

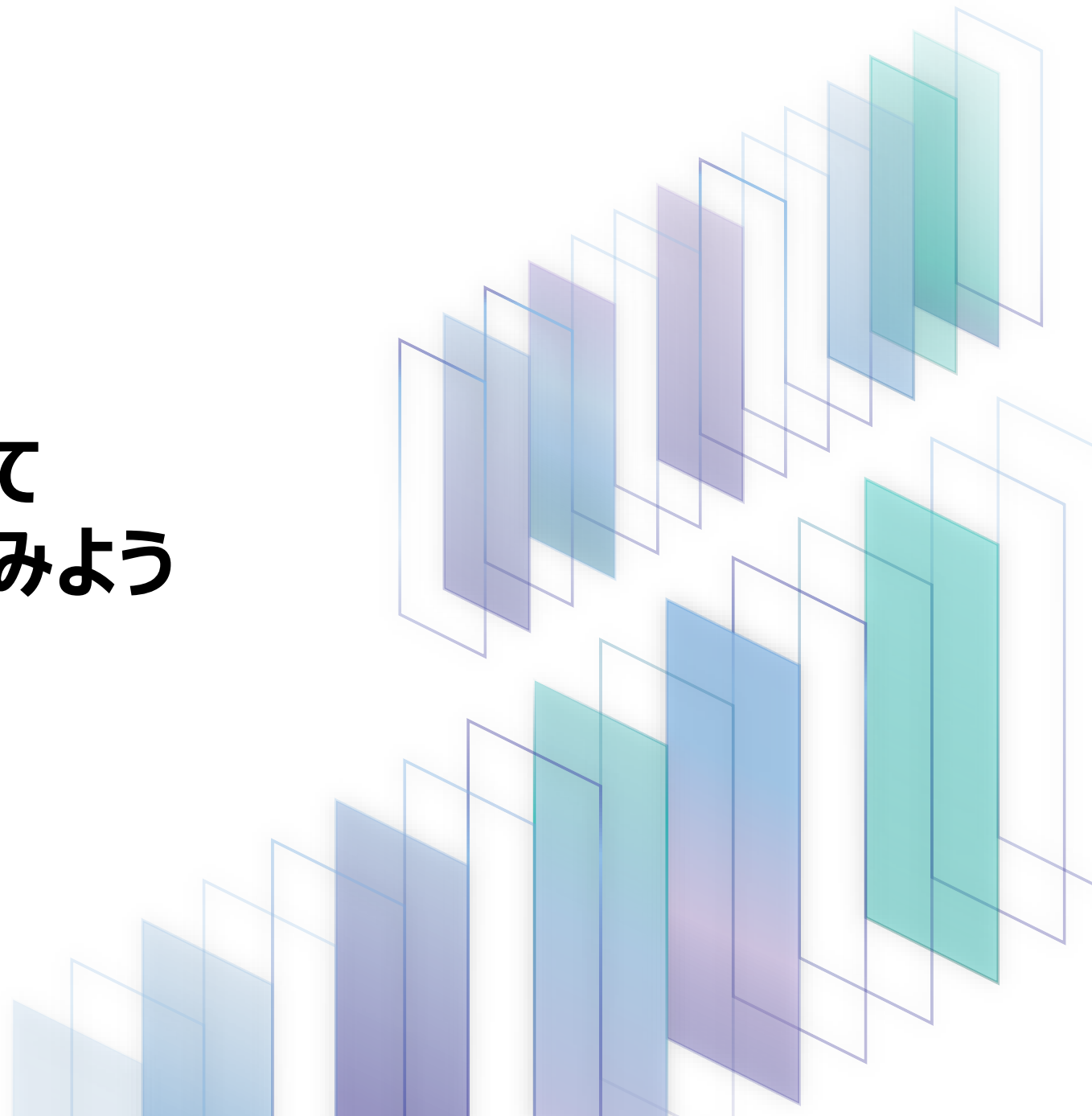
開発者向けオンラインセミナー

IRISのベクトル検索を使って テキストから画像を検索してみよう

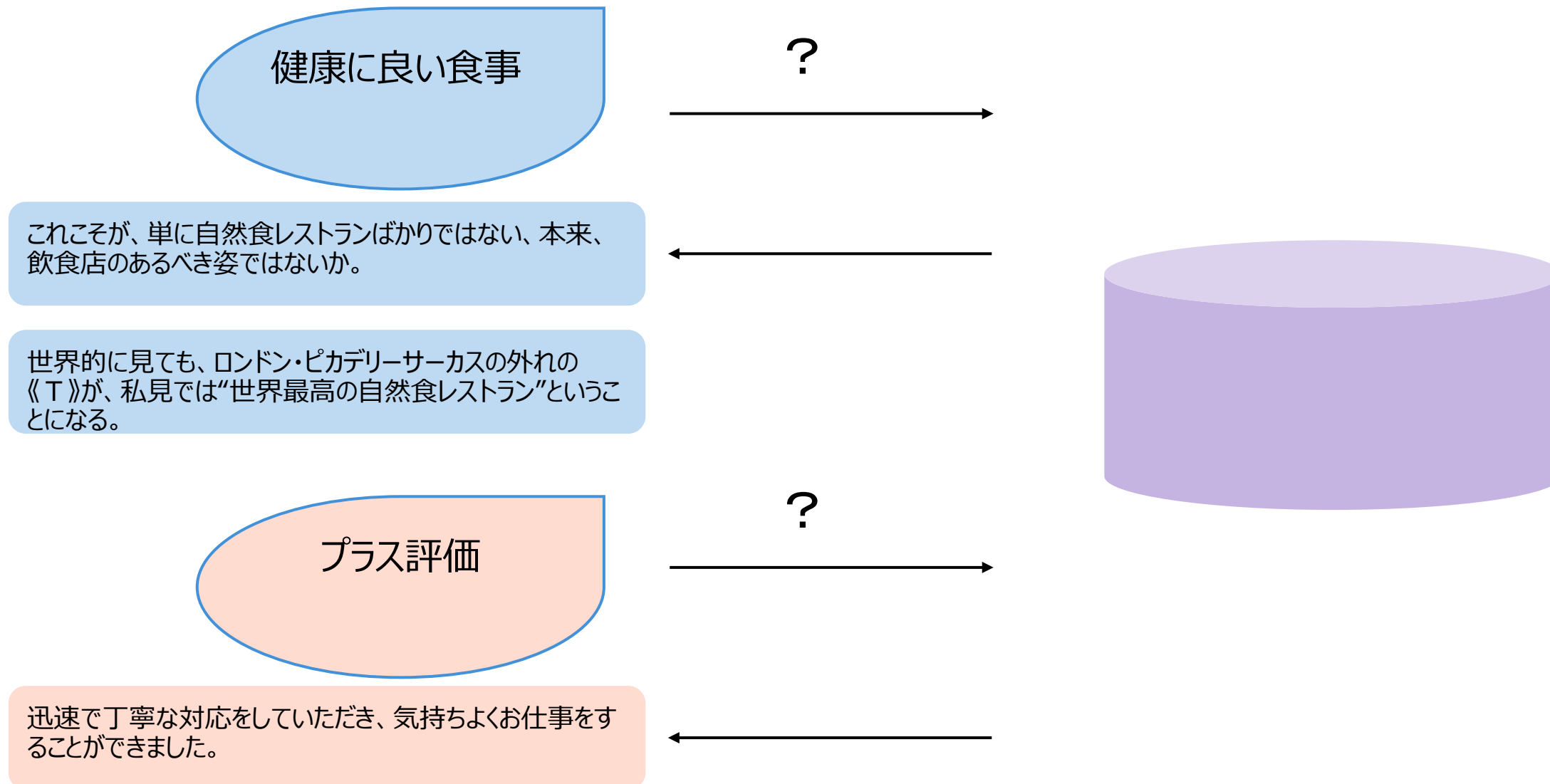
堀田 稔

インターシステムズジャパン株式会社

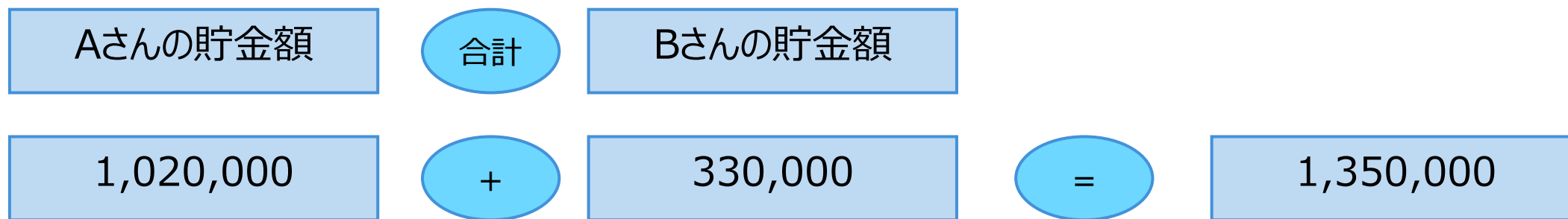
2025年6月10日



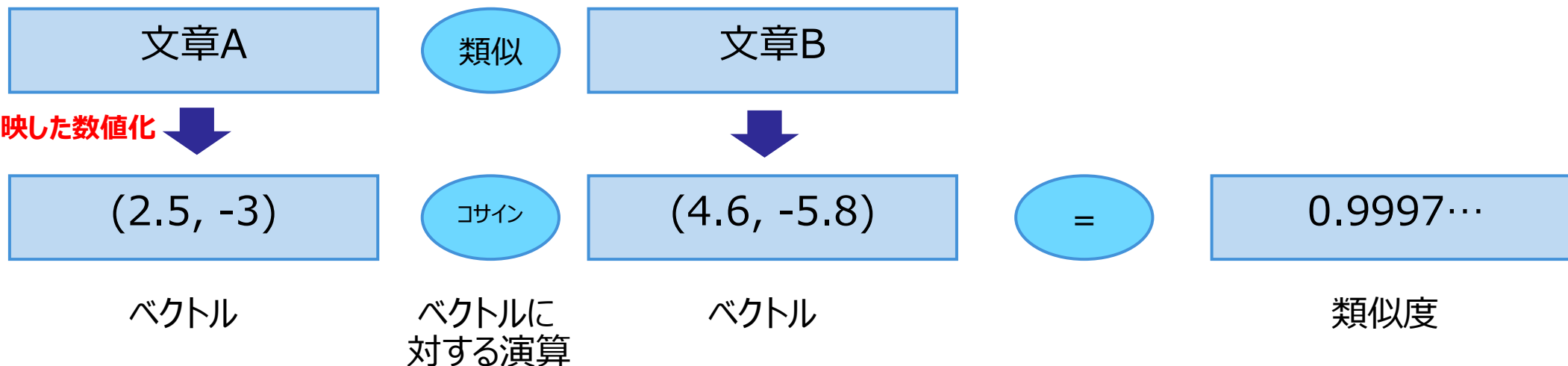
「高度な」テキスト検索 - 類似する文章、関連の深い文章を調べる



「高度な」テキスト検索はどのようにして可能になるか - 文章に対する「計算」



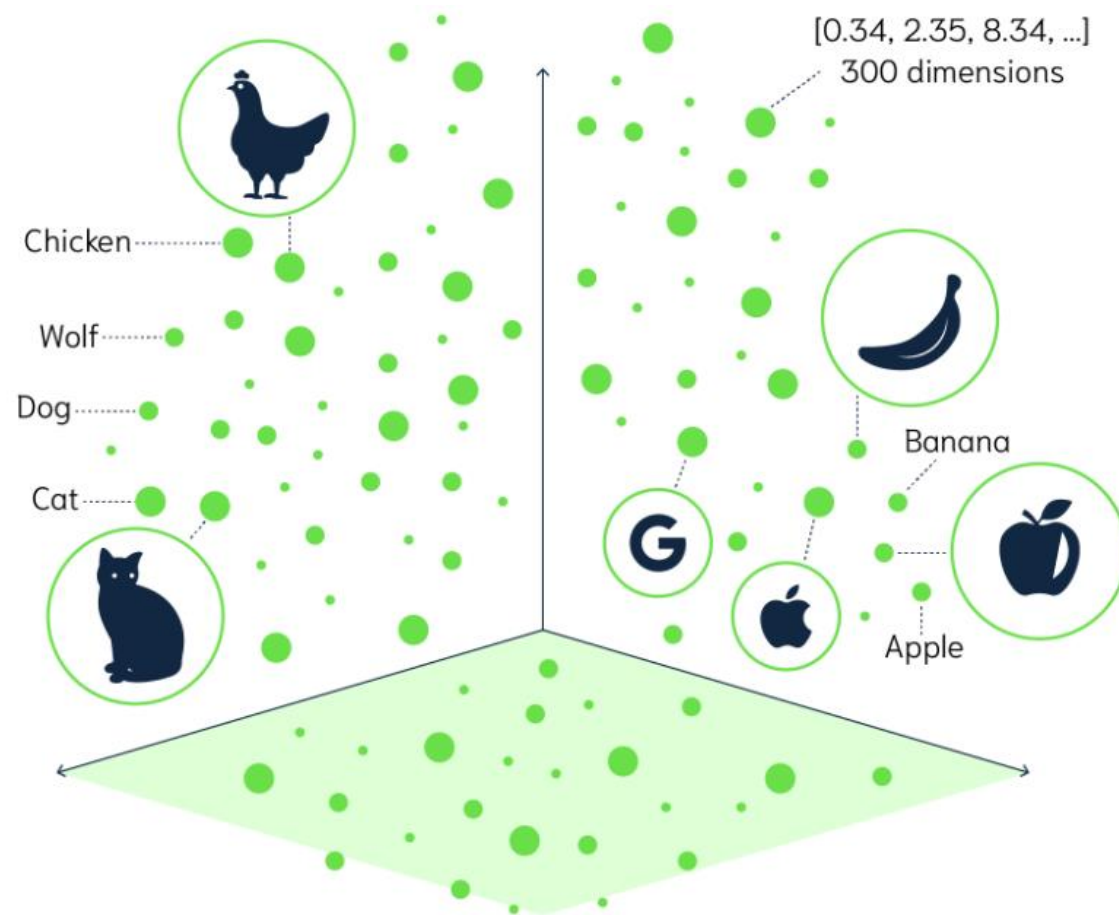
文の意味を反映した数値化 ↓



どのように適切なベクトルを得るのか？ - 埋め込み(Embedding)



- テキストデータに対して、その意味を反映する形でベクトルを割り当てる。
 - その割り当てを「埋め込み(Embedding)」と呼ぶ。
 - 得られたベクトルを「埋め込みベクトル」と呼ぶ。
- 「近い意味」を持つテキストが、「近い」ベクトルに割り当てられる。
 - ベクトルにテキストの持つ意味に関する情報が「詰め込まれている」
 - ベクトル間の距離 \approx テキストの意味の間の距離 と考えられる。
- 近さを計算するために、内積やコサイン類似度が利用できる。
- Word2Vec, BERT, GPTなどの機械学習アルゴリズム/モデルが用いられる。(BERTやGPTは、Transformerと呼ばれるモデルに基づく)
- 数百～数万次元のベクトルが用いられる。

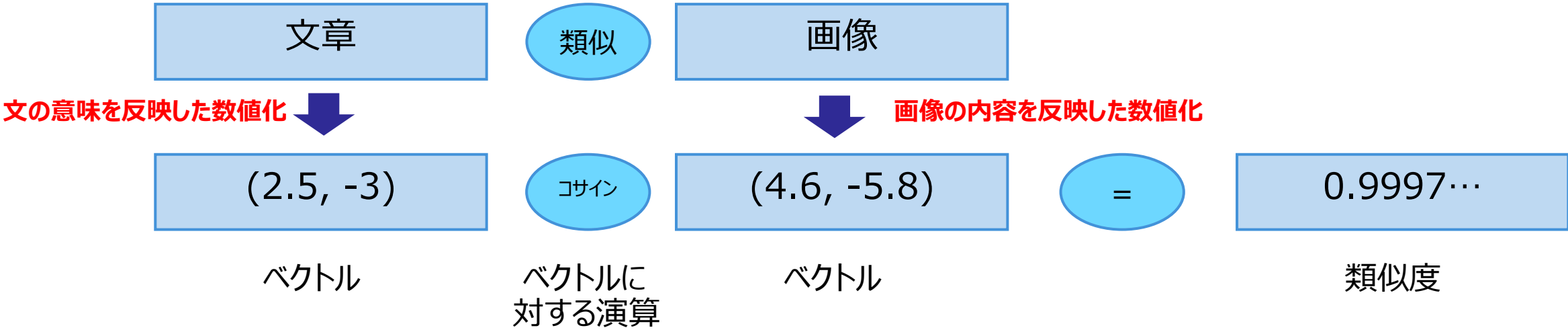


<https://opendatascience.com/a-gentle-introduction-to-vector-search/>

文章と画像は比べられるか？

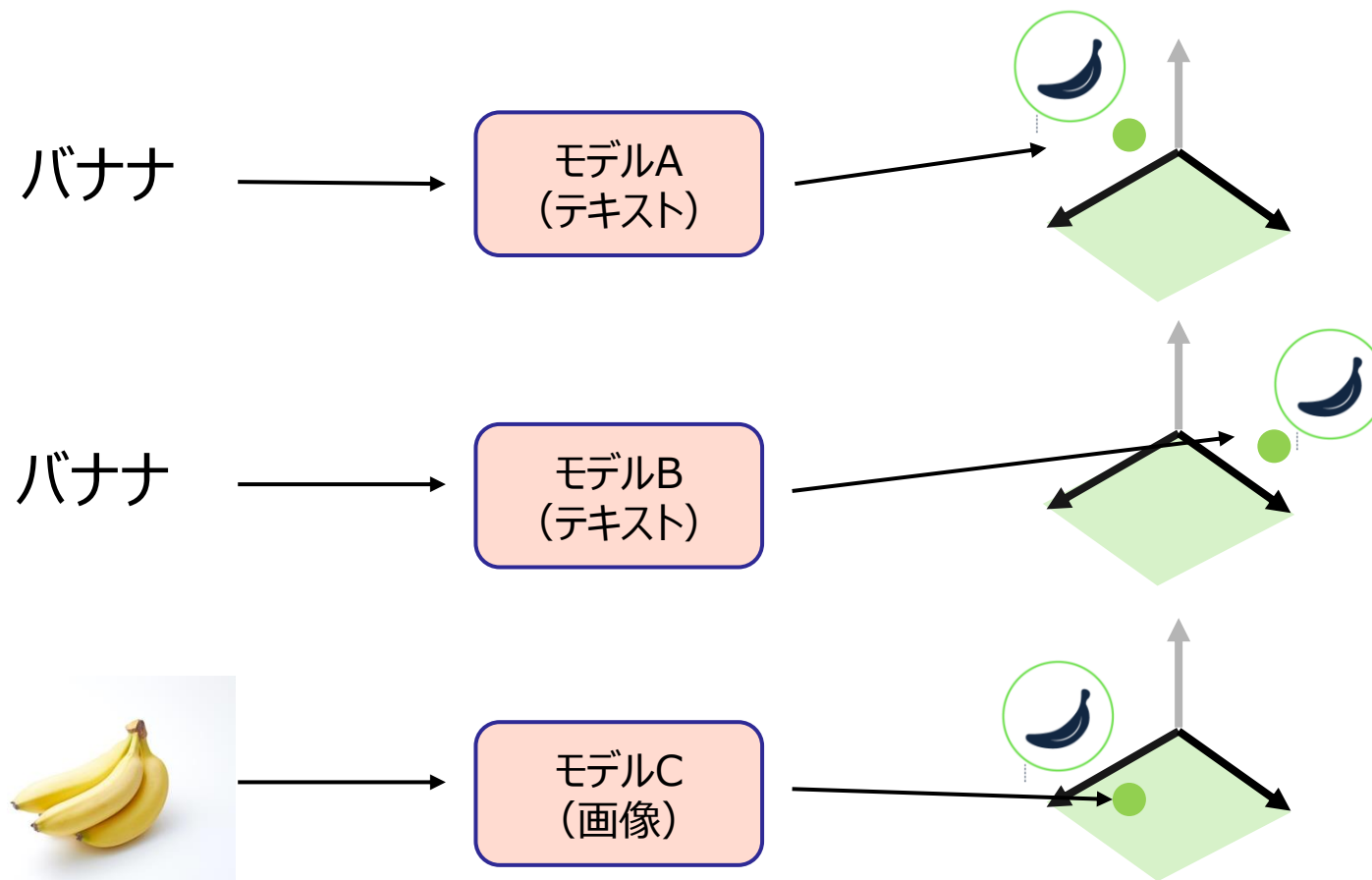


黄色い花



ベクトルの比較

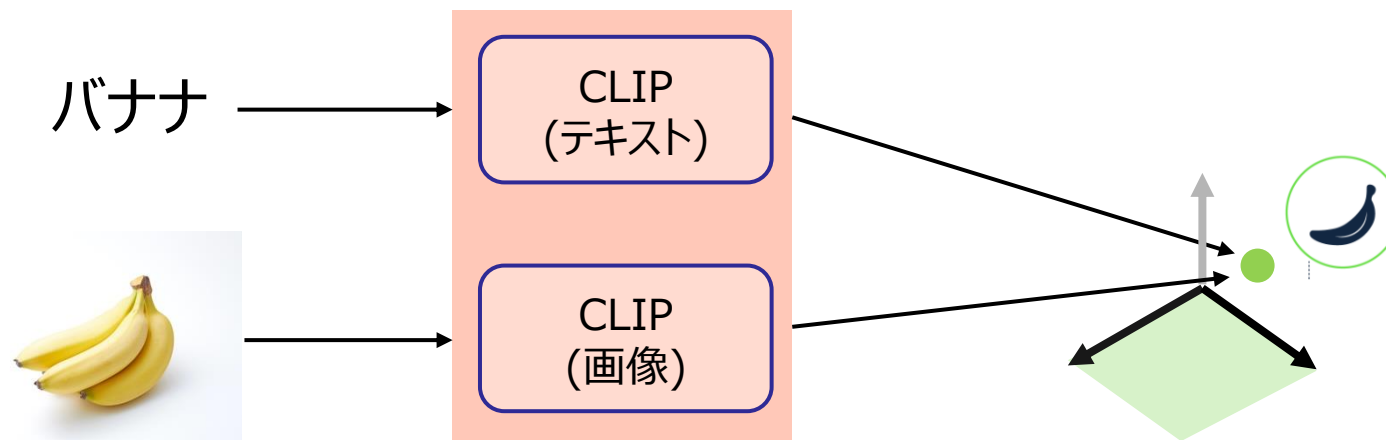
- 異なるモデルが生成した埋め込みベクトル同士は比較できない。



CLIP



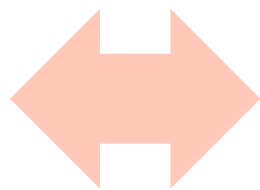
- CLIP – Contrastive Language-Image Pre-training
- Open AIが開発したマルチモーダル基盤モデル。
- 画像とその説明文（テキスト）のペアを予測できるように学習したもの。その結果、与えられたテキストに最も合致する画像を見つけたり、逆に、与えられた画像を最もよく説明するテキストを見つけたりすることが可能になった。



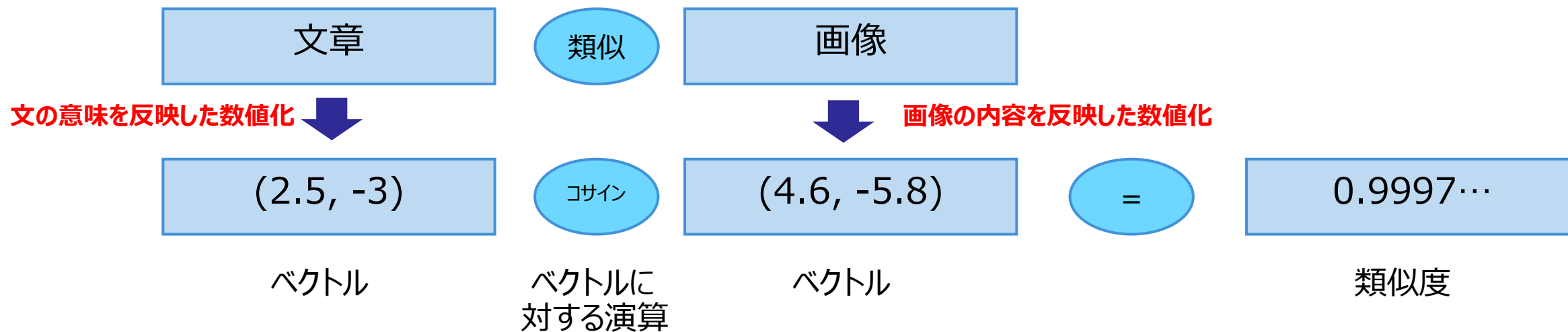
文章と画像は比べられるか？



黄色い花



CLIPにより可能！





IRISのベクトル検索機能

- IRIS 2024.1の実験的機能として提供。IRIS 2024.3で正式機能としてリリース。
- ベクトル型のデータを、INSERT文でIRISデータベースに保存。
- SELECT文で「類似する」ベクトルを検索。
- IRISネイティブのベクトル型により、高速な処理が可能。
 - (ご注意) IRIS 2024.1では、サポートされるCPUの命令セットに「最低ライン」が設定されました。
 - Intel: Haswell以降
 - AMD: Steamroller以降
- IRIS 2025.1から、ANN (Approximate Nearest Neighbor: 近似近傍探索)アルゴリズムに基づいたインデックスがサポートされ、ベクトル検索のパフォーマンスが劇的に向上している。

デモ

recruit-jp/japanese-image-classification-evaluation-dataset



⋮

約7000枚

画像モデル
clip-ViT-B-32

IRIS

ファイル名	ベクトル
"a.jpg"	[0.32, -0.065, 0.04]
"b.jpg"	[-0.58, -0.12, 0.42]
"c.jpg"	[0.89, 0.107, -0.08]

テキストモデル
clip-ViT-B-32-multilingual-v1

[0.59, -0.15, 0.08]

料理

SELECT TOP 5 ... FROM tbl
ORDER BY **VECTOR_COSINE()**



⋮

コンテスト／ミートアップ
に向けた
ウェビナーシリーズ
この後も続きます！

釣り人ビギナー Webinar サポートAIを作ろう！

ベクトル検索やRAGの活用で答えの精度を向上

第1回 ベクトル検索入門

← 本日開催！

第2回 ベクトルであそぼう！

第3回 RAGと生成AIであそぼう！

