

2021年度 第4回 SWIM研究会

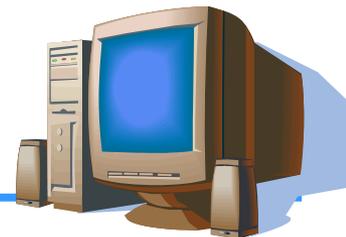
# ビッグデータとフォグノードの間の 循環型データモデルの提案

2022年 2月 18日

静岡理工科大学

情報学部 コンピュータシステム学科

工藤 司, 山本 剛大, 渡邊 智紀



## ■ 狙い: **ビッグデータ**を活用した, **個別業務システム**の**マスタ整備効率化**

1. まえがき
  2. 関連研究と本研究の狙い
  3. **巡回型データモデル**の提案
  4. **適用事例: 英語穴埋め問題自動生成システム**
  5. 実験と評価
  6. 考察
  7. まとめ
- 参考文献**

# 1. はじめに

## ■ 研究の背景:

### ➤ ビッグデータの状況

- ✓ 現在, 多様で膨大なデータが公開・共有
- ✓ データは個々の業務システムと整合しない

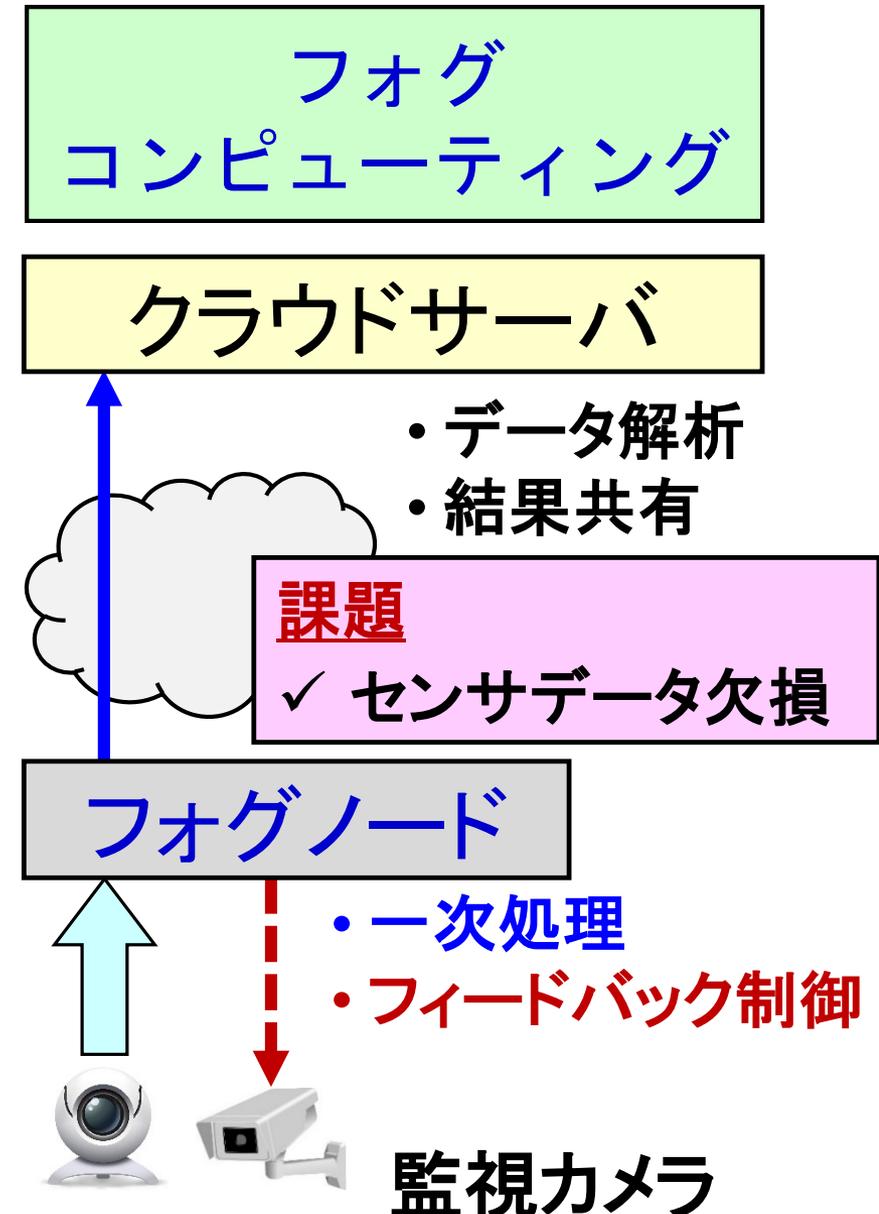
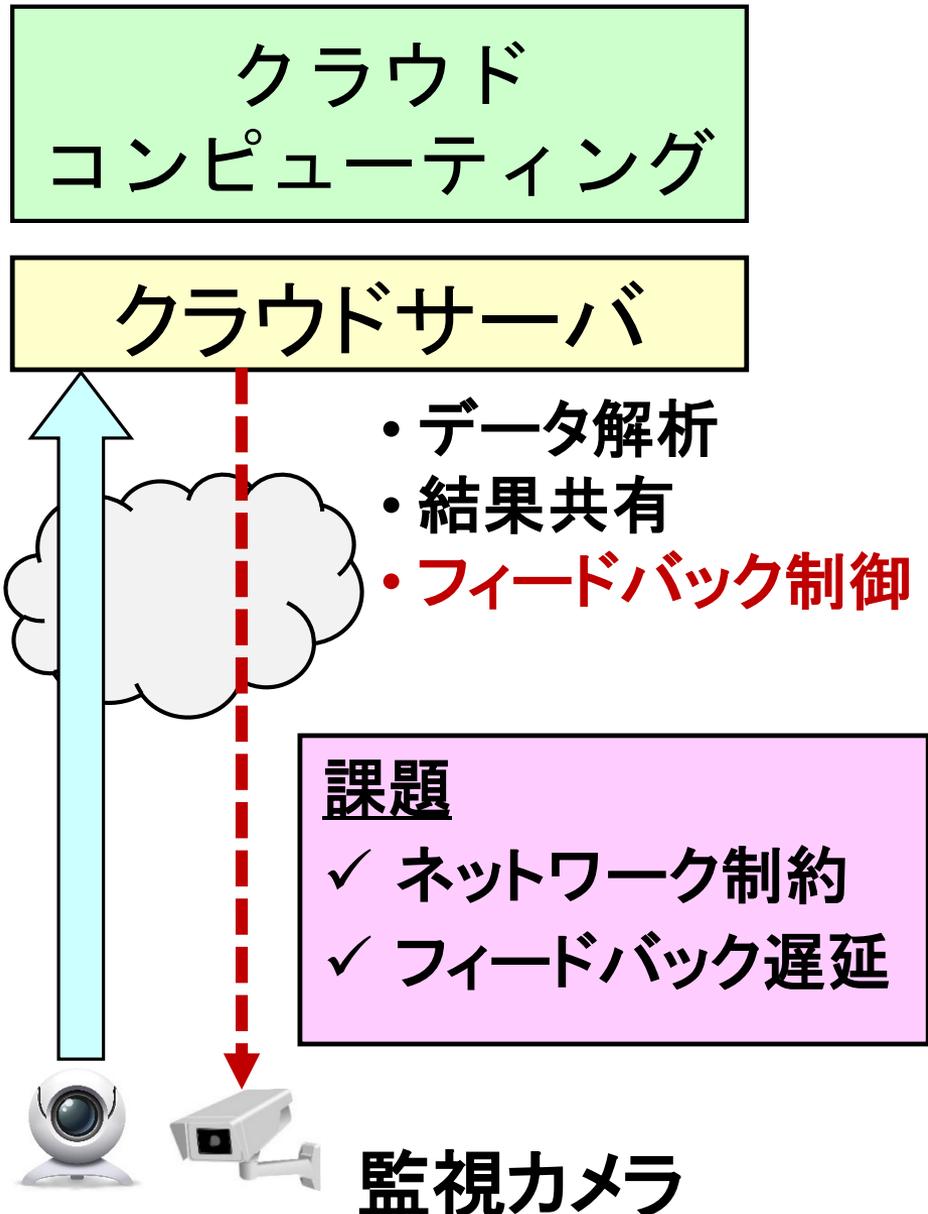
### ➤ 個々の業務システム開発における課題

- ✓ マスタ整備に大きなコスト

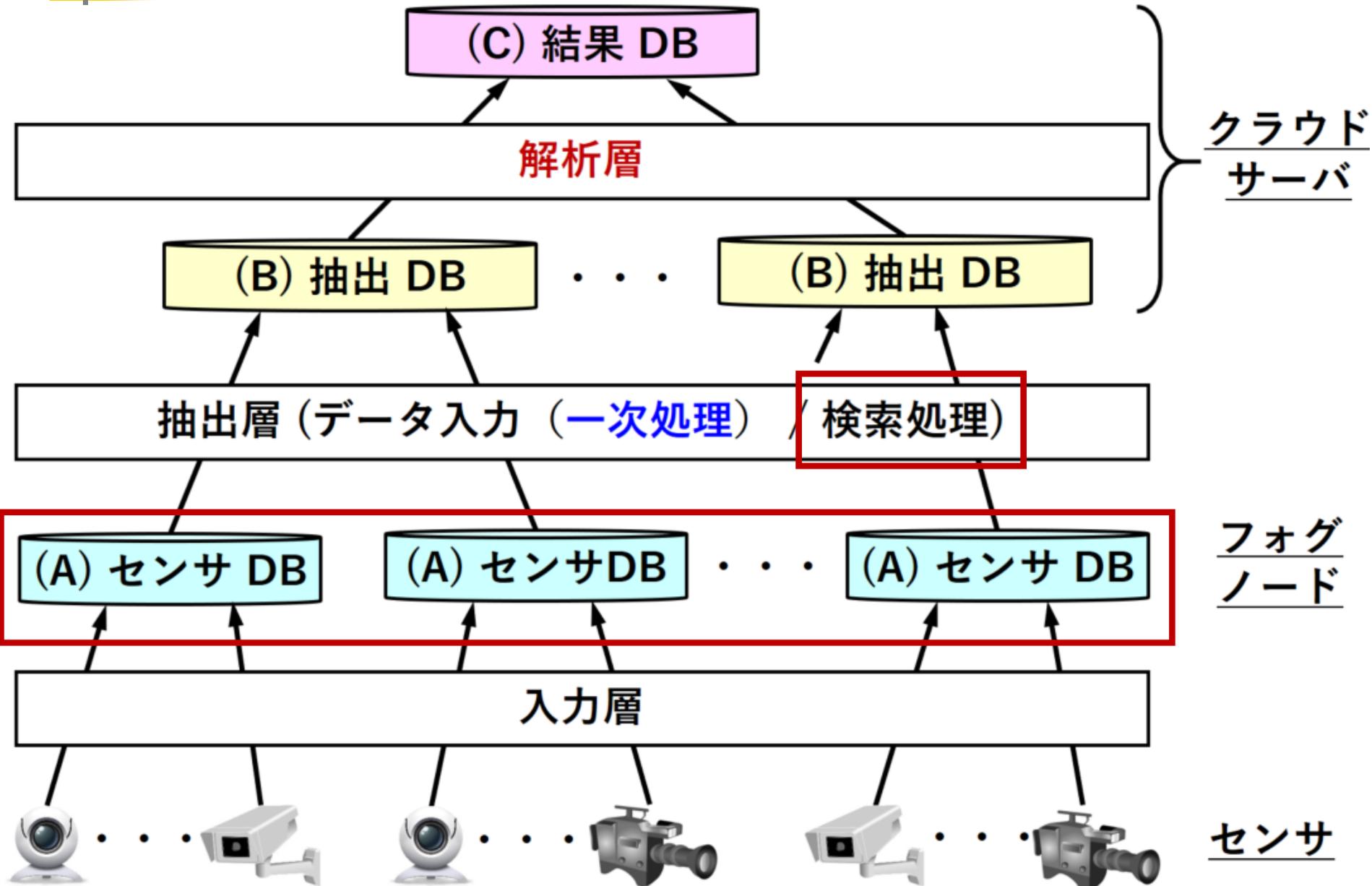
## ■ 研究の狙い

- ビッグデータ活用によるマスタ整備の効率化
- データの相互活用によるマスタの精度向上

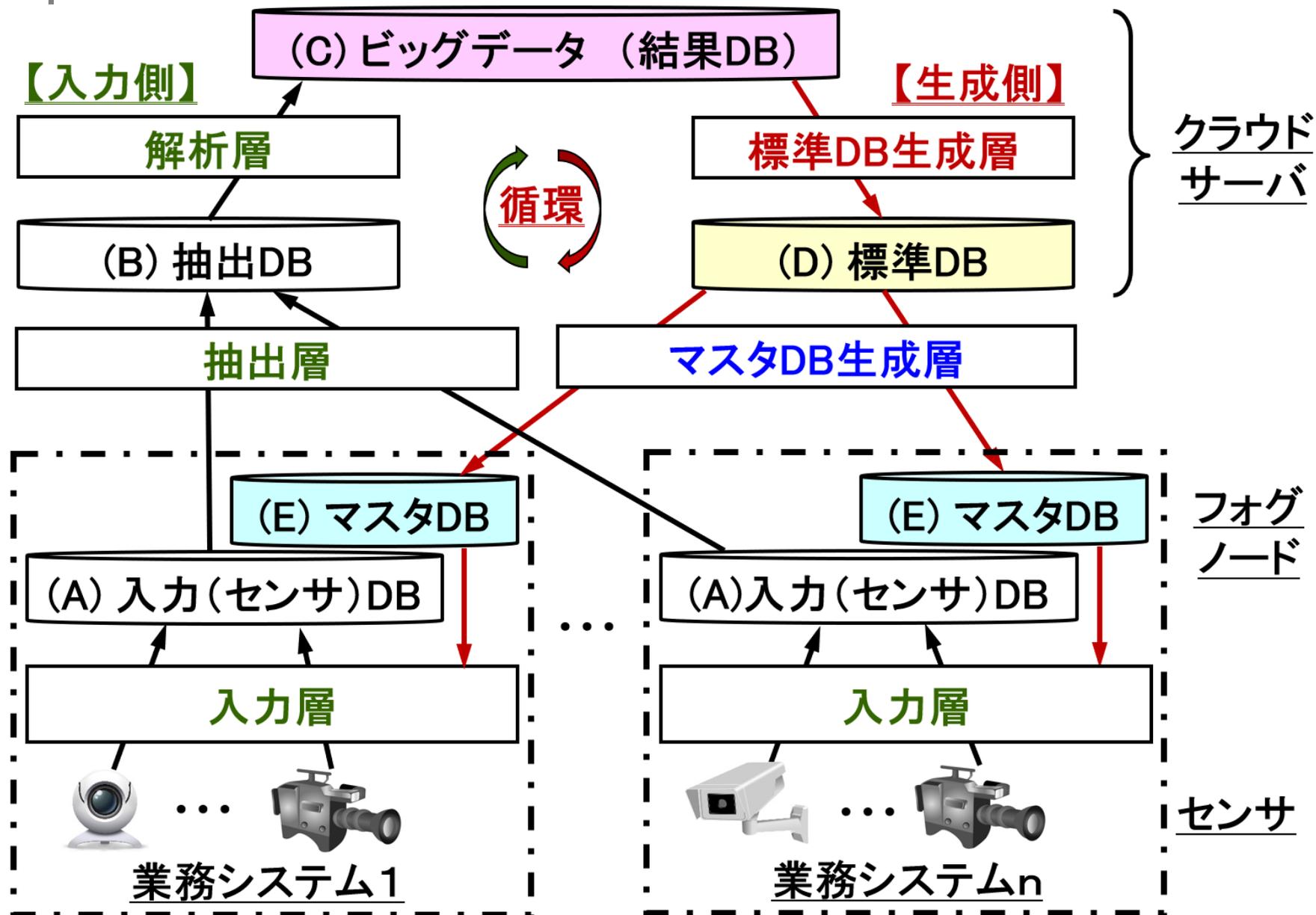
## 2. 関連研究: ビッグデータとフォグコンピューティング



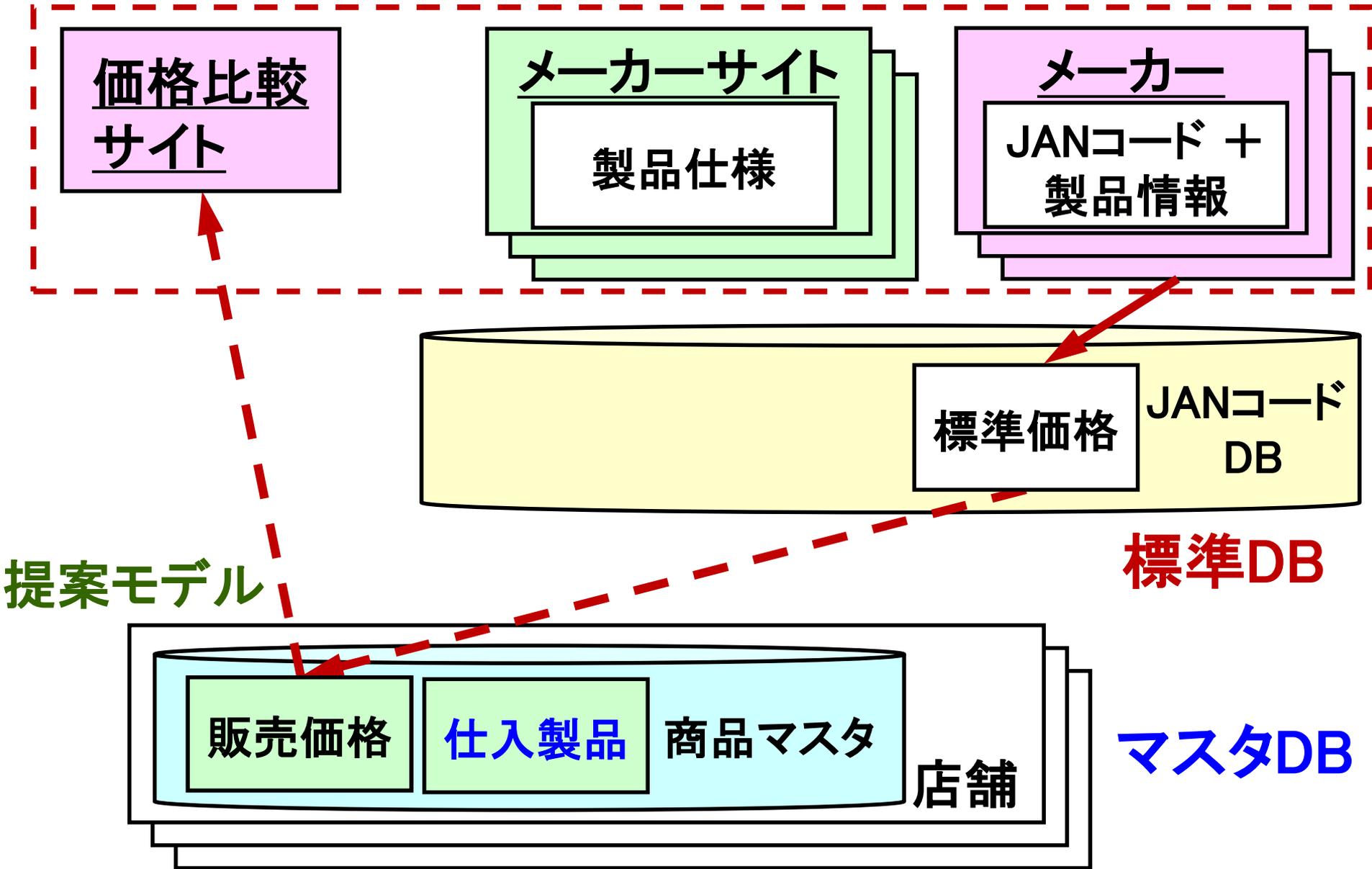
# フォグコンピューティング 3階層データモデル



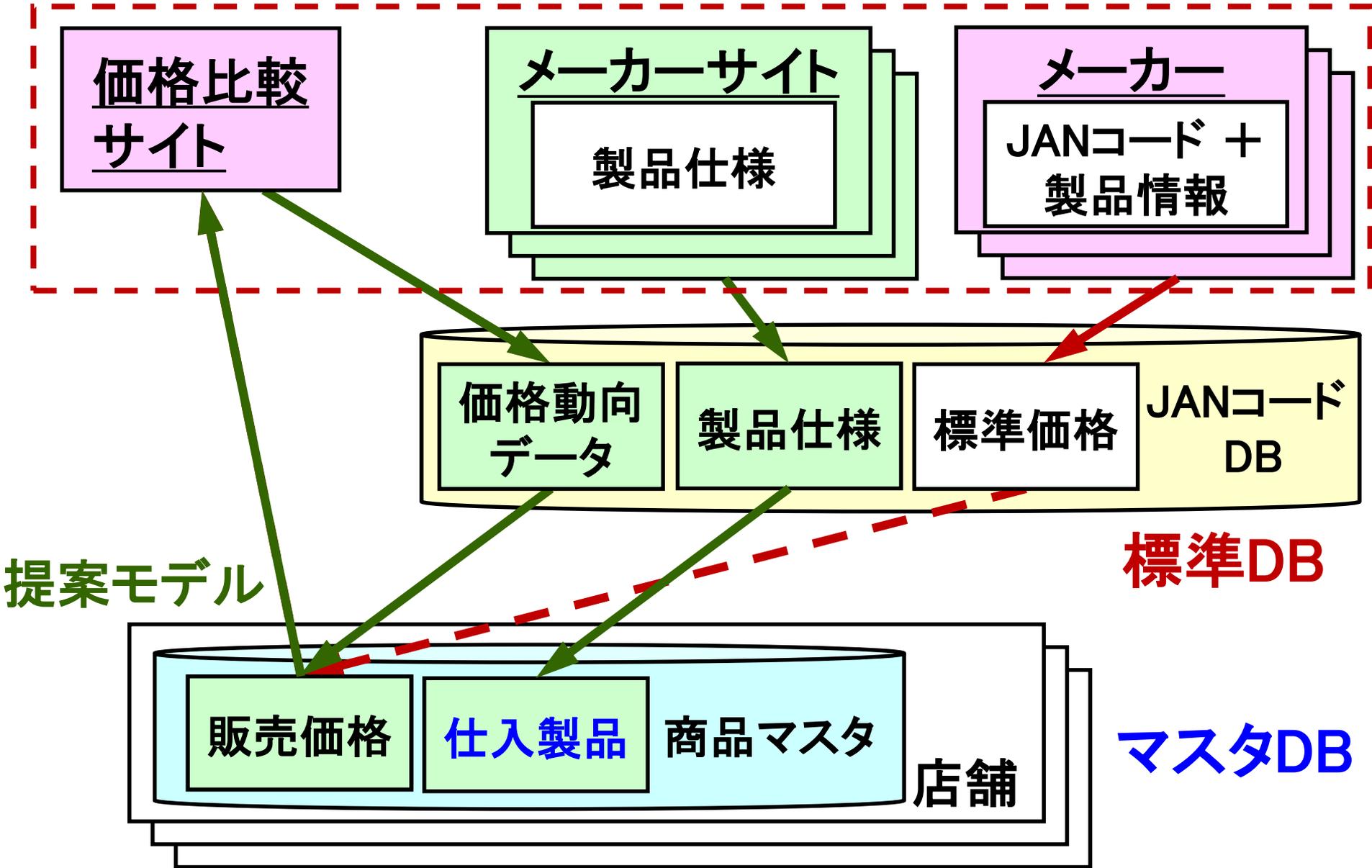
### 3. 循環型データモデルの提案



# オープン価格移行にともなうデータの流れ



# オープン価格移行にともなうデータの流れ



## 4. 適用事例：英語穴埋め問題自動生成システム

### ■ e ラーニング用 英語穴埋め問題の自動生成

- (1) 効率的な問題マスタの生成
- (2) 学習者のレベルに応じた問題の出題
- (3) 循環型データモデルによる, 問題レベルの精度向上

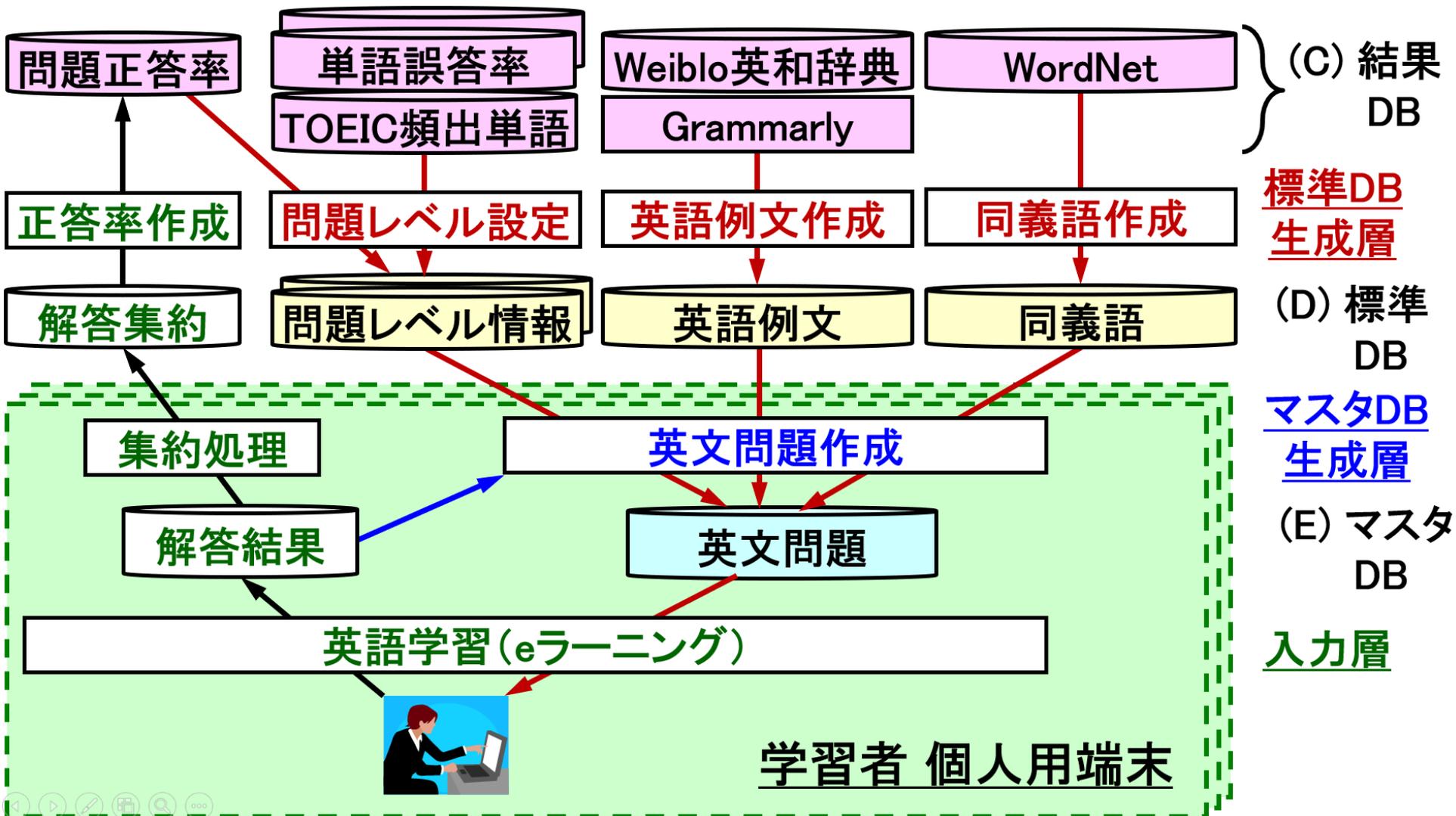
#### 【事例】

- **問題**： Will they  the people?
- **意味**： それらは人々に影響しますか？
- **選択肢**： like, affect, see, save

【参考情報】 正解：affect; **レベル 2 (準二級以上)**

# 英語例文作成の実装

## Webサイトのサービス



## ■プログラム言語

Python Ver. 3.7.10

## ■スクレイピング

Google Chrome Ver. 95.0.4638.69

Selenium Ver. 3.141.0 + Chrome Driver

## ■データベース

MySQL Ver. 8.0.22

MySQL Connector/Python Ver. 8.0.18



# Weblio英和辞典からの例文抽出(1)

**weblio** 英和辞典・和英辞典 1175万語収録!

英和和英辞典 英語例文 英語類語 共起表現 英単語帳 英語力診断 英語翻訳 オンライン英会話 スピーキング

ターボ検索 OFF ON  と一致する 項目を

Weblio 辞書 > 英和辞典・和英辞典 > 英和辞典 > brainの意味・解説

シェアする いいね! ツイート 広告を消す

意味 例文 (999件) 類語 共起表現

**brain** とは  
意味・読み方・使い方

発音を聞く プレーヤー再生 ピン留め 単語を追加 英会話で使う

意味・対訳 脳、脳髄、(食物としての)脳、(知性の中心としての)頭脳、知力、頭のいい人、秀才、知的指導者、ブレイン、(ミサイルなどの)頭脳部

音節 brain 発音記号・読み方 / bréɪn (米国英語), breɪn (英国英語) /

brainの  
変形一覧

動詞 : braining(現在分詞) brained(過去形) brained(過去分詞) brains(三人称単数現在)  
名詞 : brains(複数形)

レベル : 2 英検 : 準2級以上の単語

# Weblio英和辞典からの例文抽出 (2)

## 名詞

## 品詞

- ① (器官としての) 脳, 脳髄
- ② ((しばしば～s)) 頭脳, 知力

• a clear brain

明晰 (めいせき) な頭脳

例文

意味

• She has both beauty and a good brain.

彼女は才色兼備だ

• Use your brain.

頭を使いなさい

# Grammarlyによる英文チェック

- Grammarly: **英文校正サービス**
- 目的: **問題文として不適切な例文の除外;**  
**英語例文作成処理の検証.**
- 実装: Google Chromeのアドオンとして組み込み,  
Seleniumで操作を自動化
- 適切な例文の事例: ページにエラー情報を含まない.

The screenshot shows the Grammarly web interface. At the top, there are language selection menus for '言語を検出する' (Detect language) and '言語' (Language). The '英語' (English) menu is selected on the left, and the '日本語' (Japanese) menu is selected on the right. Below the menus, there are two text input fields. The left field contains the English text 'Will they affect the people?' and has a green 'G' icon below it. The right field contains the Japanese translation '彼らは人々に影響を与えますか?' and a star icon. Below the Japanese text, the romanized version 'Karera wa hitobito ni eikyō o ataemasu ka?' is visible. At the bottom of the interface, there are icons for microphone, speaker, and a character count '28 / 5,000'.

# Grammarlyによる英文チェック

## ■エラーを検知した事例: エラー情報の表示

言語を検出する 英語 日本語 韓国語 ↔ 日本語 英語 韓国語

be lacking in affect ×

感情に欠けている ☆

Kanjō ni kakete iru

20 / 5,000

**weblio** 1174万語収録! 英和和英辞典 英語例文 英語類語 共起表現 英単語帳 英語力診断 英

ターボ検索  OFF ON

**affect** ×

「affect」を含む例文一覧

例文

**be lacking in affect**

情動に欠ける。 - 研究社 新英和中辞典

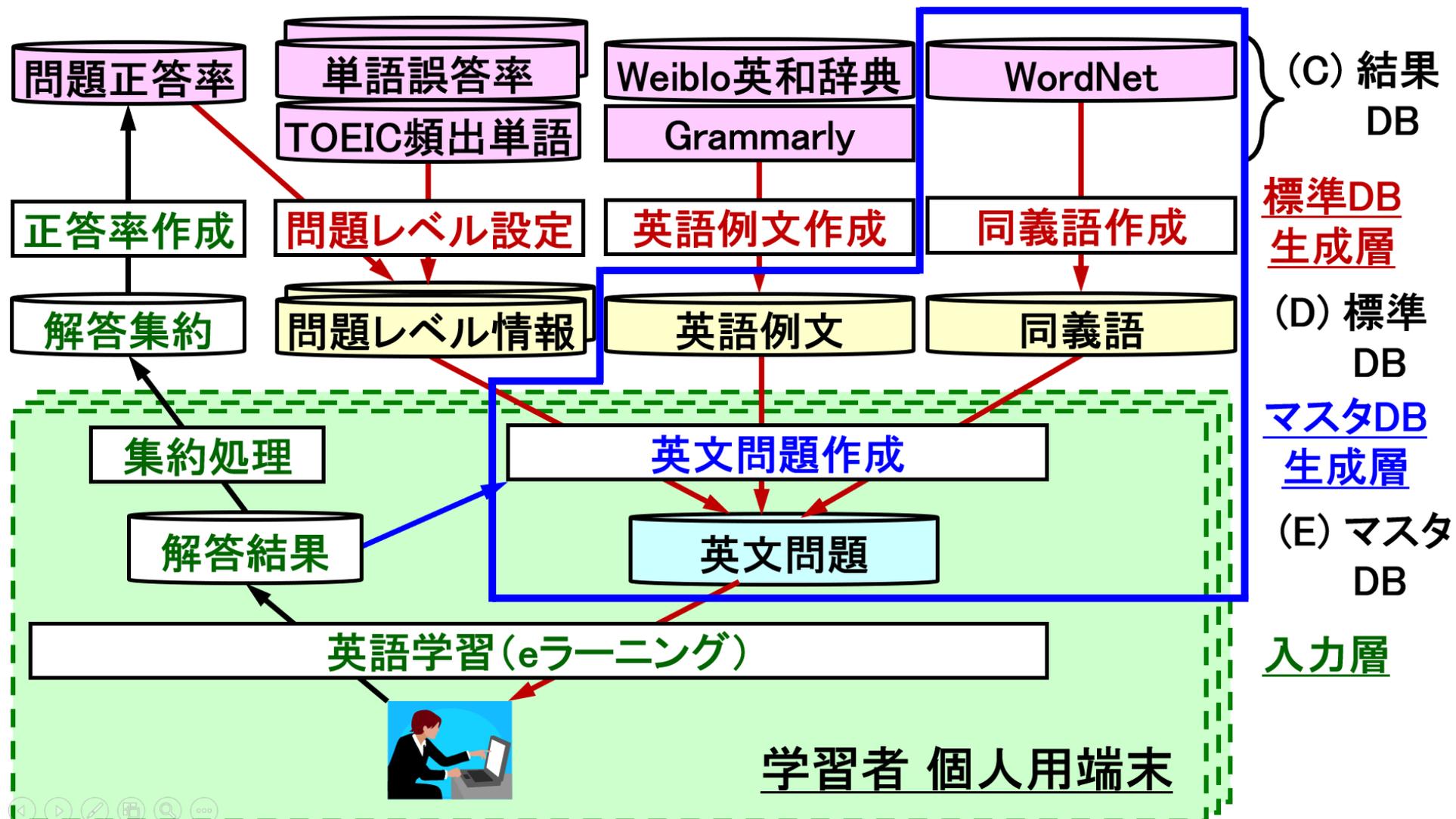
# 英語例文テーブルの構成

## ■ 英文例題テーブル (標準DB)

問題番号	単語	品詞	難易度	英文	和訳
14240	brain	1	2	a clear brain	明晰（めいせき）な頭脳
14241	brain	1	2	She has both beauty and a good brain.	彼女は才色兼備だ
14242	brain	1	2	Use your brain.	頭を使いなさい

1. 全てのユーザについて、共通のデータとして使用.
2. 英文を穴埋めにし、選択肢を付加して問題を作成.

## 2. 英文問題作成



# テーブルの構成

## ■ 英文例題テーブル(標準DB)

問題番号	単語	品詞	難易度	英文	和訳
14242	brain	1	2	Use your brain.	頭を使いなさい



難易度, 問題数

## ■ 英文問題テーブル (マスタDB: ユーザ毎に生成)

問題番号	単語	難易度	問題
14242	brain	2	Use your ____ .(頭を使いなさい)

(下行へ続く)

選択肢 1 (正解)	選択肢 2	選択肢 3	選択肢 4	字数
brain	overdue	bays	worthwhile	22

1. 指定の「難易度」以下の**英文例題**を, ランダムに選択.
2. 「**選択肢**」を追加 (単語以下の難易度), **同義語は除外**.

# 同義語の選択肢からの除外方法

■ WordNet: 英語語彙データベース, 同義語グループを定義.

■ 実装: MySQLで, 同義語検索ビューとして実装

➤ 単語(選択肢)をキーに, 同義語を検索.

単語	グループ	品詞
brain	1315218	2
brain	1389468	2
brain	5413129	1
brain	5541806	1
brain	5548525	1
brain	7547610	1
brain	9976200	1

単語	同義語	グループ	品詞
brain	encephalon	5413129	1
brain	head	5541806	1
brain	mind	5541806	1
brain	nous	5541806	1
brain	psyche	5541806	1
brain	brainpower	5548525	1
...			

■ WordNetの同義語分類

■ ビューとしての実装

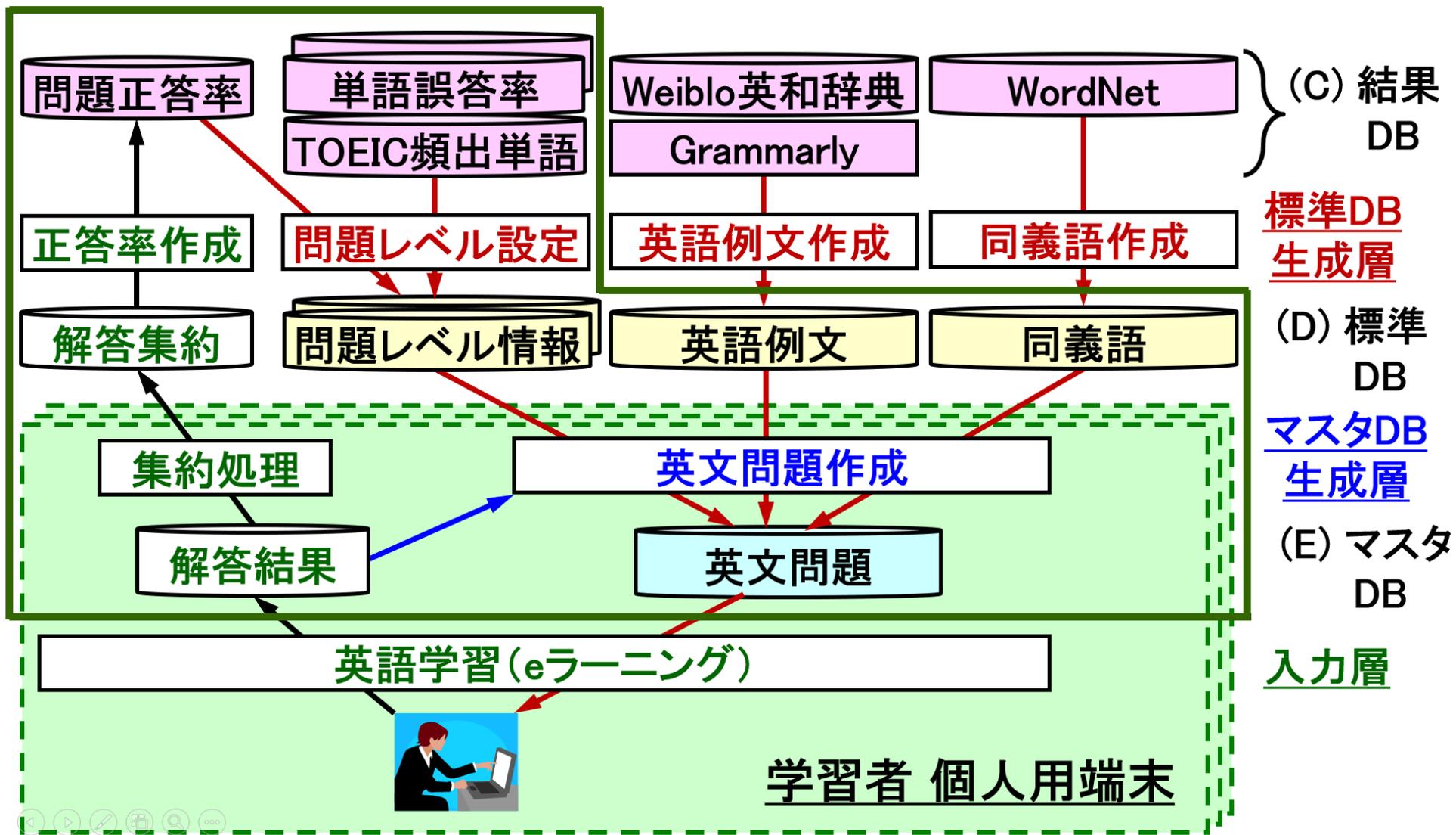
# 入力層の実装:eラーニング

- 英語学習支援システム(プログラム言語: HSP Ver. 3.6)
  - シューティング・ゲームとして実装, 学習意欲向上を狙う.



# 循環型データモデルの実装

## 3. 循環型データモデル



# テーブルの構成

- 仮説: 英文問題の難易度は、単語の難易度と整合しない。
  - 問題ごとの正解実績反映で、難易度の精度を向上

## ■ 解答結果テーブル (個々の正答／誤答)

学生番号	実行日時	出題番号	問題番号	難易度	結果
1	20210706100000	1	3199	2	1
1	20210706100000	2	3054	4	0
1	20210706100000	3	12278	5	1

正答  
誤答  
正答



問題ごとに集計

## ■ 問題正答テーブル (ビッグデータへの反映)

問題番号	レベル	単語	品詞	難易度	正答数	誤答数
14242	1	brain	1	2	528	956
14242	2	brain	1	2	413	202
13591	2	active	1	2	614	677
26638	3	basis	1	2	76	188

## 5. 実験と評価

### ■ 以下の実験により、提案方式の有効性を評価

#### 1. マスタ生成効率：英語問題マスタの生成

- ✓ 手作業と効率を比較評価

#### 2. マスタの有効性：

- ✓ 34,550問の英語問題を作成
- ✓ 英語学習のシューティングゲームで使用

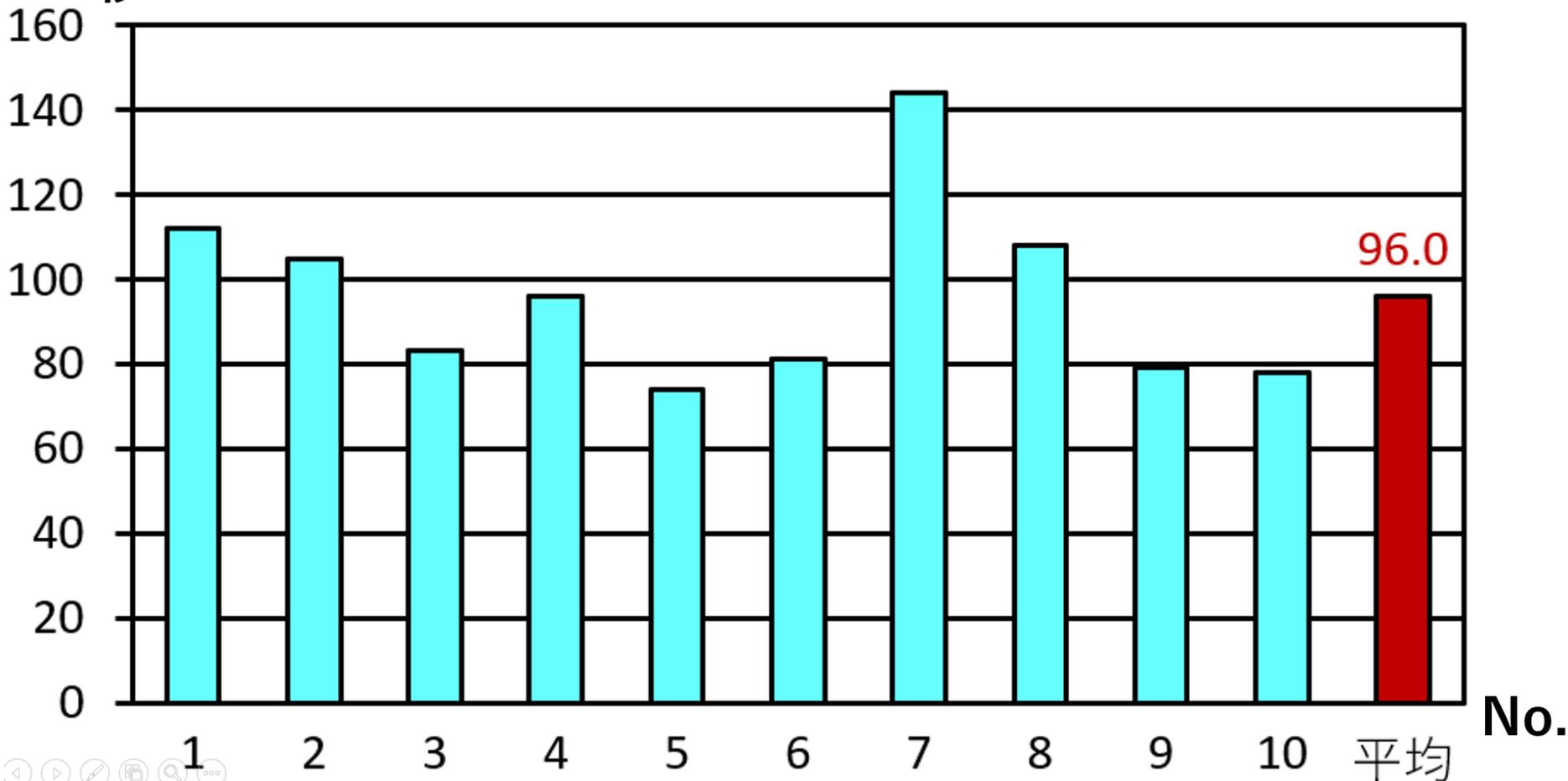
#### 3. 循環型データモデル：

- ✓ 英語問題と単語難易度の相関を評価

# 手作業による英語問題作成時間の評価

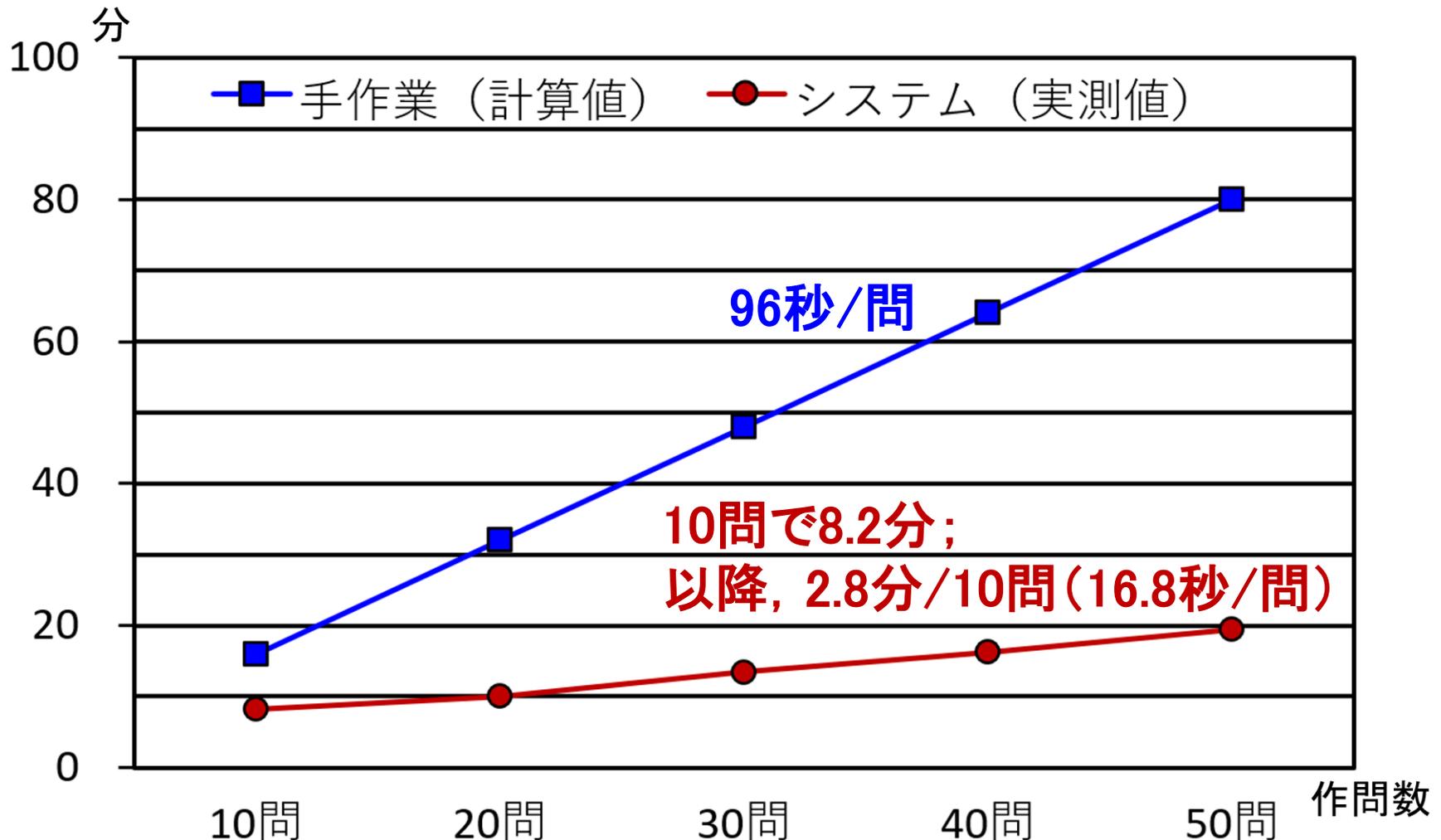
- 提案方式(自動)と同様の手順を手作業で実行
- 問題作成時間(例題+選択肢): 平均 96秒/1問

秒



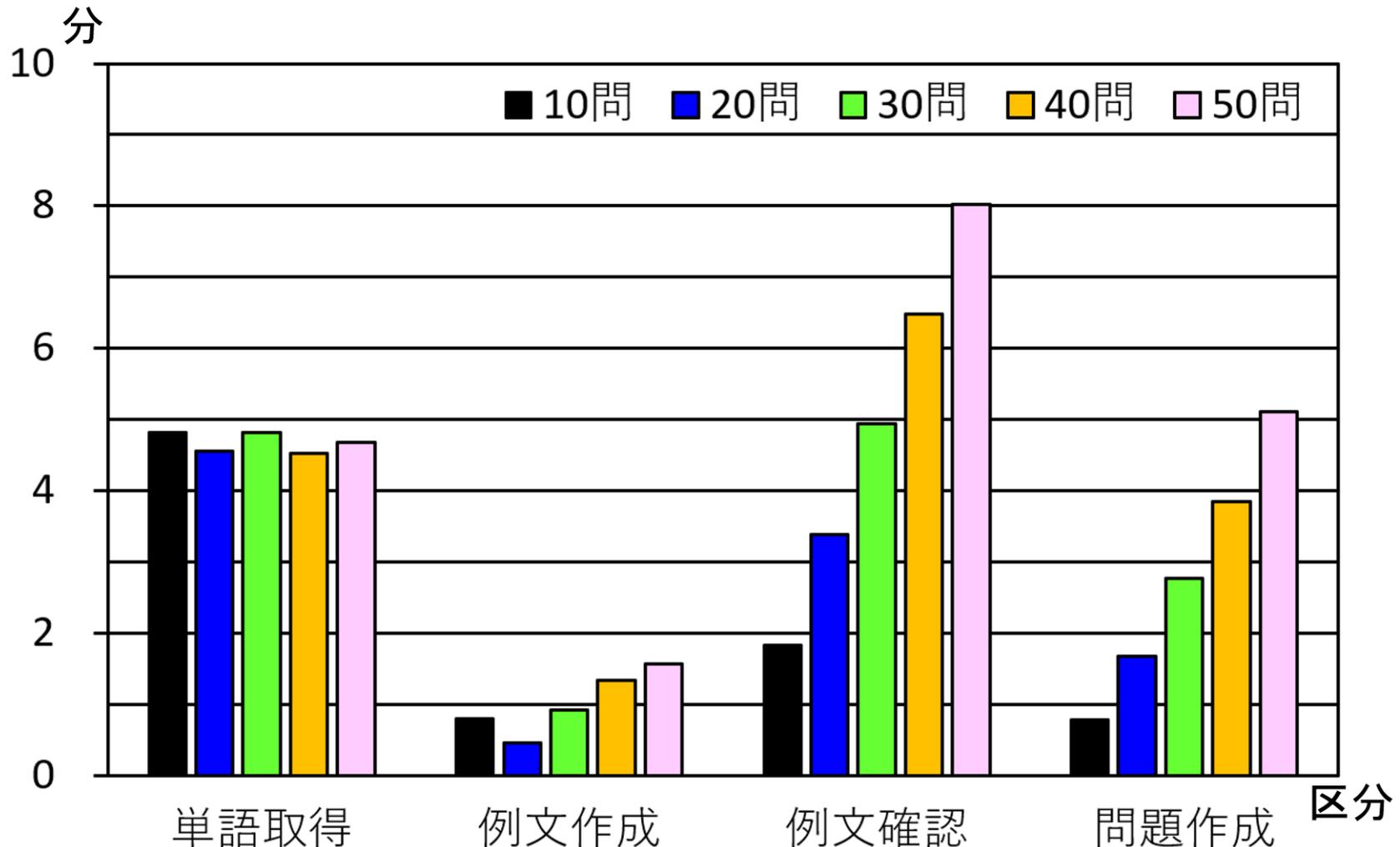
# 手作業と提案方式の問題作成時間 比較評価

- **提案方式**は**手作業**に比較し、**5.7倍**の効率.
- Webサイト負荷抑制のため待ち時間. 起動した後は**自動処理**.



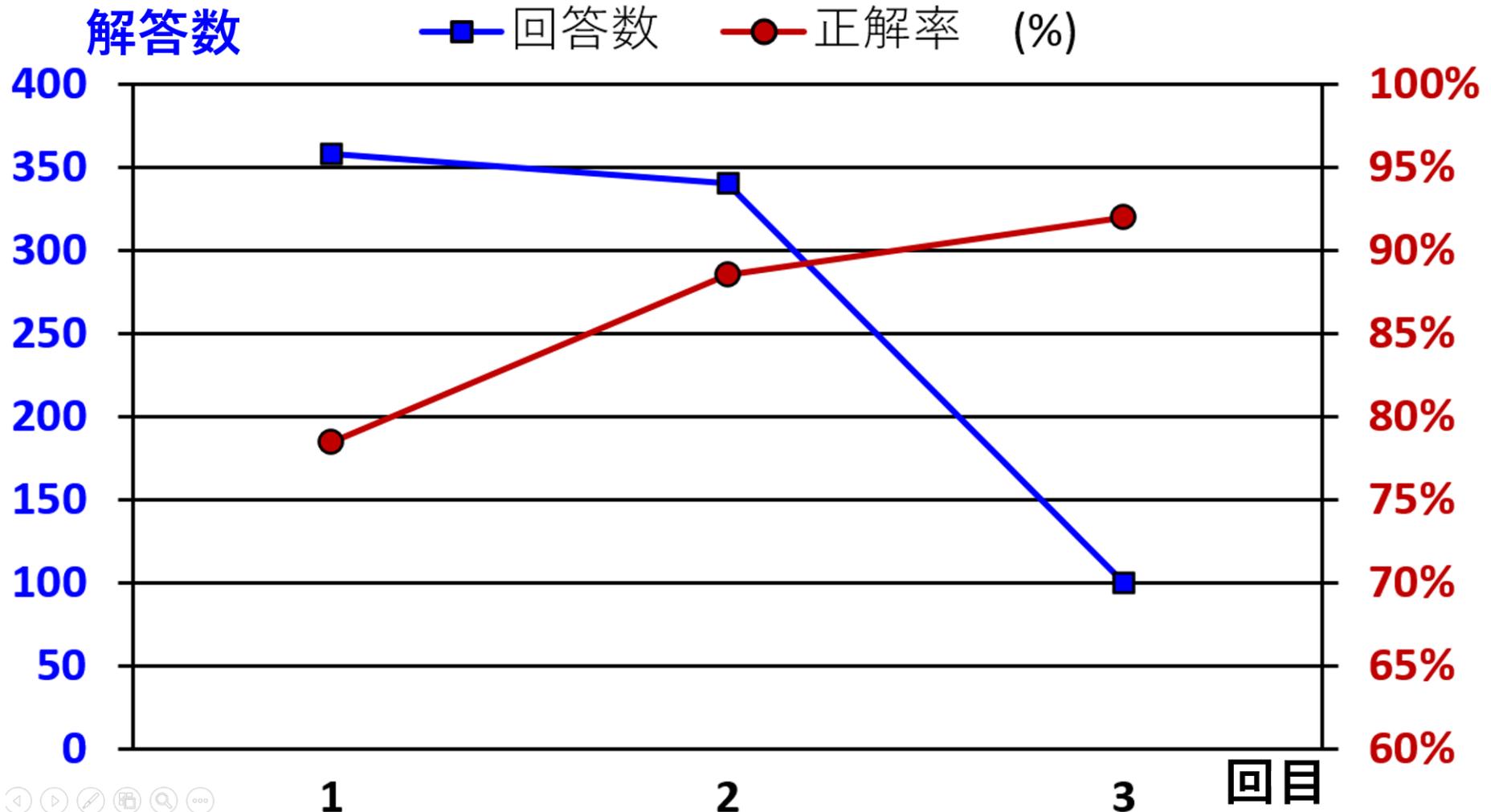
# 提案方式の処理区分ごとの時間評価

- 待ち時間：例文作成：1秒/件； 例文確認：9秒/件（応答待ち）
- 問題作成はデータベース操作のみ：分散処理による短縮可能



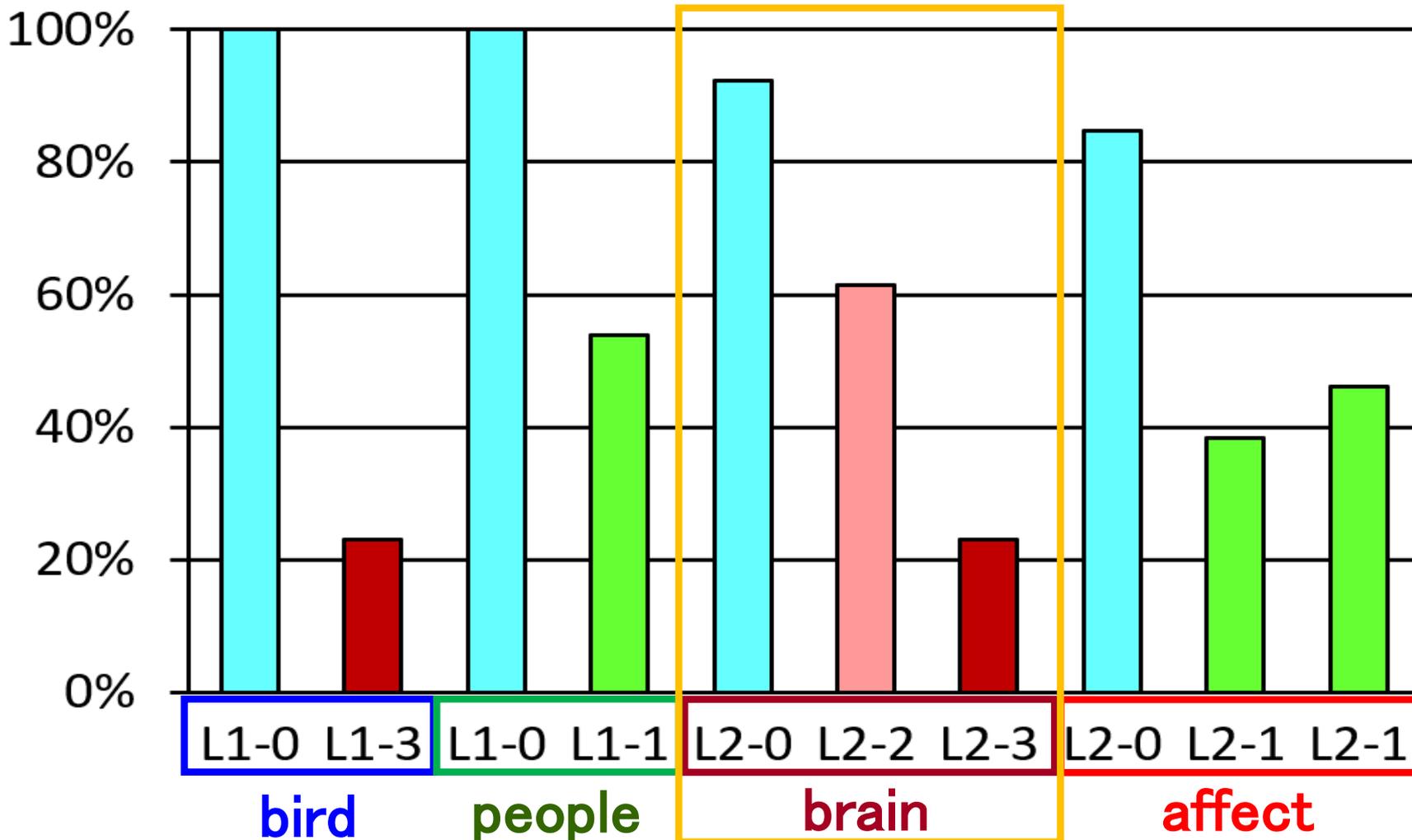
# シューティングゲームによるマスタ有効性評価

- 繰り返し学習することで、正解率は向上；生成した問題は有効
- 解答数は、ゲームの習熟度やコンディションに依存.



# 循環型データモデルの評価

■ 問題の難易度は、辞書のレベル表記だけでは規定できない。



レベル - 区分 (0:常用; 1:異なる意味; 2:熟語; 3:ことわざ)

# 問題の難易度の事例

■ 単語: brain (レベル 2 (準二級以上))

■ L2-0: 常用 92.3%

頭を使いなさい

Use your .

■ L2-3: 熟語 61.5%

彼女は才色兼備だ

She has both beauty and a good .

■ L2-4: ことわざ 23.1%

小人閑居して不善を為す

An idle  is the devil's workshop.

## 6. 考察

### ■ 2段階で、ビッグデータから各業務マスタを生成

#### 1. 標準DB(英語例文)生成

- ✓ スクレイピングに時間を要するが、自動実行可能
- ✓ 生成後は、各ユーザで共通使用が可能

#### 2. 各マスタ(英文問題)生成はデータベース操作のみ

- 提案方式により、効率的な業務マスタ生成が可能

### ■ 循環型データモデル

- 「英語問題」と「単語」の難易度は必ずしも相関しない
  - ✓ 業務側データのフィードバックで、精度向上が可能

## 7. まとめ

### ■ 研究の背景

- ✓ システム導入で業務マスタの整備コストが発生
- ✓ 一方, ビッグデータとして膨大なデータが公開

### ■ 循環型データモデルの提案

1. 2段階で, ビッグデータから各業務マスタを生成
2. 業務データのビッグデータ反映により, 精度向上

### ■ 英文問題生成システムへの適用

1. 英語問題の効率的な自動生成が可能
2. 循環型データモデルは, 難易度の精度向上に有効

### □ 将来研究: 他の英文問題評価サイト活用の有効性評価

本研究はJSPS科研費19K11985の助成を受けたものです。

# ビッグデータとフォグノードの間の 循環型データモデルの提案

ご清聴ありがとうございました

2022年 2月 18日

静岡理工科大学

情報学部 コンピュータシステム学科

工藤 司, 山本 剛大, 渡邊 智紀

