

解答検証を利用した選択式問題への自動解答

佐藤 航[†] 酒井 哲也[†]

[†] 早稲田大学基幹理工学部情報理工学科 〒169-8555 東京都新宿区大久保 3-4-1

E-mail: ^{††}nn.ks-11897@ruri.waseda.jp, ^{†††}tetsuyasakai@acm.org

あらまし 質問応答 (Question Answering) において、日常生活に近いような質問文の内容のみだけでなく、その背景に必要とされる前提知識や質問文の前後の文脈を考慮する必要性のある状況での自動解答にはまだ課題が残されている。そのような質問応答において、日本の大学入試の世界史科目を利用し調査していくことを目的とする NTCIR QA Lab [2] というタスクがある。本研究では、QA Lab タスクにおける選択肢が与えられる質問に注目する。質問に対する解答の選択肢が単語もしくは文章で与えられた場合に、正しいものを自動選択するために各選択肢の解答検証を行うシステムの作成を行う。そのシステムの評価実験を行うために NTCIR QA Lab タスクで提供されるセンター試験の世界史の教科書データと問題文・選択肢データを利用して提案手法を検証する。

キーワード 質問応答, 解答検証, 情報検索

1. はじめに

質問応答 (Question Answering) とは、与えられる質問に対してコンピュータが自動・または半自動で解答を生成し、その生成された解答の評価を行う分野である。情報検索において、ユーザがいくつかのキーワードを入力して検索する時、ユーザはその検索結果の上から十数件のドキュメントしか参照しない状況がよくある。この方法で検索を行うと、適合性が最も高いものがユーザの最も求めている情報に必ずしも一致しない状況が発生しうる。そのような精度よりも再現率が重要視される状況は例えば医療・金融・知識検索などの専門的な知識を要求される場面で発生する。再現率の高いシステムの実現によって、専門家がより適切な判断を下すための助けとすることが出来る [4]。

これに対して、日本の大学入試の世界史科目を利用し調査していくことを目的とする NTCIR QA Lab [2] というタスクがある。このタスクは 2014 年に第 1 回が開催され、2017 年には第 3 回が開催される予定である。このタスクへの参加者に対して、大学入試の世界史科目のセンター試験と 5 大学 (東京大学・京都大学・北海道大学・早稲田大学・中央大学) の二次試験の問題データと解答データが配布される。参加者はセンター試験、二次試験の問題に対して自動解答をするシステムを作成し、ある特定の年の本番の試験で出題された問題を実験データとして使用し、各システムの評価を行う [9]。

センター試験の問題は選択肢式になっており、問題文と選択肢が与えられ、解答としてふさわしい選択肢を 1 つ選択するシステムを作成するタスクになっている。二次試験の問題は記述式になっており、問題文やその問題に解答するための条件が設けられているので、その条件に従って解答を作成するシステム、または部分的なモジュールを作成するタスクとなっている。センター試験ではただ選択肢を眺めるだけで正解に結びつくことの出来る問題はそれほど多くない。高得点を取るためには、問題の背景にあるコンテキストや時代背景などについての理解も

要求される。教科書の知識をそのまま適用するのではなく、知識から推論することによって初めて結論が得られる問題も出題されており、より現実世界に即した問題設定になっていると考えられる。また、二次試験ではそれらに加えて持ち合わせた知識を正しい論理構造で組み合わせて自ら解答を作成することが求められる。

本稿では、このタスクで提供されているデータを利用して、センター試験タスクに取り組む。問題文内に登場するキーワードと選択肢内に登場するキーワードを抽出し、教科書データを検索したその結果を利用することで、その選択肢の解答の正当性の評価を行う。

以下、2. 節でセンター試験タスクの概要、3. 節で関連研究、4. 節で提案手法について紹介し、5. 節で実験とその結果についてまとめ、最後に 6. 節で考察・結論について述べる。

2. センター試験タスクの概要

本節では、NTCIR QA Lab タスクで提供されているセンター試験タスクの概要について説明する。

センター試験タスクでは、本試験の 8 年分の問題とそれに対応する正解データが XML の形式で提供されている [8]。また、提供される問題データについて、問題文ごとに問題の種類や解答の形式を判別するための question format が割り当てられており、センター試験タスクにおいては以下の 4 つに大別される。

Factoid

Factoid 型では、問題に対して選択肢は単語 1 つで与えられるシンプルな選択肢になっている。

Slot-Filling

Slot-Filling 型では、問題文で空欄が与えられ、それに当てはまるものの候補が選択肢として与えられる問題になっている。

True-or-False

True-or-False 型では、選択肢が文で与えられるので正しいものを、もしくは誤っているものを選ぶ問題になっている。

Unique

Unique 型では、画像や地図上の位置などが選択肢になっており、今回はこの画像が関連する問題に関しては考慮しないことにする。

3. 関連研究

本節では、本研究に関連する研究について述べる。

Kobayashi ら [6] は、2016 年に開催された本タスクについて、まずセンター試験の世界史科目について問題にどのような特徴があるかを分析を行い、以下の 3 つの仮説を立てその正当性について検証した。

- (1) センター試験世界史科目においては、1 つの問題に対して章や節などをまたいで手がかりを探す必要性はなく、問題を解くために必要な手がかりは一段落内に収まっているということ。
- (2) 問題文や選択肢の中では教科書で記述されている時代的な記述よりもより大雑把な記述がなされる (例えば、教科書内では 1453 年の出来事として書かれているが、問題文には 15 世紀として書かれる場合がある) ということ。
- (3) 上記 (2) に関して、場所についても同様のことが言える (例えば、教科書のフランスという表現が選択肢内ではヨーロッパと書かれる場合がある) ということ。

以上の考えに基づき 4 つの異なる Solver について提案をしている。また、彼らは質問応答のアプローチとして大別するとこれまでに 3 種類のアプローチが取られてきたことについて紹介し、それぞれの Solver はこの 3 種類のアプローチのいずれか 1 つをベースとした構成になっている。最終的にはそれぞれの Solver が出した答えを統合して 1 つの選択肢を決定するというシステムを提案している。

上で述べたように、複数の Solver を用意してそこから得られる結果を組み合わせることで 1 つの答えを得る手法の有用性は、Ferrucci ら [4] によって考案された IBM Watson^(注1) によっても示されている。彼らは、文書検索エンジンを用いた情報検索に加え、知識データベースを利用した検索も組み合わせることによって再現率を高めることに成功している。

Kimura ら [5] は、センター試験の世界史の問題において、下線部情報が問題文に必要となる場合とそうでない場合の分析を行った。センター試験の世界史の問題においてよく見られる問題の形式として、出題の前提となる文章中の、特定の下線部に関連する事柄について問うものがある。このような出題の例を図 1 に示す。ここで、問 2 では問題を解くために下線部の情報が必要であるのに対し、問 3 では下線部の情報は必要なく、問題文と選択肢のみで完結していることが分かる。

(注1) : <https://www.ibm.com/watson/>

第 2 問 学校のあり方は、時代と地域によって様々に異なっていた。世界史における学校・教育に関する次の文章 A～C を読み、下の問い(問 1～9)に答えよ。

(配点 25)

A ヨーロッパ世界では、学校教育は古代ギリシア以来の伝統を持っている。プラトンが開いた学園であるアカデメイアは、**ア**の皇帝①ユスティニアヌスによって 6 世紀に異教の温床として閉鎖されるまで、約 900 年間存続した。中世になると、今日のヨーロッパの大学の原型が出現している。イスラーム世界では、**イ**と呼ばれる機関で法学を中心とした教育が行われた。イスラーム世界の文化は、中世のヨーロッパに大きな影響を与え、②12 世紀ルネサンスの源の一つとなった。

問 2 下線部①の人物について述べた文として誤っているものを、次の①～④のうちから一つ選べ。 **11**

- ① 聖ソフィア(ハギア=ソフィア)聖堂を建設した。
- ② 北アフリカのヴァンダル王国を滅ぼした。
- ③ イタリア半島の領土をすべて失った。
- ④ 『ローマ法大全』を編纂させた。

問 3 下線部②に関連して、12 世紀に起こった出来事について述べた文として正しいものを、次の①～④のうちから一つ選べ。 **12**

- ① イベリア半島では、ナスル朝が滅んだ。
- ② 中央アジアでは、ホラズム朝が倒された。
- ③ タイでは、アユタヤ朝が成立した。
- ④ 日本では、鎌倉幕府が開かれた。

図 1 下線部に関連したことを問う問題と問わない問題の例

また、Magnini ら [7] は質問応答システムの解答の検証方法に関して、相互情報量(Pointwise Mutual Information)をベースに Corrected Conditional Probability (CCP) という評価指標を提案した。CCP は、質問クエリ内に含まれているキーワードと、解答の候補となるキーワードによるインターネット上の検索結果のドキュメント数に基づいている。そして彼らは、TREC 2001 QA Track [3] のデータを利用して CCP に有用性があることを示している。

4. 提案手法

本節では、本研究で提案するシステムの手法について説明する。

以下では今回提案するシステムについて、キーワードの抽出、検索クエリの生成、選択肢のスコアリングと解答の決定手法の三つに分けて説明する。また、システムについてシステム全体の挙動の概要を図 2 に示す。

4.1 キーワード抽出

まずは、問題データを読み込み XML をパースする。問題文と選択肢をそれぞれパースしたのち、はじめにそれぞれの文に対して形態素解析を行う。形態素解析は、工藤ら [10] による MeCab を使用した。図 1 でも触れた通り、その問題を解くために問題文と選択肢だけではなく出題の前提となる文章中に引かれている下線部の情報が必要となる場合がある。これだと XML をパースした段階では問題文の内容が不十分である。そこで、前提となる文章をパースしている最中で下線部が表れた

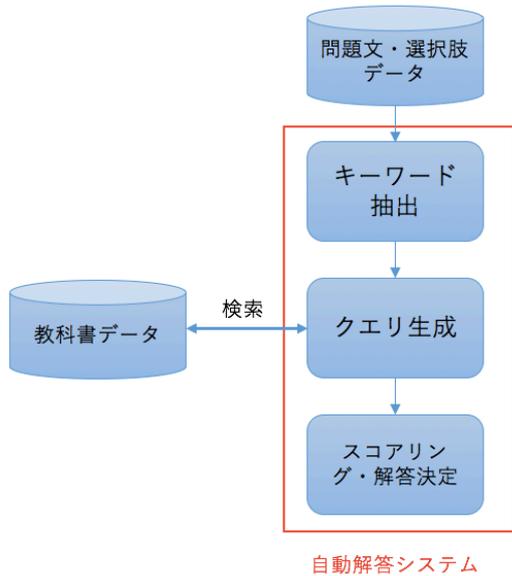


図2 提案システムの概要

らその内容を問題文データの XML の target ID 属性によって紐付けて下線部の単語や文を保存しておく。そして、問題文中に下線部がある場合、その保存した情報を問題文の下線部に補完する。この補完を終えたら形態素解析を行い、キーワード抽出をする。

キーワード抽出では形態素解析の結果を利用して名詞を抽出し、更に名詞の中でも固有名詞は問題を解くために重要になると考えられるので固有名詞とそれ以外の名詞を別に扱うことにする。形態素解析の結果、問題文において固有名詞と判定された名詞の集合を *question_proper_nouns*、それ以外の名詞の集合を *question_general_nouns*、選択肢において固有名詞と判定された名詞の集合を *choice_proper_nouns*、それ以外の名詞の集合を *choice_general_nouns* としてキーワード抽出を完了する。

4.2 クエリ生成

次に、4.1 節によって抽出された名詞集合を基に、教科書データを検索するためのクエリを作成する。問題文の形態素解析から得られる固有名詞、一般名詞のそれぞれの集合 *question_proper_nouns* と *question_general_nouns* から生成される検索クエリを問題文クエリ (*question_query*) とする。同様に、選択肢の形態素解析から得られる固有名詞、一般名詞のそれぞれの集合 *choice_proper_nouns* と *choice_general_nouns* から生成される検索クエリを選択肢クエリ (*choice_query*) とする。

もし、与えられた問題に対して、その選択肢の主張していることが正しいとする。その時に問題文内に現れる名詞集合と、選択肢内に現れる名詞集合を連結したクエリを生成すれば、そのクエリに対して検索結果は後のスコアリングが高くなると考えられる。これに従って、その選択肢を検証するための検索クエリとして *question_query* と *choice_query* を組にした検証クエリ (*validation_query*) を作成する。よって、検索クエリは

表1 年ごとのシステムの得点

年	1997	2001	2003	2005	2007	2009
得点	22 / 85	24 / 87	40 / 98	17 / 100	22 / 95	40 / 88

次のような形になる。

$$validation_query = (question_query, choice_query) \quad (1)$$

つまり、1つの問題文に対して検証クエリは選択肢の個数ぶんだけ生成されることになる。そして、その生成されたクエリによって、教科書データ内を検索する。検索をするにあたり、全文検索システムとして Elasticsearch [1] を使用した。Elasticsearch によりインデックスを作成する際には、Kobayashi ら [6] の考え (3. 節) に基づき、教科書データの文章を句点で区切った上で、各文をドキュメントとして扱った。

4.3 スコアリング・解答決定

教科書データ内のドキュメントの類似度と検証クエリとの類似度の高さを $relativity_score_{document, validation_query}$ とする。4.2 節で紹介した方法によって生成された検証クエリを、全文検索システムに入力すると、全ドキュメントのうち $relativity_score$ の高い順にドキュメント集合が結果として出力される。

次に、検証クエリによって得られるドキュメント集合とその $relativity_score$ に基づき、この検証クエリのクエリ得点 (*query_score*) を算出する。センター試験の世界史の問題を解くためには、ドキュメントの数よりも検索結果上位の $relativity_score$ の方が重要であると仮定し、クエリ得点を $relativity_score$ の上位 k 件の平均とする。

$$query_score = \frac{\sum_{i=1}^k relativity_score_i}{k} \quad (2)$$

今回は $k = 5$ とした。

システムが最終的に選ぶ選択肢は、この検証クエリから出されるクエリ得点によって決定する。「正しいもの」を選択する問題であればクエリ得点が最高のものを選択し、逆に「誤っているもの」を選択する問題であればクエリ得点は最低のものを選択するようにする。

5. 実験と結果

本節では、実験方法とその評価方法について説明し、実験を行った結果を示す。

今回は、センター試験本試験6年分 (1997, 2001, 2003, 2005, 2007, 2009) のデータを評価のために利用する。今回作成したシステムの得点と満点を年ごとに以下の表1に示す。

今回の実験で使用したセンター試験世界史6年分 (1997, 2001, 2003, 2005, 2007, 2009) のデータに対するシステムの評価値を、試験が実施された年の昇順に図3～図8にグラフとして示す。

また、グラフの各項目についての説明を以下に示す：

ID

問題の解答番号を表す。

p/n

その問題について、「正しいもの」を選ぶ問題なのか、「誤っているもの」を選ぶ問題なのかを示す。問題ごとに、p ならば「正しいもの」、n ならば「誤っているもの」を選ぶ問題であるということを表す。

ans

その問題について、正解の選択肢の番号を表す。また、番号が丸で囲われている時は、システムの解答が正解に一致していることを表す。

1 ~ 6

その問題について、選択肢ごとにシステムが出力した *query_score* の値を表す。

次に、不正解だった問題に関して、正解である選択肢のクエリ得点と、システムが選択した選択肢のクエリ得点の差の分布を図 9 に示す。横軸は得点の差を切り下げた値を表し、その範囲に対して該当する問題の個数を縦軸で表す。

6. 考察・結論

本節では、5. 節で得られた結果を基に失敗分析、考察をする。

6.1 キーワード抽出

キーワード抽出を行う時、形態素解析を行う段階において、一単語として取得されるべき単語が分割して抽出されてしまうという問題があることが分かった。この問題によって、一単語として取得されるべき単語が分割され、より一般的で広い意味を持つ単語となってしまう。その影響を受け、後に行う検索の結果にその単語と全く関係の無い文が高い得点で評価されてしまうという問題が発生する。

例として、2001 年本試験第 1 問 問 1 の問題文と選択肢を図 10 に示す。

この問題において、問題文冒頭に「下線部①の帝国」とある。これは図 10 において問題文の上に記されている出題の前提となる文章中の「神聖ローマ帝国」に対応している。システムがこの対応付けをできていれば、この問題を解くために必要な条件は揃っていると言える。

そこで、問題文と各選択肢についてのキーワード抽出の結果を表 2 に示す。

表 2 2001 年本試験第 1 問 問 1 の問題文と選択肢に対するキーワード抽出の結果

問題文 / 選択肢	抽出されたキーワード
問題文	ローマ帝国, 帝国
選択肢 1	帝国, 五, 賢, 帝, 時代, 最盛
選択肢 2	フランス, 14, 世紀, 帝国, 百, 戦争
選択肢 3	ナポレオン, 帝国, 復活
選択肢 4	ハインリヒ, ローマ, 帝国, 皇帝, 4, 教皇, 対立

この表より、上でも述べたように一単語として取得されるべき単語（神聖ローマ帝国、五賢帝、百年戦争、ハインリヒ 4 世）が一単語として取得されていないことが分かる。一般的な情報

検索においては、検索結果の再現率を上げるために検索クエリを細かく分割するという手法があるが、精度を下げる要因となりうる。実際に、例で挙げた問題では選択肢 4 が正解となるが、システムは 1 を選択肢している。そこで、実際に選択肢 1 に対して生成された検証クエリに対して *relativity_score* の高いドキュメントを参照してみると、「最盛」や「時代」というキーワードに共起して神聖ローマ帝国は別の国の事柄が書かれているドキュメントに高いスコアが付けられていることが確認できた。

6.2 クエリ生成

キーワード抽出を終えると、本システムでは次にクエリ生成を行う。このクエリ生成の段階において明らかとなった問題として、抽象的な単語との共起が挙げられる。この節でも、1 つ例を挙げて分析を行う。例として、2003 年本試験第 2 問 問 2 の問題文と選択肢を図 11 に示す。

この問題の正解は 3 であり、システムは 4 を選択している。まず、この問題文と選択肢に対するキーワード抽出の結果を表 3 に示す。

表 3 2003 年本試験第 2 問 問 2 の問題文と選択肢に対するキーワード抽出の結果

問題文 / 選択肢	抽出されたキーワード
問題文	ルネサンス
選択肢 1	ヴェネツィア, メディチ, ルネサンス, 文化, 保護
選択肢 2	ボッカチオ, ルネサンス, 代表, 画家, 一
選択肢 3	ドン, キホーテ, スペイン, セルバンテス, ルネサンス, 代表, 文学, 一
選択肢 4	ラブレール, 愚, 礼, 讃, 教会, 聖職, 腐敗, 風刺

この問題において、教科書データ内に 3 の記述が見られないというわけではない。確かに教科書データ内には「スペインのセルバンテスは『ドン=キホーテ』を著して、没落する騎士たちを風刺し、エル=グレコは宗教改革に対抗して幻想的な宗教画を描いた。」という文があり、これがこの問題を解くための手がかりである。ここで、それぞれの選択肢に対して抽出されたキーワードに注目すると、文化、代表、文学、教会、一といったキーワードが抽出できていることが分かる。このような抽象的な単語が含まれることによって、検証クエリに対する検索結果のドキュメント数が非常に多くなることが分かった。各選択肢に対して生成された検証クエリに対してその検索結果のドキュメント数を表 4 に示す。

表 4 2003 年本試験第 2 問 問 2 において、各選択肢に対して検証クエリとその検索結果の適合文書数

生成した検証クエリ	適合文書数
(問題文, 選択肢 1)	592
(問題文, 選択肢 2)	1494
(問題文, 選択肢 3)	1494
(問題文, 選択肢 4)	152

また、各検証クエリに対する *relativity_score* の高いドキュ

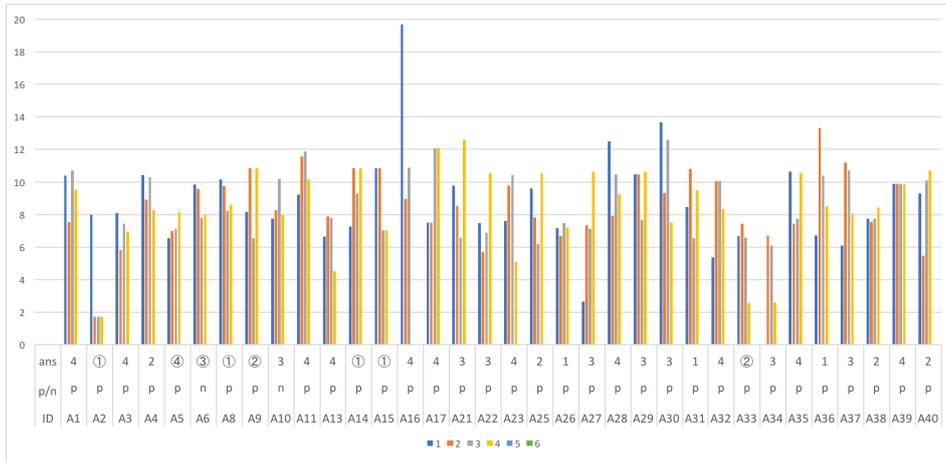


図3 1997年本試験の結果

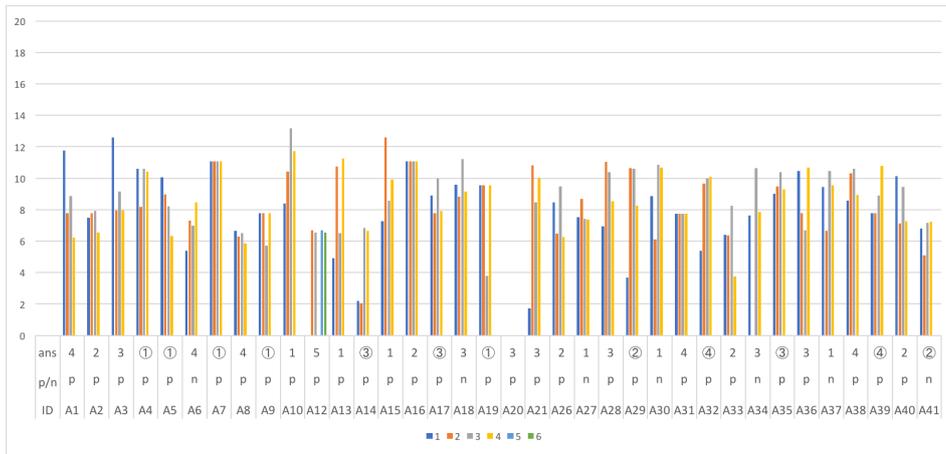


図4 2001年本試験の結果

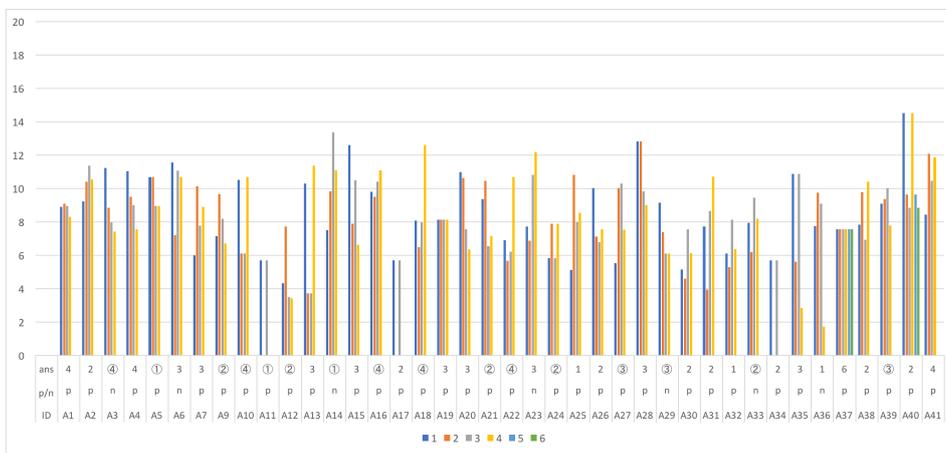


図5 2003年本試験の結果

メントを参照すると、各選択肢を評価する上で欠かせないと考えられる単語（選択肢ごと）に、メディチ、ボッカチオ、セルバンテス、ラブレー）に共起しているわけではなく、抽象的な単語に共起している文書のスコアが高くなっていることが確認できた。

6.3 不正解だった問題のクエリ得点

5. 節の図9 に不正解だった問題に対して、正解に対するクエ

リ得点とシステムが選んだ選択肢に対するクエリ得点の差を示したが、不正解だった問題の計 125 問のうち、74 問がクエリ得点の差は 2 点未満であったことが分かった。このグラフから、正解である選択肢に対するクエリ得点が低くなってしまっているわけではないと考えられる。4.2 節でも述べた通り、適合文書数が多くなると誤答に対するクエリ得点が大きくなる場合がある例を示したが、それと合わせて誤答に対するクエリ得点が

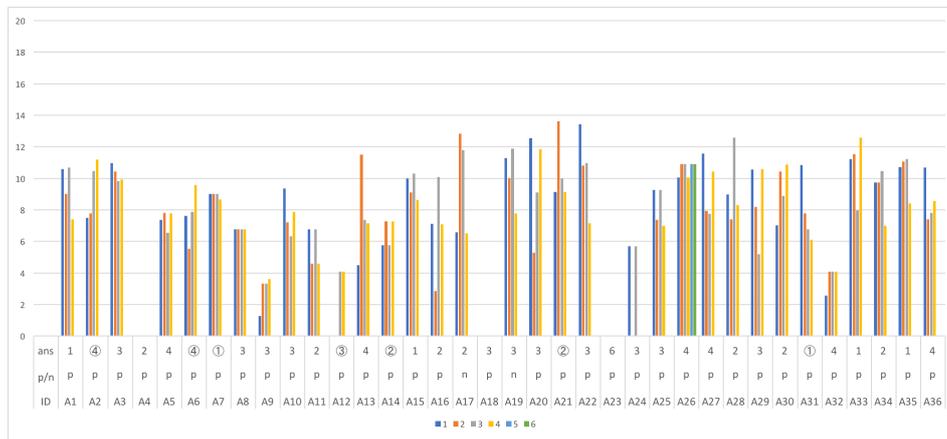


図 6 2005 年本試験の結果

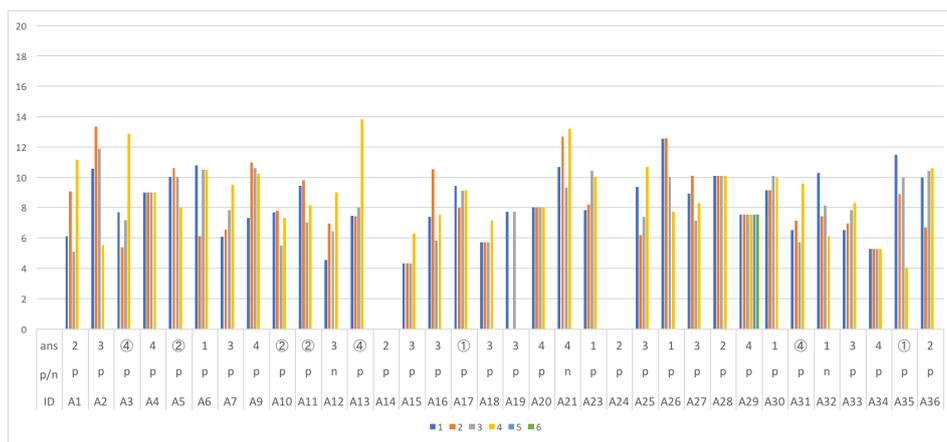


図 7 2007 年本試験の結果

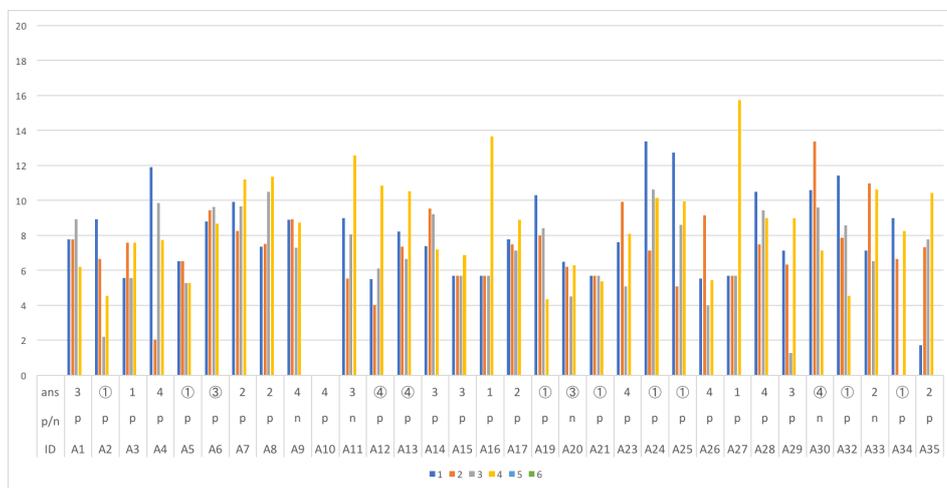


図 8 2009 年本試験の結果

大きくなっていることが確認できた。

今回、教科書データからインデックスを作成するにあたって、1つの問題に対して解くために必要な手がかりは一段落内に収まっているという考えのもとに一文を区切っていたので、適合文書数が多くなることは望ましくない結果であると言える。

6.4 結 論

本研究では、選択式問題への自動解答を実現するために、問題文から抽出されるキーワードと、選択肢から抽出されるキー

ワードをペアとしたクエリ生成によって教科書データに対して検索をする自動解答システムの作成を試みた。問題文と選択肢のそれぞれから抽出されるキーワード集合のペアによって生成される検証クエリによって検索し、その検索結果が良い選択肢が選択肢の正当性が高いと考えてシステムを構成した。しかし、5.節でも触れたように、そもそもキーワードとして抽出されるべき単語が抽出できなかったり、キーワードとしてふさわしくない単語をキーワードとして選択した場合、精度の低い検索結

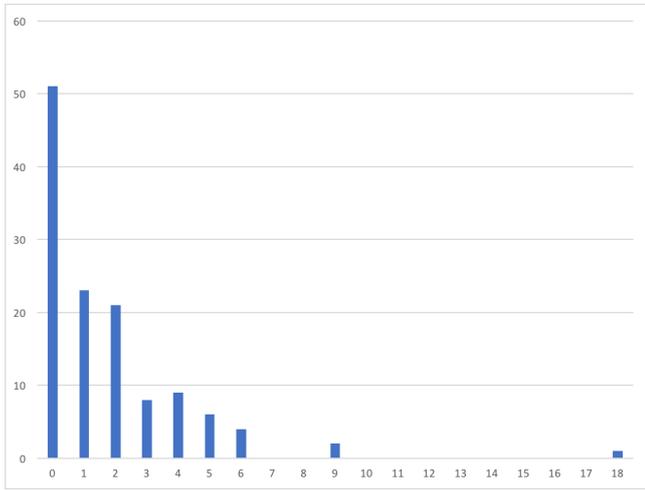


図9 正解である選択肢のクエリ得点と、システムが選択した選択肢のクエリ得点の差の分布

第1問 世界史上では様々な形で帝国が形成されてきた。そのような帝国や、それに組み込まれた地域について述べた次の文章A～Cを読み、下の問い(問1～11)に答えよ。(配点 27)

A ① 神聖ローマ帝国の権威は、理想的に西方キリスト教世界全体に及ぶべきものであったが、実質的な支配領域は現在のドイツ・オーストリアを中心とする地域に限られた。ヨーロッパの大半を支配する帝国を実現したのは、市民革命の中から台頭したナポレオンである。だが、そうした帝国の存立は、やがて各地のナショナリズムの高揚に伴い不可能になった。② ビスマルクの實現したドイツ帝国は国民国家を目指したものであったし、オーストリアは多民族国家の維持に苦心した。他方、③ ヨーロッパの外では列強の植民地帝国が形成される。これに遅れたドイツとイタリアで、④ 「第三帝国」のヨーロッパ支配を夢想したヒトラーと、地中海を支配する「帝国」建設を叫ぶムッソリーニが現れた。

問1 下線部①の帝国について述べた文として正しいものを、次の①～④のうちから一つ選べ。 1

- ① この帝国は、五賢帝時代に最盛期を迎えた。
- ② 14世紀にこの帝国は、フランスに対して百年戦争を起こした。
- ③ ナポレオンが、この帝国を復活させた。
- ④ この帝国の皇帝ハインリヒ4世は、ローマ教皇と対立した。

図10 2001年本試験第1問問1の問題文と選択肢

果となってしまった。

今後の課題として、大きく分けると4.1節と4.2で挙げた2つの課題が残されている。

まず、一単語として取得できることが望ましい単語が分割されて取得される問題については、世界史に登場する単語に特化した辞書を用意し、その辞書を併用して形態素解析を行う方法が考えられる。また、形態素解析ではない方法による単語の分割も考えられる。その一例を挙げると、文字種の切り替わる位置で単語を切るか切らないかを決定するという方法である。例えば、カタカナから漢字に切り替わる位置、漢字からカタカナに切り替わる位置を切らないという方法を取れば、神聖ローマ帝国という単語は取得可能であると考えられる。この文字種による切り方は、本来は一単語ではない単語がつながってしまう

第2問 地図は、それが作られた当時の人々の活動範囲や世界観をよく反映している。地図の作製やその時代背景に関して述べた次の文章A～Cを読み、下の問い(問1～10)に答えよ。(配点 26)

A 古代ギリシア・ローマでは、できる限り正確な世界地図を作ろうとする試みがなされた。下图は、2世紀ごろ活躍したアレクサンドリア出身の①に由来するとされる地図である。中世に入ると、①の地図は忘れられ、キリスト教の世界観を投影した「世界図」が作られるようになった。しかし、13世紀になるとイタリアでは、商業の発達に伴い、実用的な地図が作製されるようになる。その後②ルネサンス期に①の地図が再発見され、それが印刷術によって普及したこと、そして③大航海時代が到来したことなどによって、ヨーロッパにおいて正確な地図を作る本格的な試みが始まった。



問2 下線部②について述べた文として正しいものを、次の①～④のうちから一つ選べ。 13

- ① ヴェネツィアのメディチ家は、ルネサンス文化を保護したことで知られている。
- ② ボッカチオは、ルネサンスの代表的な画家の一人である。
- ③ 『ドン・キホーテ』を著したセルバンテスは、スペインにおけるルネサンスの代表的な文学者の一人である。
- ④ ラブレールは、『愚神礼讃』で教会や聖職者の腐敗を風刺した。

図11 2003年本試験第2問問2の問題文と選択肢

ということもあり得るが、辞書を作成する際の単語の候補抽出の方法としても利用可能である。

また、クエリ生成に関しては、シンプルな方法としては検索結果の精度を上げるために抽象的な単語の排除が考えられる。具体的に、どのような単語を抽象的な単語とするかについては考察を要するが、例えばキーワード抽出によって得られた単語集合内の単語を使用し1単語ずつの検索結果を利用する方法が考えられる。1単語での検索した結果、適合文書数が多いものはキーワードとしてふさわしくない可能性があると考えられるので、クエリ生成の前にその単語を排除することができる。考えられる。

5.章で挙げた課題に加え、インデックスを作成する段階にも課題が残されている。今回のインデックスの作成方法は、教科書データ内の文章を句点によって区切り、一文をインデックスとして扱った。しかし、例えば指示語を含む文章の場合に、その文のみでは主語が何なのかが分からなくなることが発生する。インデックスが元の文章と同じ情報を保持するために、インデックスを作成する際に内容を補完する方法についても模索したい。

文 献

- [1] Elasticsearch: RESTful, Distributed Search & Analytics - Elastic. <https://www.elastic.co/jp/products/elasticsearch>.
- [2] QA Lab - NTCIR プロジェクト. <http://ntcir.nii.ac.jp/jp/QA-Lab/>.
- [3] Eric Brill, Jimmy J Lin, Michele Banko, Susan T Dumais, Andrew Y Ng, et al. Data-Intensive Question Answering. In *TREC*, Vol. 56, p. 90, 2001.
- [4] David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A Kalyanpur, Adam Lally, J William Murdock, Eric Nyberg, John Prager, et al. Building Watson: An overview of the DeepQA project. *AI magazine*, Vol. 31, No. 3, pp. 59–79, 2010.
- [5] Tasuku Kimura, Ryosuke Nakata, and Hisashi Miyamori. KSU Team s Multiple Choice QA System at the NTCIR-12 QA Lab-2 Task. In *Proceeding of the 12th NTCIR Conference*, pp. 437–444, 2016.
- [6] Mio Kobayashi, Hiroshi Miyashita, Ai Ishii, and Chikara Hoshino. NUL System at QA Lab-2 Task. In *Proceeding of the 12th NTCIR Conference*, pp. 413–420, 2016.
- [7] Bernardo Magnini, Matteo Negri, Roberto Prevete, and Hristo Tanev. Is it the right answer?: exploiting web redundancy for Answer Validation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pp. 425–432. Association for Computational Linguistics, 2002.
- [8] Hideyuki Shibuki, Kotaro Sakamoto, Madoka Ishioroshi, Akira Fujita, Yoshinobu Kano, Teruko Mitamura, Tatsunori Mori, and Noriko Kando. Task Overview for NTCIR-12 QA Lab-2. In *Proceeding of the 12th NTCIR Conference*, pp. 392–408, 2016.
- [9] Hideyuki Shibuki, Kotaro Sakamoto, Yoshinobu Kano, Teruko Mitamura, Madoka Ishioroshi, Kelly Y Itakura, Di Wang, Tatsunori Mori, and Noriko Kando. Overview of the NTCIR-11 QA-Lab Task. In *Proceeding of the 11th NTCIR Conference*, pp. 518–529, 2014.
- [10] 工藤拓, 山本薫, 松本裕治. Conditional Random Fields を用いた日本語形態素解析. 情報処理学会研究報告自然言語処理 (NL) , Vol. 2004, No. 47, pp. 89–96, may 2004.