

# 信州大学シラバス検索用 ChatBotの開発

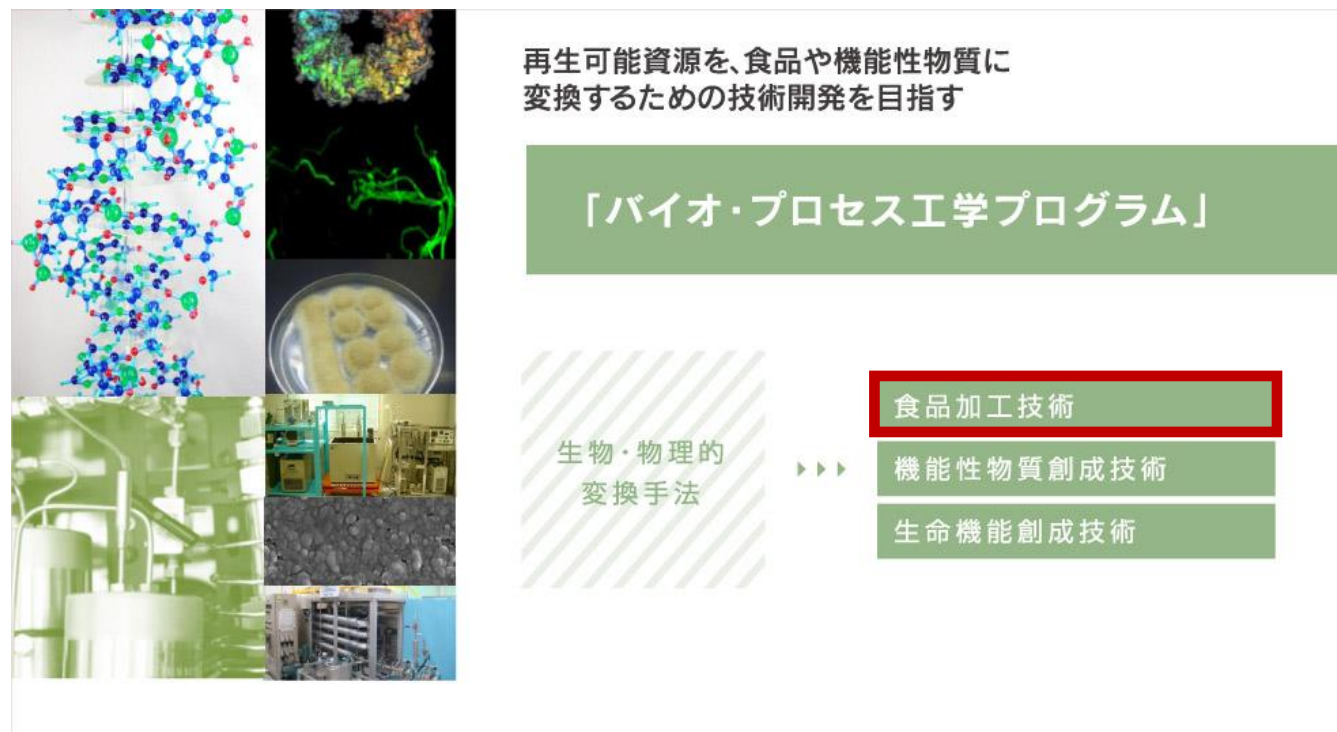
20T2095H 筑山 夢希翔

# シラバス検索システムの課題

## 直接シラバスに書かれている単語しか検索にヒットしない

例：食品加工技術について学べる科目が知りたい

A. 遺伝子工学、生物化学工学、酵素利用学など



# シラバス検索システムの課題

## 直接シラバスに書かれている単語しか検索にヒットしない

例：食品加工技術について学べる科目が知りたい

A. 遺伝子工学、生物化学工学、酵素利用学など

シラバス検索システムだと、「食品加工技術」で検索してもヒットしない

授業の内容・計画

食品加工技術

(授業の内容・計画の一部で検索します)



開講年度	開講期間	コード	授業名	教員氏名	曜日・時限	開講部局	開講場所	ダウンロード
------	------	-----	-----	------	-------	------	------	--------

この条件に該当するシラバスデータはありません。

[←検索TOPページへ](#)

[↑このページのトップへ](#)

## シラバス検索システムをChatBotとして実装すると・・・

- ・ あいまい検索ができる

→ シラバスに直接書かれていない単語でも、関連する科目を検索できる

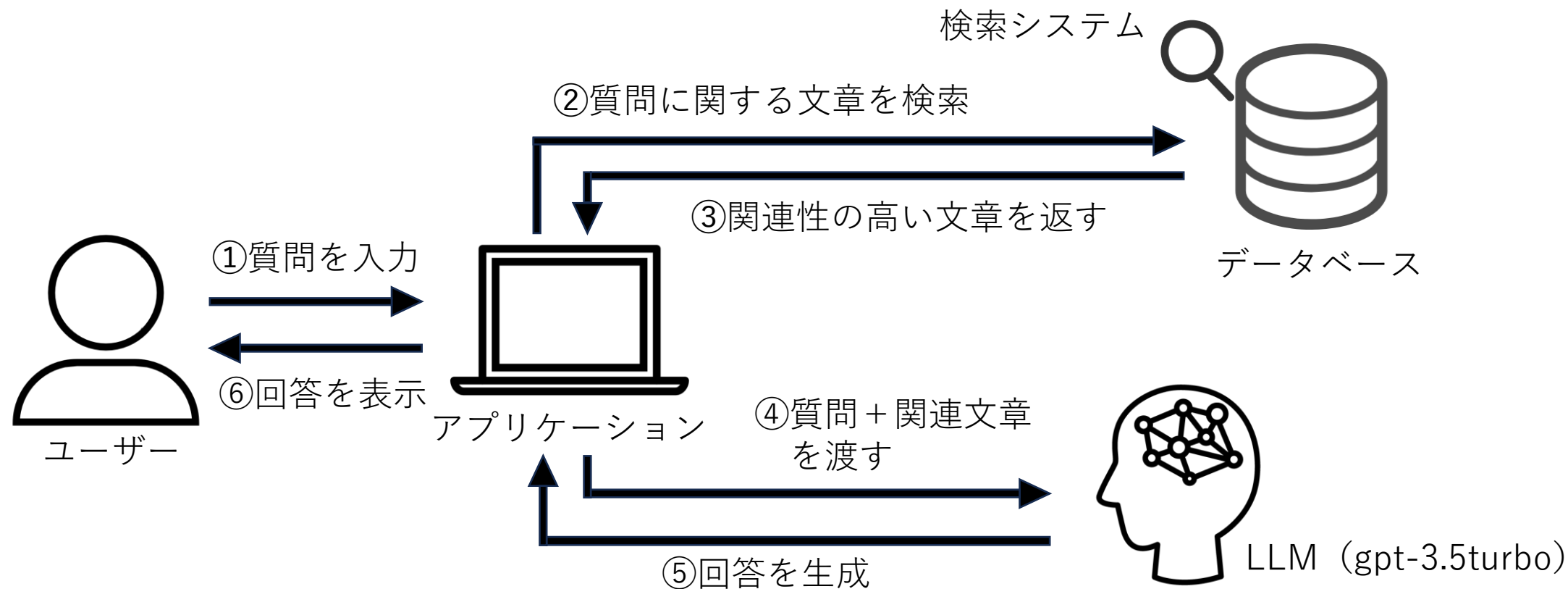
- ・ 情報取得の効率化が期待できる

→ 情報を探す時間を短縮できる

- ・ 対話形式の情報提供が可能になる

→ 自然な言葉で質問をして情報を取得できるようになる

# ChatBot



② エンベディングにかかる料金：0.02ドル/100万トークン

④ 入力にかかる料金：0.50ドル/100万トークン

⑤ 出力にかかる料金：1.5ドル/100万トークン

※ ひらがな：1～2トークン  
漢字：1～3トークン

# データベースの作成

シラバスから信州大学工学部の425科目のテキストデータを取得



「授業科目名」「登録コード」「単位数」「対象学生」  
「授業の概要」「授業計画」「成績評価の方法」  
「履修上の注意」  
のみ記述したテキストデータに整形



エンベディングでベクトル化し、各テキストデータの  
トークン数とベクトルデータをCSVファイルに保存

整形したシラバステキストデータ (例: データベース)

```
I【授業科目】
データベース(16T以降)
【登録コード】
T2059300
【単位数】
3
【対象学生】
電子情報システム工学科3年生
【授業の概要】
コンピュータを活用した高度な情報処理において中心的な役割を担うデータベースの理論を講義すると共に、オープンソースデータベースの基本概念のみならず、データベースの利用方法を対象としたデータ分析に取り組む。
【授業計画】
授業は以下のように進行する。
第1回: ガイダンス
第2回: 正規形と直積
第3回: 内部結合
第4回: Inline View
第5回: グループ化と集計
第6回: 外部結合
第7回: Case
第8回: 再帰
第9回: 相関クエリ
第10回: CAD/ACID
第11回: 試験の練習
第12回: 試験
第13回: データ分析の説明
第14回: データ分析の演習
第15回: データ分析のまとめと授業アンケートの回答
【成績評価の方法】
試験60%, データ分析レポート40%で成績を評価する。
【履修上の注意】
演習のため、学内ネットワークに無線接続可能なノートパソコンを1台以上持参すること。
```

# データベースの作成

シラバスから信州大学工学部の425科目のテキストデータを取得



「授業科目名」「登録コード」「単位数」「対象学生」  
「授業の概要」「授業計画」「成績評価の方法」  
「履修上の注意」  
のみ記述したテキストデータに整形



エンベディングでベクトル化し、各テキストデータの  
トークン数とベクトルデータをCSVファイルに保存

作成したデータベース

	A	B	C	D	E	F
1	content	n_tokens	embeddings			
2	【授業科目名】	1130	[0.004745927639305592, 0.033772598952			
3	【授業科目名】	1184	[0.0006407003966160119, 0.03626687452			
4	【授業科目名】	663	[-0.044434159994125366, 0.04823917895			
5	【授業科目名】	1680	[-0.010290670208632946, 0.06725294888			
6	【授業科目名】	1058	[-0.029775824397802353, 0.01779208146			
7	【授業科目名】	1166	[-0.004016796592622995, 0.05417417734			
8	【授業科目名】	751	[-0.01627153716981411, 0.076269485056			
9	【授業科目名】	1212	[-0.035571739077568054, 0.04550809785			...
10	【授業科目名】	1333	[-0.007260095328092575, 0.04621823504			
11	【授業科目名】	1100	[0.0059289224445819855, 0.03274044767			
12	【授業科目名】	1315	[-0.006091572344303131, 0.06194613873			
13	【授業科目名】	1318	[-0.005632823798805475, 0.06198139488			
14	【授業科目名】	1299	[-0.0032956176437437534, 0.0639334619			
15	【授業科目名】	1304	[-0.0042739855125546455, 0.0636999458			
16	【授業科目名】	1311	[-0.003163097659125924, 0.06409461051			
17	【授業科目名】	1315	[-0.005081960000097752, 0.06059921160			
18	【授業科目名】	1318	[-0.00434818584471941, 0.061185006052			
19	【授業科目名】	1299	[-0.0026867161504924297, 0.0613817162			
20	【授業科目名】	1339	[-0.0024684746749699116, 0.0631196945			

# 検索機能の実装

質問内容をエンベディングによりベクトル化し、データベースの各データとの距離をコサイン類似度で計算



合計3000トークンを越えるまで距離が近い順にコンテキストにテキストを追加していく



ChatGPTのAPIに作成したコンテキストを渡して回答を得る

作成したコンテキスト例(質問: データベースではどんなことを学びますか?)

```
【授業科目】
データベース(16T以降)
【登録コード】
T2059300
【単位数】
3
【対象学生】
電子情報システム工学科3年生
【授業の概要】
コンピュータを活用した高度な情報処理において中心的役割を果たすデータベースの基本概念のみならず、データベースの利用法について実習する。さらに、デ
【授業計画】
授業は以下のように進行する。
第1回: ガイダンス
第2回: 正規形と変換
第3回: 内部結合
第4回: Inline View
第5回: グループ化と集計
第6回: 外部結合
第7回: Case
第8回: 再帰
第9回: 相関クエリ
第10回: CAD/ACID
第11回: 試験の練習
第12回: 試験
第13回: データ分析の説明
第14回: データ分析の演習
第15回: データ分析のまとめと授業アンケートの回答
【成績評価の方法】
試験60%。データ分析レポート40%で成績を評価する。
【履修上の注意】
演習のため、学内ネットワークに無線接続可能なノートパソコンを毎回持参すること。
###
【授業科目】
データエンジニアリング基礎
【登録コード】
T8002200
【単位数】
1
【対象学生】
学部2・3年生
【授業の概要】
本授業では、データエンジニアリングに関する基礎的項目を学修範囲とする。具体的
(情報通信技術)の進展とビッグデータ、コンピュータでデータを扱うためのデータ
タ抽出方法、収集したデータの加工方法について学習する。これらデータ・AIの活用
【授業計画】
授業はすべてeALPSコース上でオンライン(非同期型)で実施する。
インターネット(クラウド)上で利用できる演習環境(Google Colab)も用意する。
☆:コア学修項目 ※:教養・データサイエンス・AIを学ぶ上で基礎となる学修項目。
単元1. ビッグデータとデータエンジニアリング(☆)
単元2. データ表現(☆)(演習)
単元3. データ収集(演習)
単元4. データベース(演習)
単元5. データ加工(演習)
単元6. ITセキュリティ(演習)
単元7. プログラミング基礎 前半(※)(演習)
単元8. プログラミング基礎 後半(※)(演習)
【成績評価の方法】
・講義内容に基づくオンライン型の課題(CBT)を課す。成績は、オンライン型の課題
【履修上の注意】
・講義・課題(CBT)・練習問題はすべてオンライン・非同期型で実施する。冒頭のガ
・各単元にはeALPSの小テスト機能を用いた課題(CBT)が設定されていて、かつ課題
###
```

```
【授業科目】
データサイエンス基礎
【登録コード】
T8001200
【単位数】
1
【対象学生】
【授業の概要】
本授業では、データサイエンスに関する基礎的項目を学習する。具体的には、デ
学やアルゴリズムの基礎も取り上げる。
本授業はすべてeALPSコース上でオンライン(非同期型)で実施する。インター
ム、および参考書などを用いて学習し、eALPSコースのすべての課題を受験する。
【授業計画】
☆:コア学修項目 ※:教養・データサイエンス・AIを学ぶ上で基礎となる学修項目
単元1 データ駆動型社会とデータサイエンス(☆)(演習)
単元2 分析設計(☆)(演習)
単元3 データ観察(演習)
単元4 データ分析(演習)
単元5 データ可視化(演習)
単元6 数学基礎(前半)(※)(演習)
単元7 数学基礎(後半)(※)(演習)
単元8 アルゴリズム(演習)
【成績評価の方法】
成績は、eALPSコース上の課題の受験により評価する。
【履修上の注意】
###
【授業科目】
データマイニング(16T以降)
【登録コード】
T2065400
【単位数】
2
【対象学生】
電子情報システム工学科3年生
【授業の概要】
本講義では、データマイニングに関するさまざまな解析方法の基礎について解説
【授業計画】
1. ガイダンス: 授業の進め方と導入
2. データマイニングとは?
3. 統計的手法 (1): 国庫分析
4. 統計的手法 (2): ロジスティック国庫分析
5. 統計的手法 (3): 主成分分析
6. クラス分類 (1): k近傍法
7. クラス分類 (2): 決定木分析
8. クラス分類 (3): サポートベクターマシン
9. クラスタリング (1): 群平均法(階層的)
10. クラスタリング (2): k-means法(非階層的)
11. クラスタリング (3): 自己組織化マップ
12. 機械学習: ニューラルネットワーク
13. データマイニングの応用 (1): テキストマイニング
14. データマイニングの応用 (2): ウェブマイニング
15. まとめ、授業アンケート
【成績評価の方法】
出席およびレポート課題の内容の総合点で評価する。90点以上をS、80~89点をA
【履修上の注意】
レポート課題に取り組むために、PCの準備をして講義に出席すること、
```



# 評価実験

実験目的：シラバス検索システムとオリジナルチャットボットの性能比較を行うために、それぞれの方法を使った場合の回答時間と正誤数を評価する

実験内容：同研究室の所属学生12人に協力してもらい、作成した8問の問題をシラバス検索システムとチャットボットを使って解いてもらった

Aグループ（6人）とBグループ（6人）に分かれて、下の表のように各問題を指定した方法を使って解いてもらう

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
シラバス検索システム	A	B	A	B	A	B	A	B
ChatBot	B	A	B	A	B	A	B	A

説明内容

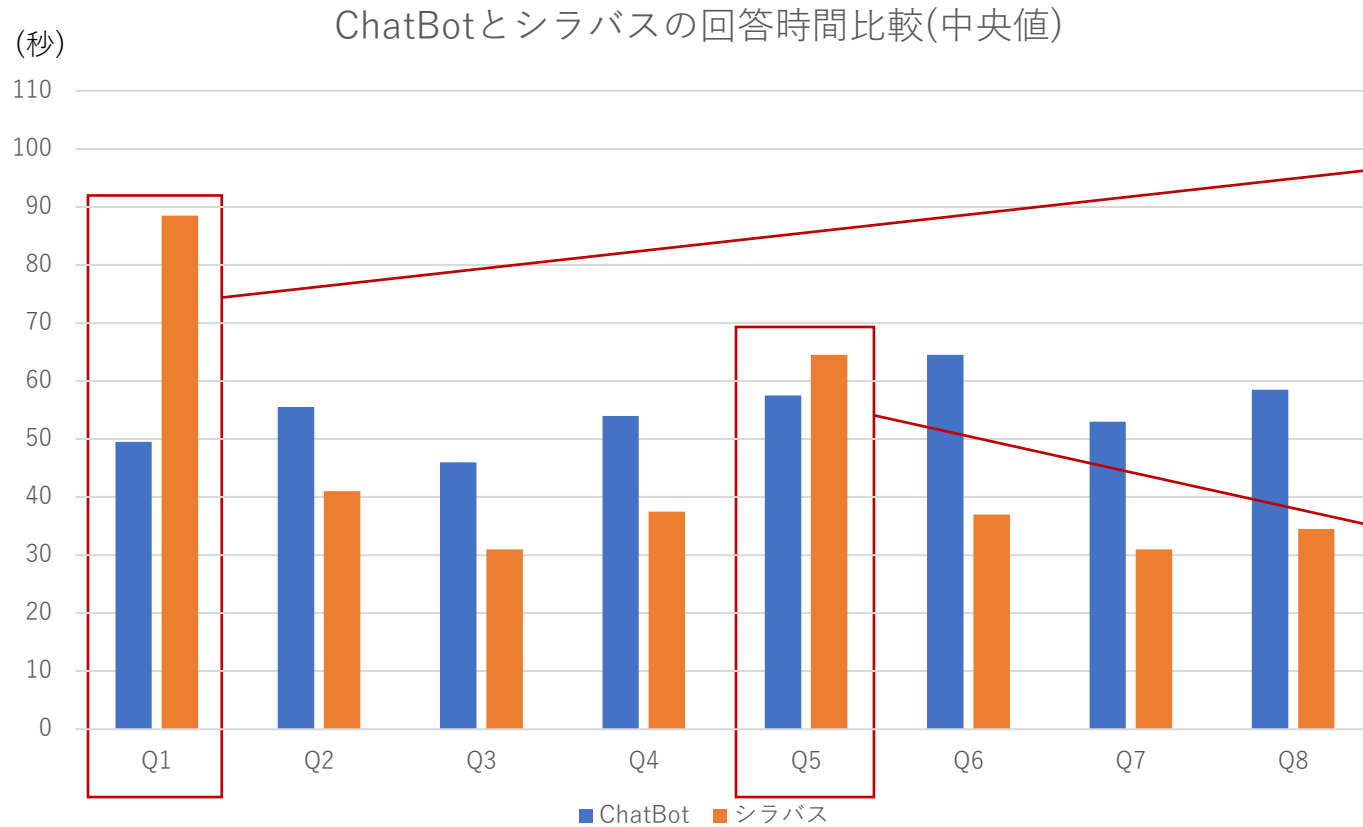
- ・ グーグルフォームを使う
- ・ 問題文のコピペ禁止
- ・ 問題に分からない言葉があったらブラウザで調べて良い
- ・ 時間の測定は、ストップウォッチを使う

# 問題

- Q1. Net Zero Energy Houseについて学べる科目は？
- Q2. 無機化学実験で持参するものは？
- Q3. 建築法規の成績評価方法は？
- Q4. 無機材料物性を履修する前に履修すべき科目は？
- Q5. 光化学の演習内容は？
- Q6. 化学工学と対で開講する科目は？
- Q7. 上下水道工学の担当教員の先生はどこで実務経験を積んでいる？
- Q8. 地盤の力学の出欠の確認は何で行う？

電子情報システム工学科以外の4学科で開講されている科目についての問題を作成した

# 回答時間比較



Q1

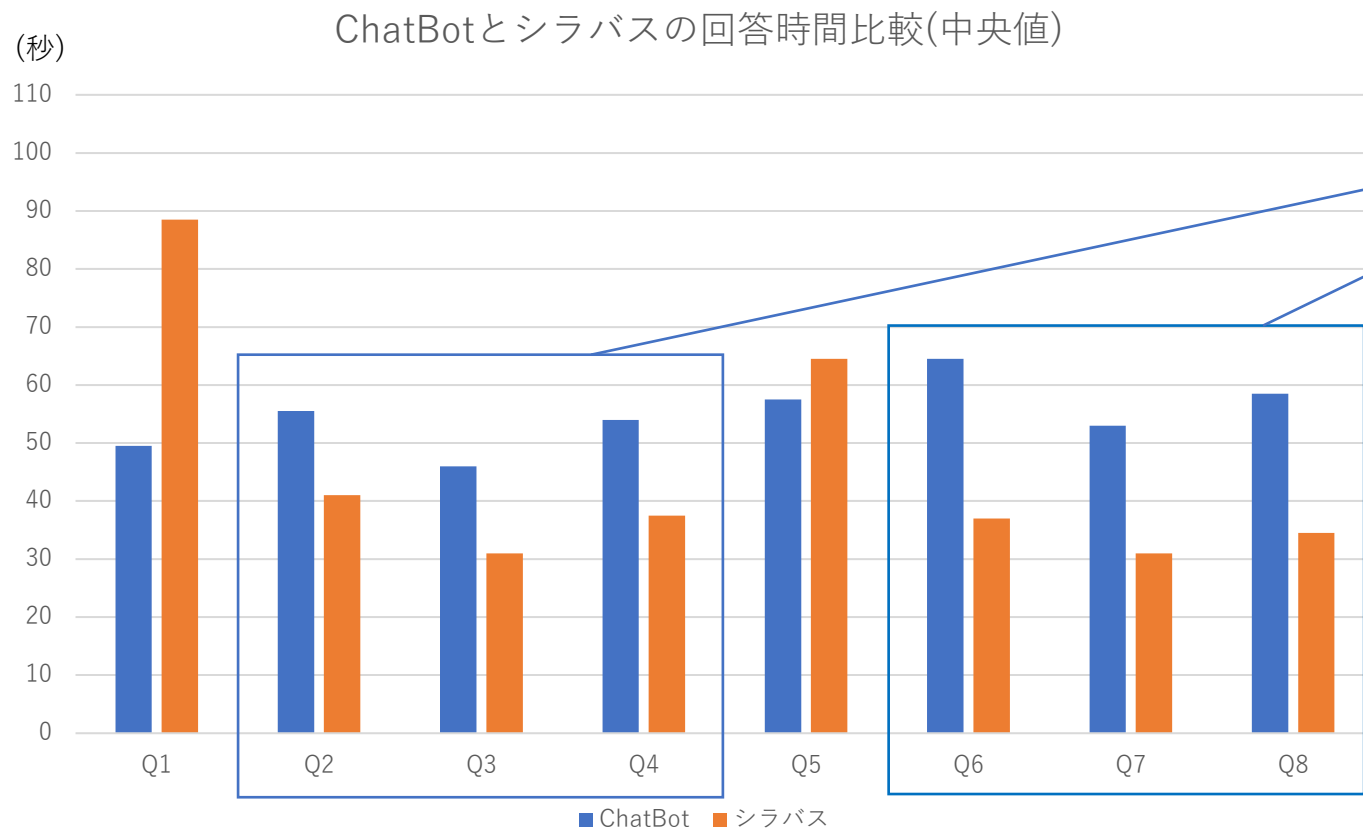
シラバスで検索しても  
ヒットしない単語がある  
問題

Q5

複数の回答が求められ、  
探すのが大変な問題

Q1, Q5ではChatBotの方が早く回答できている

# 回答時間比較



Q2~4, 6~8  
答えが一つで、  
見つけるのが簡単な問題

Q1, Q5以外ではシラバスの方が早く回答できている

# 正誤数比較

Q1	ChatBot	シラバス	Q2	ChatBot	シラバス	Q3	ChatBot	シラバス	Q4	ChatBot	シラバス
	○	○		×	○		○	○		○	○
	○	○		○	○		×	○		○	○
	○	○		○	○		×	×		○	○
	○	○		○	×		○	○		○	○
	×	×		○	○			○		○	○
	○	○		○			○	○		○	○
Q5	ChatBot	シラバス	Q6	ChatBot	シラバス	Q7	ChatBot	シラバス	Q8	ChatBot	シラバス
	×	×		○	○		○	○		○	○
	○	○		○	○		○	×		×	○
	×	×		○	○		○	○		○	○
	○	×		○	○		○	○		○	○
	×	×		○	○		○	×		○	○
	○	×		×	○		○	○		○	○

# 正誤数比較

Q1	ChatBot	シラバス	Q2	ChatBot	シラバス	Q3	ChatBot	シラバス	Q4	ChatBot	シラバス
	○	○		×	○		○	○		○	○
	○	○		○	○		×	○		○	○
	○	○		○	○		×	×		○	○
	○	○		○	○		○	○		○	○
	×	×		○	○		○	○		○	○
	○	○		○	○		○	○		○	○
Q5	ChatBot	シラバス	Q6	ChatBot	シラバス	Q7	ChatBot	シラバス	Q8	ChatBot	シラバス
	×	×		×	×		○	○		○	○
	○	○		○	○		○	×		○	○
	×	×		○	○		○	○		○	○
	○	×		○	○		○	○		○	○
	×	×		○	○		○	×		○	○
	○	×		×	○		○	○		○	○

ChatBotの方が正答数が多い問題

複数の回答が求められる問題で、見落としがあった

問題文の解釈を誤った

# 考察

## ・ 回答時間について

Q1に関して...

→ ChatBotを使うことでシラバスに書かれていない単語に対して曖昧な検索ができ、迅速な情報取得が可能になる

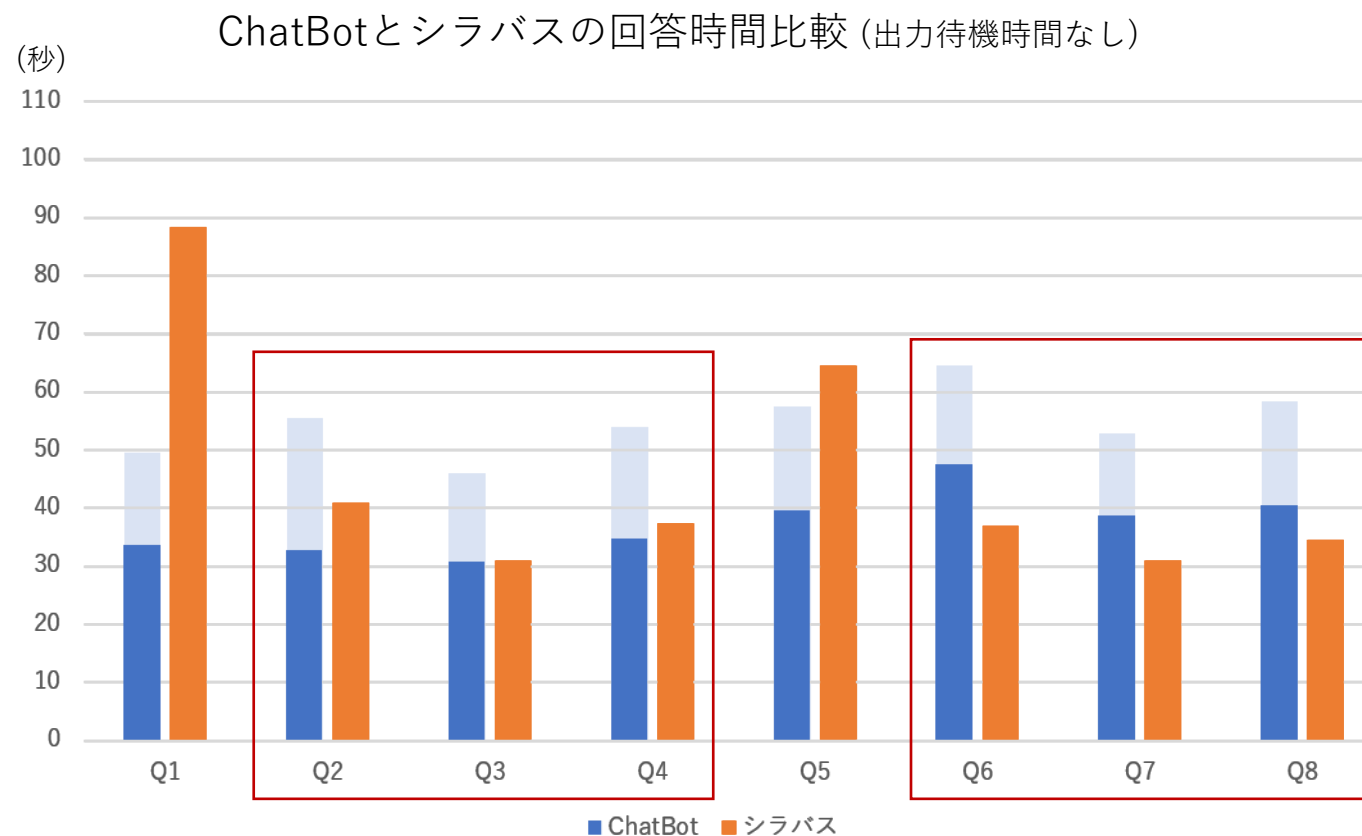
Q5に関して...

→ ChatBotは複数の情報を統合して提供できるため、情報を探す手間を省ける

→ ChatBotはシラバス検索システムに比べて回答時間が長くなる傾向があるが、出力待機時間(中央値16.2秒)が生じるためである

# 考察

- 回答時間について



出力待機時間を考慮しなければ、同じくらいの早さ



# 考察

- ・ 正誤数について

- ChatBotは機械的に処理を行うため、人為的ミスを削減できる

- 検索機能の性能が低いため、関連する単語や科目が多い場合、意図しない科目がヒットする可能性がある

# まとめ

## ChatBotを使用するのに適した状況

- ・ 曖昧な情報を基に検索する必要があるとき
- ・ 対話形式で情報を取得したいとき
- ・ 複数の情報源から集めた総合的な回答が必要なとき

## シラバス検索システムを使用するのに適した状況

- ・ 具体的な科目名で迅速に検索したいとき
- ・ 確実に正確な情報が必要なとき

## 1回の出力にかかるコストについて

$$\begin{array}{ccccccc} \text{入力料金} & & \text{出力料金} & & \text{入力料金} & \text{出力料金} & \\ 0.50 \text{ドル} / 1\text{Mトークン} & & 1.50 \text{ドル} / 1\text{Mトークン} & & & & 1 \text{ドル} = 154 \text{円} \\ 3500 \times 0.00000005 + 140 \times 0.00000015 = 0.00175 + 0.00021 = 0.00196 \text{ドル} = \underline{\underline{\text{約}0.30 \text{円}}} \end{array}$$

入力トークン数                      出力トークン数