

初学者向け論文読解支援システム

Paper reading support system for beginning students

吉田 蒼葉[†], 宮治 裕[†]
Aoba Yoshida[†] Yutaka Miyaji[†]

[†] 青山学院大学 社会情報学部

[†] School of Social Informatics, Aoyama Gakuin University.

要旨

初学者が特定の分野を学びたい場合、知識や情報を得るための一般的な情報源として学術論文が考えられる。しかし、学術論文の読解には専門的な予備知識が必要であるため、初学者にとっては難しいという問題がある。本研究では、専門用語の理解に焦点を当て、専門用語抽出モジュールである「termextract」を利用して専門用語を抽出し、WikipediaAPIを利用して注釈をつけることで論文の理解を支援するシステムを提案した。実験では、一定の専門用語抽出精度が確認されたとともに、スムーズな論文読解を支援できたことが確認された。一方で、専門用語への注釈の内容を改善する必要性が確認された。

1. はじめに

初学者が特定の分野を学びたい場合、知識や情報を得るための一般的な情報の取得手段として学術論文が挙げられる。しかし、初学者にとって学術論文読解は予備知識が必要な困難な課題である [1]。その中でも、初学者にとっては専門用語や独特の表現の理解が困難である [1]。

そのほかの情報の取得手段として、ChatGPT[2]をはじめとしたAIチャットサービスが挙げられる。しかし、学生を対象にした先行研究において、学生にとって難易度の高い問題や未知の問題においてChatGPTの回答への誤認識が高くなるという結果[3]が出ており、初学者にとって困難な課題である学術論文読解にAIチャットサービスを用いることは同様の問題が想定される。

以上のことから、初学者の専門的な語彙を補うことで論文読解を支援するシステムが必要であると考え、本研究では論文中から専門用語と考えられる語を抽出し注釈をつけることで初学者の専門的な語彙を補うシステムを構築する。

2. システム

本システムは、初学者が学術論文読解の際に使用することを想定して構築したWebアプリケーションである。本システムを使用することで、専門的な語彙を補い、学術論文読解を支援することを目的としている。システム構成図を図1に示す。PDFのテキスト化にPDFMiner[4]、形態素解析にMeCab[5]、専門用語抽出にtermextract[6]、データベースに日本語教育語彙表[7]および松下達彦(2011)「日本語を読むための語彙データベース (VDRJ) Ver. 1.0 (教師用)」[8]、注釈にWikipediaAPI[9]を利用した。

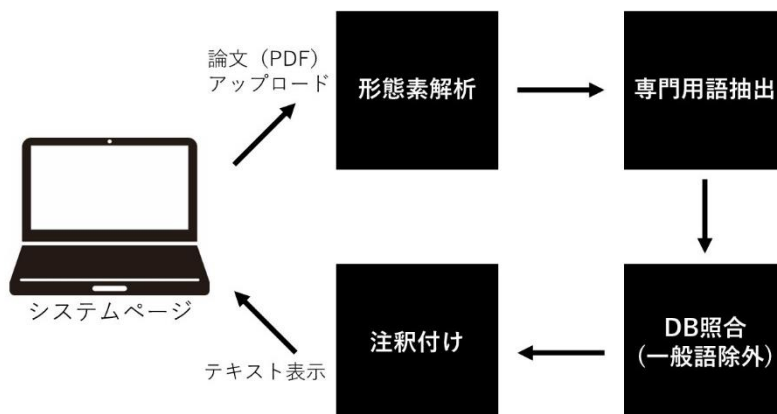


図1 システム構成図

3. 実験

3.1. 目的

本実験の目的は、本研究で構築したシステムを用いて、専門用語の適切な抽出をすることと、専門的な語彙を補うことで論文読解を支援することができるか否かを確認することである。

3.2. 方法

実験は、研究用に開発したシステムを使用した学術論文読解と、実験参加者の従来どおりの方法での学術論文読解を論文 A と B でそれぞれおこなった後に読解後アンケート（自由記述）に回答してもらい、実験終了後に実験後アンケート（4段階,自由記述）に回答してもらうという方法で実施した。実験参加者は大学生10名で、5名1組のチームに分かれておこなった。また、順序効果を相殺するために実験の順序をチームごとに変更した。

3.3. 結果

読解後アンケート調査の結果は図2のようになった。研究用システムの専門用語の抽出精度に関しては、調べる必要があると回答された語のうち、論文 A では 68.2%，論文 B では 100%の抽出精度であった。また、論文 A,B ともに注釈の内容が悪かったと回答されたのは 4.0%であり、注釈がなかった（造語扱いだった）と回答されたのは論文 A では 37.8%，論文 B では 33.0%であった。また、従来どおりの論文読解後のアンケートでは論文 A では理解できた語が 58.3%，理解できなかった語が 41.7%，論文 B では理解できた語が 61.7%，理解できなかった語が 38.3%であった。また、従来どおりの手法での読解と研究用システムを使用しての論文 B の読解は同様の結果であった。

実験後アンケート調査の結果は表1のようになった。設問 5,6,9,10 では平均値が 3.50 以上であり、研究用システムは学術論文読解に役立っているという結果であった。自由記述欄では、「単語を調べる手間が省けた」「システムがなかったらなんとなくの理解で調べずに読み進めてしまっていた」といった記述が見られた。このことから、専門用語への着目によってあいまいな知識で読み進めることを防ぐことができたとともに、単語を調べる手間を省けたことでスムーズな学術論文読解を支援できたと考えられる。

しかし、論文 A では 73.6%，論文 B では 37.0%の単語が理解できなかったという結果であったこと、設問 7,8 の平均値が 3.00 であったことに加えて、自由記述欄での「端的な注釈にするとより使いやすくなる」「注釈がついていない語があって残念だった」という意見から、抽出した語への注釈の内容を今後改善する必要があると考えられる。

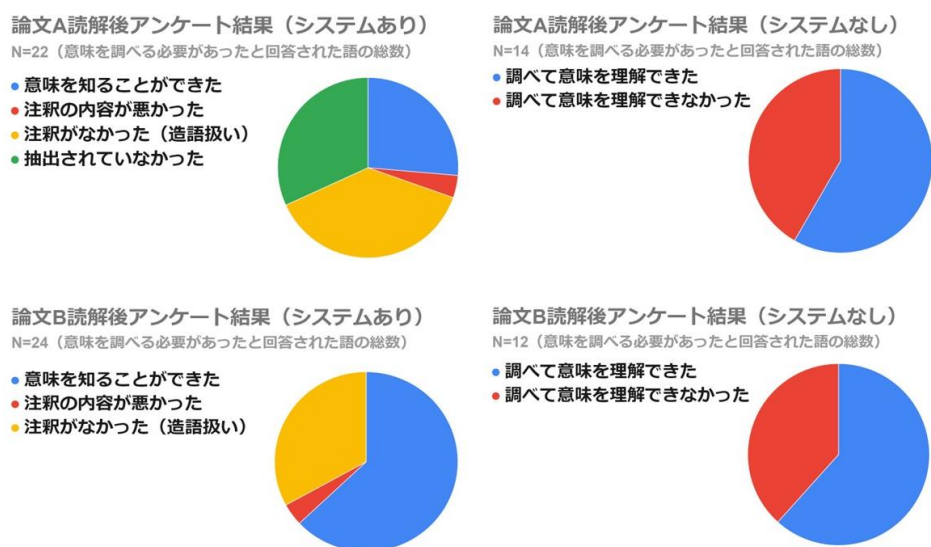


図2 読解後アンケート結果

表1 実験後アンケート結果

番号	設問	平均値	分散値
1	論文を読むことはその分野を知るために必要であると思いますか	3.90	0.10
2	論文読解において専門用語の理解は重要だと思いますか	3.80	0.18
3	論文読解において専門用語は障壁になっていると思いますか	3.40	0.71
4	論文読解において専門用語を調べる頻度はどの程度ですか	2.60	1.16
5	今回の実験で専門用語へ着目したことは読解に役立ちましたか	3.50	0.50
6	専門用語に着目するために本システムは役立ちましたか	3.70	0.23
7	本システムで抽出されていた語は適切であったと思いましたか	3.00	0.44
8	本システムで抽出されていた語の意味をどの程度知ることができましたか	3.00	0.22
9	本システムは問題なく使用できましたか	3.60	0.27
10	本システムは論文読解に役立ちましたか	3.80	0.18

4. おわりに

初学者にとって学術論文の読解は専門的な予備知識が必要であるため困難な課題である。本研究では専門用語抽出と注釈によって初学者の専門的な語彙を補い、論文読解を支援することを目的としたシステムを提案した。

実験では、本システムの有用性を検証した。実験の結果から、専門用語の誤認識を防ぐことができたとともに、単語を調べる手間を省けたことでスムーズな学術論文読解を支援できたと考えられ、本システムは有用であったといえる。一方で、抽出した語への注釈内容に問題があると判明した。今後、注釈を端的にすること、注釈に対応した語を増やすこと等の改善が必要であると考えた。

参考文献

- [1] 犬塚美輪, “相互説明を用いた学術論文読解の指導”, *The Science of Reading*, Vol.53, No.3, 2010, pp.83-94.
- [2] ChatGPT <https://chat.openai.com/> (参照 2023-10-01)
- [3] Merten Nikolay Dahlkemper, Simon Zacharias Lahme, and Pascal Klein, “How do physics students evaluate artificial intelligence responses on comprehension questions? A study on the perceived scientific accuracy and linguistic quality”, *Phys. Rev. Phys. Educ. Res.*, 19, 2023, Article 010142
- [4] PDFMiner <https://github.com/pdfminer/pdfminer.six> (参照 2023-10-01)
- [5] MeCab. <https://taku910.github.io/mecab/> (参照 2023-10-01)
- [6] termextract <http://gensen.dl.itc.u-tokyo.ac.jp/pytermextract/> (参照 2023-10-01)
- [7] 日本語教育語彙表 <http://jhlee.sakura.ne.jp/JEV/> (参照 2023-10-10)
- [8] 日本語を読むための語彙データベース (VDRJ) Ver. 1.0 (教師用)
<http://www17408ui.sakura.ne.jp/tatsum/list.html> (参照 2023-10-10)
- [9] WikipediaAPI https://www.mediawiki.org/wiki/API:Main_page/ja (参照 2023-10-01)