

Title	Web調べ学習における課題関連度の提案 -興味の変移に対応したLODに基づく指標の算定-
Author(s)	山内, 拓磨
Citation	
Issue Date	2020-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/16427
Rights	
Description	Supervisor: 長谷川 忍, 先端科学技術研究科, 修士 (情報科学)

修士論文

Web 調べ学習における課題関連度の提案

-興味の変移に対応した LOD に基づく指標の算定-

1810187 山内 拓磨

主指導教員 長谷川 忍
審査委員主査 長谷川 忍
審査委員 東条 敏
白井 清昭
池田 心

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和 2 年 3 月

Abstract

In recently, development of ICT is changing our daily life. In the educational field, researchers and practitioners aim to enhance the quality of education through ICT in education. One of the components of ICT in education is to cultivate the information literacy of learners. Web-based investigative learning is one of the ways to achieve this goal. In such learning, learners select resources suitable for knowledge construction from the Web by themselves, which has a vast number of resources. Then, learners extract and connect information related questions that learners want to learn. However, Web resources in the form of hyperspace are often unstructured. Therefore, learners must create learning scenarios by themselves, which consist of items and order that they should learn. It is not so easy for learners to create their own scenarios concurrently with the knowledge construction process.

In order to address this issue, Kashihara et al. proposed a model of Web-based investigative learning with Web resources. Based on this model, they also developed the interactive learning scenario builder (iLSB) to support Web-based investigative learning. The model supposes that learners set a question which they want to learn at first and learn about it. However, learners often find information that has no relation to the question. At that time, learners' interest may shift to other topics. Of course, learners can also learn another question instead of learning the first question. To deal with this issue, Sato et al. diagnosed learning scenario creation by using linked open data (LOD).

When we focus on learners' self-initiative, the system should follow the learning process from not only the original question but also the new question that they are newly interested. In these cases, learners may create multiple learning scenarios in their learning process. When learners connect these scenarios or questions, learners are expected to obtain effective learning outcomes. However, it is not easy for learners to connect with different knowledge. Therefore, we propose indexes for question relevance, which are criteria based on LOD to represent the strength of the relationship among the questions belonging to different learning scenarios.

The proposed method computes question relevance using related words connected to each question by LOD. If two questions have some relations, they would have common related words in LOD. Thus, the method calculates the distance of such a combination of questions on the LOD and takes question distance, question similarity, and question coupling. Question distance is defined by the number of nodes that appear in the shortest path to connect two questions on LOD. Question similarity is calculated by Simpson's coefficient between two sets, which are consisted of related words of each

question. Question coupling is determined by the number of found elements connecting with question keyword in both directions on LOD. Finally, question relevance is defined by using these parameters.

In this research, we had a case study to classify the combination of questions following each strength of the relation. We compute question relevance using seven learning scenarios created in previous research and obtain 542 combinations of questions. We analysed each tendency of question distance, question similarity, and question coupling using these combinations. As a result, we set thresholds of each parameter. Question distance and question coupling are divided into four levels, and question similarity is divided into three levels. Next, we combined question distance with question coupling and question similarity each, and two planes were constructed. The plane made by a combination of question distance and question coupling was divided into 16 areas. The plane made by a combination of question distance and question similarity was divided into 12 areas. Then, we classify these areas based on the tendency of obtained combinations of questions in each area.

In addition, we had an experiment to analysis the tendency of question relevance. We computed question relevance using nine learning scenarios created in previous research and obtained 566 combinations of questions. We sampled 40 combinations of questions from these combinations. Further, we divided the sampled combinations into four groups. One of these groups had ten combinations. Each group were rearranged in the order of strength of relation by two teachers. Then, we compared the results with rearranged groups by question relevance.

As a result, we found that question relevance focused on question similarity tended to be correlated with subjective scoring. Moreover, we thought that question relevance focused on question distance had a different tendency from the associations by subjects. Such relation seems to have a possibility to indicate the relationship that the subjects did not notice.

In conclusion, learners would be expected to obtain awareness regarding the relationship between two questions by using question relevance in Web-based investigative learning. It is a important topic of future research to consider how to suggest question relevance for learners and develop a system to support learners using question relevance.

目次

第1章 はじめに	1
1.1 研究の背景と目的	1
1.2 本論文の構成	2
第2章 関連研究	4
2.1 Web 調べ学習モデル	4
2.1.1 Web リソース探索フェイズ	5
2.1.2 Navigational Learning フェイズ	5
2.1.3 学習シナリオ作成フェイズ	5
2.2 interactive Learning Scenario Builder	6
2.3 課題展開の診断	6
第3章 課題関連度の算出方法	8
3.1 想定する学習活動	8
3.2 Linked Open Data	8
3.3 関連のある学習課題	10
3.4 課題関連度の算出	10
3.4.1 課題間距離	14

3.4.2 課題間類似度.....	14
3.4.3 課題間結合度.....	14
第4章 ケーススタディ	16
4.1 閾値の設定.....	16
4.1.1 課題間距離の傾向.....	16
4.1.2 課題間類似度の傾向	17
4.1.3 課題間結合度の傾向	17
4.2 領域の分類.....	18
4.3 エラー分析.....	21
第5章 実験.....	23
5.1 実験の目的.....	23
5.2 実験の内容.....	23
5.3 実験結果	26
5.4 考察	40
第6章 おわりに	43
6.1 まとめ.....	43
6.2 今後の課題.....	44

図目次

図 2.1 : Web 調べ学習モデル	4
図 2.2 : 学習シナリオの一例	5
図 2.3 : iLSB の UI	6
図 3.1 : RDF データのトリプル	8
図 3.2 : 入次数と出次数	9
図 3.3 : 課題関連度算出の処理フロー	11
図 3.4 : DBpedia Japanese における入次数の傾向	12
図 3.5 : DBpedia Japanese における出次数の傾向	13
図 3.6 : 算出する 3 つのパラメータ	14
図 4.1 : 課題間距離の傾向	17
図 4.2 : 課題間類似度の傾向	18
図 4.3 : 課題間結合度の傾向	18
図 4.4 : 課題間距離-課題間結合度の各領域の分類	20
図 4.5 : 課題間距離-課題間類似度の各領域の分類	21

表目次

表 3.1 : 地球温暖化の関連語句の一部とその入次数・出次数.....	9
表 4.1 : ケーススタディに用いた学習シナリオ一覧.....	16
表 4.2 : 課題間距離-課題間距離の各領域の事例.....	19
表 4.3 : 課題間距離-課題間類似度の各領域の事例	19
表 4.4 : 分類が適当でない事例の一部.....	21
表 4.5 : 展開元のキーワードの入次数・出次数.....	22
表 4.6 : 課題キーワードの入次数・出次数.....	22
表 5.1 : 実験に用いた学習シナリオ一覧	23
表 5.2 : 実験に用いた関連 (グループ 1)	24
表 5.3 : 実験に用いた関連 (グループ 2)	24
表 5.4 : 実験に用いた関連 (グループ 3)	25
表 5.5 : 実験に用いた関連 (グループ 4)	25
表 5.6 : 人手によるラベル付けの基準.....	26
表 5.7 : グループ 1 の並べ替えの結果.....	28
表 5.8 : グループ 2 の並べ替えの結果.....	28

表 5.9：グループ 3 の並べ替えの結果.....	29
表 5.10：グループ 4 の並べ替えの結果.....	29
表 5.11：人手による並べ替えのスピアマンの順位相関係数	30
表 5.12：課題間距離-課題間結合度に対するスピアマンの順位相関係数	30
表 5.13：課題間距離-課題間類似度に対するスピアマンの順位相関係数	30
表 5.14：課題間距離-課題間結合度のラベル付けに対する混同行列(A) ..	30
表 5.15：課題間距離-課題間結合度のラベル付けに対する混同行列(B) ..	31
表 5.16：課題間距離-課題間類似度のラベル付けに対する混同行列(A) ..	31
表 5.17：課題間距離-課題間類似度のラベル付けに対する混同行列(B) ..	31
表 5.18：課題間距離-課題間結合度の再現率・適合率・F 値(A)	32
表 5.19：課題間距離-課題間結合度の再現率・適合率・F 値(B)	32
表 5.20：課題間距離-課題間類似度の再現率・適合率・F 値(A)	32
表 5.21：課題間距離-課題間類似度の再現率・適合率・F 値(B)	32
表 5.22：3 クラス分類の課題間距離-課題間結合度の再現率・適合率・F 値 (A).....	33
表 5.23：3 クラス分類の課題間距離-課題間結合度の再現率・適合率・F 値 (B).....	33
表 5.24：3 クラス分類の課題間距離-課題間類似度の再現率・適合率・F 値 (A).....	33

表 5.25：3 クラス分類の課題間距離-課題間類似度の再現率・適合率・F 値 (B).....	33
表 5.26：グループ 1 の並べ替え結果(フィルタリングなし).....	34
表 5.27：グループ 2 の並べ替えの結果(フィルタリングなし).....	34
表 5.28：グループ 3 の並べ替えの結果(フィルタリングなし).....	35
表 5.29：グループ 4 の並べ替えの結果(フィルタリングなし).....	35
表 5.30：課題間距離-課題間結合度に対するスピアマンの順位相関係数 (フ ィルタリングなし).....	36
表 5.31：課題間距離-課題間類似度に対するスピアマンの順位相関係数 (フ ィルタリングなし).....	36
表 5.32：フィルタリングなしの課題間距離-課題間結合度の ラベル付けに 対する混同行列(A).....	36
表 5.33：フィルタリングなしの課題間距離-課題間結合度の ラベル付けに 対する混同行列(B).....	36
表 5.34：フィルタリングなしの課題間距離-課題間類似度の ラベル付けに 対する混同行列(A).....	37
表 5.35：フィルタリングなしの課題間距離-課題間類似度の ラベル付けに 対する混同行列(B).....	37

表 5.36：フィルタリングなしの課題間距離-課題間結合度の 再現率・適合率・F 値(A).....	37
表 5.37：フィルタリングなしの課題間距離-課題間結合度の 再現率・適合率・F 値(B).....	38
表 5.38：フィルタリングなしの課題間距離-課題間類似度の 再現率・適合率・F 値(A).....	38
表 5.39：フィルタリングなしの課題間距離-課題間類似度の 再現率・適合率・F 値(B).....	38
表 5.40：フィルタリングなしの 3 クラス分類の課題間距離-課題間結合度の 再現率・適合率・F 値(A).....	39
表 5.41：フィルタリングなしの 3 クラス分類の課題間距離-課題間結合度の 再現率・適合率・F 値(B).....	39
表 5.42：フィルタリングなしの 3 クラス分類の課題間距離-課題間類似度の 再現率・適合率・F 値(A).....	39
表 5.43：フィルタリングなしの 3 クラス分類の課題間距離-課題間類似度の 再現率・適合率・F 値(B).....	39
表 5.44：各パラメータと人手による関連付けの間の スピアマンの順位相関係数(A).....	40

表 5.45：各パラメータと人手による関連付けの間の スピアマンの順位相 関係数(B).....	40
---	----

第1章 はじめに

1.1 研究の背景と目的

近年、社会の情報化が急速に進み、携帯電話やパソコンといった情報機器が個人にも広く普及している。また Web の発展により、ネットワークに接続されたこれらの情報機器と Web ブラウザさえあれば、誰もがいつでもどこからでも容易に情報にアクセスすることが可能となっている。こういった背景から Web ブラウザを利用してインターネット上で学習を行う Web-based Learning (WBL) が注目されている。WBL は従来の学習・教育環境における時間的・空間的な制約を軽減し、学びたい人が学びたいことを自由に学ぶことができる「開かれた学び」を実現するための重要な基盤になりつつある [1]。WBL は主に Web を用いた調べ学習(Web 調べ学習)と Web 上の学習教材やシステムを用いて行う e-Learning に大別することができる。Web 調べ学習では学習者自身が様々な Web ページから学習目的に沿った情報を抽出し学習を行うのに対して、e-Learning では学習するためのコンテンツが用意されておりそれを活用して学習する。

また急速な情報化の進展に伴い、教育においても情報化を通じて教育の質の向上が目指されている [2]。こうした教育の情報化を構成する要素の一つとして、情報活用能力の育成が挙げられている。この情報活用能力の育成における目標は、①情報活用の実践力、②情報の科学的な理解、③情報社会に参画する態度、の3つの観点に整理されている [3]。このうち情報活用の実践力とは、課題や目的に応じて情報手段を適切に活用することを含めて、必要な情報を主体的に収集・判断・表現・処理・創造し、受け手の状況などを踏まえて発信・伝達できる能力のことである。この情報活用能力における小中学生の課題として、複数の Web ページから特定の情報を見つけ出し関連付けること、複数の Web ページの情報を整理・解釈すること、受け手の状況に応じて情報発信することなどが挙げられる [4]。このうち複数の Web ページから特定の情報を見つけ出し関連付けることや、複数の Web ページから得られた情報を整理・解釈する能力は、膨大なリソースを持つ Web 空間から学習者自らが学習するリソースを選択し、情報の抽出・関連付けを行う Web 調べ学習において育成することが可能である。

Web 調べ学習では Web 空間に存在する多種多様な学習リソースを利用して主体的かつ網羅的な学習を行うことが期待されるが、こうした学習リソースは非構造的な情報である。そのため従来のテキスト教材などとは異なり、学習する

項目やその学習順序(学習シナリオ)はあらかじめ定められておらず, 学習者は学習と並行して, 学習シナリオの作成を行う必要がある. これは学習者への認知的負荷が大きく, 学習が進むにつれて何をどのように学んできたかが不明瞭になりやすく, その結果として知識構造に行き詰まりが生じることがある [5].

この問題に対して, 先行研究において Web 調べ学習のモデルが提案され, そのモデルに基づいた Web 調べ学習支援システムが開発された [6]. この Web 調べ学習モデルは初めに学習する課題(初期課題)を一つ設定し, それをより細かい課題に分解して学習を進めることで, 初期課題について網羅的かつ体系的に学習を行うトップダウン型の学習であり, 一貫して初期課題について学習することが想定されている. しかし Web 上には様々な情報が存在するため, Web 調べ学習を進めていく中で検索エンジンの検索結果や Web ページのハイパーリンクなどから初期課題とは関連性の少ない情報も表示され, それによって初期課題とは関連性の少ないものの, 学習者が新たに興味を持った課題について調べ始めるといったことが起こりうる. この問題に対して, 先行研究では学習者の主体性を損なわずに初期課題に沿った学習を促すために, 学習者の課題展開に対して Linked Open Data(LOD)を用いて課題展開の診断を行った [7].

一方でより学習者の主体性を重視した場合, 新たに興味を持った課題を別の初期課題として展開することが考えられる. これを許容した場合, 異なる初期課題についてそれぞれ独立した複数の学習シナリオが作成されることになる. 人間はある概念についてそれを取り巻く他の概念と関連付けて認識・記憶しつつ学習する [8]ため, 複数の学習シナリオに対して学習課題間での知識構造を関連付けることでより高い学習効果が期待できる. しかし異なる学習課題から得た知識構造を関連付けることは学習者にとって容易なことではない.

そこで, 本研究の最終的な目標は学習者に対して異なる学習課題から得られた知識構造の関連付けを促すような気づきを与える支援を行うことである. これを実現するために本研究では, 異なる学習課題間の関連の強さの指標として LOD に基づいた課題関連度を提案する.

1.2 本論文の構成

本論文の構成は以下のとおりである.

- 第1章 はじめに
本研究の背景と目的, 本論文の構成について述べる.
- 第2章 関連研究
関連研究について述べ, 本研究の位置づけを示す.
- 第3章 課題関連度の算出方法

提案する課題関連度の算出方法について述べる.

- 第4章 ケーススタディ

提案する課題関連度について, その傾向を分析するために行ったケーススタディの結果を示す.

- 第5章 実験

提案する課題関連度と人手による関連付けの間の関係性を分析するために行った実験の概要と結果を示し, それに対する考察を述べる.

- 第6章 まとめ

本研究のまとめを記す.

第2章 関連研究

2.1 Web 調べ学習モデル

Web 調べ学習では学習者自身が学習シナリオを作成しつつ学習を進めていく必要があり、その認知的負荷が高いという問題に対して、先行研究において Web 調べ学習における学習活動を 3つのフェイズに分解したモデルが提案された [6].

Web 調べ学習は図 2.1 のように Web リソース探索フェイズ, Navigational Learning フェイズ, 学習シナリオ作成フェイズの 3つのフェイズからなり、学習者は部分課題が生成されなくなるまでこの 3 フェイズを繰り返すことで初期課題について学習することが想定されている. その結果として図 2.2 のような初期課題を根とする木構造で表現された学習シナリオが作成される. 図 2.2 は地球温暖化を初期課題として学習を行った場合の学習シナリオの一例である.

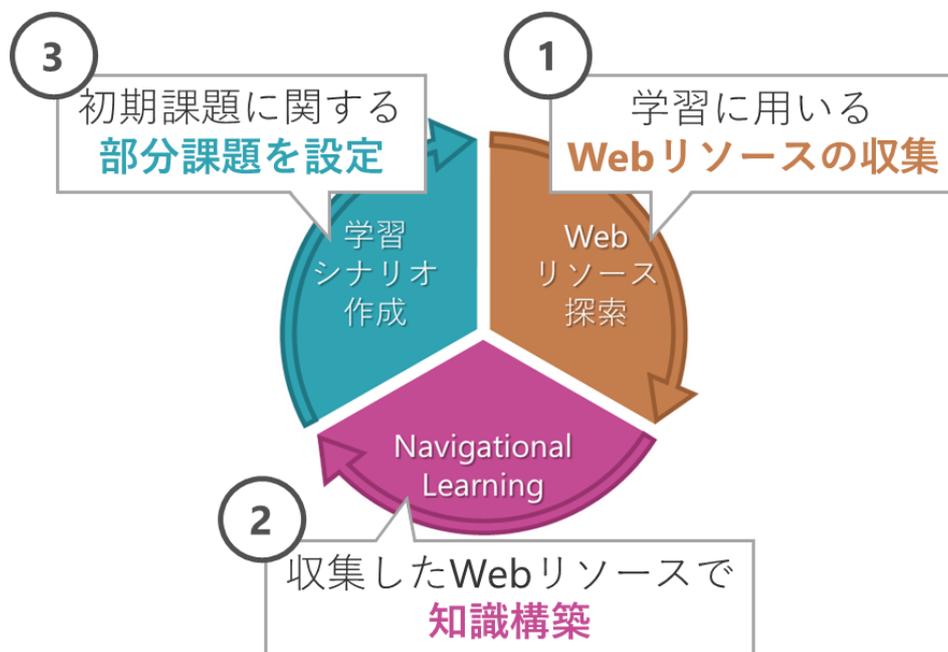


図 2.1 : Web 調べ学習モデル

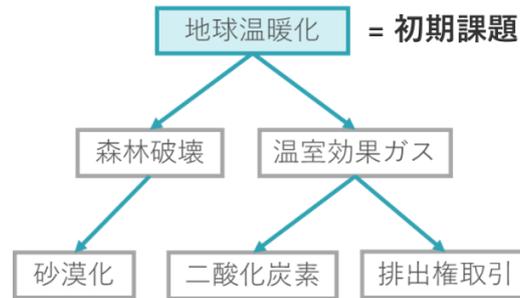


図 2.2：学習シナリオの一例

次に Web 調べ学習モデルを構成する 3つのフェイズそれぞれについて述べる。

2.1.1 Web リソース探索フェイズ

Web リソース探索フェイズは学習に用いるための学習リソースとなる Web リソースを探索・収集するフェイズである。検索エンジンを用いて課題を端的に表すキーワードに関する Web リソースを探索・収集する。

2.1.2 Navigational Learning フェイズ

Navigational Learning フェイズでは Web リソース探索フェイズで収集したリソースから情報を抽出し、それらを関連付けることにより課題についての知識構築を行うフェイズである。

2.1.3 学習シナリオ作成フェイズ

学習シナリオ作成フェイズは Navigational Learning フェイズで構築した知識から、より深く学習する必要がある項目を抽出し部分課題として展開するフェイズである。展開した部分課題についても、Web 調べ学習モデルに沿った学習を行い、部分課題が発生しなくなるまで学習を行う。

このモデルは初期課題を設定し、それをより細かい課題に分解して学ぶことで網羅的かつ体系的に学ぶトップダウン型の学習である。そのため初期課題に沿わないような学習課題の展開は想定されていない。

一方で本研究ではある学習課題についての学習はこのモデルに沿って行われることを想定するが、初期課題に沿わないが学習者が興味を持った学習課題については、別の新たな初期課題として学習することを許容している。

2.2 interactive Learning Scenario Builder

前節で述べた Web 調べ学習モデルに沿った学習を行うための学習環境として、interactive Learning Scenario Builder (iLSB)が Firefox のアドオンとして開発された [6]. 図 2.3 のように iLSB は Web リソース探索フェイズのための検索エンジン、Navigational Learning フェイズのための Keyword Repository, 学習シナリオ作成フェイズのための Question Tree Viewer によって構成されている. iLSB はそれぞれの学習課題ごとに独立した Keyword Repository を持ち, 学習者は自身が抽出したある学習課題に関するキーワード群を, 対応する Keyword Repository 上で関連付けることにより, その学習課題についての知識構造を構築する. このシステムにより学習者の知識構築と課題展開を明示することで, Web 調べ学習の足場を築くことで課題展開の活性化を期待する.

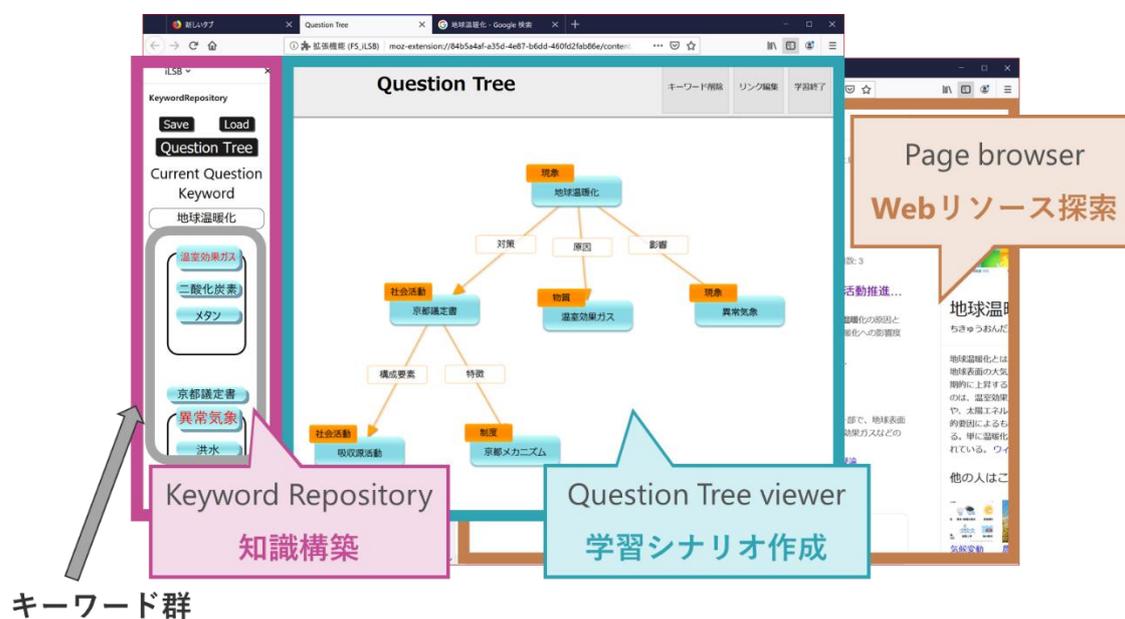


図 2.3 : iLSB の UI

本研究でもある初期課題を設定して学習する際にこの iLSB を使用して学習し, 学習シナリオを作成することを想定する.

2.3 課題展開の診断

Web 調べ学習モデルでは学習の中で行われる課題展開は初期課題に沿ったも

のであることが想定されている。しかし課題展開が初期課題に沿ったものであるかどうかの妥当性の評価は学習者に委ねられているため、必ずしも妥当な課題展開が行われるとは限らない。そのため学習シナリオの妥当性を診断する必要があるが、Web 調べ学習は課題解決のためのプロセスは多様であるため、解となる学習シナリオを用意することは難しい。そこで先行研究では Linked Open Data(LOD)を用いて課題を表すキーワード(課題キーワード)間の関連度や類似度を算出し、この 2 つの指標を用いることで学習者の主体性を維持しつつ課題展開を診断する手法を提案した [7]。この提案手法では展開した学習課題について、その学習課題の展開元となった学習課題と初期課題それぞれと間の関係を調べ、それをもとに課題展開が初期課題に沿ったものであるかを診断した。この提案手法による診断結果と人手による診断結果との一致率は 77.8%であり、高い精度で課題展開の診断を行うことが可能であることが示された。

しかし Web 調べ学習を進める中において、Web ページのハイパーリンクや検索エンジンの検索結果などで目にした、初期課題とは関連性の少ない情報へ学習者の興味が遷移することが起こりうる。そしてそれを初期課題に関する新たな部分課題として展開した場合に、提案された課題展開の妥当性を診断すると不適当な課題展開であると判定されると考えられる。これは初期課題について学習するという目的からは外れているためである。一方で学習者の主体性をより重視した場合にはこの課題を別の初期課題として展開することが考えられる。

本研究ではこれを許容し、複数の学習シナリオが作成されるような学習が行われることを想定する。その際、人間はある概念についてそれを取り巻く他の概念と関連付けて認識・記憶しつつ学習する [8]ため、異なる学習シナリオの学習課題間に対して、関連付けを行いながら学習することで、より高い学習効果が期待される。しかし、学習者自身が関連付けを行うためには複数の視点から知識構造をとらえる必要があり、これは学習者にとって容易ではなく、関連を見落としたこと自体に気付くことも難しい。この問題に対して本研究では、異なる学習シナリオの学習課題間の関連の強さを表す指標(課題関連度)を、LOD に基づいて算出することを目指す。

第3章 課題関連度の算出方法

3.1 想定する学習活動

本研究で対象とする学習は iLSB を用いた Web 調べ学習である。その想定される学習者の行う学習活動は以下のようなものである。

- 1) 学習者が設定した、もしくは事前に与えられた初期課題に対して、iLSB を用いて Web 調べ学習モデルに沿った学習を行う。
- 2) 学習者が興味を持ったが初期課題に沿わないような課題が現れた場合には、別の初期課題として展開し、学習する。
- 3) 部分課題展開が生起しなくなるまで学習した学習シナリオが作成されたとき、その学習シナリオに属する学習課題に対して、他の学習シナリオの学習課題との関連付けを行う。

3.2 Linked Open Data

課題関連度の算出に用いる Linked Open Data(LOD)について述べる。LOD は Web 上でデータ同士を結び付けて公開する仕組みのことである。本研究では Wikipedia 日本語版から情報を抽出した LOD である DBpedia Japanese [9] を用いる。

DBpedia Japanese は RDF と呼ばれる形式のデータ(RDF データ)から構成されるものである。RDF データは図 3.1 のように主語、述語、目的語の 3 つの要素からなる組(トリプル)でリソース間の関係を表現する。図 3.1 の例は「地球温暖化」が「二酸化炭素」に対して Wikipedia 内でリンクを持つことを表現している。



図 3.1 : RDF データのトリプル

DBpedia Japanese は RDF データを検索するためのクエリ言語である SPARQL を用いてクエリを送信することで、特定のデータを取り出すことが可能である。これにより指定した条件を満たすトリプルを DBpedia Japanese から抽出することができる。例えば「地球温暖化」を主語に持ち「wikiPage WikiLink」を述語に持つようなトリプルを抽出した場合、「二酸化炭素」などが条件を満たすトリプルの目的語として得られる。このとき本研究では「二酸化炭素」のことを「地球温暖化」というキーワードに対する関連語句と呼ぶこととする。

また DBpedia Japanese ではリソースに対応する Wikipedia のページについて、そのページが他の Wikipedia のページからリンクされている数(入次数)と、そのページ内にある他の Wikipedia のページへのリンクの数(出次数)の情報を得ることができる。図 3.2 の場合は A の入次数は 3、出次数は 2 である。

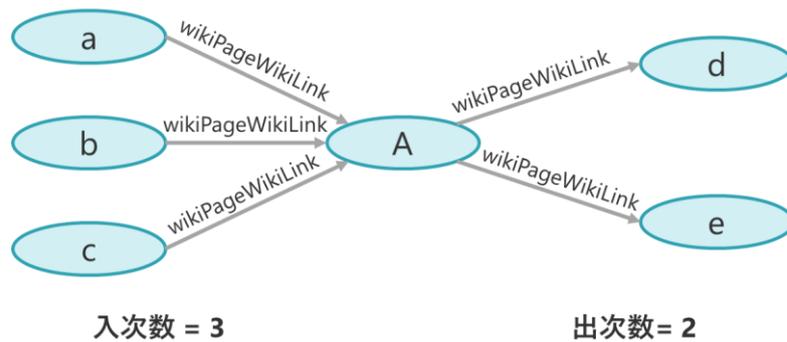


図 3.2：入次数と出次数

表 3.1：地球温暖化の関連語句の一部とその入次数・出次数

語句	入次数	出次数
日本	135478	2324
2008 年	75227	893
京都議定書	220	243
化石燃料	195	197

しかし DBpedia Japanese 上には圧倒的多数のリソースにつながる少数のリソースと、それほど多くのノードと繋がらない大多数のリソースが存在する。例として表 3.1 に DBpedia Japanese 上での「地球温暖化」の関連語句の一部とその入次数・出次数を示す。「日本」のような国名や地域名、「2008 年」のよ

うな西暦などのリソースは前者にあたり、これらのリソースは一般性が高い情報である傾向にある。これらのリソースは多くのリソースと繋がるがゆえに、実際には関連性の弱い 2 つのリソースの間に入って、それらを DBpedia Japanese 上において短い距離で結びつけてしまうことがある。これに対し入次数と出次数それぞれについて閾値を設定し、それ以上のものは関連語句として扱わないことで対処する必要がある。

3.3 関連のある学習課題

本研究においてどのような学習課題の組み合わせを関連のあるものとして扱うかについて述べる。それぞれの学習課題について iLSB 上で関連付けられたキーワード群の情報から LOD を用いて、それぞれの学習課題についての関連語句群を取得した時に、もし 2 つの学習課題に関連があるのならば、その 2 つの関連語句群について積集合を取ると 2 つの学習課題の接点となる要素が積集合に現れることが考えられる。このことからそれぞれの学習課題についての関連語句群の積集合が空集合でないものを本研究では直接的な関連のある学習課題として扱うこととし、その関連の強さを積集合の要素をもとに算出する。

3.4 課題関連度の算出

この節では LOD を用いて iLSB の Keyword Repository に蓄えたキーワードの情報からその学習課題に対する関連語句集合を取り出し、他の学習課題の関連語句群と比較することで課題関連度を決定するパラメータを算出する方法について述べる。

図 3.3 は課題関連度算出の処理フローである。 Q_A は iLSB 上の課題ノードで、課題キーワード q_A とその Keyword Repository に蓄えられたキーワード群 $key_{A1}, key_{A2}, \dots, key_{An}$ からなるキーワード集合である。 Q_B も同様に、課題キーワード q_B とその Keyword Repository に蓄えられたキーワード群 $key_{B1}, key_{B2}, \dots, key_{Bm}$ からなるキーワード集合である。DBpedia Japanese におけるキーワード x の関連語句集合を $R(x)$ と定義し、キーワード集合 $X = \{x_1, x_2, \dots, x_n\}$ の関連語句集合を

$$R(X) = R(x_1) \cup R(x_2) \cup \dots \cup R(x_n)$$

で定義する。

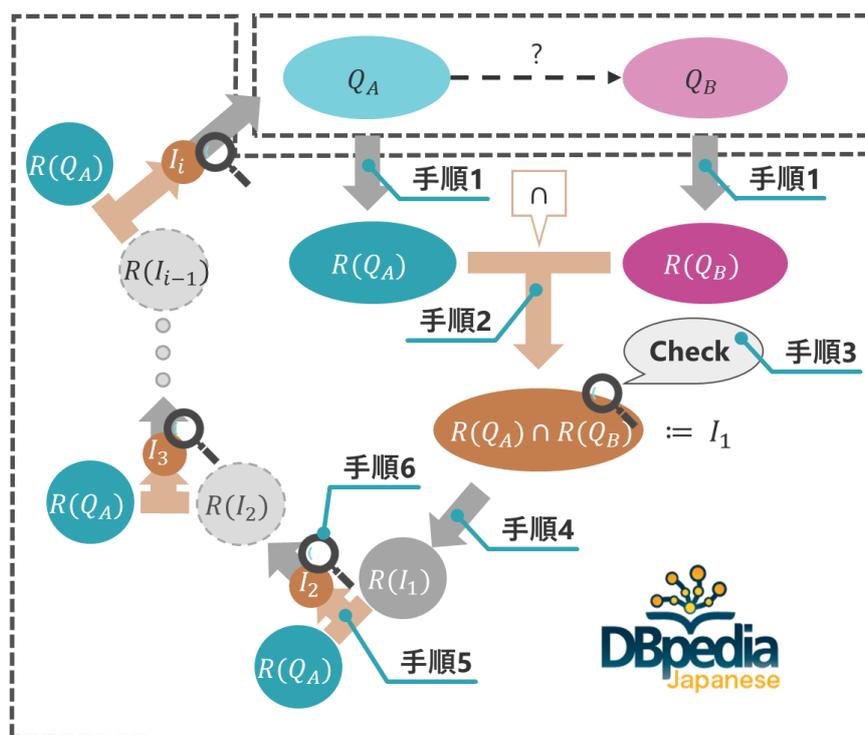


図 3.3 : 課題関連度算出の処理フロー

図 3.3 が示す Q_A から見た Q_B の課題関連度を算出手順について説明する。

- 手順1. 関連の有無を調べる課題 Q_A , Q_B についてそれぞれ関連語句集合 $R(Q_A)$, $R(Q_B)$ を抽出するためのクエリを送信する。
- 手順2. $R(Q_A)$ と $R(Q_B)$ の積集合 (I_1) をとる。
- 手順3. I_1 が空集合でない場合, I_1 に DBpedia Japanese 上で Q_A の課題キーワード q_A と相互につながる語句(相互接続する要素)が含まれているかを調べる。
- 手順4. 相互接続する要素が含まれていない場合, I_1 の関連語句集合 $R(I_1)$ を抽出するためのクエリを送信する。
- 手順5. 得られた $R(I_1)$ と $R(Q_A)$ の積集合 (I_2) をとる。
- 手順6. I_2 が空集合でない場合, I_2 に対して手順 3. の操作を行う。相互接続する要素が含まれていない場合, この手順 6. で扱った積集合について手順 4-6. の操作を行う。これを相互接続する要素が現れる, もしくは関連語句集合を取得した回数(ステップ数)が一定回数に達するまで繰り返す。

これによって算出する課題関連度は 2 つの学習課題の直接的な関連を示すものである。これは関連語句群の積集合から注目している学習課題 Q_A の課題キーワード q_A に結び付くまでの DBpedia Japanese 上の距離を求めている。これに

より 2 つの学習課題の接点となっている語句群が注目している学習課題 Q_A によって重要度が高いかどうかを調べる。

前節で述べた通り手順 2.の操作を行った際に接点となる語句群が現れた学習課題の組み合わせを関連のある学習課題として扱うため、 I_1 が空集合となった場合はその課題間に関連はないものと判定し、処理を終える。 I_1 が要素を持つとしても、この一連の操作の中で相互接続する要素が現れる以前に $I_i (i \in \mathbb{N})$ が空集合となる場合、もしくは相互接続する要素が現れないまま処理を終えた場合は関連がないものとする。

また DBpedia Japanese を用いて取得した関連語句群には、3.2 節で述べた通り一般性の高いと思われる語句も含まれている。これらにより関連性の薄いものがつながることを防ぐために、入次数と出次数それぞれについて閾値を設定し、フィルタリングを行う。閾値は DBpedia Japanese 上のリソースの入次数と出次数の傾向を調査し、経験的に決定した。

入次数とリソース数の関係は図 3.4 に示すようになった。入次数の高い側から見ていくと 1000 を下回ったあたりで、リソース数が増加を始める傾向が見られた。また 2008 年といった西暦などの一般性の高いリソースは入次数 1000 未満ではあまり現れなくなった。これらのことから入次数についての閾値は 1000 とした。

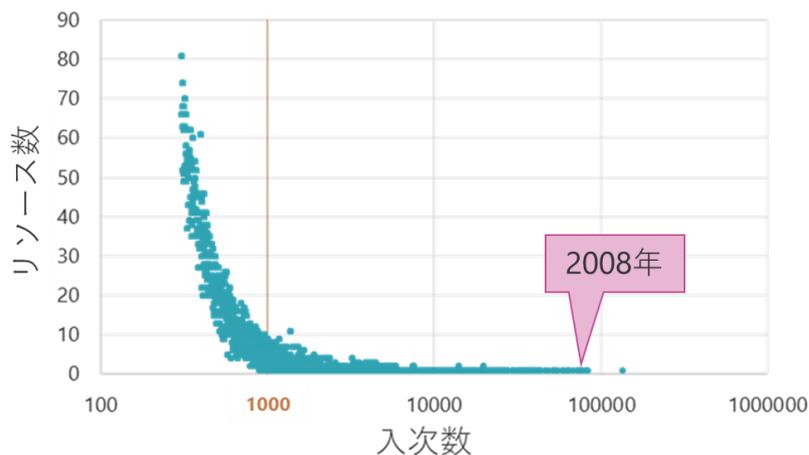


図 3.4 : DBpedia Japanese における入次数の傾向

出次数とリソース数の関係は図 3.5 に示すようになった。出次数が 300 の前後でリソース数が 100 を超えていることや、化学の要素を含むリソースの間で強力なハブとなってしまいう素(出次数:305)などの元素をある程度取り除けることから、出次数についての閾値は 300 とした。

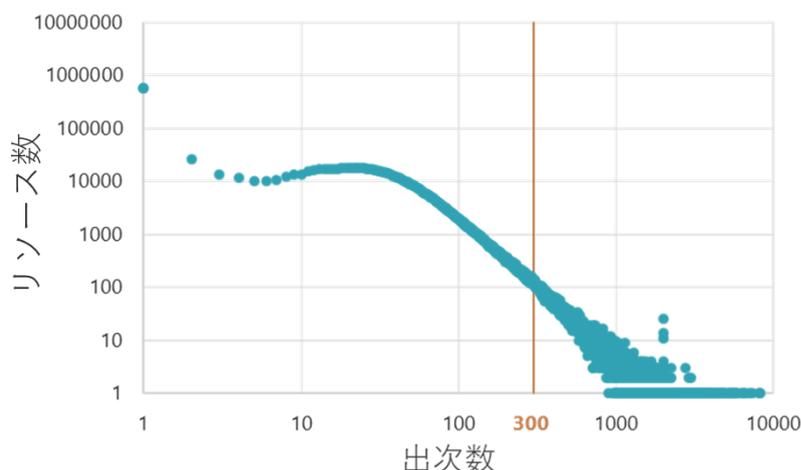


図 3.5 : DBpedia Japanese における出次数の傾向

また、iLSB である学習課題から部分課題展開を行う際にはその学習課題に関連付けられたキーワードから選択したものを部分課題として展開する。そのため部分課題の課題キーワードは必ず展開元の学習課題に含まれる。これらの学習課題に対して課題関連度を求める際に、展開した課題キーワードの関連語句を介してつながるような学習課題がある場合には、部分課題の課題関連度がより強くなることが望ましいが、展開元の学習課題も同じキーワードを包含するためどちらの学習課題も同程度の課題関連度として現れる可能性がある。これを防ぐため、Keyword Repository に蓄えられたキーワードのうち部分課題として展開したものは、課題関連度を算出する際のキーワード集合から除外する。

手順 3.もしくは手順 6.で相互接続する要素が現れた場合は、関連の強さを測るための以下で述べる 3 つのパラメータを取り出し、処理を終える。これらのパラメータの組み合わせから関連の強さを求め、それを課題関連度とする。

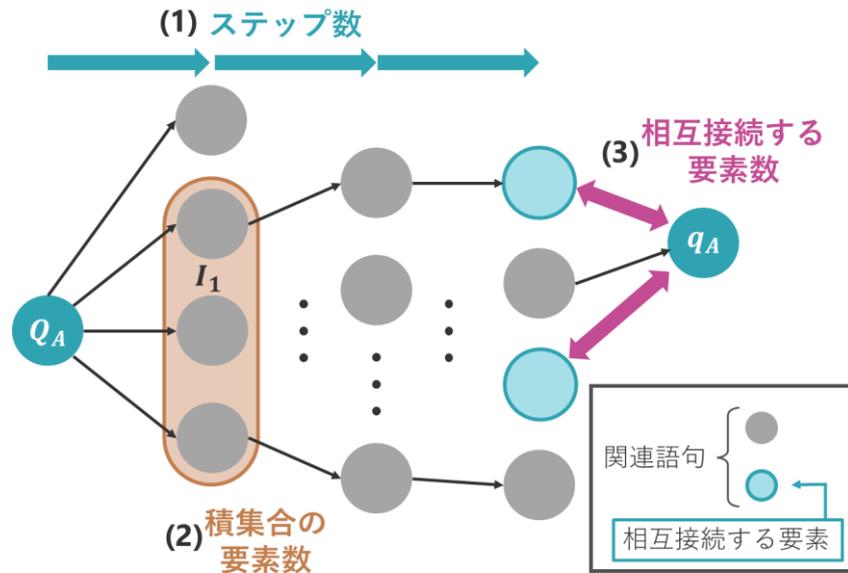


図 3.6：算出する 3 つのパラメータ

3.4.1 課題間距離

課題間の距離を表すパラメータとして、相互接続する要素が現れるまでにかかったステップ数を課題間距離として定義する。図 3.6 の場合、課題間距離は 3 である。課題間距離が小さいほど関連度は強いと考えられる。

3.4.2 課題間類似度

2 つの課題の関連語句集合の類似度を表すパラメータとして $R(Q_A)$ と $R(Q_B)$ の積集合 I_1 の要素数を扱う。このパラメータを Simpson 係数によって正規化したものを課題間類似度として定義する。Simpson 係数は 2 つの集合の類似度を測る指標としてよく用いられるもので(1)式で求められる。

$$Simpson(X, Y) = \frac{|X \cap Y|}{\min(|X|, |Y|)} \quad (1)$$

図 3.6 の場合、 $R(Q_A) \leq R(Q_B)$ とすると課題間類似度は 0.75 である。課題間類似度は 0 から 1 の間の値をとり、課題間類似度が大きいほど関連度は強いと考えられる。

3.4.3 課題間結合度

2 つの課題がいくつの要素を通じてつながっているのかを表すパラメータとして、現れた相互接続する要素数を課題間結合度として定義する。図 3.6 の場合、

課題間結合度は 2 である。課題間結合度が大きいほど関連度は強いと考えられる。

第4章 ケーススタディ

本章では前章で述べた算出手順によって得られた学習課題の組み合わせについて、3つのパラメータの傾向の分析と、それを基に課題関連度を段階的に分類するために行ったケーススタディについて述べる。

4.1 閾値の設定

表 4.1 に示す先行研究で得られた 7 つの異なる初期課題を持つ学習シナリオを用いてケーススタディを行った。この 7 つの学習シナリオから双方向に関連のある課題の組み合わせとして 271 組が抽出され、これをそれぞれの方向に分解すると 542 の関連が得られた。3 つのパラメータを用いて課題関連度ごとに分類するために、この 542 の関連についてそれぞれのパラメータの傾向を分析し、閾値を設定した。なおここで取得した関連はステップ数 15 未満で相互接続する要素が現れた学習課題の組み合わせである。

表 4.1：ケーススタディに用いた学習シナリオ一覧

初期課題	学習課題数	平均キーワード数
地球温暖化	6	8.3
電気自動車	7	6.1
オゾン層破壊	5	4.0
栄養素	15	4.7
喫煙	23	2.6
生命保険	11	11.4
原子力	13	1.6

4.1.1 課題間距離の傾向

課題間距離の傾向は図 4.1 のようになった。課題間距離が 1,2,3 である関連の数はそれぞれ 160, 191, 133 で、この合計は全体の約 9 割を占めており、課題間距離が 4 以上である関連の数はこれらに比べ大きく減少する傾向にあった。この結果から課題間距離は関連の強い順に A:1, B:2, C:3, D:4 以上、の 4 領域に分割した。

またこのケーススタディにおいては課題間距離が 8 以上となる関連は見つか

らなかった。これをもとに算出時の終了条件であるステップ数は10に設定した。

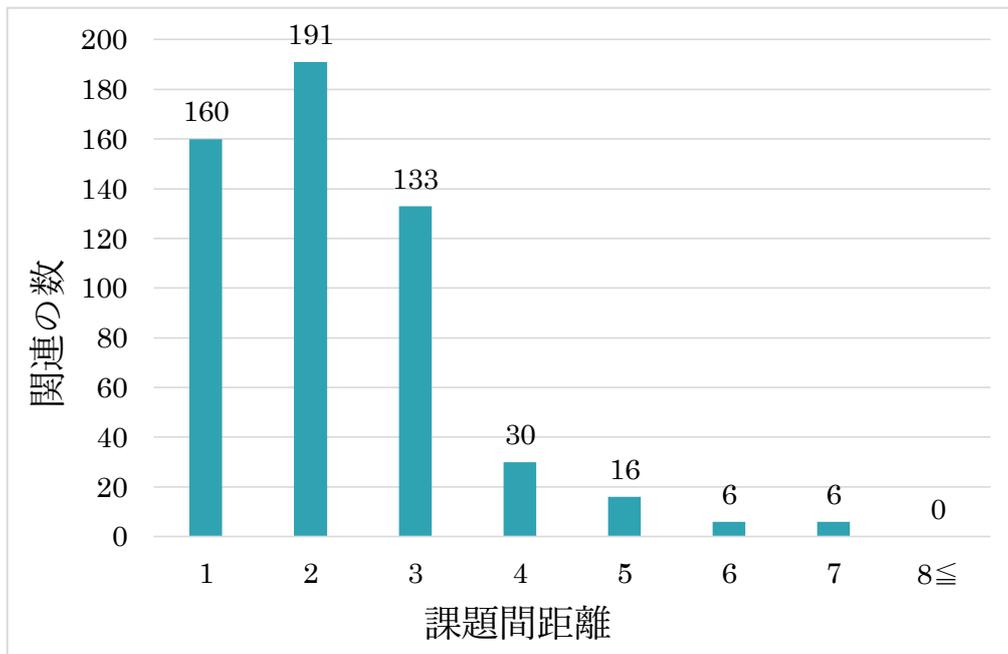


図 4.1：課題間距離の傾向

4.1.2 課題間類似度の傾向

課題間類似度の傾向は図 4.2 のようになった。図 4.2 は課題間類似度が 0.01 以下の関連の数が 47, 0.01 より大きく 0.02 以下の関連の数が 46, といったように計数をおこない、その分布を見たものである。その結果, 0.03 と 0.04 の間と, 0.07 と 0.08 の間でそれぞれ関連の数に大きく差が現れた。このことから課題間類似度を関連の強い順に A: 0.07 より大きい, B: 0.03 より大きく 0.07 以下, C: 0.03 以下, の 3 領域に分割した。

4.1.3 課題間結合度の傾向

課題間結合度の傾向は図 4.3 のようになった。課題間結合度が 1 である関連の数は 274 と半数以上を占めており, 次いで多い 2 である関連の数は 111, 3 である関連の数は 59 と課題間結合度が強くなるにしたがって関連の数は減少する傾向が見られた。この結果から課題間結合度を関連の強い順に A: 5 以上, B: 3 もしくは 4, C: 2, D: 1, の 4 領域に分割した。

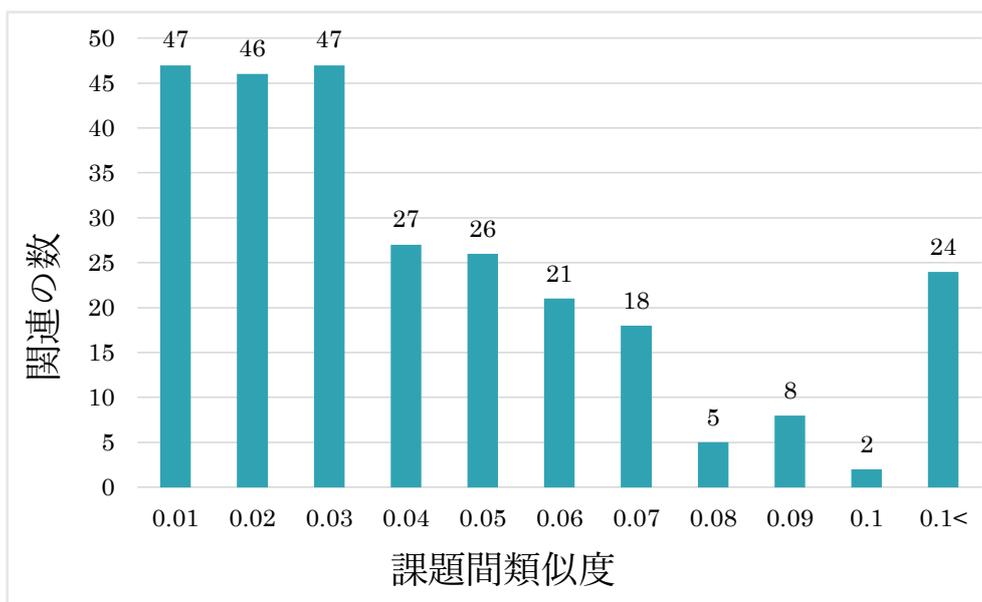


図 4.2 : 課題間類似度の傾向

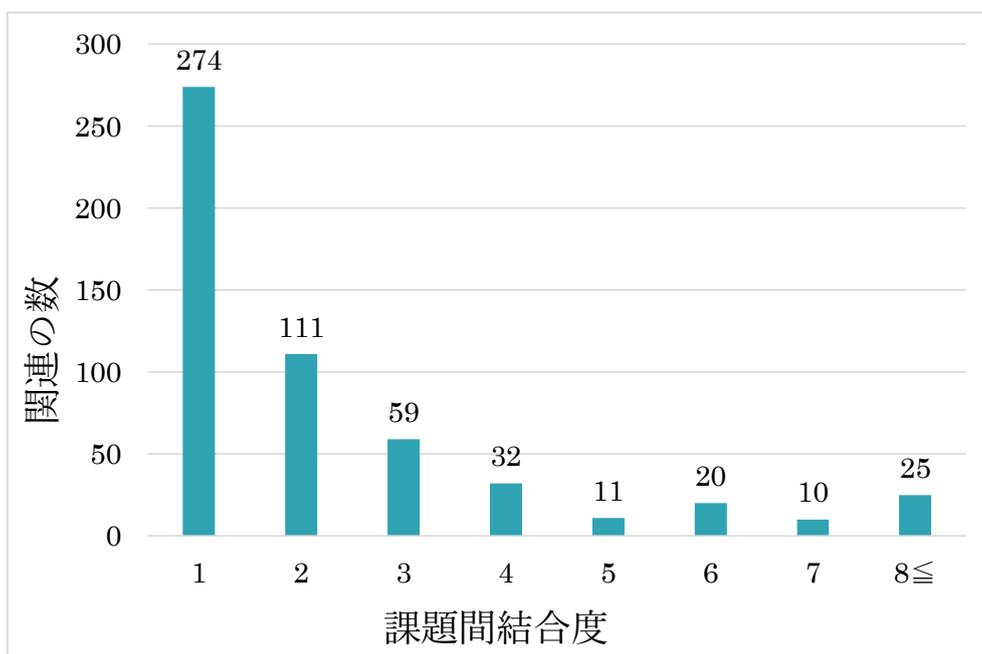


図 4.3 : 課題間結合度の傾向

4.2 領域の分類

前節で各パラメータについて閾値を設定し段階的に分割した。これをより細分化するために課題間距離と課題間結合度，課題間距離と課題間類似度をそれぞれ組み合わせ，平面を構成した。課題間類似度と課題間結合度の組み合わせ

については、課題間結合度は課題間距離との結びつきが強く、課題間結合度を比較する際には課題間距離をそろえる必要があると考えたため、今回は考慮しないこととした。これらの平面は各パラメータの閾値を用いることで課題間距離-課題間結合度は16、課題間距離-課題間類似度は12の領域に分割される。ここで課題間距離-課題間結合度は特に課題間距離、課題間距離-課題間類似度は特に課題間類似度に注目した課題関連度とする。表4.2, 4.3にそれぞれの平面の各領域に現れた関連の事例を示す。表4.2, 4.3ともに左上にあるほど関連が強く、右下に行くほど関連が弱くなると考えられる。

表 4.2：課題間距離-課題間距離の各領域の事例

		課題間距離			
		A	B	C	D
課題間結合度	A	化石燃料 → 地球温暖化	ビタミン → 喫煙	化石燃料 → 禁煙	地球温暖化 → 原子力事故
		地球温暖化 → 気候変動	不飽和脂肪酸 → 電気自動車	ビタミン → がん保険	生命保険 → 化石燃料
		呼吸器疾患 → 化石燃料	不飽和脂肪酸 → 喫煙	ビタミン → 原子力事故	火災 → 不飽和脂肪酸
	B	酸性雨 → 電気自動車	地球温暖化 → オゾンホール	京都議定書 → 脂質	地球温暖化 → 受動喫煙
		原子力発電 → 電気自動車	地球温暖化 → 栄養素	京都議定書 → 脂溶性ビタミン	脂溶性ビタミン → 生命保険
		喫煙 → 火災	タンパク質 → 温室効果ガス	生命保険 → 地球温暖化	原子力発電 → 喫煙
	C	地球温暖化 → 電気自動車	回生ブレーキ → 京都議定書	脂質 → 地球温暖化	温室効果ガス → タンパク質
		温室効果ガス → 電気自動車	気候変動 → 電気自動車	脂溶性ビタミン → 地球温暖化	電気自動車 → ニコチン
		地球温暖化 → モントリオール議定書	地球温暖化 → アミノ酸	脂質 → 温室効果ガス	原子力 → ABS
	D	地球温暖化 → 回生ブレーキ	京都議定書 → 回生ブレーキ	回生ブレーキ → 地球温暖化	たばこ病 → 地球温暖化
		京都議定書 → 電気自動車	地球温暖化 → 脂質	地球温暖化 → タンパク質	たばこ病 → 温室効果ガス
		温室効果ガス → オゾンホール	アミノ酸 → 地球温暖化	地球温暖化 → ニコチン	たばこ病 → 京都議定書

表 4.3：課題間距離-課題間類似度の各領域の事例

		課題間距離			
		A	B	C	D
課題間類似度	A	呼吸器疾患 → 化石燃料	化石燃料 → 呼吸器疾患	地球温暖化 → 一酸化炭素	がん保険 → たばこ病
		飽和脂肪酸 → 喫煙	京都議定書 → ウィーン条約	喫煙 → 酸性雨	
		ウィーン条約 → 京都議定書	化石燃料 → ウィーン条約	一酸化炭素 → 酸性雨	
	B	オゾンホール → 酸性雨	酸性雨 → オゾンホール	たばこ病 → 脂溶性ビタミン	がん保険 → 喫煙
		栄養素 → 酸性雨	脂溶性ビタミン → たばこ病	電気自動車 → 一酸化炭素	電気自動車 → ニコチン
		モントリオール議定書 → 地球温暖化	一酸化炭素 → 電気自動車	ニコチン → 栄養素	がん保険 → ニコチン
	C	ビタミン → 京都議定書	温室効果ガス → 不飽和脂肪酸	京都議定書 → ビタミン	がん保険 → タンパク質
		タンパク質 → 火災	タンパク質 → がん保険	ABS → 原子力	原子力 → ABS
		京都議定書 → 原子力発電	火災 → タンパク質	地球温暖化 → 不飽和脂肪酸	たばこ病 → 不飽和脂肪酸

課題間距離-課題間結合について述べる際に、課題間距離が d かつ課題間結合度が c の領域を D_{dc} と呼称する。 D_{AA} , D_{AB} , D_{AC} には「呼吸器疾患→化石燃料」、「酸性雨→電気自動車」、「温室効果ガス→モ電気自動車」などの妥当だと考えられる関連が比較的多く挙げられた。 D_{AD} には「京都議定書→電気自動車」のような妥当だと考えられる関連も含まれているが、全体としてはそういった関連の割合

は高くない結果となった。また D_{BA} , D_{BB} でも「ビタミン→喫煙」, 「地球温暖化→オゾンホール」のような妥当だと考えられる関連が含まれているが, D_{AD} と同様に全体としての割合は高くない。 D_{BC} や D_{CA} などの残りの領域ではより弱い関連の割合が高くなる。 これらのことから課題間距離-課題間結合度を用いた課題関連度は図 4.4 のように各領域を分類した。 これによって分類した課題間距離-課題間結合度の各領域の課題関連度は $[1]>[2]=[3]>>[4]$ のような関係になっていると考えられる。

課題間距離-課題間類似度について述べる際に, 課題間類似度が s かつ課題間距離が d の領域を S_{sd} と呼称する。 S_{AA} , S_{AB} には「呼吸器疾患→化石燃料」, 「京都議定書→ウィーン条約」のような妥当だと考えられる関連が比較的多く現れた。 また S_{AC} , S_{AD} では「地球温暖化→一酸化炭素」, 「がん保険→たばこ病」のような「呼吸器疾患→化石燃料」などと比べると直接的ではないが妥当だと考えられる関連が現れた。 S_{BA} には「モントリオール議定書→地球温暖化」のような妥当だと考えられる関連が含まれていたが, 全体としてその割合は高くない結果となった。 一方, S_{BB} や S_{CA} などの残りの領域ではより弱い関連の割合が高くなる。 これらのことから課題間距離-課題間類似度を用いた課題関連度は図 4.5 のように各領域を分類した。 これによって分類した課題間距離-課題間類似度の各領域の課題関連度は $[1]>[3] \geq [2] >> [4]$ のような関係になっていると考えられる。

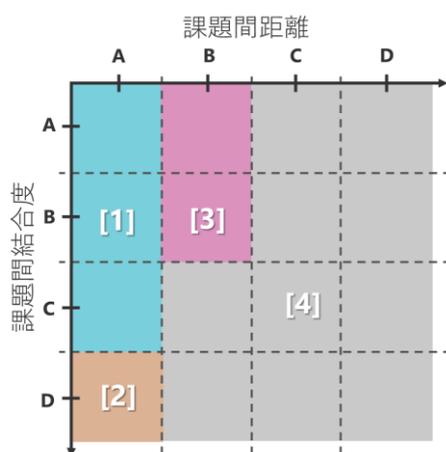


図 4.4 : 課題間距離-課題間結合度の各領域の分類

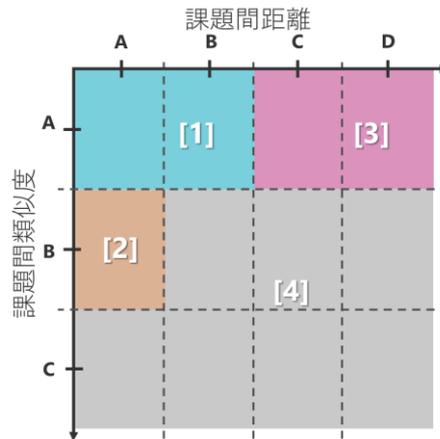


図 4.5：課題間距離-課題間類似度の各領域の分類

4.3 エラー分析

前節で課題間距離-課題間結合度，課題間距離-課題間類似度についてそれぞれ4つの領域に分類を行ったところ，関連性は少ないと思われるが[1]，[2]，[3]の領域に属している課題の組み合わせも含まれていていた．特に課題間距離-課題間結合度の[2]，[3]や課題間距離-課題間類似度の[2]では，[1]に比べてこれらの割合が高かった．これらの分類が適当でない事例の一部とその分類を表 4.4 に示す．

表 4.4：分類が適当でない事例の一部

Q_A	Q_B	D_{dc}	S_{sd}
地球温暖化	火災	[2]	[2]
ビタミン	化石燃料	[1]	[2]
気候変動	生命保険	[2]	[2]

分類が適当でない事例を調べたところ，「地球温暖化」，「ビタミン」，「火災」，「生命保険」などの一部の課題との組み合わせのものが多く現れていることが分かった．また表 4.4 で示した事例における相互接続する要素を見ると，「地球温暖化→火災」が「化石燃料」，「ビタミン→化石燃料」が「食品添加物」など，「気候変動→生命保険」が「気候」であった．これらの展開元となった， Q_B に含まれるキーワードはそれぞれ「火」，「環境問題」，「人間」であった．これらのキーワードについての入次数と出次数を表 4.5 に，それぞれの課題キーワードについての入次数と出次数を表 4.6 に示す．

表 4.5：展開元のキーワードの入次数・出次数

語句	入次数	出次数
火	432	254
環境問題	702	403
人間	1878	217

表 4.6：課題キーワードの入次数・出次数

語句	入次数	出次数
地球温暖化	850	311
ビタミン	472	515
気候変動	281	63
火災	959	485
化石燃料	195	197
生命保険	2048	209

表 4.5 が示す通り「ビタミン→化石燃料」をつなげる要因となった「環境問題」の出次数は 403, 「気候変動→生命保険」をつなげる要因となった「人間」の入次数は 1878 であった。これらは DBpedia Japanese から得られたキーワードの入次数や出次数に対して設定した閾値を超えている。また「地球温暖化→火災」をつなげる要因となった「火」はどちらもその閾値未満であったが、2つの課題キーワードである「地球温暖化」と「火災」はどちらも出次数が閾値を超えていた。

このことから、これらのハブとなるキーワードを通じて関連性の少ない課題同士が結び付けられたと考えられる。前章で述べた課題関連度の算出過程において、DBpedia Japanese から得られたキーワードは入次数や出次数を制限することで一般性の高いと思われるキーワードを取り除いていたが、iLSB の Keyword Repository 内にあるキーワードを用いる際にはそれらを考慮していなかった。そのため、Keyword Repository 内のキーワードにも入次数や出次数について制限を設けることが対策として挙げられる。

第5章 実験

本章では提案した課題関連度と人手による関連付けの間の関係性を分析するために行った実験について述べる。

5.1 実験の目的

本実験は提案した課題関連度による関連付けと人手による関連付けの間どのような関係性があるのかを分析するために行った。また課題関連度の算出において関連語句の入次数と出次数によるフィルタリングを行わなかった場合との比較も行った。

5.2 実験の内容

表 5.1 に示す先行研究で得られた 9 つの学習シナリオに対して課題関連度を算出し、それによって得られた関連について、課題関連度の強い順に並べ替えたものと人手によって関連の強い順に並べ替えたものの比較を行った。またそれぞれの関連についてどういった傾向の関連であるかラベルを付け、その比較も行った。

表 5.1：実験に用いた学習シナリオ一覧

初期課題	学習課題数	平均キーワード数
地球温暖化	6	8.3
電気自動車	7	6.1
オゾン層破壊	5	4.0
栄養素	15	4.7
喫煙	23	2.6
生命保険	11	11.4
原子力	13	1.6
民事訴訟	16	2.0
確定申告	10	5.0

表 5.1 に示す 9 つの学習シナリオからは計 566 の関連が得られた。この 566 の関連から課題間距離-課題間結合度、課題間距離-課題間類似度の分類において

[1]から[4]のそれぞれの領域に現れた関連を重複がないようにランダムに5つずつ、計40をサンプリングした。これらからさらに10ずつ取り出し4つのグループに分割した。その際に各グループに課題間距離-課題間結合度か課題間距離-課題間類似度のどちらかにおいて[1]から[4]の各レベルが最低1つは含まれるように分割した。この4グループについてそれぞれ課題関連度と人手による並べ替えを行った。表5.2, 5.3, 5.4, 5.5に各グループの関連とその関連の課題間距離 d , 課題間結合度 c , 課題間類似度 s と、属する領域 D_{dc} と S_{sd} のレベルを示す。

表 5.2 : 実験に用いた関連 (グループ 1)

ID	Q_A	Q_B	d	c	s	D_{dc}	S_{sd}
1	化石燃料	回生ブレーキ	1	1	0.01471	[2]	[4]
2	飽和脂肪酸	喫煙	1	3	0.42132	[1]	[1]
3	原子力	回生ブレーキ	1	1	0.03676	[2]	[2]
4	オゾンホール	化石燃料	1	1	0.07018	[2]	[1]
5	保険法	確定申告	2	3	0.02454	[3]	[4]
6	オゾンホール	温室効果ガス	2	3	0.07018	[3]	[1]
7	一酸化炭素	化石燃料	1	3	0.06612	[1]	[2]
8	タンパク質	温室効果ガス	2	9	0.00424	[3]	[4]
9	電気自動車	不飽和脂肪酸	3	3	0.02712	[1]	[4]
10	所得税	死亡	1	1	0.01550	[2]	[4]

表 5.3 : 実験に用いた関連 (グループ 2)

ID	Q_A	Q_B	d	c	s	D_{dc}	S_{sd}
1	喫煙	栄養素	1	1	0.06881	[2]	[2]
2	呼吸器疾患	化石燃料	1	8	0.85106	[1]	[1]
3	地球温暖化	所得税	4	3	0.00775	[4]	[4]
4	がん保険	たばこ病	5	2	0.24400	[4]	[3]
5	禁煙	確定申告	1	1	0.00990	[2]	[4]
6	脂溶性ビタミン	生命保険	4	4	0.00281	[4]	[4]
7	電気自動車	温室効果ガス	2	3	0.11864	[3]	[1]
8	喫煙	タンパク質	3	10	0.08300	[4]	[3]
9	地球温暖化	一酸化炭素	3	2	0.11570	[4]	[3]
10	タンパク質	化石燃料	2	5	0.01581	[3]	[4]

表 5.4 : 実験に用いた関連 (グループ 3)

ID	Q_A	Q_B	d	c	s	D_{dc}	S_{sd}
1	地球温暖化	モントリオール 議定書	1	2	0.06502	[1]	[2]
2	ニコチン	がん保険	1	1	0.05128	[2]	[2]
3	オゾンホール	京都議定書	1	1	0.07018	[2]	[1]
4	電気自動車	アミノ酸	2	1	0.04202	[4]	[4]
5	タンパク質	一酸化炭素	1	2	0.02479	[1]	[4]
6	タンパク質	たばこ病	1	2	0.03162	[1]	[2]
7	がん保険	法人税	3	1	0.01681	[4]	[4]
8	電気自動車	京都議定書	2	6	0.10484	[3]	[1]
9	一酸化炭素	地球温暖化	1	7	0.11570	[1]	[1]
10	酸性雨	一酸化炭素	3	4	0.09302	[4]	[3]

表 5.5 : 実験に用いた関連 (グループ 4)

ID	Q_A	Q_B	d	c	s	D_{dc}	S_{sd}
1	火災	地球温暖化	2	2	0.06250	[4]	[4]
2	原子力事故	タンパク質	2	1	0.00851	[4]	[4]
3	一酸化炭素	温室効果ガス	1	7	0.12397	[1]	[1]
4	禁煙	がん保険	1	1	0.03960	[2]	[2]
5	京都議定書	不飽和脂肪酸	2	1	0.02712	[4]	[4]
6	酸性雨	受動喫煙	1	1	0.02326	[2]	[4]
7	所得税	火災	1	1	0.00775	[2]	[4]
8	たばこ病	ビタミン	3	1	0.07143	[4]	[3]
9	電気自動車	原子力事故	3	4	0.01702	[4]	[4]
10	生命保険	所得税	1	2	0.06977	[1]	[2]

課題関連度による並び替えは課題間距離-課題間結合度, 課題間距離-課題間類似度それぞれの分類を用いて 2 通り行った. この時にある関連が属する領域 D_{dc} , S_{sd} によって分類されるレベルをその関連のラベルとして扱った. 課題間距離-課題間結合度の順位付けの基準は以下の手順で行った.

- 手順1. 属する領域 D_{dc} のレベルが[1]>[2]>[3]>[4]の順に並べ替える.
- 手順2. 同じレベルの中で課題間距離について昇順に並べ替える.
- 手順3. 手順2後に同順位のもの、それらを課題間結合度について降順に並べ替える.
- 手順4. 手順3後に同順位のもの、それらを課題間類似度について降順に並べ替える. これによっても差がつかないものは同順位として扱う.

課題間距離-課題間類似度の順位付けの基準は以下の手順で行った.

- 手順1. 属する領域 S_{sd} のレベルが[1]>[3]>[2]>[4]の順に並べ替える.
- 手順2. 同じレベルの中で課題間類似度について降順に並べ替える.
- 手順3. 手順2後に同順位のもの、それらを課題間距離について昇順に並べ替える.
- 手順4. 手順3後に同順位のもの、それらを課題間結合度について降順に並べ替える. これによっても差がつかないものは同順位として扱う.

人手による関連の並べ替えは教員2名が個々に行った. 同時に表5.6に示す基準をもとに並べ替えた関連にラベル付けを行った. ここでいう文脈とは2つの学習課題の接点となるキーワード(群)を考えたときに、その2つの学習課題が同じような脈絡で現れるかどうかである. このラベルのi, ii, iii, ivがそれぞれ課題関連度の[1], [2], [3], [4]に対応している.

表 5.6 : 人手によるラベル付けの基準

ラベル	基準
i	文脈も距離も近い
ii	距離は近いが、文脈は異なる
iii	距離は遠いが、文脈は近い
iv	距離が遠く、文脈も異なる もしくは接点を持たない

5.3 実験結果

表5.2, 5.3, 5.4, 5.5に示した関連に対して人手と課題間距離-課題間結合度, 課題間距離-課題間類似度による並び替えを行ったところ, それぞれ表5.7, 5.8, 5.9, 5.10のようになった. ここでAとBは人手による結果である.

まず人手による関連付けがどの程度一致するのを見るために各グループの A と B の結果の間のスピアマンの順位相関係数を求めた。スピアマンの順位相関係数 r_s は n 対の変数 X と Y のデータがあり、 X と Y それぞれに順位がついているとき、 n 対あるうちの i 対目のデータ X_i と Y_i の順位の差を d_i とすると(2)式で求められる。

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2)$$

データに同順位のものが含まれる場合には(3)式で求められる。

$$r_s = \frac{T_x + T_y - \sum_{i=1}^n d_i^2}{2\sqrt{T_x T_y}} \quad (3)$$

ここで変数 X の同順位の組の数を X_R 、変数 Y の同順位の組の数を Y_R 、変数 X の j 番目の同順位の組のデータ数を t_j 、変数 Y の k 番目の同順位の組のデータ数を t_k としたとき、 T_x 、 T_y はそれぞれ(4)式、(5)式で求まる。

$$T_x = \frac{n^3 - n - \sum_j^{X_R} (t_j^3 - t_j)}{12} \quad (4)$$

$$T_y = \frac{n^3 - n - \sum_k^{Y_R} (t_k^3 - t_k)}{12} \quad (5)$$

(2)式を用いて求めた各グループの A と B の結果の間のスピアマンの順位相関係数を表 5.11 に示す。

次に各グループの A, B の結果と課題間距離-課題間結合度, 課題間距離-課題間類似度による結果の間のそれぞれのスピアマンの順位相関係数を求めた。(2)式を用いてスピアマンの順位相関係数を求めた結果を表 5.12, 5.13 に示す。

表 5.7: グループ 1 の並べ替えの結果

順位	A		B		課題間距離- 課題間結合度		課題間距離- 課題間類似度	
	ID	ラベル	ID	ID	ID	ラベル	ID	ラベル
1	1	i	6	i	2	i	2	i
2	6	i	7	i	7	i	4	i
3	4	ii	4	ii	4	ii	6	i
4	7	ii	2	iii	3	ii	7	ii
5	5	ii	5	iii	10	ii	3	ii
6	10	iii	8	iii	1	ii	10	iv
7	2	iii	3	iv	8	iii	1	iv
8	3	iii	1	iv	6	iii	5	iv
9	8	iv	9	iv	5	iii	8	iv
10	9	iv	10	iv	9	iv	9	iv

表 5.8: グループ 2 の並べ替えの結果

順位	A		B		課題間距離- 課題間結合度		課題間距離- 課題間類似度	
	ID	ラベル	ID	ID	ID	ラベル	ID	ラベル
1	4	i	9	i	2	i	2	i
2	7	i	7	i	1	ii	7	i
3	9	ii	2	i	5	ii	9	iii
4	2	ii	1	ii	10	iii	8	iii
5	1	iii	10	ii	7	iii	4	iii
6	8	iii	4	iii	8	iv	1	ii
7	3	iii	8	iii	9	iv	5	iv
8	5	iv	5	iv	6	iv	10	iv
9	6	iv	6	iv	3	iv	3	iv
10	10	iv	3	iv	4	iv	6	iv

表 5.9：グループ 3 の並べ替えの結果

順位	A		B		課題間距離- 課題間結合度		課題間距離- 課題間類似度	
	ID	ラベル	ID	ID	ID	ラベル	ID	ラベル
1	1	i	1	i	9	i	9	i
2	10	i	9	i	1	i	3	i
3	3	i	3	i	6	i	8	i
4	2	ii	8	ii	5	i	10	iii
5	8	ii	5	ii	3	ii	1	ii
6	9	ii	6	ii	2	ii	2	ii
7	5	iii	10	iii	8	iii	6	ii
8	6	iii	2	iii	4	iv	5	iv
9	7	iv	7	iv	10	iv	4	iv
10	4	iv	4	iv	7	iv	7	iv

表 5.10：グループ 4 の並べ替えの結果

順位	A		B		課題間距離- 課題間結合度		課題間距離- 課題間類似度	
	ID	ラベル	ID	ID	ID	ラベル	ID	ラベル
1	4	i	3	i	3	i	3	i
2	10	i	9	i	10	i	8	iii
3	1	ii	6	ii	4	ii	10	ii
4	3	ii	1	ii	6	ii	4	ii
5	9	iii	10	ii	7	ii	6	iv
6	8	iii	4	iii	1	iv	7	iv
7	6	iv	8	iii	5	iv	1	iv
8	7	iv	2	iv	2	iv	5	iv
9	5	iv	5	iv	9	iv	2	iv
10	2	iv	7	iv	8	iv	9	iv

表 5.11：人手による並べ替えのスピアマンの順位相関係数

	グループ 1	グループ 2	グループ 3	グループ 4
r_s	0.45	0.60	0.60	0.53

表 5.12：課題間距離-課題間結合度に対するスピアマンの順位相関係数

	グループ 1	グループ 2	グループ 3	グループ 4
A	0.16	-0.15	0.24	0.52
B	0.31	0.41	0.77	0.33

表 5.13：課題間距離-課題間類似度に対するスピアマンの順位相関係数

	グループ 1	グループ 2	グループ 3	グループ 4
A	0.44	0.76	0.65	0.48
B	0.65	0.79	0.73	0.22

また A, B の人手によるラベル付けの結果と課題間距離-課題間結合度, 課題間距離-課題間類似度のそれぞれのラベル付けの結果に対する混同行列を表 5.14, 5.15, 5.16, 5.17 に示す.

表 5.14：課題間距離-課題間結合度のラベル付けに対する混同行列(A)

		課題間距離-課題間結合度			
		i	ii	iii	iv
A	i	2	3	2	2
	ii	4	2	2	2
	iii	3	3	0	4
	iv	0	3	2	6

表 5.15：課題間距離-課題間結合度のラベル付けに対する混同行列(B)

		課題間距離-課題間結合度			
		i	ii	iii	iv
B	i	5	1	2	2
	ii	3	3	2	1
	iii	1	2	2	4
	iv	0	5	0	7

表 5.16：課題間距離-課題間類似度のラベル付けに対する混同行列(A)

		課題間距離-課題間類似度			
		i	ii	iii	iv
A	i	3	3	2	1
	ii	5	2	1	2
	iii	1	3	2	4
	iv	0	0	0	11

表 5.17：課題間距離-課題間類似度のラベル付けに対する混同行列(B)

		課題間距離-課題間類似度			
		i	ii	iii	iv
B	i	6	2	1	1
	ii	2	3	0	4
	iii	1	2	4	2
	iv	0	1	0	11

これらをもとにそれぞれの再現率，適合率，F 値を計算した結果を表 5.18, 5.19, 5.20, 5.21 に示す. なおここで示すマクロ平均 F 値はマクロ平均再現率とマクロ平均適合率の調和平均である.

表 5.18：課題間距離-課題間結合度の再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.22	0.22	0.22
ii	0.20	0.18	0.19
iii	0.00	0.00	-
iv	0.55	0.43	0.48
マクロ平均	0.24	0.21	0.22

表 5.19：課題間距離-課題間結合度の再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.50	0.56	0.53
ii	0.33	0.27	0.30
iii	0.22	0.33	0.27
iv	0.58	0.50	0.54
マクロ平均	0.41	0.42	0.41

表 5.20：課題間距離-課題間類似度の再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.33	0.33	0.33
ii	0.20	0.25	0.22
iii	0.20	0.40	0.27
iv	1.00	0.61	0.76
マクロ平均	0.43	0.40	0.42

表 5.21：課題間距離-課題間類似度の再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.60	0.67	0.63
ii	0.33	0.38	0.35
iii	0.44	0.80	0.57
iv	0.92	0.61	0.73
マクロ平均	0.57	0.61	0.59

また比較的これらの値が小さい ii と iii を一つのクラスにまとめて再現率・適合率・F 値を計算した結果を表 5.22, 5.23, 5.24, 5.25 に示す.

表 5.22 : 3 クラス分類の課題間距離-課題間結合度の再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.22	0.22	0.22
ii or iii	0.35	0.41	0.38
iv	0.55	0.43	0.48
マクロ平均	0.37	0.35	0.36

表 5.23 : 3 クラス分類の課題間距離-課題間結合度の再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.50	0.56	0.53
ii or iii	0.50	0.53	0.51
iv	0.58	0.50	0.54
マクロ平均	0.53	0.53	0.53

表 5.24 : 3 クラス分類の課題間距離-課題間類似度の再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.33	0.33	0.33
ii or iii	0.40	0.62	0.48
iv	1.00	0.61	0.76
マクロ平均	0.58	0.52	0.55

表 5.25 : 3 クラス分類の課題間距離-課題間類似度の再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.60	0.67	0.63
ii or iii	0.50	0.69	0.58
iv	0.92	0.61	0.73
マクロ平均	0.67	0.66	0.66

次に課題関連度の算出において関連語句を入次数と出次数によるフィルタリングを行わなかった場合の並べ替えの結果を表 5.26, 5.27, 5.28, 5.29 に示す.

表 5.26 : グループ 1 の並べ替え結果(フィルタリングなし)

順位	課題間距離-課題間結合度		課題間距離-課題間類似度	
	ID	ラベル	ID	ラベル
1	7	i	2	i
2	2	i	6	i
3	6	ii	7	i
4	5	ii	5	i
5	3	ii	9	iii
6	1	ii	4	iii
7	10	ii	3	ii
8	8	ii	1	iv
9	4	iii	10	iv
10	9	iv	8	iv

表 5.27 : グループ 2 の並べ替えの結果(フィルタリングなし)

順位	課題間距離-課題間結合度		課題間距離-課題間類似度	
	ID	ラベル	ID	ラベル
1	2	i	2	i
2	9	i	9	i
3	10	i	5	i
4	5	ii	8	i
5	8	ii	1	i
6	1	ii	4	ii
7	3	iii	7	iii
8	7	iii	10	iii
9	4	iv	3	iv
10	6	iv	6	iv

表 5.28：グループ 3 の並べ替えの結果(フィルタリングなし)

順位	課題間距離-課題間結合度		課題間距離-課題間類似度	
	ID	ラベル	ID	ラベル
1	9	i	9	i
2	1	i	1	i
3	5	i	3	i
4	6	i	8	i
5	3	ii	10	i
6	2	ii	4	i
7	8	iii	6	ii
8	10	iii	2	ii
9	4	iv	5	ii
10	7	iv	7	iv

表 5.29：グループ 4 の並べ替えの結果(フィルタリングなし)

順位	課題間距離-課題間結合度		課題間距離-課題間類似度	
	ID	ラベル	ID	ラベル
1	3	i	3	i
2	10	i	10	i
3	7	ii	7	i
4	4	ii	1	i
5	6	ii	8	i
6	1	iii	4	ii
7	5	iii	6	ii
8	9	iii	9	iv
9	2	iv	5	iv
10	8	iv	2	iv

これらと人手による並べ替えの結果を比較し、スピアマンの順位相関係数を求めた結果を表 5.30, 5.31 に示す。

表 5.30：課題間距離-課題間結合度に対するスピアマンの順位相関係数
(フィルタリングなし)

	グループ 1	グループ 2	グループ 3	グループ 4
A	0.35	-0.08	0.35	0.48
B	0.59	0.47	0.82	0.27

表 5.31：課題間距離-課題間類似度に対するスピアマンの順位相関係数
(フィルタリングなし)

	グループ 1	グループ 2	グループ 3	グループ 4
A	0.20	0.41	0.55	0.58
B	0.67	0.54	0.76	0.28

また A, B の人手によるラベル付けの結果とフィルタリングを行った場合の課題間距離-課題間結合度, 課題間距離-課題間類似度のそれぞれのラベル付けの結果に対する混同行列を表 5.32, 5.33, 5.34, 5.35 に示す.

表 5.32：フィルタリングなしの課題間距離-課題間結合度のラベル付けに対する混同行列(A)

		課題間距離-課題間結合度			
		i	ii	iii	iv
A	i	2	4	2	1
	ii	5	2	3	0
	iii	3	4	2	1
	iv	1	4	1	5

表 5.33：フィルタリングなしの課題間距離-課題間結合度のラベル付けに対する混同行列(B)

		課題間距離-課題間結合度			
		i	ii	iii	iv
B	i	6	2	2	0
	ii	4	2	3	0
	iii	1	5	1	2
	iv	0	5	2	5

表 5.34：フィルタリングなしの課題間距離-課題間類似度の
ラベル付けに対する混同行列(A)

		課題間距離-課題間類似度			
		i	ii	iii	iv
A	i	5	2	1	1
	ii	8	1	1	0
	iii	4	3	0	3
	iv	3	1	2	5

表 5.35：フィルタリングなしの課題間距離-課題間類似度の
ラベル付けに対する混同行列(B)

		課題間距離-課題間類似度			
		i	ii	iii	iv
B	i	8	0	1	1
	ii	4	3	2	0
	iii	5	3	0	1
	iv	3	1	1	7

これらをもとにそれぞれについて再現率・適合率・F 値を求めた結果を表 5.36, 5.37, 5.38, 5.39 に示す.

表 5.36：フィルタリングなしの課題間距離-課題間結合度の
再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.22	0.18	0.20
ii	0.20	0.14	0.17
iii	0.20	0.25	0.22
iv	0.45	0.71	0.56
マクロ平均	0.27	0.32	0.29

表 5.37：フィルタリングなしの課題間距離-課題間結合度の
再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.60	0.55	0.57
ii	0.22	0.14	0.17
iii	0.11	0.13	0.12
iv	0.42	0.71	0.53
マクロ平均	0.34	0.38	0.36

表 5.38：フィルタリングなしの課題間距離-課題間類似度の
再現率・適合率・F 値(A)

	再現率	適合率	F 値
i	0.56	0.25	0.34
ii	0.10	0.14	0.12
iii	0.00	0.00	-
iv	0.45	0.56	0.50
マクロ平均	0.28	0.24	0.26

表 5.39：フィルタリングなしの課題間距離-課題間類似度の
再現率・適合率・F 値(B)

	再現率	適合率	F 値
i	0.80	0.40	0.53
ii	0.33	0.43	0.38
iii	0.00	0.00	-
iv	0.58	0.78	0.67
マクロ平均	0.43	0.40	0.41

またフィルタリングを行った場合の結果と同様に ii と iii を一つのクラスにまとめて再現率・適合率・F 値を計算した結果を表 5.40, 5.41, 5.42, 5.43 に示す。

表 5.40：フィルタリングなしの3クラス分類の課題間距離-課題間結合度の
再現率・適合率・F値(A)

	再現率	適合率	F 値
i	0.22	0.18	0.20
ii or iii	0.55	0.50	0.52
iv	0.45	0.71	0.56
マクロ平均	0.41	0.47	0.44

表 5.41：フィルタリングなしの3クラス分類の課題間距離-課題間結合度の
再現率・適合率・F値(B)

	再現率	適合率	F 値
i	0.60	0.55	0.57
ii or iii	0.61	0.50	0.55
iv	0.42	0.71	0.53
マクロ平均	0.54	0.59	0.56

表 5.42：フィルタリングなしの3クラス分類の課題間距離-課題間類似度の
再現率・適合率・F値(A)

	再現率	適合率	F 値
i	0.56	0.25	0.34
ii or iii	0.25	0.45	0.32
iv	0.45	0.56	0.50
マクロ平均	0.42	0.42	0.42

表 5.43：フィルタリングなしの3クラス分類の課題間距離-課題間類似度の
再現率・適合率・F値(B)

	再現率	適合率	F 値
i	0.80	0.40	0.53
ii or iii	0.44	0.73	0.55
iv	0.58	0.78	0.67
マクロ平均	0.61	0.64	0.62

また課題間距離，課題間類似度，課題間結合度の 3 つのパラメータそれぞれについて人手による関連付けとの間の相関を計算した結果を表 5.44，5.45 に示す．なお課題間距離，課題間結合度については同順位のものを含むデータであるため，(3)式を用いてスピアマンの順位相関係数を求めた．

表 5.44：各パラメータと人手による関連付けの間の
スピアマンの順位相関係数(A)

	課題間距離	課題間類似度	課題間結合度
グループ 1	-0.07	-0.47	0.18
グループ 2	-0.28	-0.13	0.83
グループ 3	-0.26	0.32	0.58
グループ 4	-0.01	0.39	0.60

表 5.45：各パラメータと人手による関連付けの間の
スピアマンの順位相関係数(B)

	課題間距離	課題間類似度	課題間結合度
グループ 1	-0.53	0.24	0.68
グループ 2	0.29	0.00	0.73
グループ 3	0.20	0.59	0.56
グループ 4	-0.29	0.57	0.44

5.4 考察

表 5.11 に示した人手による並べ替えの結果のスピアマンの相関係数は 0.45 から 0.60 で中程度の正の相関がみられた．この結果から人手による関連付けはある程度一致はするものの，ばらつきもみられることが分かった．そのため課題関連度による関連付けと人による関連付けの関係においても，完全に一致する必要はないと考え，それぞれの課題関連度がどのような傾向を持つかという点に注目する．

表 5.12 に示した課題間距離・課題間結合度に対するスピアマンの順位相関係数は，B がグループ 3 に対して行った並び替えの結果との間で 0.77 と強い正の相関を示したものや，A がグループ 1，グループ 2 に対して行った並び替えの結果との間でそれぞれ 0.16，-0.15 とほとんど相関のみられないものもあったが，全体としては弱い正の相関がある結果となった．これらを入次数・出次数によるフィルタリングを行わなかった場合の結果である表 5.30 と比較すると，全体

としてはフィルタリングを行うことで比較的相関が弱くなる傾向が見られた。これはフィルタリングを行うことによって、LOD 上で二つの学習課題の間を直接的につなぐノードが、一般性の高いものとして取り除かれたことが原因だと考えられる。

また表 5.13 に示した課題間距離-課題間類似度に対するスピアマンの順位相関係数は、B がグループ 4 に対して行った並び替えの結果との間で 0.22 と弱い正の相関を示したものもあったが、全体としては中程度の正の相関がある結果となった。これらを入次数・出次数によるフィルタリングを行わなかった場合の結果である表 5.31 と比較すると、全体としてはフィルタリングを行うことで比較的相関が強くなる傾向が見られた。課題間距離-課題間類似度の場合と異なり、フィルタリングによってスピアマンの相関係数が下がらなかった原因としては、課題間類似度はほかの 2 つのパラメータに比べ LOD の特定のノードに対する比重が比較的 low、一般性の高いものが除かれても全体として共有する要素の割合には変化があまりなかったためだと考えられる。

これらの結果から、課題間距離-課題間結合度と課題間距離-課題間類似度を比較すると、課題間距離-課題間類似度が比較的人の行う関連付けに近いと考えられる。また課題間距離-課題間類似度は第 4 章で行ったケーススタディや評価実験の結果を見ると比較的近い分野同士の学習課題を関連付ける傾向にあると考えられる。一方で課題間距離-課題間結合度は人の行う関連付けとは異なる傾向が見られたが、これを利用することで人があまり意識しない関連を見つけることができる可能性がある。またその傾向としては課題間距離-課題間類似度と比べて異なる分野の学習課題も関連付ける傾向が見られた。

一方で表 5.18, 5.19 に示した課題間距離-課題間結合度の再現率・適合率・F 値については、全体的に高いとは言えない結果となった。これらを入次数・出次数によるフィルタリングを行わなかった場合の結果である表 5.36, 5.37 と比較すると、全体として差はあまりなく、一定の傾向はみられなかった。

また表 5.20, 5.21 に示した課題間距離-課題間類似度の再現率・適合率・F 値については、課題間距離-課題間類似度と比べると良いものの、十分に高いとは言えない結果となった。しかしラベル iv については A, B どちらに対しても再現率は高い値となったことから、関連の弱いものについてはある程度排除することができたと考えられる。これらを入次数・出次数によるフィルタリングを行わなかった場合の結果である表 5.38, 5.39 と比較すると、フィルタリングを行うことで全体的に多少向上したといえる。特にラベル iv については、適合率は少し低下したものの再現率が大きく向上した。

一方でフィルタリングの有無にかかわらず課題間距離-課題間結合度、課題間距離-課題間類似度どちらにおいてもラベル ii, iii の再現率・適合率・F 値がラ

ベル i, iv に比べて低い傾向が見られた。ii と iii は関連の傾向は違うものの関連付けの優先度としてはあまり差がないと考えられる。そこでラベル ii, iii を一つのクラスとしてまとめた再現率・適合率・F 値を求めたところ表 5.22, 5.23, 5.24, 5.25, 5.40, 5.41, 5.42, 5.43 に示すように向上する傾向が見られたが十分に高いといえる値にはならなかった。

表 5.44, 5.45 に示した各パラメータと人手による関連付けの間のスピアマンの順位相関係数は、課題間距離については弱い負の相関、課題間類似度については中程度の正の相関、課題間結合度については弱い正の相関がみられた。比較的相関の強い課題間類似度は、課題キーワードから連想されるキーワードの一致率であり、これを用いることで注目している学習課題と分野の近い学習課題を結び付けることが期待できる。一方で、課題間距離と課題間結合度については相関が弱く、人が関連付けを行う際に比較的考慮されていない情報だと考えられ、これらを活用することで人の意識しづらい関連に対する気づきを促すことが可能であると考えられる。課題間距離は DBpedia Japanese 上で二つの学習課題が結びつくまでの距離であり、これが小さいほど結びつきが強いと考えられる。課題間距離が小さい関連の中には異なる分野の学習課題の組み合わせも含まれることがあり、課題間類似度だけでは見つけることが難しい異なる分野間の関連を見つけていくことが期待できる。また課題間結合度は二つの学習課題を双方向につなぐ要素の数であるため、このパラメータが高いものは注目する課題を入れ替えた場合にも強い関連があると考えられ、これを提示することで学習時の視点の切り替えを促すことが期待できる。

第6章 おわりに

6.1 まとめ

本研究では、Web 調べ学習を行う中で作成された、複数の学習シナリオに対して学習課題間での知識構造の関連付けを行うことが、学習者にとって容易ではないという問題に対して、課題間の関連を提示するために LOD に基づく課題関連度を提案した。提案した課題関連度では iLSB 上で学習課題について関連付けたキーワードをもとに、DBpedia Japanese を用いてその学習課題の関連語句集合を取得し、これをもとに学習課題間の関連の強さを算出した。この関連語句集合で共通する要素を持つような学習課題の組み合わせを関連のある学習課題として、その学習課題の組み合わせについて課題間距離、課題間類似度、課題間結合度の 3 つのパラメータを算出した。この 3 つのパラメータをもとに関連の強さに応じた分類を行うため、ケーススタディを行い各パラメータに対する閾値を設定し、課題間距離-課題間結合度、課題間距離-課題間類似度をそれぞれ組み合わせた際の領域に対して関連の強さに応じて 4 段階に分類を行った。

また、提案した課題関連度と人手による関連付けの間の関係性を分析するため、これに基づいた関連付けの優先度と人手による関連付けの優先度がどの程度一致するのか実験を行った。その際にそれぞれの関連に対して人手によるラベル付けも行い、このラベルが課題関連度による領域の分類とどの程度一致するのかも分析した。

その結果人手による順位付けに対して、課題間距離-課題間結合度については弱い正の相関がみられ、課題間距離-課題間類似度については中程度の正の相関がみられた。このことから比較的人の関連付けと近い課題関連度は課題間距離-課題間類似度であり、課題間距離-課題間結合度は人の意識が向きにくい関連を見つける可能性があると考えられる。またラベル付けの結果を比較し、その再現率・適合率・F 値を求めたところ、課題間距離-課題間結合度、課題間距離-課題間類似度どちらにおいても全体的に十分に高いとは言えない結果となったが、課題間距離-課題間類似度については人手によって関連が弱いと分類されたものに対する再現率は高い値を示した。また課題関連度を算出する際に取得する関連語句を入次数・出次数によってフィルタリングしなかった場合に得られた各パラメータを用いて並べ替えとラベル付けについても比較を行った。フィルタリングを行わなかった場合の課題関連度による並べ替えの結果と人手による並べ替えの結果との順位相関係数を求めたところ、課題間距離-課題間結合度につ

いてはフィルタリングを行った場合よりも比較的相関が強くなり、課題間距離-課題間類似度についてはフィルタリングを行った場合よりも比較的相関が弱くなった。またラベル付けの結果に対する再現率・適合率・F 値は、課題間距離-課題間結合度については一定の傾向はみられなかったが、課題間距離-課題間類似度については全体として良化する傾向が見られた。

6.2 今後の課題

本研究の今後の課題として、学習課題の関連語句集合を取得する際に iLSB の Keyword Repository 内に含まれるキーワードの情報を用いているがその中に入次数や出次数が高いキーワードが含まれている場合にその影響を抑制する方法の検討や課題関連度の妥当性を向上させるための各パラメータの閾値や各領域の分類の再検討、LOD に基づいた学習課題の抽象度を推定から比較する学習課題の絞り込みの検討、課題関連度をもとに学習者へ推薦を行う際の提示方法の検討、また提案した課題関連度を用いた支援システムの実装とその評価実験が挙げられる。

参考文献

- [1] 長谷川忍, “Web における主体的学びとリフレクション支援,” *主体的学び 創刊号 特集パラダイム転換「教育から学習へ、ICT 活用へ」*, pp. 61-71, 2014.
- [2] 文部科学省, “教育の情報化のビジョン ~21 世紀にふさわしい学びと学校の創造を目指して~, ” 28 4 2011. [オンライン]. Available: https://www.mext.go.jp/component/a_menu/education/micro_detail/__icsFiles/afieldfile/2017/06/26/1305484_01_1.pdf. [アクセス日: 22 1 2020].
- [3] 情報化の進展に対応した初等中等教育における情報教育の推進などに関する調査研究協力者会議, “第 I 章 情報化に対応した教育について,” 1998. [オンライン]. Available: https://www.mext.go.jp/b_menu/shingi/chousa/shotou/002/toushin/980801e.htm#1-1. [アクセス日: 22 1 2020].
- [4] 文部科学省, “情報活用能力調査の結果概要,” 3 2015. [オンライン]. Available: https://www.mext.go.jp/component/a_menu/education/detail/__icsFiles/afieldfile/2015/03/24/1356195_1.pdf. [アクセス日: 22 1 2020].
- [5] 柏原昭博, 坂本雅直, 長谷川忍, 豊田順一, “ハイパー空間における主体的学習プロセスのリフレクション支援,” *人工知能学会論文誌*, Vol.18, No.5,

- pp.245-256, 2003.
- [6] A. Kashihara , N. Akiyama, “Learning Scenario creation for Promoting Investigate Learning on the Web,” Journal of information and systems in education, Vol.15, No.1, pp.62-72, 2007.
- [7] 佐藤禎紀, 柏原昭博, 長谷川忍, 太田光一, 鷹岡亮, “Web 調べ学習における主体的学習プロセスの診断手法,” 教育システム情報学会第 6 回研究会, Vol.32, No.6, pp.99106 , 2018.
- [8] G. Siemens, “Connectivism: A learning theory for the digital age,” International Journal of Instructional Technology and Distance Learning, 2(1), 3-10, 2005.
- [9] “DBpedia Japanese,” [オンライン]. Available: <http://ja.dbpedia.org/>. [アクセス日: 22 1 2020].

対外発表

・山内拓磨，太田光一，長谷川忍，柏原昭博：Web 調べ学習における課題関連度の提案 -興味の変移に対応した LOD に基づく指標の算定-, 教育システム情報学会研究報告 Vol.32, No.51, pp.83-90 (2020).

謝辞

本研究を進めるにあたり指導教員である長谷川忍准教授には様々な助言，御指導を賜りましたことに深く感謝いたします。また太田光一助教にも様々な助言，御指導を賜りましたことに深く感謝いたします。また電気通信大学大学院情報理工学研究科の柏原昭博教授にも助言いただきましたことに感謝いたします。