボットのベストプラクティス

対話型 AI の機能



人間からの質問・指示



意図を理解



回答を生成


対話の継続

**ユーザは、対話を通じてより良い回答が得られるように
AIに働きかける**

最初の会話

人間

そうめんとうどんの違いを、一言で教えて

 Copilot

そうめんは細くて冷やして食べることが多く、うどんは太くて温かく食べることが多い麺です。

AI

会話の継続

人間

どちらがおすすめ？

Copilot

おすすめは、**季節や気分によって変わります！**

- 🌞 **暑い日**には、冷たくてさっぱりした「そうめん」がおすすめ。
- ❄️ **寒い日**には、温かくてもちもちの「うどん」がぴったり。

どんなシーンで食べたいか教えてくれたら、もっとぴったりのおすすめができますよ！

AI

対話型 AI を理解する



- ChatGPT や Google Gemini や Microsoft 365 Copilot などの対話型AI（人間と自然な会話ができるAI）が広く利用
- 対話型AIは、5つの働きを統合して実現
 1. 見る・聞く・感じる力（知覚と理解）
 2. 覚える・調べる力（知識処理と記憶）
 3. 考える・推理する力（論理推論と数学的思考）
 4. 作る・創作する力（生成と創造）
 5. 計画する・実行する力（計画と実行）

対話 AI で何ができる？ & 超重要注意点

できることの例

- 情報提供、質問応答、文章作成支援、要約、翻訳、プログラミング支援など。
- 文字、画像、音声でのコミュニケーション

超重要！使う上での注意点

- **鵜呑みにしない！** 必ず自分で考えて、試して、理解する（AIは古い情報、間違った情報を提供する可能性がある）
- **丸写しはダメ！** AIが作ったものをそのまま自分の作品として提出するのは禁止
- **個人情報・秘密はAIに与えない！**

Microsoft 365 Copilot 始め方ガイド (大学アカウント利用)

① アクセス方法

パソコンのWebブラウザで以下のURLにアクセス

<https://m365.cloud.microsoft/chat/?auth=2>

② ログイン

大学のIDとパスワード

ログイン後、**チャット画面**が出るので確認

何かお手伝いできることはありますか?

Copilot へメッセージを送る



Microsoft 365 Copilot 始め方ガイド（大学アカウント利用）

チャット画面の基本操作

ファイルアップロード

画面下の「+」ボタンをクリック

→ Word、Excel、PDF、画像などをアップロード可能

マイク入力

マイクが接続されている場合

→ マイクボタンから音声入力が可能



+ ボタン

マイクボタン

テキスト入力

質問や指示をテキストで入力 → Copilotが応答



14-1. 自然言語処理と 人工知能

自然言語処理と人工知能

自然言語処理は、人間が普段使う言語（日本語、英語など）を**コンピュータが理解**したり、**生成**したりする技術

- AI による**人間の言語の理解**
- AI と人間の**自然な対話**
- **言語間の翻訳**
- **文書類の分析と活用**

Microsoft 365 Copilot を用いたインターネット情報リサーチ

- キーワードや文章による検索が可能

例：「サステナビリティ」の Microsoft 365 Copilot によるリサーチ

「サステナビリティ (Sustainability)」は、環境・社会・経済の3つの側面で持続可能な発展を目指す考え方です。日本語では「持続可能性」とも訳されます。

サステナビリティの主な3つの柱：

1. 環境的側面

- 自然資源の保護（森林、水、空気など）
- 気候変動への対応（CO₂削減、再生可能エネルギーの活用）
- 生物多様性の保全

2. 社会的側面

- 公平な教育や医療へのアクセス
- ジェンダー平等や人権の尊重
- 地域社会の活性化と文化の継承

3. 経済的側面

- 長期的な経済成長の確保
- 責任ある消費と生産（SDGsの目標12）
- グリーン経済や循環型経済の推進

関連するキーワード：

Microsoft 365 Copilot を用いたインタ...

- SDGs（持続可能な開発目標）
- ESG投資（環境・社会・ガバナンスを重視する投資）
- カーボンニュートラル
- サーキュラーエコノミー（循環型経済）

もし、特定の分野（例えば「企業のサステナビリティ戦略」や「教育におけるサステナビリティ」など）について詳しく知りたい場合は、教えてください。どんな切り口で深掘りしましょうか？

Microsoft 365 Copilot を用いた要約

- AIからの回答、人間が与えたファイルなどの**要約**が可能

例：Microsoft 365 Copilot に「**日本国憲法の全文を確認して。それから、400文字以内に要約して。**」

日本国憲法の全文は、以下の公式サイトで確認できます：

 [衆議院公式サイト：日本国憲法全文](#) ①

日本国憲法の400字以内の要約：

日本国憲法は、国民主権、基本的人権の尊重、平和主義を柱とする日本の最高法規です。天皇は象徴であり、国政に関与しません。戦争の放棄を明記し、戦力と交戦権を否認しています。国民には自由と権利が保障され、公共の福祉のためにそれを行行使する責任があります。国会、内閣、司法などの統治機構の役割が定められ、憲法改正には厳格な手続きが必要です。

特定の条文や章について詳しく知りたい場合は、お知らせください！ |

Microsoft 365 Copilot を用いた推敲

- 人間が与えた文章やファイルなどの**推敲**が可能

例：Microsoft 365 Copilot に「私は今日は朝ご飯は卵で昼ご飯はお弁当で途中でお茶を買って、そのあとでお茶をもう1本買った**という文章を推敲して。**」

✦ 推敲例 1（自然で簡潔）

「今日は朝ご飯に卵を食べ、昼はお弁当を食べました。途中でお茶を買い、そのあともう1本買いました。」

✦ 推敲例 2（少し丁寧な表現）

「今日は朝食に卵を食べ、昼食はお弁当でした。外出中にお茶を1本買い、その後、もう1本追加で購入しました。」

✦ 推敲例 3（カジュアルな口語調）

「今日は朝ごはんに卵、昼はお弁当。途中でお茶を買って、さらにもう1本買った。」

どの雰囲気が好みですか？あるいは、もっとフォーマルにしたい、日記風にしたいなど、ご希望があればそれに合わせて調整します！

プログラミング支援 (Windsurf を使用)

The screenshot displays the Windsurf IDE interface. On the left, a code editor shows Python code for calculating vector directions. Lines 193 and 194 are highlighted with a grey selection box. The code is as follows:

```
188 # 矢印の方向を計算
189 dx = x - cx
190 dy = y - cy
191 norm = np.sqrt(dx**2 + dy**2)
192 if norm > 0:
193     dx = dx / norm * 30
194     dy = dy / norm * 30
195     cv2.arrowedline(result, (cx, cy), (cx+dx, cy+dy), cv2_AA_8BGR)
```

Below the code editor is a terminal window showing the command prompt: `PS C:\Users\user\Documents>`. On the right side, an AI chat interface is visible. A message from the user asks: "プログラムのこの部分はなぜ必要なの" (Why is this part of the program necessary?). The AI's response is: "Let me check the code around line 194 in your file to help explain why this part is necessary." Below the chat, a status bar indicates "Analyzed a.py#L174-214".

範囲を選んでいる

AI に質問、相談

Windsurf は簡単にインストール、利用可能

翻訳（日本語から英語，英語から日本語）

The screenshot shows the DeepL translator interface. At the top, there are two buttons: "テキスト" (Text) and "ドキュメント" (Document). Below these, the source language is set to "日本語 - 自動検出" (Japanese - Auto-detect) and the target language is "英語" (English). The input text is "白い雲と青い空が美しい" (Shiroi kumo to aoi sora ga utsukushi). The output text is "Beautiful white clouds and blue sky". There are icons for voice input/output, a character count of 11/5000, and a star icon for favorites. A "フィードバックを送信" (Send feedback) link is visible at the bottom right.

The screenshot shows the DeepL translator interface. At the top, there are two buttons: "テキスト" (Text) and "ドキュメント" (Document). Below these, the source language is set to "英語" (English) and the target language is "日本語" (Japanese). The input text is "Once upon a midnight dreary, while I pondered, weak and weary, Over many a quaint and curious volume of forgotten lore, |". The output text is "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,". There are icons for voice input/output, a character count of 120/5000, and a star icon for favorites.

翻訳（ファイルの使用）

ワード、パワーポイントファイルの中から、
文字のデータを取り出すことも当たり前

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>

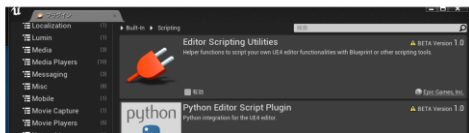
金子邦彦



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

翻訳結果

元のパワーポイントファイル

自然言語処理とAIの応用例

AIとの対話：AIアシスタントとの対話や指示

リサーチ：インターネットからの情報検索（検索エンジンの活用）

要約：長文からのポイント抽出

文書作成・推敲：文章の作成と改善

プログラミング支援：コード作成，デバッグ，最適化

翻訳：言語間の翻訳（例：動画の自動翻訳，字幕付け）



演習 1. AIを用いたり サーチ

情報検索の具体的な手順

1. トピックとキーワードの設定

- トピック例：**脱炭素**
- キーワード例：**再生可能エネルギー，温室効果**

2. Microsoft 365 Copilot を用いたインターネット情報リサーチ

<https://m365.cloud.microsoft/chat/?auth=2>

3. 情報の評価

- 結果の信頼性の確認
- 当初の目的に沿うものか？ 新たな発見があるか？

AI アシスタントの活用



AIアシスタントの主な用途

- 音声による家庭内装置の操作
- スマートフォンの音声コマンド
- スマートスピーカーの操作
- 日常的なタスク支援

マルチモーダルAI（人間の言葉、画像の連携）

- ・ 人間の言葉と画像の両方を理解
- ・ 画像を言葉で説明、言葉に合う画像を生成することも可能

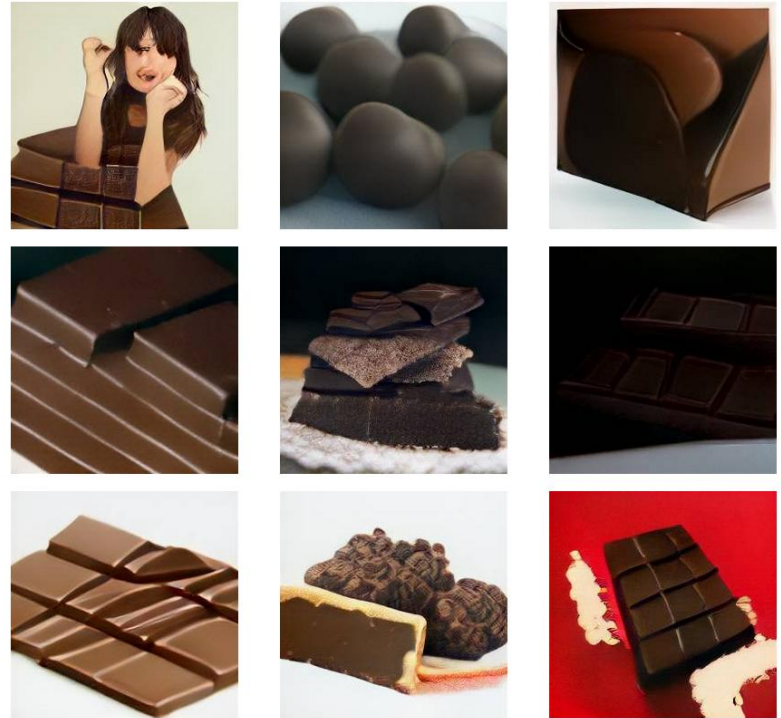
DALL-E mini

Generate images from text

What do you want to see?

I like chocolate

I like chocolate



Again!

AI が生成した画像

DALL-E mini のオンラインデモ

<https://huggingface.co/spaces/dalle-mini/dalle-mini>



14-2. 単語や文章の類似性

語の類似性を扱う際の課題

私の母は**親切だ**

私の母は**優しい**

私の母は**思いやりがある**

これらの単語の類似性を，データに基づいて裏付けることが重要（AI の有効分野の1つ）

単語と形態素解析



- **単語**： 言葉の最小単位 「私」、「は」 など
- **形態素解析**： 文章を単語に自動分割、単語それぞれの品詞（名詞、動詞など）を判定する技術

日本国憲法の第40条を形態素解析してください

対象文：

「何人も、抑留又は拘禁された後、無罪の裁判を受けたときは、法律の定めるところにより、国にその補償を求めることができる。」

表層形	品詞（推定）	意味・役割
何人も	名詞	「誰でも」
、	記号	読点
抑留	名詞	拘束
又は	接続詞	または
拘禁	名詞	拘束
された	動詞	「する」の過去受け身形

単語ベクトルの概要



単語ベクトルは、自然言語処理の基礎技術

- 単語の数値化
- 意味的類似性の計算を可能にする

単語の数値化

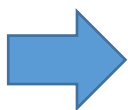


単語を特徴ベクトルに変換。それにより、単語間の類似度を計算可能にする。

```
[ -0.47851562 -0.23242188 -0.14748094 0.06787109 -0.23632812 0.11669922
0.14550781 -0.09765825 0.15136719 0.11865234 -0.30859375 -0.14648438
-0.19726562 -0.40039062 0.17578125 0.20507812 0.28710938 0.10449219
0.08300781 -0.23925781 0.08349609 -0.03857422 0.285625 0.0534068
-0.31445312 0.30664062 -0.25195312 0.34980938 -0.00582888 -0.25195312
-0.02368164 -0.06591797 -0.13574219 0.26953125 -0.06298828 0.05493164
-0.4609375 -0.06787109 0.15429688 0.0546875 -0.01312256 0.06982422
0.08984375 0.21386719 0.07470703 -0.21972656 0.203125 -0.33007812
-0.20410156 -0.30859375 -0.08349609 0.02246094 0.20703125 0.12597656
0.07226562 0.1640625 -0.15039062 -0.0534668 -0.0534668 0.10791016
0.05004883 -0.03100586 -0.17480489 0.13574219 -0.41601562 0.12597656
-0.30273438 -0.04492188 -0.25976562 0.01599121 0.09277344 0.14941406
-0.10058594 -0.0030395 -0.33203125 -0.20996094 0.04980469 0.11035156
-0.10498047 -0.37890625 -0.31445312 0.328125 -0.24804688 -0.140625
-0.08789062 -0.10791016 0.04174905 0.34375 -0.21679688 0.10693359
-0.29882812 0.20507812 -0.25195312 -0.28125 0.22753906 -0.16894531
-0.14160156 -0.07421875 0.6640625 0.14550781 -0.02331543 -0.19042969
-0.01928711 0.19042969 0.13476562 0.20996094 -0.00512695 -0.140625
0.27734375 -0.34765625 -0.2734375 0.13769531 0.08886719 0.09960938
-0.00854492 0.13769531 -0.15917969 -0.75 0.12792969 0.03222656
0.07763672 -0.33203125 -0.36914062 -0.04443359 0.5234375 0.19335938
-0.20800781 0.14355469 -0.30859375 0.14746094 -0.11279297 -0.02392578
-0.13789531 0.0435703 -0.07470703 -0.07617188 0.06225586 0.09277344
0.09375 -0.12109375 0.23925781 -0.24902344 -0.31054688 -0.14453125
-0.12695312 0.0320312 0.07080078 -0.02528955 -0.3046875 -0.38671875
0.19140625 -0.53125 -0.03100586 0.05249023 -0.16308594 -0.20800781
-0.11621094 0.1640625 -0.13789531 0.22070312 -0.3984375 -0.05981445
-0.26367188 -0.29101562 -0.08984375 -0.25 0.16308594 -0.06884766
-0.27539062 0.15039062 0.02197266 0.01696777 -0.36914062 -0.11962891
-0.3984375 -0.0145874 0.515625 -0.55859375 0.12597656 -0.31054688
-0.31839538 0.01104736 -0.07470703 0.06152344 0.24023438 0.10009766
-0.24804688 0.13183594 -0.10205078 -0.19824219 -0.09912109 -0.0402832
-0.65234375 -0.08203125 0.02368164 0.4765625 -0.37890625 0.03100586
-0.11132812 -0.29882812 -0.41796875 -0.17089844 -0.34960938 0.05517578
0.03417969 0.08300781 -0.09082031 -0.421875 -0.11425781 0.24707031
0.08056641 0.28710938 -0.17578125 -0.2265625 -0.38671875 0.05981445
-0.02502441 -0.1640625 0.12158203 -0.16210938 -0.26757812 0.16308594
0.09667969 0.2820312 0.0534668 -0.140625 -0.20605469 0.16210938
-0.18652344 -0.04248047 -0.08447266 -0.16796875 0.04467773 -0.29296875
0.00671387 -0.13378906 -0.04003906 -0.28710938 0.18457031 0.21484375
0.328125 0.15429688 -0.234375 -0.25585938 -0.19433594 -0.00405894
-0.16992188 0.33007812 -0.21679688 -0.01403809 -0.14257812 0.265625
0.20800781 0.18066406 0.25195312 0.00765991 -0.02868547 -0.046875
-0.00811768 -0.0320312 -0.05737305 -0.20117188 0.20703125 -0.02026367
-0.29882812 0.296875 -0.23632812 -0.05737305 -0.05712891 -0.26367188
0.125 0.09423828 -0.16796875 0.2109375 0.09130859 -0.03881836
0.34570312 -0.00436401 -0.28515625 0.13671875 -0.14257812 0.25976562
0.03833008 0.22070312 -0.20800781 -0.02478027 -0.05932617 -0.0320312
0.12353516 0.13476562 0.08642578 0.22460938 -0.25195312 -0.20899438
0.22265625 -0.08991406 -0.359375 -0.07763672 -0.23242188 0.1875 ]
```

単語

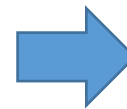
iPhone



数値化

単語

iPad



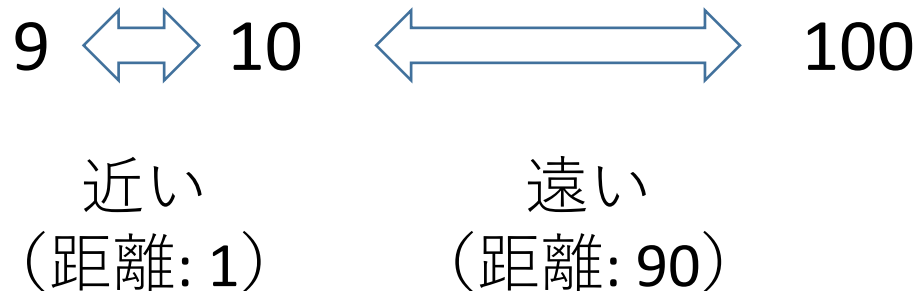
数値化

```
[ -0.01696777 -0.11914062 -0.32226562 -0.02441406 -0.26171875 0.03515625
0.11572266 0.28710938 0.28320312 0.01565396 0.00153351 0.15722656
0.24707031 -0.23535156 0.13671875 0.16796875 0.2890625 0.03735352
0.19042969 -0.13867188 0.00442505 -0.171875 0.22851562 0.0625
-0.11035156 0.26953125 0.16015625 0.31445312 -0.04443359 -0.32421875
-0.17578125 -0.00479126 -0.30664062 0.17871094 -0.02783203 -0.08203125
-0.3046875 0.01062012 0.046875 0.2734375 0.10986328 0.08007812
0.09521484 0.01611328 -0.00872803 -0.03955078 -0.24023438 0.12451172
0.09619141 -0.31445312 -0.01031494 -0.04443359 -0.18945312 -0.03710938
0.07861328 0.19042969 -0.0703125 -0.01599121 0.04956055 0.11962812
0.09570312 0.10351562 -0.33203125 -0.09716797 -0.38671875 0.16210938
-0.20996094 0.11425781 -0.00158691 0.15332031 -0.12695312 -0.12988281
-0.00750732 0.02038574 -0.04663086 -0.20507812 -0.23828125 -0.04614258
0.10595703 -0.26757812 -0.03613281 0.13867188 -0.2578125 0.15136719
-0.19140625 -0.26367188 0.4140625 0.45507812 -0.14257812 -0.15136719
-0.1953125 -0.14160156 -0.22851562 -0.16503906 0.02929688 -0.26953125
0.16699219 0.18066406 0.30664062 0.07910156 0.00387573 -0.30859375
-0.02111816 -0.03857422 0.31835938 0.17292969 0.00805664 -0.03222656
0.38867188 -0.16113281 -0.05053711 0.02197266 0.1328125 0.23046875
0.02941895 -0.02576584 -0.33984375 -0.4609375 -0.05395508 0.02624512
-0.07568359 -0.375 -0.2759062 0.09814453 0.3203125 -0.0255127
-0.11669922 0.07714844 -0.51953125 0.06689453 -0.00631006 0.04307117
-0.23046875 0.08203125 -0.0378418 0.18457031 0.13867188 0.10302734
0.2109375 -0.10839944 0.13476562 0.00933594 -0.06640625 -0.26367188
0.06933594 0.05175781 0.3359375 0.01000977 -0.37695312 -0.3515625
0.0625 -0.3359375 -0.10742188 0.12792969 0.10449219 -0.36914062
-0.27734375 -0.13054688 -0.30078125 0.39648438 -0.3515625 -0.31640625
-0.0534668 -0.16796875 -0.27539062 0.13574219 0.2890625 0.125
0.00653076 0.04833984 -0.04052734 -0.13789531 -0.41796875 0.09912109
-0.29492188 0.27148438 -0.29492188 -0.3828125 -0.25 -0.12890625
-0.3671875 -0.171875 -0.08154297 -0.26171875 0.1953125 0.02148438
-0.18554688 0.04345703 -0.09716797 0.04223633 0.09228516 0.14550781
-0.37109375 -0.0402832 -0.17480469 0.19824219 -0.16503906 0.03540039
-0.06542969 -0.31640625 -0.14160156 0.02600098 -0.27148438 0.11669922
0.06347656 -0.22753906 -0.02502441 -0.16503906 0.13378906 0.08837891
0.05810547 0.32617188 -0.19726562 -0.10595703 -0.09228516 -0.01647949
0.17480469 -0.08984375 -0.1328125 -0.27734375 -0.13671875 0.31445312
-0.05957031 0.08105469 -0.02294922 0.06445312 -0.29882812 0.12207031
0.08007812 -0.00300598 -0.28710938 -0.06889453 0.00604248 -0.2734375
0.12695312 -0.06640625 0.01611328 -0.14160156 0.06640625 0.24121094
0.12255859 -0.00909424 -0.12255859 -0.38523438 -0.203125 -0.22851562
-0.2734375 0.10839944 0.00668335 0.00747681 -0.31835938 0.09033203
0.11816406 0.09765625 0.125 0.03442383 0.02038574 0.12158203
-0.23925781 0.06933594 0.13183594 -0.07373047 0.08251953 0.09667969
-0.13769531 0.23535156 -0.20996094 -0.2890625 -0.04638672 -0.15620312
0.21972656 0.29492188 -0.06347656 0.03222656 0.16992188 -0.09619141
0.40234375 -0.01806641 -0.0612793 0.11962891 -0.00196838 0.25
-0.15234375 0.18164062 0.02453613 0.04858398 -0.20507812 0.00245667
0.13671875 0.5 -0.14746094 -0.13476562 -0.265625 -0.1953125
0.4375 -0.00512695 -0.12890625 -0.04907227 -0.2421875 0.19335938 ]
```

例：iPhoneを300次元の
数値の組に変換

例：iPadを300次元の
数値の組に変換

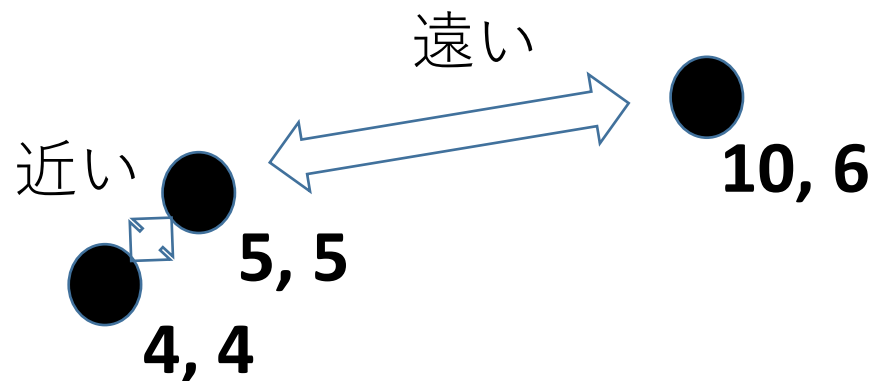
数値と距離



この考え方を多次元（例：300次元）に
拡張して単語の類似度を表現する

ベクトルの距離

2次元（2個の数値）での例



- 近い点 : $(4, 4), (5, 5)$
- 遠い点 : $(5, 5), (10, 6)$

点同士の距離で類似度を表現することができる。

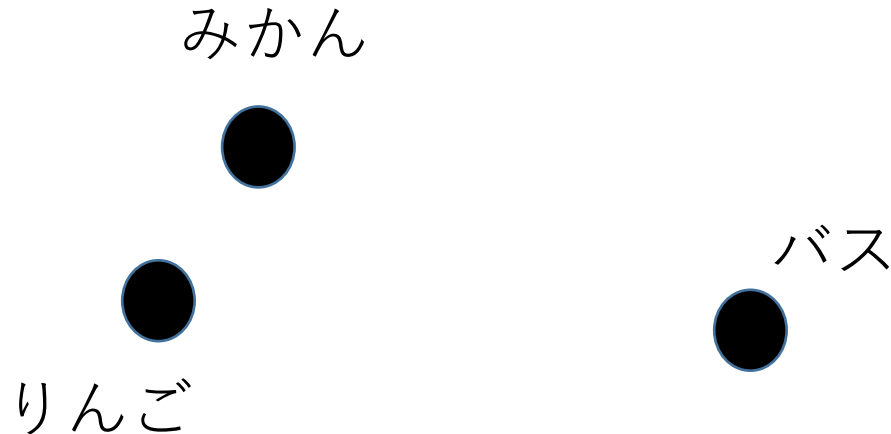
単語ベクトルの性質

主な特徴：

- **単語を特徴ベクトル（複数の数値の組）で表現**
- **意味の近い単語は空間的に近く配置**

例：「みかん」と「りんご」は近い

例：「バス」はこれらから遠い位置に配置





演習 2. 単語の類似度


① Microsoft Copilot に次のプロンプトを与える

次の単語について、似ているものでグループを作ってください

手、飛行機、新幹線、指、みかん、バナナ、顔、列車、りんご、タクシー、肩、ひざ、髪の毛、レモン、フェリー


乗り物グループ

- ・ 飛行機
- ・ 新幹線
- ・ 列車
- ・ タクシー
- ・ フェリー

 これらはすべて人や物を運ぶための交通手段です。

果物グループ

- ・ みかん
- ・ バナナ
- ・ りんご
- ・ レモン

 食べられる果物で、日常的に親しまれているものです。

人体の部位グループ

- ・ 手
- ・ 指
- ・ 顔
- ・ 肩
- ・ ひざ
- ・ 髪の毛

チャットボットの利用上の注意点

 人の体の一部を表す言葉です。動作や感覚に関係するものも含まれています。

② Microsoft Copilot に次のプロンプトを与える

次の単語について、似ているものでグループを作ってください

光、星、電球、太陽、月、ライト、針、時計、注射、裁縫、松、竹、梅、桜、杉、板、机、紙、本、ノート

💡 光・照明・天体グループ

- ・ 光 (自然・人工の光)
- ・ 星 (天体)
- ・ 太陽 (恒星)
- ・ 月 (衛星)
- ・ ライト (人工照明)
- ・ 電球 (照明器具)

👉 これらは「光」や「明るさ」に関係するものです。天体と人工物が混在していますが、共通点は「光を発する・反射する」ことです。

🔧 道具・機能グループ

- ・ 針 (細い金属製の道具)
- ・ 時計 (時間を測る道具)
- ・ 注射 (医療行為)
- ・ 裁縫 (針を使う作業)

👉 「針」を中心に、医療や手芸、時間計測など、道具や機能に関連する言葉です。

🌳🌸 植物グループ (木・花)

- ・ 松
- ・ 竹
- ・ 梅
- ・ 桜
- ・ 杉

11-4. チャットボットの概要

👉 すべて植物で、特に日本文化に深く関わる木や花です。「松竹梅」は縁起物としても有名ですね。

📁🪑 文房具・家具グループ

- ・ 板
- ・ 机
- ・ 紙
- ・ 本
- ・ ノート

👉 学習や作業に使う道具・素材です。机は家具ですが、他は文房具や書類関連です。

14-3. 文書データを活用する基本技術

自然言語処理の3つの重要タスク



① 文書分類

文書を、事前に定めたカテゴリに自動で分類

② 感情極性判定

文書に表現された意見や感情が、ポジティブかネガティブかを判定

③ トピックモデリング

大量の文章から、隠れている可能性のある主題（トピック）を発見

- 主な用途

記事の振り分け、問い合わせやアンケートの振り分け、メールの有害・無害分類、多量書類の整理

- 主要技術

ナイーブベイズ（単語の出現頻度を使って分類）

例： **お得、今すぐ、割引** という単語が多ければ
「広告メール」に分類

- 主な用途

商品レビューの分析（満足度の抽出など）、改善意見の発見、炎上の発見、

- 主要技術

辞書ベース（感情語辞書を用いる判定）

例： 困った、良い、楽しい などの感情語をカウント、ポジティブ後8個、ネガティブ後4語 → ポジティブと判定

トピックモデリング



- 主な用途

ニュースの主要意見、トピックの変化の分析、図書などの分野別の振り分け、新しいトピックの発見、気づきにくい本音の発見

- 主要技術

LDA

例) この記事は「政治」60%、「海外」40%
のように、複数のトピックの混合として考える
手法

演習 3. 文書分類



演習 3. 文書分類



目的

- 文書の自動分類体験
- 特徴抽出の理解

使用データ

タイトル,内容,カテゴリ

テキサス州洪水で106人死亡,テキサス州中部で7月4日に発生した大洪水により106人が死亡。140人以上が依然として行方不明,災害

トランプ大統領がウクライナ支援発表,トランプ大統領がウクライナへの武器供給とロシアへの50日間の和平期限を発表,政治

ケニア反政府デモで31人死亡,ケニアで反政府デモが発生し警察との衝突により31人が死亡。500人以上が逮捕される,国際

国家安全保障顧問マイク・ウォルツ氏が辞任,マイク・ウォルツ国家安全保障顧問の職務が宙に浮いた状態。国家安全保障会議を率いることはなくなると判明,政治

イスラエルがシリアを空爆,ダマスカスでイスラエルによる空爆が発生。ドゥルーズ派住民保護を目的とした攻撃,国際

コカ・コーラがアメリカでサトウキビ糖使用へ,トランプ大統領の発表によりコカ・コーラが米国での製品にサトウキビ糖の使用に合意,経済

インドで学校給食に死んだヘビ混入,インドで学校給食に死んだヘビが混入し数百人の児童が体調不良。人権当局が調査開始,国際

シークレットサービス職員が停職処分,2024年のトランプ集会警備に関わったシークレットサービス職員らが停職処分を受ける,政治

銅価格が記録的高値を更新,銅価格が記録的な高値に急騰。さらなる価格上昇の可能性も指摘される,経済

24万年前の化石からタンパク質を回収,科学者らが2400万年前のサイの化石から記録的に古いタンパク質の回収に成功,科学技術

① Microsoft Copilot に次のプロンプトを与える

以下のニュース記事データを使って、文書分類を行ってください

1. 各記事の特徴（キーワード、文章パターン）を抽出してください

2. それぞれの記事の分類の根拠について説明してください

タイトル,内容,カテゴリ

テキサス州洪水で106人死亡,テキサス州中部で7月4日に発生した大洪水により106人が死亡。140人以上が依然として行方不明,災害

トランプ大統領がウクライナ支援発表,トランプ大統領がウクライナへの武器供給とロシアへの50日間の和平期限を発表,政治

ケニア反政府デモで31人死亡,ケニアで反政府デモが発生し警察との衝突により31人が死亡。500人以上が逮捕される,国際

国家安全保障顧問マイク・ウォルツ氏が辞任,マイク・ウォルツ国家安全保障顧問の職務が宙に浮いた状態。国家安全保障会議を率いることはなくなると判明,政治

イスラエルがシリアを空爆,ダマスカスでイスラエルによる空爆が発生。ドゥルーズ派住民保護を目的とした攻撃,国際

コカ・コーラがアメリカでサトウキビ糖使用へ,トランプ大統領の発表によりコカ・コーラが米国での製品にサトウキビ糖の使用に合意,経済

インドで学校給食に死んだヘビ混入,インドで学校給食に死んだヘビが混入し数百人の児童が体調不良。人権当局が調査開始,国際

シークレットサービス職員が停職処分,2024年のトランプ集会警備に関わったシークレットサービス職員らが停職処分を受ける,政治

銅価格が記録的高値を更新,銅価格が記録的な高値に急騰。さらなる価格上昇の可能性も指摘される,経済

24万年前の化石からタンパク質を回収,科学者らが2400万年前のサイの化石から記録的に古いタンパク質の回収に成功,科学技術

分類結果一覧

タイトル	カテゴリ	キーワード例	分類根拠
テキサス州洪水で106人死亡	災害	洪水、死亡、行方不明	災害に関する語（洪水、死亡、行方不明）が用いられており、自然災害報道であるため
トランプ大統領がウクライナ支援発表	政治	[タイトルなし] トランプ、ウクライナ、武器供給	政治家（トランプ）の外交政策に関する記事であるため
ケニア反政府デモで31人死亡	国際	ケニア、反政府、デモ、死亡	国際的な事件（デモ、衝突）に関する記事であり、複数国に関与しているため
国家安全保障顧問マイク・ウォルツ氏が辞任	政治	国家安全保障、辞任、マイク・ウォルツ	政府機関や政治家に関する記事があるため
イスラエルがシリアを空爆	国際	イスラエル、シリア、空爆、住民保護	国際的な軍事行動に関する記事があり、複数国に関与しているため
コカ・コーラがアメリカでサトウキビ糖使用へ	経済	コカ・コーラ、サトウキビ糖、製品	企業活動と製品変更に関する記事があり、経済に関連しているため

特徴抽出の方法

- **キーワード抽出**：記事のタイトルと内容を統合し、頻出語を抽出。
- **文書パターン**：政治・国際・災害・経済・科学技術などのカテゴリに応じて、登場人物・事件・対象物の種類を分析。
- **分類根拠**：キーワードと文脈から、どのカテゴリに属するかを論理的に判断。



演習 4 . 感情極性判定

演習 4 . 感情極性判定

目的

- テキストの感情分析理解
- 感情語の抽出理解

使用データ

レビューID,商品,レビュー内容

- 1,Apple AirPods Pro 2,ノイズキャンセリング機能が素晴らしい！通話品質も良く、バッテリーの持ちも予想以上でした
- 2,Owala FreeSip ウォーターボトル,デザインがおしゃれで使いやすい。飲み口が2種類あるのが便利だが、少し重いかも
- 3,Stanley IceFlow タンブラー,保温性能は確かに良いけど、蓋の開け閉めが少し固い。でも人気なのは納得
- 4,Apple Watch Series 10,健康管理機能が充実していて満足。ただし値段が高すぎる。もう少し安ければ最高
- 5,Fitbit Inspire 3,シンプルで使いやすいフィットネストラッカー。睡眠トラッキングが正確で重宝してます
- 6,Amazon Echo Dot (最新モデル),音質が以前より格段に良くなった。セットアップも簡単でスマートホーム初心者におすすめ
- 7,Roku Streaming Stick 4K Plus,4K画質がきれいで動作もサクサク。リモコンの音声機能も便利で満足度高い
- 8,Apple iPad 11インチ (A16チップ),処理速度が速く画面も美しい。ただしApple Pencilが別売りなのが残念
- 9,Wireless Earbuds Bluetooth 5.4,この価格でこの音質は驚き！バッテリー持続時間も長くてコスパ最高です
- 10,Blink Video Doorbell,設置が簡単でアプリも使いやすい。画質は良いが夜間の映像がもう少し明るければ完璧

① Microsoft Copilot に次のプロンプトを与える

以下の商品レビューデータを使って、感情極性判定を実施してください。

1. 各レビューの感情語（ポジティブ・ネガティブ語）を抽出してください
2. 各レビューをポジティブ・ネガティブ・ニュートラルに分類してください
3. 各レビューの判定の根拠を示してください
4. レビューの中に皮肉があれば説明してください

レビューID,商品,レビュー内容

1,Apple AirPods Pro 2,ノイズキャンセリング機能が素晴らしい！通話品質も良く、バッテリーの持ちも予想以上でした

2,Owala FreeSip ウォーターボトル,デザインがおしゃれで使いやすい。飲み口が2種類あるのが便利だが、少し重いかも

3,Stanley IceFlow タンブラー,保温性能は確かに良いけど、蓋の開け閉めが少し固い。でも人気なのは納得

4,Apple Watch Series 10,健康管理機能が充実していて満足。ただし値段が高すぎる。もう少し安ければ最高

5,Fitbit Inspire 3,シンプルで使いやすいフィットネストラッカー。睡眠トラッキングが正確で重宝してます

6,Amazon Echo Dot (最新モデル),音質が以前より格段に良くなった。セットアップも簡単でスマートホーム初心者におすすめ

7,Roku Streaming Stick 4K Plus,4K画質がきれいで動作もサクサク。リモコンの音声機能も便利で満足度高い

8,Apple iPad 11インチ (A16チップ),処理速度が速く画面も美しい。ただしApple Pencilが別売りなのが残念

9,Wireless Earbuds Bluetooth 5.4,この価格でこの音質は驚き！バッテリー持続時間も長くてコスパ最高です

10,Blink Video Doorbell,設置が簡単でアプリも使いやすい。画質は良いが夜間の映像がもう少し明るければ完璧

レビューID	商品名	分類	ポジティブ語	ネガティブ語	皮肉
1	Apple AirPods Pro 2	ポジティブ	素晴らしい, 良く, 予想以上	なし	なし
2	Owala FreeSip ウォーターボトル	ポジティブ	おしゃれ, 使いやすい, 便利	重い	あり
3	Stanley IceFlow タンブラー	ニュートラル	納得	固い	あり
4	Apple Watch Series 10	ニュートラル	満足, 最高	高すぎる, もう少し安ければ	あり
5	Fitbit Inspire 3	ポジティブ	使いやすい, 正確, 重宝	なし	なし
6	Amazon Echo Dot (最新モデル)	ポジティブ	良く, 格段に良くなった, 簡単, おすすめ	なし	なし
7	Roku Streaming Stick 4K Plus	ポジティブ	便利, 満足, きれい, サクサク	なし	なし
8	Apple iPad 11インチ (A16チップ)	ポジティブ	速く, 美しい	残念	あり
9	Wireless Earbuds Bluetooth 5.4	ポジティブ	驚き, 最高	なし	なし
10	Blink Video Doorbell	ポジティブ	使いやすい, 簡単, 完璧	もう少し明るければ	あり



演習 5. トピックモデリング

演習 5. トピックモデリング

目的

・ トピックモデリングの体験

使用データ

タイトル,内容,カテゴリ

文書ID,内容

1. ハデュール大学の学生が作ったロボットが、ルービックキューブを0.103秒で解いて世界記録を更新
2. ベラ・ルービン天文台が宇宙の美しい画像を初公開し、天文学者たちが新発見に期待
3. チンパンジーが賢い道具を使って虫を捕まえる技術を、タンザニアの研究者が発見
4. 土星の衛星数が274個に増え、「衛星の王様」の称号をさらに確固たるものに
5. イルカのお母さんが赤ちゃんと「赤ちゃん言葉」で会話していることが最新研究で判明
6. カラスの知能が5~7歳の子どもと同じレベルで、道具を使いパズルを解けることが証明
7. 地球に「三日月」が一時的に現れる珍しい天体現象が2024年9月に観測される
8. 豚から人間への腎臓移植が2ヶ月間成功し、医学界で画期的な成果として注目
9. 古代の砂漠アートをAIが発見し、人類の歴史を解明する新たな手がかりを提供
10. 蛾が植物の「会話」を聞いている驚きの能力を科学者が発見
11. フィラデルフィア動物園でガラパゴス象亀の絶滅危惧種の赤ちゃんが4匹誕生
12. サメが実は音を出すことができる可能性を研究者が発見し、従来の常識を覆す
13. ミミズが作物の収量を増やし、マイクロプラスチックから植物を守る働きを解明
14. 宇宙の木材が建築材料として使用される革新的な技術が開発される
15. 地球温暖化で若いアスリートの熱中症対策がより重要になっていることが判明
16. NASA/ハッブル宇宙望遠鏡が新しい球状星団ESO 591-12の美しい画像を撮影
17. 国際宇宙ステーションでNASAのCrew-11ミッション宇宙飛行士が科学実験を実施
18. スロバキアの高校生が新しい薬の製造方法を発明し、国際科学フェアで優勝
19. チンパンジーが仲間にケガの応急処置をする行動をウガンダで研究者が観察
20. 科学者が患者一人のために世界初の遺伝子編集治療薬を開発し、治療に成功
21. オランダの研究者が空気で動く脳のないソフトロボットを開発
22. ハワイで発見された新種の蛾の幼虫が、クモの巣に住んで昆虫の死骸を食べる
23. 青いサメの皮膚に隠された色を変える能力を科学者が発見
24. 量子物理学の新しい研究で、材料をねじることで特殊な量子状態を作る方法を発見
25. ドリチョウタンパク質が植物の水分保持に重要な役割を果たすことが判明
26. 組織の形成に関する「組織コード」と呼ばれる5つのルールを科学者が発見
27. フィジーの熱帯雨林で、アリの戦争を防ぐ特殊な住居を作る植物を発見
28. バッパという名前の猫が新しいウイルスの発見に貢献
29. 老化したラットの脳で200種類以上の異常タンパク質を発見
30. カリフォルニア大学バークレー校が「olo」という新しい色を発見
31. 南カリフォルニア地震で動物たちが事前に異常行動を見せる様子を撮影
32. アースデイ2025が55周年を迎え、環境保護の重要性を再認識
33. ファストファッションの廃棄物問題を解決するため、古着から包装材料を作る技術を開発
34. SpaceXが史上初めてロケットブースターのキャッチに成功
35. タコが狩りのリーダーシップを発揮し、協力しない魚を「パンチ」する行動を観察
36. 土星の輪が2025年に一時的に見えなくなる珍しい天体現象が発見
37. ムームードウンとベストという2匹の動物がインターネットで人気者になる
38. ミミズが土壌を改善し、作物の成長を促進する仕組みを詳しく解明
39. クモが獲物を捕まえるために偽のホタルを使う巧妙な戦略を発見
40. 月が従来考えられていたより4000万年古い可能性があることが判明
41. ベレグリン月面着陸船が任務失敗後に燃え尽きる
42. 2023年に発見された恐竜化石の重要な発見が複数報告される
43. ペンジャミン・フランクリンが12歳で水泳用の手の蹼を発明した歴史的事実
44. 11歳のジニー・ローがキッチンで使える折りたたみ椅子を発明
45. 15歳のルイ・ブライコが点字システムを発明
46. 8歳の子エルシー・ラノンが赤ちゃん用おしりふきポケット付きおむつの特許取得
47. 子どもの発明家精神を育む教育の重要性と好奇心の役割
48. 科学者になるために必要な資質と発明家になるための条件の違い
49. 音楽家が楽器に関連する製品を発明することで科学的思考を活用する例
50. 詩人と天文学者の異なる視点が新しい発明につながる可能性について

① Microsoft Copilot に次のプロンプトを与える

以下の文書データを使って、トピックモデリングを実施してください。

1. 文書群から潜在的なトピック（主要なテーマ）を3つ発見してください
2. 各トピックを代表するキーワードを抽出してください
3. 各文書がどのトピックに最も関連するかを判定し、説明してください

使用データ

タイトル,内容,カテゴリ

文書ID,内容

1. ハドラー大学の学生が作ったロボットが、ルービックキューブを0.103秒で解いて世界記録を更新
2. ベラ・ルービン天文台が宇宙の美しい画像を初公開し、天文学者たちが新発見に期待
3. チンパンジーが賢い道具を使って虫を捕まえる技術を、タンザニアの研究者が発見
4. 土星の衛星数が274個に増え、「衛星の王様」の称号をさらに確固たるものに
5. イルカのお母さんが赤ちゃんと言葉で会話していることが最新研究で判明
6. カラスの知能が5~7歳の子どもの同じレベルで、道具を使いパズルを解けることが証明
7. 地球に「ミニ月」が一時的に現れる珍しい天体現象が2024年9月に観測される
8. 豚から人間への腎臓移植が2ヶ月間成功し、医学界で画期的な成果として注目
9. 古代の砂漠アートをAIが発見し、人類の歴史を解明する新たな手がかりを提供
10. 蛾が植物の「会話」を聞いている驚きの能力を科学者が発見
11. フィラデルフィア動物園でガラパゴス象亀の絶滅危惧種の赤ちゃんが4匹誕生
12. サメが実は音を出すことができる可能性を研究者が発見し、従来の常識を覆す
13. ミミズが作物の収量を増やし、マイクロプラスチックから植物を守る働きを解明
14. 宇宙の木材が建築材料として使用される革新的な技術が開発される
15. 地球温暖化で若いアスリートの熱中症対策がより重要になっていることが判明
16. NASAハッブル宇宙望遠鏡が新しい球状星団ESO 591-12の美しい画像を撮影
17. 国際宇宙ステーションでNASAのCrew-11ミッション宇宙飛行士が科学実験を実施
18. スロバキアの高校生が新しい薬の製造方法を発明し、国際科学フェアで優勝
19. チンパンジーが仲間にケガの応急処置をする行動をウガンダで研究者が観察
20. 科学者が患者一人のために世界初の遺伝子編集治療薬を開発し、治療に成功
21. オランダの研究者が空気中で動く脳のないソフトロボットを開発
22. ハワイで発見された新種の蛾の幼虫が、クモの巣に住んで昆虫の死骸を食べる
23. 青いサメの皮膚に隠された色を変える能力を科学者が発見
24. 量子物理学の新しい研究で、材料をねじることで特殊な量子状態を作る方法を発見
25. ドリチョウタンパク質が植物の水分保持に重要な役割を果たすことが判明
26. 組織の形成に関する「組織コード」と呼ばれる5つのルールを科学者が発見
27. フィジーの熱帯雨林で、アリの戦争を防ぐ特殊な住居を作る植物を発見
28. ペッパーという名前の猫が新しいウイルスの発見に貢献
29. 老化したラットの脳で200種類以上の異常タンパク質を発見
30. カリフォルニア大学バークレー校が「olo」という新しい色を発見
31. 南カリフォルニア地震で動物たちが事前に異常行動を見せる様子を撮影
32. アースデイ2025が55周年を迎え、環境保護の重要性を再認識
33. ファストファッションの廃棄物問題を解決するため、古着から包装材料を作る技術を開発
34. SpaceXが史上初めてロケットブースターのキャッチに成功
35. タコが狩りのリーダーシップを発揮し、協力しない魚を「パンチ」する行動を観察
36. 土星の輪が2025年に一時的に見えなくなる珍しい天体現象が発見
37. ムームードウンとベストという2匹の動物がインターネットで人気者になる
38. ミミズが土壌を改善し、作物の成長を促進する仕組みを詳しく解明
39. クモが獲物を捕まえるために偽のホタルを使う巧妙な戦略を発見
40. 月が従来考えられていたより4000万年古い可能性があることが判明
41. ペレグリン月面着陸船が任務失敗後に燃え尽きる
42. 2023年に発見された恐竜化石の重要な発見が複数報告される
43. ベンジャミン・フランクリンが12歳で水泳用の手の蹼を発明した歴史的事実
44. 11歳のジーニー・ローがキッチンで使える折りたたみ椅子を発明
45. 15歳のルイ・ブライユが点字システムを発明
46. 8歳のチェルシー・ラングが赤ちゃん用おしりふきポケット付きおむつの特許取得
47. 子どもの発明家精神を育む教育の重要性と好奇心の役割
48. 科学者になるために必要な資質と発明家になるための条件の違い
49. 音楽家が楽器に関連する製品を発明することで科学的思考を活用する例
50. 詩人と天文学者の異なる視点から新しい発明につながる可能性について

抽出されたトピックと代表キーワード

トピック 0: 科学技術・医療・発明

- 代表キーワード例：
 - 世界記録、宇宙、衛星、ロボット、腎臓移植、遺伝子編集、赤ちゃん、発明、ミミズ、教育

トピック 1: 環境・自然・生物の行動

- 代表キーワード例：
 - 恐竜化石、蛾、量子状態、サメ、植物、薬、宇宙飛行士、ファッション、詩人、科学実験

トピック 2: 宇宙・動物・自然現象

- 代表キーワード例：
 - ハッブル望遠鏡、地震、チンパンジー、熱中症、熱帯雨林、蛾の幼虫、AI、ウイルス、老化

📄 文書のトピック分類（例）

文書内容（抜粋）	割り当てられたトピック
パデュー大学のロボットが世界記録	トピック 0
ベラ・ルービン天文台が宇宙画像公開	トピック 0
チンパンジーが道具で虫を捕獲	トピック 1
土星の衛星数が274個に増加	トピック 0
イルカが赤ちゃん言葉で会話	トピック 0
カラスがパズルを解く	トピック 2
古代の砂漠アートをAIが発見	トピック 2
蛾が植物の会話を聞く	トピック 1