

理工学部のデータサイエンス教育

- 電子情報工学科 ロボット創造工学科 ナノ物質工学科 -

機械創造工学科は、R8よりロボット創造工学科に名称変更します。

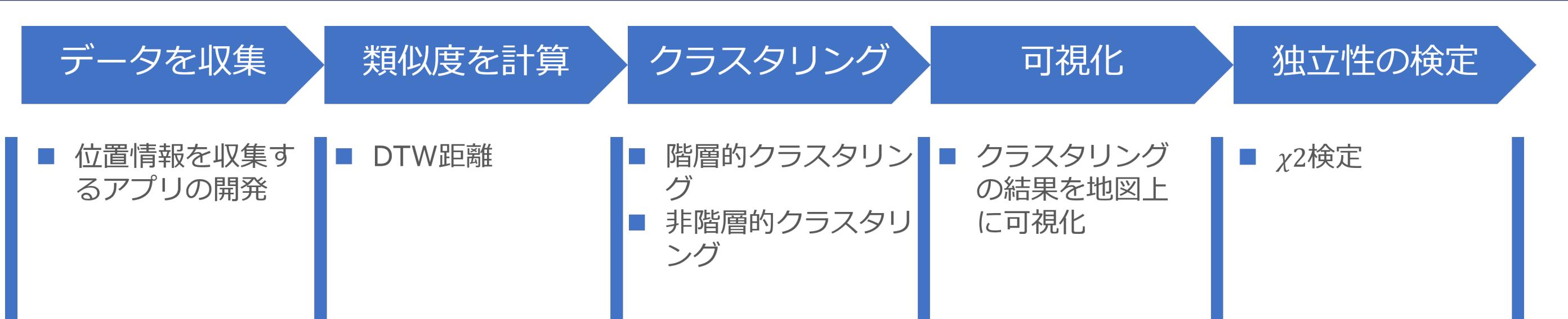
科目

- リテラシレベル: 情報処理, 文理学
- 応用基礎レベル
- AI基礎: 機械学習の基礎, 深層学習の基礎, 生成AIの基礎
- データサイエンス基礎: データ分析, データ可視化, アルゴリズム
- データエンジニアリング基礎: ビッグデータ, データ表現, データ収集, データ加工, プログラミング, 課題解決型学習 (PBL)

ソフトウェア

- 探索的データ解析, 可視化, 統計手法の適用のために, RStudio から R を使用
- 機械学習の手法を適用するために, Spyder や Jupyter Notebook から Anaconda の Python を仮想環境を作成して使用

移動経路の分析



- データの収集は、位置情報を収集するAndroidアプリケーションによって、位置情報(緯度経度)をサーバコンピュータに送信して、データベースに格納
- 移動経路は、人によって異なるため、DTW距離を非類似度としてクラスタリング
- アンケートの回答内容とクラスタの2元分割表の独立性の検定によって、傾向を分析

DTW (Dynamic Time Warping; 動的時間伸縮法)

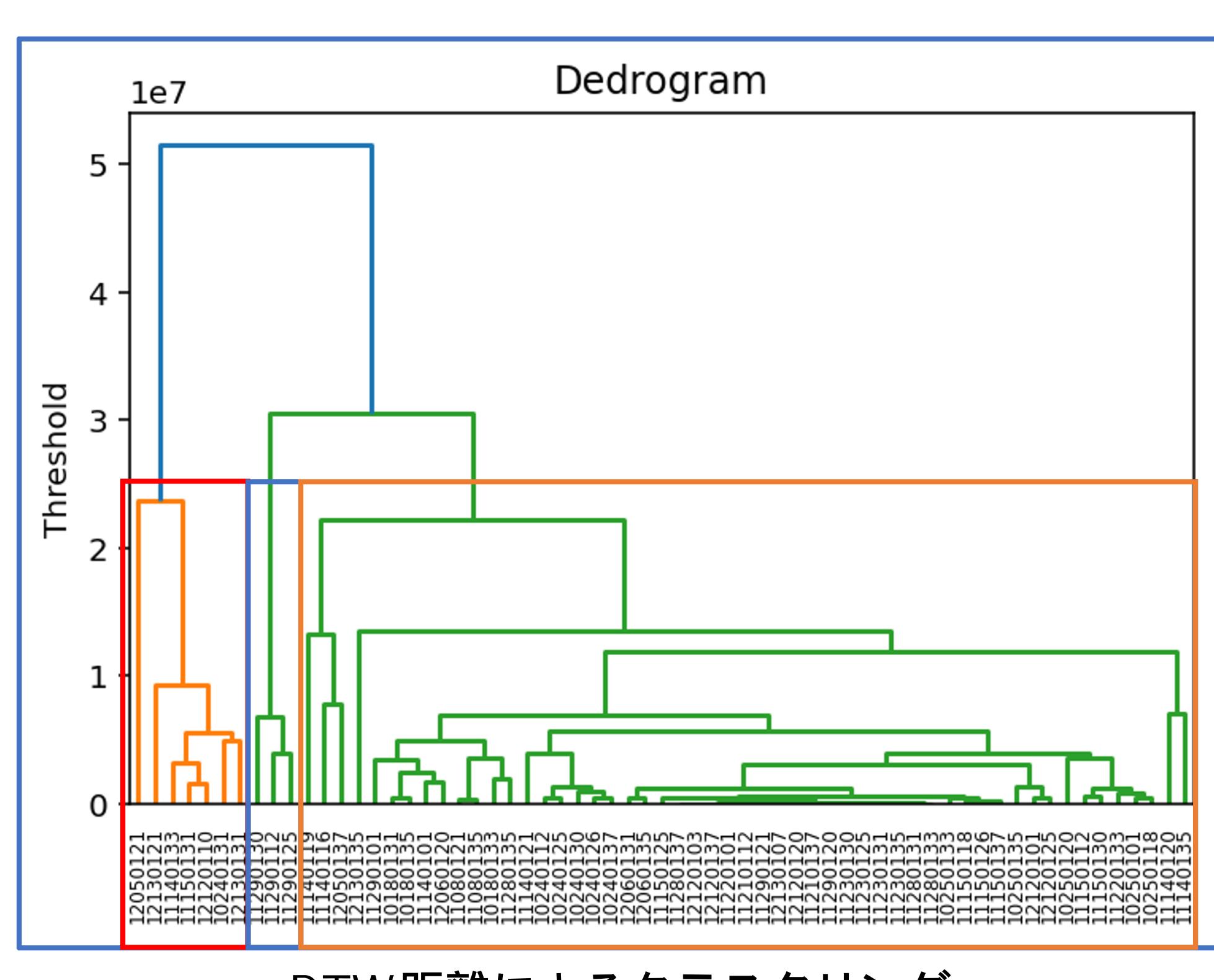
- 2つのデータ数が違っても計算できる
- 2つのデータの各点の距離を総当たりで比較して、それらの距離が最短となるパスを計算

no	ID	緯度	経度
1	1080131	34.197586	133.645817
2	1080131	34.197474	133.645887
3	1080131	34.197424	133.645861
:	:	:	:
147003	1114019	34.197180	133.645851

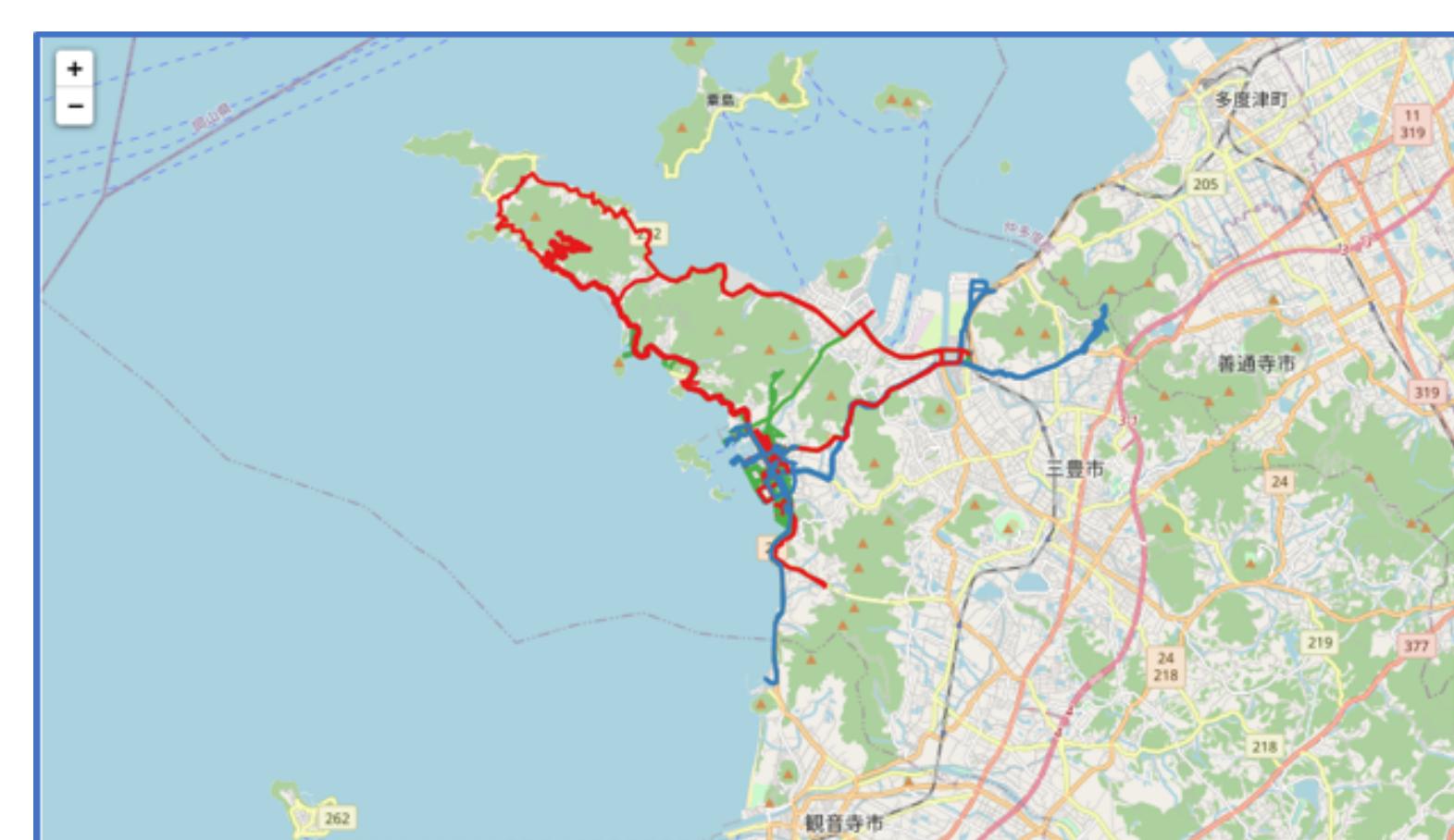
no	ID	他ID	DTW距離
1	1080131	1080113	1890970.8781618185
2	1080131	1080135	485594.30930427305
3	1080131	10240112	2049843.4412680108
:	:	:	:
1953	1114019	12120110	14116340.726371925

位置情報

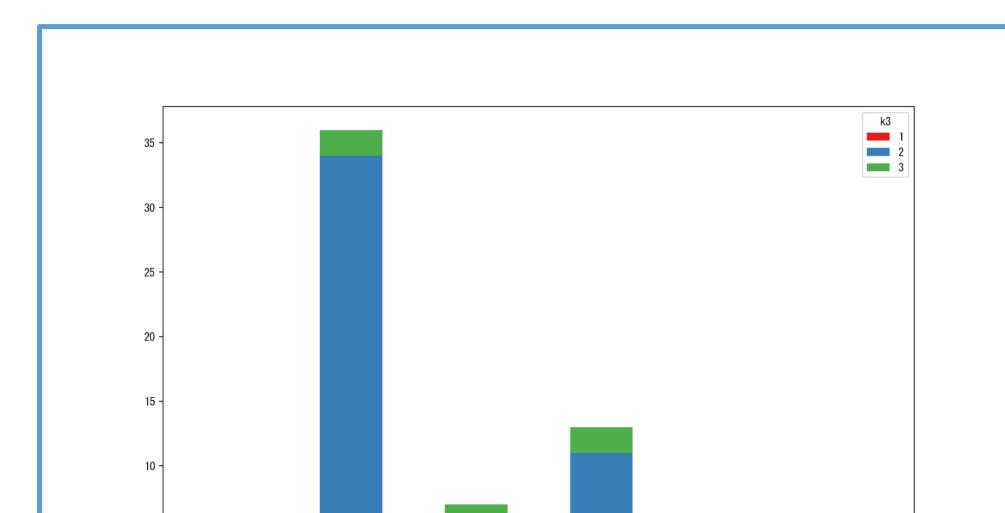
DTW距離



DTW距離によるクラスタリング



クラスタごとの移動経路



属性 P値

年代 2.7×10^{-5}

↓

年代とグループには関係がある

↓

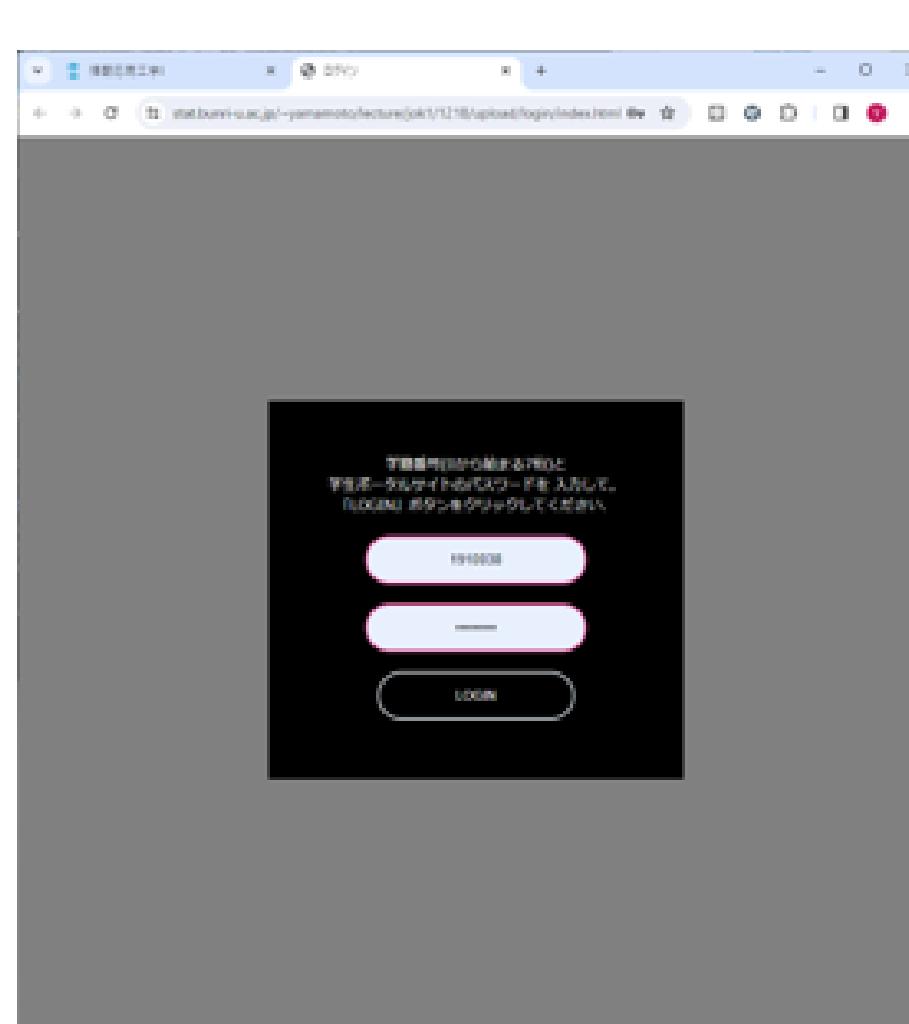
結果

	10代	20代	30代	40代	50代	60代以上
1	0	0	0	6	0	2
2	1	34	3	5	1	3
3	0	2	4	2	0	0

クラスタと属性の関係

課題解決型学習のためのWebアプリケーション

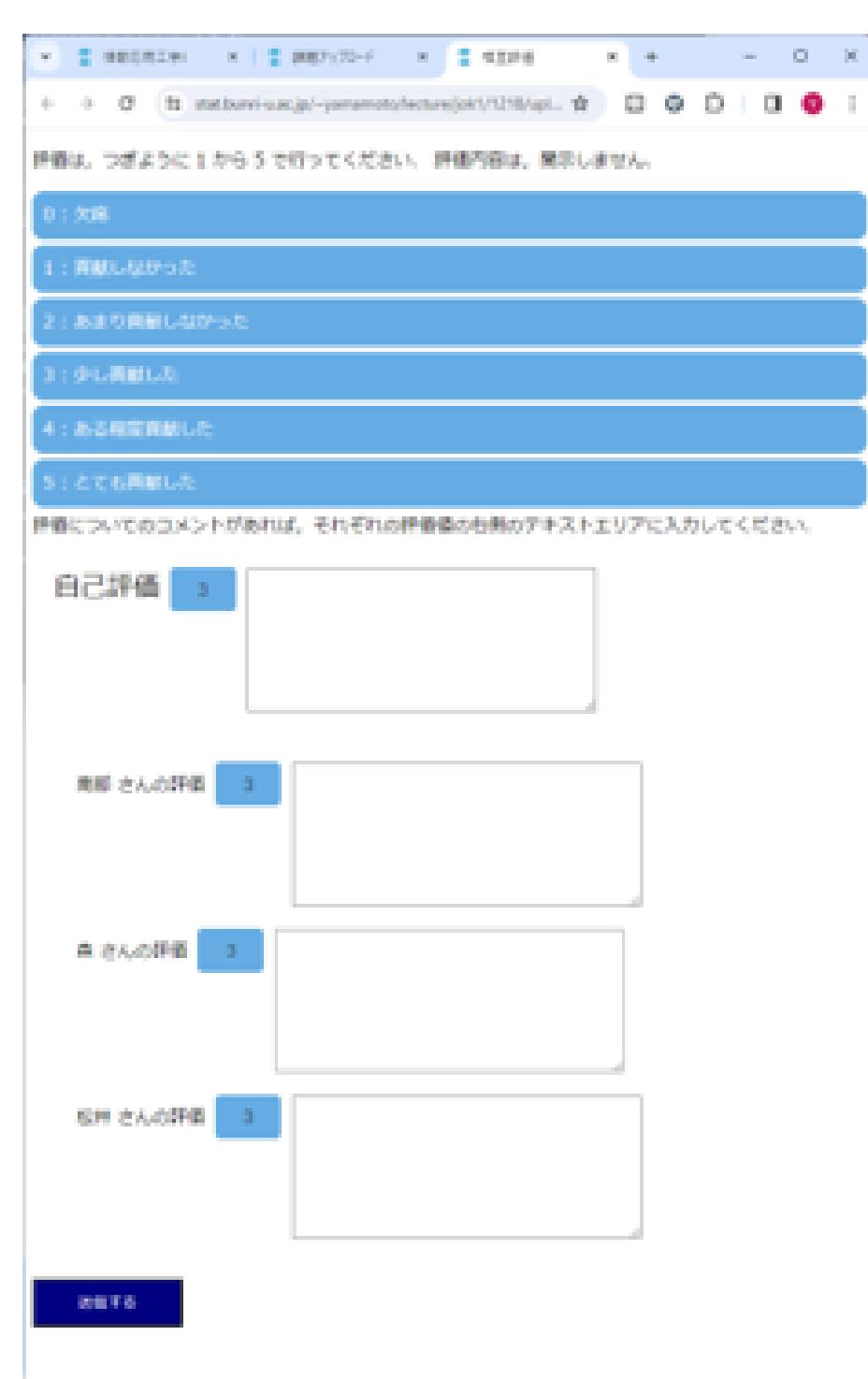
- グループでデータ分析を行い、結果を共有して、プレゼンテーション資料を作成するためのPHPとJavaScriptによるWebアプリケーション
- 機能
 - プログラム、結果、報告書の共有
 - 報告書の提出
 - 相互評価



LDAPによるログイン



報告書の共有と提出



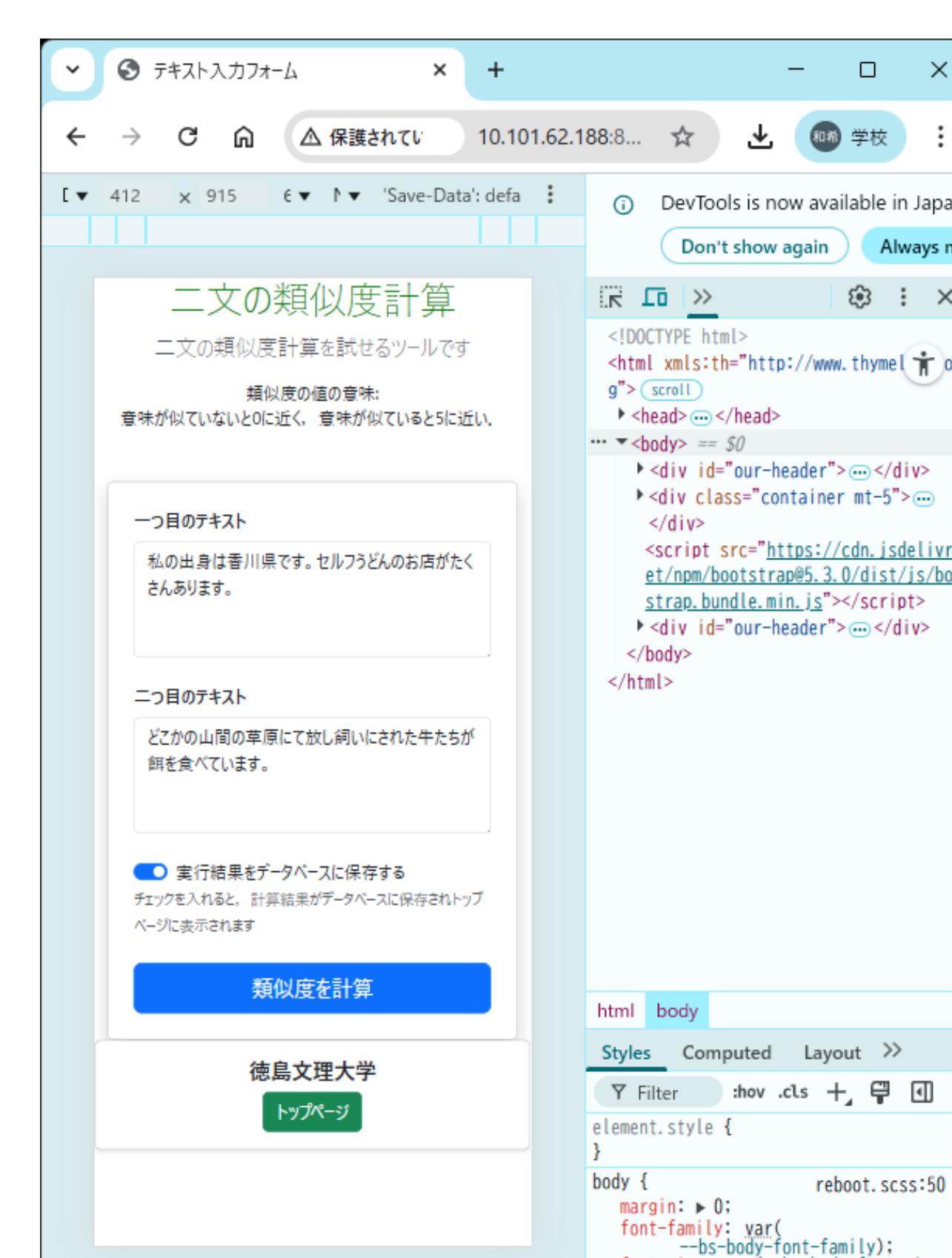
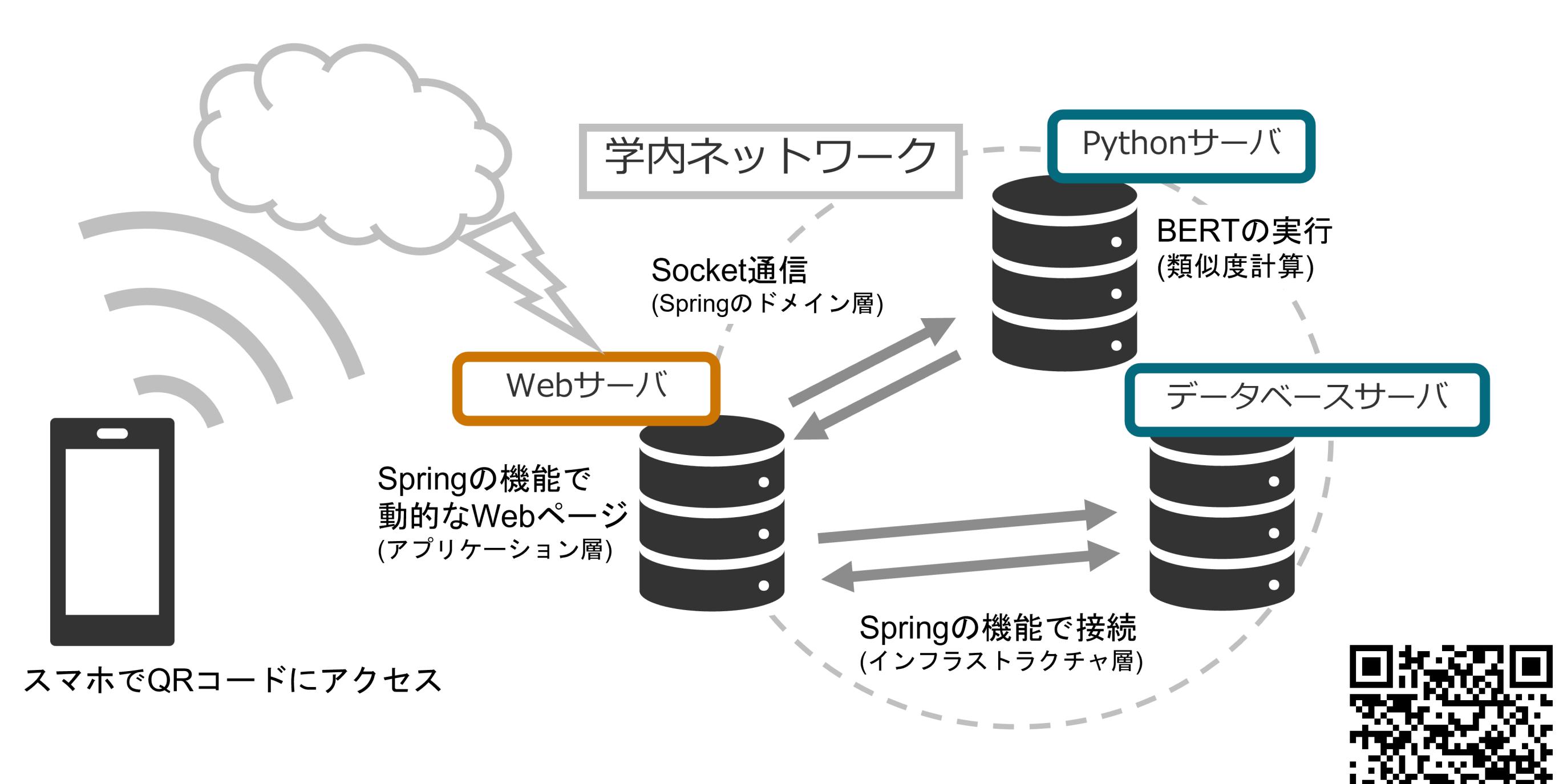
相互評価

大規模言語モデルを利用したテキストアナリティクスの例

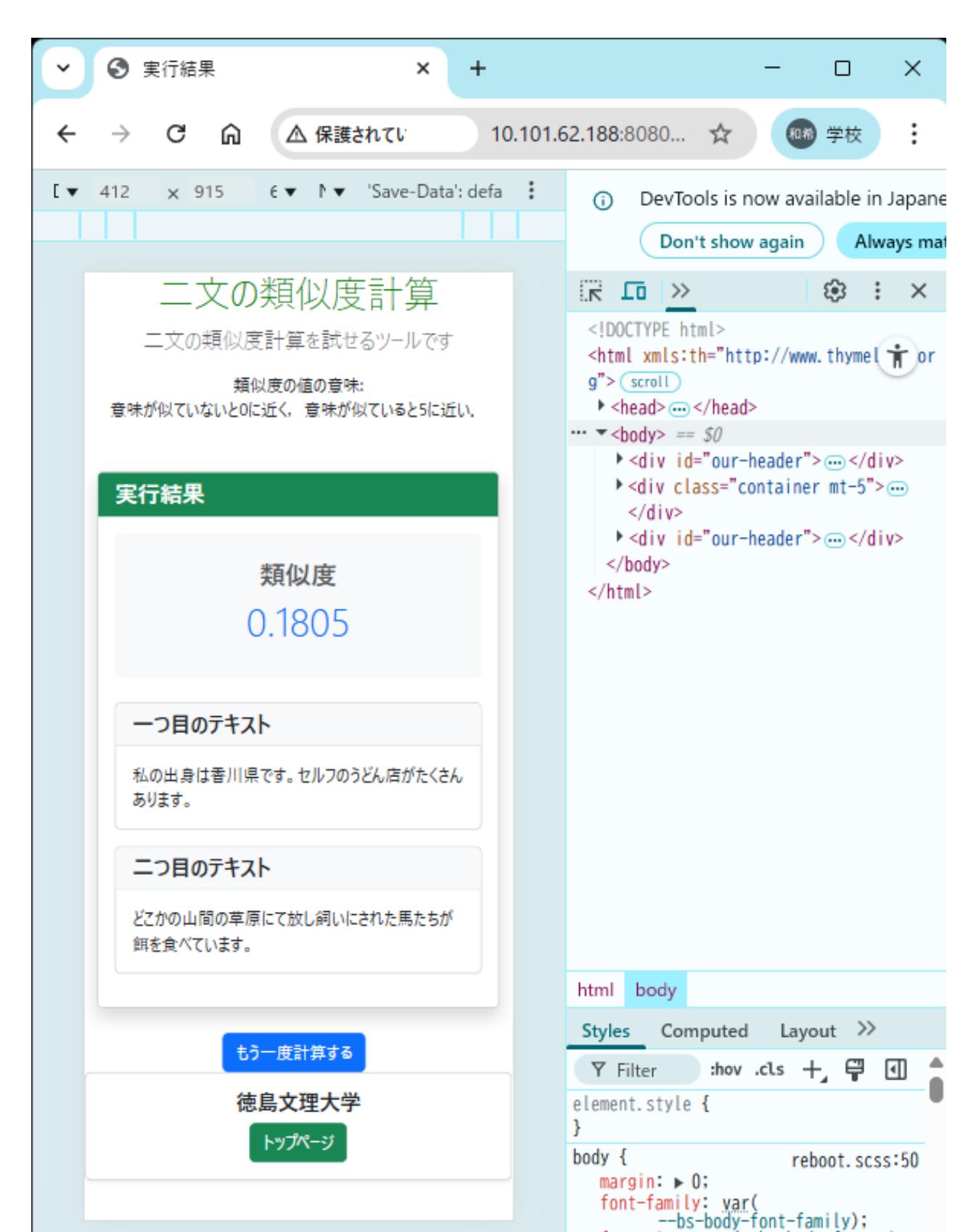
Webページに入力した2つのテキストの類似度をBERTを利用して計算
BERT (Bidirectional Encoder Representations from Transformers) は、Googleの研究者によって2018年に導入された言語モデルファミリーであり、自然言語処理 (NLP; Natural language processing) 実験で使用される。

2つのテキストの類似度を計算するソフトウェア

- Webページに2つのテキストを入力
- Spring Boot によって、入力されたテキストを取得
- Socket によって、Java のプログラムから Python サーバに2つのテキストを送信
- Python サーバで BERT によって、2つのテキストの類似度を計算
- 計算結果をデータベースサーバに保存
- Web ブラウザに結果を返す



2つのテキストの入力



類似度の表示