Vol.2014-NL-215 No.9
2014/2/6

質問応答システムを用いた多岐選択式問題の解答器の作成に関する研究

石下 円香[1,a)]　狩野 芳伸[2,1]　神門 典子[1]

概要：本稿では，多岐選択式の歴史科目の試験問題を解くために，既存の質問応答システムを用いる手法を提案する．質問応答システムとは，自然言語での質問文を入力とし，情報源から解答そのものを抽出して提示するシステムである．本稿では，多岐選択式の試験問題として，大学入試センター試験の世界史 B と日本史 B を対象とした．センター試験の問は既存の質問応答システムが想定する入力・出力形式とは異なっているため，問を解くための処理を追加する必要がある．本手法では，質問応答システムそのものには変更を加えず，問を解くのに適した質問文を生成し，質問応答システムに入力し，解候補を得たのちに，問の解答となる選択肢を選ぶ処理を行う手法を採用した．センター試験の問には，空欄に入る語を選ぶ問や，文の正誤の判定をする問など様々な形式があるが，本手法では，問の形式ごとに質問文の生成方法を変えることによって，形式の違いに対応した．センター試験の過去問題及び代々木ゼミナールが作成したセンター試験模試の結果，世界史で約 4 割，日本史で約 3 割 5 分の正解率が出せることが分かった．

1. はじめに

現在，国立情報学研究所は，大学入試問題を解く計算機プログラムを開発することを目的とした，人工知能プロジェクトを進めている [7]．このプロジェクトでは，2016 年までに大学入試センター試験 (以下，センター試験とする) において高得点をマークし，2021 年までに東大二次試験に合格することを目指している．

大学入試問題を解くような大規模な解答器は，複雑で多くの要素技術が必要となることが考えられ，一つの研究チームでの単独での開発は困難である．そこで，多くのチームが参加したり，それぞれが得意な部分を出し合って，相互に成果を活用できるようにすることを目指し，ツールやデータの共有・組み合わせ・実行を容易にする統合研究基盤の構築が行われている [9]．

試験問題を解くのに応用可能な仕組みとして，既存の質問応答システム (以下，QA システムとする) の仕組みが応用できると考えられており，ベースシステムとして MinerVA[4], [6] と Javelin[5] の 2 つの既存の QA システムのコンポーネント化を行った．システム開発者は，UIMA のフレームワークにより，これらのコンポーネントを自由に組み合わせたり，特定のコンポーネントを新規のものに入れ替えたり，改良したり，新たなコンポーネントを追加したりして用いることができる．

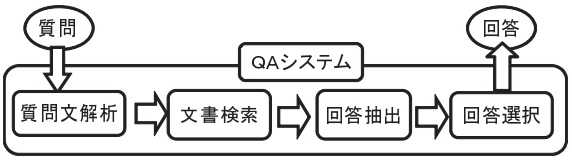


図 1　QA システムのコンポーネント

コンポーネントは，具体的には，図 1 に示すように，質問解析，文書検索，回答抽出，回答選択の 4 つである．質問解析では，入力された質問文の解析を行い，キーワードの抽出や質問が何を聞いているかを表す質問タイプの判定を行う．文書検索では，抽出されたキーワードを用いて質問に関連する文書を検索する．回答抽出では検索された文書から解候補となる文字列を抽出する．回答選択では，抽出された解候補の周りのキーワードや，質問タイプと一致するかなどの情報から，回答らしさのスコアが付けられ，回答らしさの高い解候補が回答として出力される．2 つの QA システム中のコンポーネントは，互換が可能である．本研究では，このコンポーネント化された既存の QA システムを利用して，大学入試問題に解答するベースシステムの開発を目指している．

本研究では，QA システムを利用した大学入試問題解答器 (以下，入試問題解答器，とする) を作成するにあたっ

[1] 国立情報学研究所
National Institute of Informatics
[2] 科学技術振興機構 さきがけ
PREST, Japan Science and Technology Agency (JST)
[a)] ishioroshi@nii.ac.jp

ⓒ 2014 Information Processing Society of Japan

て，まず，多岐選択式の問題に焦点を当てた．本稿では，多岐選択式の問題を自動で解く入試問題解答器について述べる．

多岐選択式の問題として，センター試験に焦点を当て，その中でも特に，受験者の知識を問う問題が多く，QA システムにもっとも合致していると考えられる歴史の科目に焦点を当て，解答器の作成を行った．センター試験の問は，既存の QA システムが想定する入力の形式とは異なっているため，問を解くのに適した質問文を生成する必要がある．本研究では，質問応答システムそのものには変更を加えず，問を解くのに適した質問文を生成し，質問応答システムに入力し，解候補を得たのちに，問の解答となる選択肢を選ぶ処理を行う手法を採用した．センター試験の問には，空欄に入る語を選ぶ問題や，文の真偽を判定する問題など様々な形式がある．問の形式ごとに質問文の生成方法を変えることによって，形式の違いに対応した．

2. 関連研究

狩野ら [10] は，センター試験の問を解く手法として，文書検索を基本とした手法を提案している．この手法では，単語を基本とする知識での解答を目指しているが，本研究では QA システムを用いることで，文章の論理構造なども考慮にいれて解答することを目指している点で異なる．

また，主に言明の真偽を問うタイプの問題に焦点を当てたアプローチとして，含意関係認識を使ったアプローチと，世界史オントロジーを用いるアプローチなどがある [11][8]．真偽を問うタイプの問題においては，これらの技術を本研究で用いる QA システムに導入することによって精度向上を狙うことが可能である．

また，金山ら [2] は，DeepQA[1] を用いて言明の真偽判定をする手法を提案している．DeepQA は，アメリカのクイズ番組 Jeopardy!のクイズに答えるシステムである．本研究でも，言明の真偽判定に同様の手法をとることを検討している．ただし，金山らの研究では，センター試験の選択肢を人手で DeepQA が受け付ける文に変更しているが，本研究では自動で変換を行っている．また，DeepQA では，Jeopardy!のクイズに対応するため，多数の質問タイプが用意されているが，本研究で用いる既存の QA システムでは人名，地名，組織名，数量といった代表的な質問タイプしか用意されていない点で異なる．

3. 使用した QA システムの概要

使用可能な QA システムとして，MinerVA と Javelin[5] があり，MinerVA には，facitoid 型質問 *1 に回答する MinerVA-N[4] と，non-factoid 型質問 *2 に回答する MinerVA-D[6] の 2 種類がある．それぞれのシステムのコンポーネントは入れ替えて使用することが可能であり，組み合わせを変えることで多数のシステムが使用できることになるが，本稿では単純に一つのシステムを用いて実験を行った．

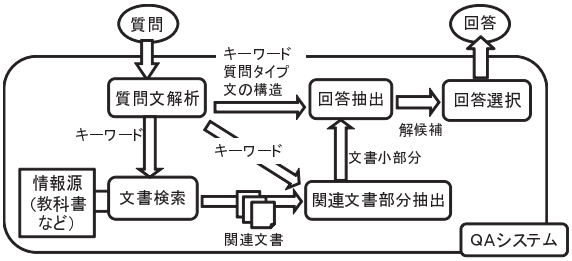


図 2 実験で用いた QA システムの構成図

実験では，MinerVA-N の文書検索部分のコンポーネントを入れ替えたものを用いた．MinerVA-N では，解答を抽出するための情報源として Web 文書が想定されており，文書検索エンジンとして Web 検索エンジンの API を用いている．実験では，試験問題に解答する観点から，歴史教科書と Wikipedia の文書を情報源として利用できるように，文書検索部分のコンポーネントを入れ替えて使用した．文書検索エンジンは，indri*3 を用いた．

実験で用いた QA システムの構成図を図 2 に示す．

まず，質問解析で入力された質問文の解析を行い，キーワードの抽出及び質問が何を聞いているかを表す質問タイプの判定を行う．文書検索では，抽出されたキーワードを用いて質問に関連する文書を検索する．次に，検索された文書全体から解候補を抽出するのは時間がかかるため，検索された文書をまず数文の塊 (パッセージ) に分割し，パッセージの形態素解析や構文解析の結果から解候補となる文字列を抽出する．抽出された解候補には，質問文との構造の一致度や解候補の周辺に現れるキーワードの数，質問タイプと一致するかなどの情報から，回答らしさのスコアが付けられる．回答らしさのスコアや，同じ解候補が違う文脈から何回出てきたかの情報を用いて解候補の最終的なスコアを求め，スコアの大きい解候補を出力する．

4. 入試問題解答器

本節では，まず，解答の対象とした大学入試問題である，センター試験問題の概要について述べる．その後，提案する入試問題解答器について述べる．

4.1 大学入試問題の概要

本節では，本研究で用いたセンター試験の問の概要を述べる．本研究では，歴史科目として，世界史 B 及び日本史 B の問題を用いた．

センター試験の世界史 B，日本史 B の問題は約 4 問の大問から成っており，各大問に約 9 問ずつの小問が含まれて

*1 人名数量などのなどの短い語句が回答となる質問
*2 定義や方法などの長い文章表現が回答となる質問
*3 http://www.lemurproject.org/indri.php

ⓒ 2014 Information Processing Society of Japan

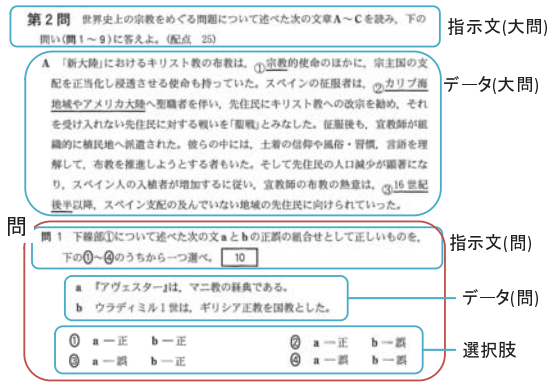


図 3 センター試験問題の例

いる．大問ごとに，「以下の文章を読んで問に答えよ.」といった指示文があり，その後文章などのデータが示されている．各小問には，指示文と選択肢が含まれており，一部の小問にはそれ以外の文章や画像のデータも含まれている．例を図 3 に示す．本稿では，選択肢が含まれている小問のことを，単純に問と呼ぶ．本研究では，XML 化されたセンター試験データ [11] を利用することを前提としている．

センター試験の主な問の形式の例を図 4 に示す．例では世界史 B の科目の例の示しているが，日本史 B においても問の形式は同様のものが多かった．図 4 にあるように，センター試験の問の指示文は QA システムが回答できる質問文の形式とはなっていないため，指示文をそのまま QA システムに入力しても，適切な解答は得られない．問を解くために適した質問文に変換し，QA システムに入力し，その出力を集約することによって問に対する解答が可能となると考えられる．

(1) 空欄に入る語を選択肢から選ぶ問
(2) 1 以外で，選択肢が単語である問
(3) 複数の出来事を年代順に並べ，正しい配列を選択肢から選ぶ問
(4) 選択肢の内容の正誤を判定し，正しい (間違った) 内容の選択肢を選ぶ問
(5) 選択肢以外に文が与えられており，文の内容の正誤を判定し，その正誤の正しい組み合わせを選ぶ問，または，正しい文の組み合わせを選ぶ問

4.2 入試問題解答器の構成

図 5 に入試問題解答器の構成図を示す．

入力として，入試問題 xml を受け取ると，試験問題解析部で問を解くのに必要な情報が抽出されるとともに，問の形式を判定する．質問文生成部では，問の形式に応じて QA システムへの入力となる質問文が生成される．解答選択部では，QA システムからの出力である解候補を使って，最終的な選択肢を決定する．

以下，4.3 節で試験問題解析部について詳しく述べる．質問文の生成方法と解答選択の方法は，問の形式ごとに違ってくる．そのため，4.4 節では，問の形式ごとに質問文生成と解答選択方法を述べる．

問 7 下線部⑦に関連して，次の文章中の空欄 ア に入れる語として正しいものを，下の①～④のうちから一つ選べ。 7
18 世紀には，アラビア半島で，ムハンマドの教えに帰ることを主張する ア の運動が始まった。 ア の運動は巡礼者を経由して，各地でイスラーム改革運動が広がるきっかけとなった。
① 十二イマーム派 ② ネストリウス派 ③ ワッハーブ派 ④ 長老派

(a) 空欄に入る語を選ぶ問

問 1 下線部①に関連して，6 世紀に台頭し，国家を築いた騎馬遊牧民として正しいものを，次の①～④のうちから一つ選べ。 10
① スキタイ ② 突 厥 ③ 月 氏 ④ 匈 奴

(b) (a) 以外で，選択肢が単語の問

問 6 下線部⑥に関連して，アメリカ合衆国の対外政策について述べた次の文 a ～ c が，年代の古いものから順に正しく配列されているものを，下の①～⑥のうちから一つ選べ。 15
a アメリカ＝メキシコ戦争が起こった。
b 門戸開放宣言(通牒)が出された。
c モンロー宣言が出された。

(c) 年代順に並び替える問 (選択肢は省略)

問 7 下線部⑦の歴史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。 7
① モールスは，電信機を発明した。
② アークライトは，無線電信を発明した。
③ 19 世紀後半に，アメリカ合衆国でラジオ放送が開始された。
④ 20 世紀前半に，インターネットが普及した。

(d) 選択肢の正誤を判定する問

問 1 下線部①について述べた次の文 a と b の正誤の組合せとして正しいものを，下の①～④のうちから一つ選べ。 10
a 『アヴェスター』は，マニ教の経典である。
b ウラディミル 1 世は，ギリシア正教を国教とした。
① a－正 b－正 ② a－正 b－誤
③ a－誤 b－正 ④ a－誤 b－誤

(e) 文の正誤を判定し，正誤の組み合わせを選ぶ問

図 4 センター試験世界史 B の問の例

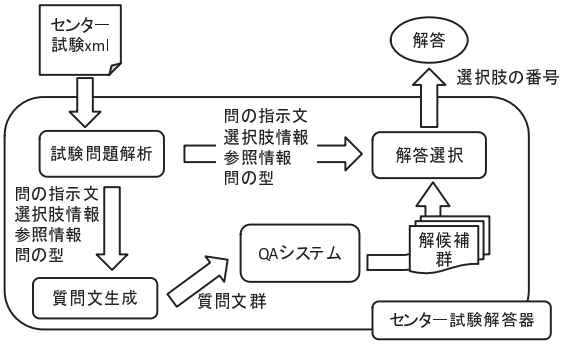


図 5 入試問題解答器の構成図

4.3 試験問題解析

試験問題解析では，問ごとに，問を解くのに必要な情報を入力された試験問題 xml から抽出するとともに，問の形

式の判定を行う．

問を解くのに必要な情報として，以下のものを抽出する．

- 問の番号
- 問の指示文
- 選択肢の内容
- 選択肢のタイプ (sectence や，term_person など)
- 指示文と選択肢以外の文章データがある場合，そのデータ *4
- 指示文や選択肢の中に参照 *5 がある場合，その参照先の記述 (空欄の参照の場合には，空欄が含まれている文章．)

また，問の形式としては，4.1 節で述べた 5 つの形式に分類した．用意した 5 つの形式に当てはまらない形式の問については，質問文が生成できないため，今回は対象外とした．今回用意した以外の問の形式にどのようなものがあったかについては，5.2 節で述べる．

4.4 質問文生成及び解答選択

本節では，質問文生成及び解答選択の処理について述べる．

試験問題解析で得られた問の内容と，問の形式を基に，QA システムの入力となる質問文を生成する．生成される質問文は，1 つの問に対して 1 つとは限らない．問を解くために必要な数だけ生成する．

すべての質問文に対する解候補を得られたら，解候補の情報を用いて解答選択をする．解答選択において，QA システムが出力した解候補と，選択肢などが一致しているかどうかの処理をする．この時，完全一致のほかに，部分一致や，wikipedia リダイレクトの情報を用いた別表記の情報も利用している．

以下に，問の形式ごとの処理を詳しく述べる．

4.4.1 空欄に入る語を選ぶ問

空欄に入る語を選ぶ問では，空欄に入る語を問うような質問文を作って QA システムに入力すれば，解候補の中に選択肢と一致するものが表れると期待した．

質問文作成では，空欄が入った文を取得し，空欄の部分を疑問詞に置き換え、文の最後に「か」をつけて質問文への変換を行った．置き換える疑問詞は，試験問題解析で抽出した選択肢のタイプの情報や，問の指示文中の情報 (「空欄に入る国名を答えよ」なら「どこ」にする，など) や，選択肢の接尾語の情報 (選択肢の語がすべて「○○派」なら，「何派」にする，など) を用いて決定した．複数の空欄に入る語の組み合わせを選ぶ問の場合には，空欄ごとに質問文を生成した．1 文中に複数の空欄が含まれている場合には，対象としている空欄のみ疑問詞に置き換えて，それ以外の空欄は，QA システムにおいてキーワードにならないような当たり障りのない語 (実験では，「あるもの」) に置き換えた．図 4(a) の例では，質問文の一つとして，「18 世紀には，アラビア半島で，ムハンマドの教えに変えることを主張する何派の運動が始まったか」が生成される．

解答選択では，解候補と一致した選択肢のうち，最も上位の解候補と一致した選択肢を 1 点，次は 1/2 点…と選択肢に点数をつけた．この処理をすべての質問文に対する解候補群で行い，合計点が最も高い選択肢を解答として選択した．

4.4.2 選択肢が単語である問

空欄に入る語を選ぶ問以外で，選択肢が単語である問では，問の指示文に質問文に相当するような表現が見られた．そこで，問の指示文から選択肢が解候補になるような質問文を作って QA システムに入力し、解候補と一致した選択肢を選択することとした．

質問文作成では，まず問の指示文を取得し，「下線部 1 について…」などの参照表現を参照先の記述で置き換えた．次に，質問として余分と考えられる部分を消し，文の最後に「は疑問詞ですか」を加えて質問文にした．質問として余分な部分としては，指示文の冒頭の「○○に関連して,」*6 や「として正しいものを…」などがある．疑問詞は，選択肢のタイプの情報などから決定する．図 4(b) の例では，質問文として，「6 世紀に台頭し，国家を築いた騎馬遊牧民は何ですか」が生成される．

解答選択では，解候補と一致する選択肢のうち，最も上位の解候補と一致した選択肢を解答として選んだ．

4.4.3 年代順に並び替える問

年代順に並び替える問では，それぞれの出来事がいつ起こったかが分かれば正しい並び替えを行えると考えられる．そこで，それぞれの出来事が起こったのはいつかを問う質問文を作って QA システムに入力し、解候補を頼りに年代順に並び替えて、同じ並びの選択肢を選択する手法をとった．

質問文生成では，それぞれの出来事を示す文ごとに，「のはいつですか」や「はいつ起こりましたか」などを追加して質問文を生成した．図 4(c) の例では，「アメリカ＝メキシコ戦争が起こったのはいつですか」などの 3 つの質問文を生成する．

解答選択では，まず，各質問文に対する最上位の解候補を，その出来事が起こった年代として取り出した．次に，年代順に出来事を並び替え，並びが一致する選択肢を解答として選択した．

4.4.4 選択肢の正誤を判定する問

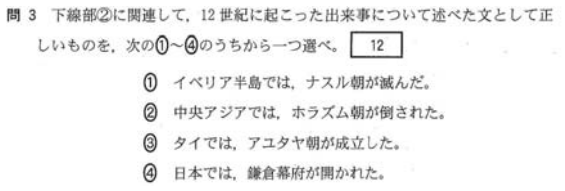
選択肢や文などの正誤を判定する場合には，金山らの手

*4 画像や表データも問を解くのに必要なデータだと考えられるが，現段階では，画像や表を含む問は処理の対象外にしているため，文章データのみを抽出する．

*5 「下線部 1」や，「文 a」など．

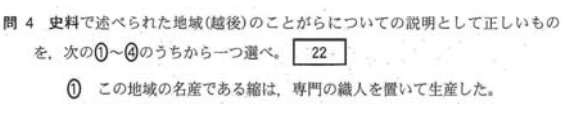
*6 2009 年の世界史 B の問題を何問か見た結果，余分な部分と判断した．

ⓒ 2014 Information Processing Society of Japan


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 6 選択肢の情報だけでは選択肢が選べない例


問 4 史料で述べられた地域(越後)のことがらについての説明として正しいものを，次の①～④のうちから一つ選べ。 22

① この地域の名産である縮は，専門の織人を置いて生産した。

図 7 選択肢の文中に参照表現が含まれている例

法 [2] と同様の手法を採用した．選択肢の文の中から名詞などを取り出し，疑問詞と置き換えて質問文を作り，QA システムに入力する．解候補に，疑問詞と置き換えた語 (以下，想定解と呼ぶ) と一致するものがあれば、その文の「正文らしさのスコア」を上げることによって，最も正文らしい (正文らしくない) 選択肢を選ぶ．

選択肢の正誤を判定する場合，選択肢の情報だけでは正確に正誤を判定できない場合がある．例えば，6 の例のように，選択肢の内容はすべて正しいが，指示文を読むと，「12 世紀に起こった出来事について，正しいものを…」とあり，選択肢の中で 12 世紀に起こった出来事は 1 つだけ，という場合が有る．このような場合に対応する手段の一つとして，本手法では，各選択肢に「12 世紀に」という記述を加えてから選択肢の真偽を判定する手法を採った．各選択肢の内容がいつ起こったのかを問う質問文を作るという手法も考えられる．しかし，時間表現を限定する記述があるときは，必ず選択肢の内容が正しいかというと，そうではわけではないということが観察の結果からわかっている．そのため，限定表現を見つけて選択肢に記述を追加する手法を採用した．

また，図 7 のように，選択肢の文中に「この地域」といった参照表現が含まれる場合もある．選択肢の文中に参照表現がある場合には，参照先になりそうな表現を下線部の情報や指示文内からできるだけ抽出するようにした．

質問文作成では，まず，上記のような理由から，指示文や下線部の情報から限定表現や参照表現の参照先になりそうなところを抽出し，各選択肢に追加した．限定表現が時間表現の場合は必ず各選択肢に追加し，それ以外の場合は，選択肢の文に主語になりそうな語がない場合にのみ主語として (「限定表現は」という記述を文頭に) 追加した．選択肢の文中に参照表現がある場合には，「この○○」と参照先の表現を入れ替えた．選択肢の文の中から，疑問詞に置き換える語を抽出し，それぞれの疑問詞に置き換える語ごとに，質問文を生成した．

疑問詞に置き換える語の判断では，まず，文中の名詞や名詞句を疑問詞に置き換える語の候補とした．固有表現注釈つきの東京書籍の教科書と，固有表現注釈つきのセンター試験問題から，固有表現タグがついた部分を抽出して，疑問詞に置き換える語の辞書とし，辞書に含まれている語を疑問詞に置き換えることとした．置き換える疑問詞は，固有表現タグのタイプを参考にして決定した．また，人名については wikipedia から人名のみを集めた人名辞書を作り，同様に利用した．もし，選択肢の文中に辞書に載っている語が一つもない場合でも，文中の名詞句を一つ選び (実験では，最後に現れる名詞句を選んだ)，疑問詞に置き換えることで，質問文を生成した．図 4(d) の例では，指示文からの限定表現は抽出されず，選択肢の 1 番に対しては，「誰は，電信意を発明したか」と「モールスは，何を発明したか」の 2 つの質問文が生成される．

選択肢の正誤を判定する問の中には，選択肢の文中の波線部の部分が正しいかどうかを判定するような問も含まれる．この場合には，波線部の部分をそのまま「疑問詞に置き換える語」とし，各選択肢について一つずつ質問文を生成した．

解答選択では，以下の方法で，それぞれの選択肢の「正文らしさのスコア」を求め，「正文らしさのスコア」を用いて選択肢を選ぶ．それぞれの質問文について，解候補に想定解が表れているかをチェックし，想定解が表れている場合，想定解と一致する解候補の順位の逆数を「正文らしさのスコア」とした．一つの選択肢に対して複数の質問文が作成されていた場合，それぞれの質問文に対する「正文らしさのスコア」の平均値を，その選択肢の「正文らしさのスコア」とした．最も「正文らしさのスコア」が高い (低い) 選択肢を，解答として選択した．

4.4.5 文の正誤の組み合わせを選ぶ問，正しい文の組み合わせを選ぶ問

問のデータとなっている文の正誤の組み合わせを選ぶ問や，正しい文の組み合わせを選ぶ問でも，文の正誤の判定には，選択肢の正誤の判定手法が使えるため，同様の手法を採用した．ただし，選択肢の正誤を判定する問では，「正文らしさのスコア」が最も高い (低い) 選択肢を選べば良いため，実際には選択肢の正誤を判定する必要がなかった．正しい文の組み合わせを選ぶ問でも，同様に「正文らしさのスコア」が高い方から 2 文を選べばよい．しかし，文の正誤の組み合わせを選ぶ問では，各文の正誤をきちんと判断しないと選択肢を選ぶことができない．そのため，「正文らしさのスコア」に閾値を設け，閾値以上の「正文らしさのスコア」を持つ文を正文，それ以外を誤文と判断することとした．

質問文生成では，選択肢の文ではなく，問中にデータとして与えられている文から質問文を生成する以外は，選択肢の正誤を判定する問と同様である．

解答選択では，まず，各文の「正文らしさのスコア」を選択肢の正誤を判定する問の場合と同様に求めた．文の正

ⓒ 2014 Information Processing Society of Japan

誤の組み合わせを選ぶ問では，「正文らしさのスコア」が閾値以上の文は正文，それ以外の文は誤文と判断し，正誤の判断が一致する選択肢を解答として選んだ．正しい文の組み合わせを選ぶ問では，「正文らしさのスコア」が高いほうから 2 文を選んで，組み合わせが一致する選択肢を解答として選んだ．

5. 評価実験

5.1 実験方法

作成した試験問題解答器の有効性を調べるため，センター試験の問題を解く評価実験を行った．

センター試験として，過去問の 2001,2005,2009 年の世界史 B 及び日本史 B の本試験のセンター試験 xml データを用いた．これらの問題の一部は，解答器を作成する際に観察したり，正誤の組み合わせを選ぶ問での正文らしさのスコアの閾値を決める際に用いている．そのため，代ゼミ模試の 2012,2013 年の世界史 B 及び日本史 B のデータでも同様の実験を行った．

センター試験 xml データでは，問を解くのに必要な知識を入力するためのタグが存在する．問を解くのに必要な知識として，画像や表の理解が必要なものは評価から除外した．

試験問題解答器で用いている QA システムでは，解候補を抽出するための情報源を切り替えることができる．実験では，情報源として教科書を使った場合と，wikipedia を使った場合の両方の結果を出した．教科書は，世界史では，東京書籍の世界史 A, 世界史 B, 新選世界史 B の教科書及び，山川出版の諸説世界史を用いた．日本史では，山川出版の諸説日本史を用いた．東京書籍の日本史の教科書は整備が間に合わなかったため，用いなかった．wikipedia は，2013 年 2 月 27 日にダウンロードした日本語 wikipedia のデータを用いた．wikipedia のデータは，世界史と日本史で共通のデータを用いた．一つの教科書を一文書とすると，一文書の文書量が膨大になってしまうため，節ごとに区切り，それぞれの節を一文書とした．wikipedia では，一つのエントリを一文書としている．

正誤の組み合わせを選ぶ問では，正文か誤文かを決める閾値をあらかじめ決めておく必要がある．2009 年の世界史 B の問の一部で予備実験をした結果を基に，実験では閾値を 0.01 とした．

この解答器では，QA システムで用いる情報源を教科書，wiki と切り替えることによって違った出力を得ることができる．既存の QA システムでは，複数の情報源がある場合，単に情報源をひとまとめにするよりも，各々の情報源を用いて解候補を出した後に，解候補をまとめる方が，精度が向上することが確認されている [3]．その為，入試問題解答器においても，QA システムで用いる情報源を変えたものの結果を解答選択部でマージした場合，解答精度はどうなるかの検討も行った．

表 1 解答できた問の数 (世界史)

	センター試験過去問	代ゼミ模試
解答できた問数	103	62
全問数	113	72

表 2 解答できた問の数 (日本史)

	センター試験過去問	代ゼミ模試
解答できた問数	90	62
全問数	108	72

表 3 世界史 B の正答率 (正答数)

	センター試験過去問	代ゼミ模試
教科書	0.368(38)	0.387(24)
wikipedia	0.359(37)	0.387(24)
組合せ	0.378(39)	0.403(25)

4.1 節で述べた 5 つの問の分類のうち，選択肢そのものに何らかのスコアがつく「空欄に入る語を選択肢から選ぶ問」「選択肢が単語である問」「選択肢の内容の正誤を判定し，正しい (間違った) 内容の選択肢を選ぶ問」に関しては，複数の結果のスコアをそのまま足し合わせて，選択肢を選びなおすことで，単純な結果のマージを行った．それ以外の問の形式については，実験では未実装であり，wikipedia を情報源としたの結果をそのまま用いることとした．「文の正誤の正しい組み合わせを選ぶ問，または，正しい文の組み合わせを選ぶ問」に関しては，それぞれの文の「正文らしさのスコア」を足し合わせることによって結果のマージが可能だと思われるが，現時点では未実装である．

5.2 実験結果

まず，実際に解答できた問の数を表 1，表 2 に示す．QA システムの情報源として教科書を使った場合と，wikipedia を使った場合では解答できる問の違いはない．

おおむね 8 割 5 分〜9 割の問には解答を出力することができた．解答できなかった問には，最初に除外した図表が必要な問以外にも，問の形式が対応できていなかった問も存在した．対応できていなかった問の形式として，「国の名とその国が建国された地域」などの語と語の組み合わせを選ぶものや，「古墳名と古墳について述べた文」などの語と文の組み合わせを選ぶものがあった．特に，語と文の組み合わせを選ぶ問は，日本史の問題で多く見られた．

QA システムで用いる情報源として教科書を使った場合と wikipedia データを使った場合と，それらの結果を組み合わせた場合の正答率を求めた．センター試験の 3 年分の過去問を用いた場合の正答率と，代ゼミ模試 2 年分を用いた場合の正答率を，表 3，表 4 に示す．

表 3 より，世界史 B では，教科書を使った場合と wikipedia を使った場合，二つの結果を組み合わせた場合のいずれも正答率に大きな差は見られなかった．また，表 4 より，日

ⓒ 2014 Information Processing Society of Japan

表 4 日本史 B の正答率 (正答数)

	センター試験過去問	代ゼミ模試
教科書	0.311(28)	0.209(13)
wikipedia	0.366(33)	0.370(23)
組合せ	0.366(33)	0.354(22)

表 5 問の形式ごとの正答率 (世界史 B)

問形式	正答数	問数	正答率
空欄	11	22	0.500
単語	5	11	0.454
並び替え	2	8	0.250
選択肢の正誤	41	108	0.379
正誤の組み合わせ	8	16	0.500
全体	67	165	0.406

表 6 問の形式ごとの正答率 (日本史 B)

問形式	正答数	問数	正答率
空欄	12	28	0.428
単語	4	7	0.571
並び替え	6	16	0.375
選択肢の正誤	22	71	0.309
正誤の組み合わせ	11	39	0.282
全体	55	161	0.341

本史 B では，教科書を使った場合，正解率が少し悪くなることが分かった．日本史 B では用いた教科書が 1 冊のみで，2 冊使った世界史 B と比較すると，量が十分でなかった可能性がある．また，世界史 B と日本史 B の正解率を比べると，世界史 B のほうが正答率が高かった．

次に，問の形式ごとの手法の有効性を見るため，問の形式ごとに分けた結果を求めた．結果を表 5 及び表 6 に示す．問の数が少ないため，センター試験の過去問と代ゼミの模試の両方を合わせた結果を示す．正答数は，二つの情報源の結果を組み合わせた場合のものを用いた．

問の形式ごとの結果では，表 5 及び表 6 より，空欄に入る語を選ぶ問や，選択肢が単語の問において，比較的高い精度で解答できることが分かった．ただし，問の形式ごとに問の数にかなりのばらつきがあるため，単純な比較はできない．

5.3 考察

表 3〜表 6 より，世界史 B ではおおむね 4 割，日本史 B ではおおむね 3 割 5 分の正解率であることが分かった．これらは試験問題に対する解答精度としては十分とは言えない．以下では，問の形式ごとに考察をする．

空欄に入る語を選ぶ問や，選択肢が単語である問では，比較的正解率は高かったが，半分程度の問で間違えている．正解できなかった問を見ると，QA システムが出力した解候補の中に，選択肢と同じものが一つも現れていない場合が多かった．正解出来た問を見ると，選択肢が人名や国名といった，QA システムが比較的高精度に解答できるタイプのものが多かったのに対し，不正解だった問では，法律名など，QA システム内で，質問タイプとして設定されていない種類のものが多かった．QA システムの質問タイプを増やすなどして精度を向上させることで，正解率の向上が期待できる．

【誤文】① 19 世紀後半，エディソンがダイナマイトを発明した。
【正文】② フォードは，自動車の大量生産方式を生み出した。

図 8 文の正誤の判定が成功した例

年代順に並び替える問では，世界史 B の正答率が特に悪い．並び替えでは，年の情報しか用いていないが，世界史では紀元前に起こった出来事などは正確に年が分からず「○○頃」と書かれていたり，事柄が起こった年の情報ではなく起こった順番のみが書かれていることがある．現状の QA システムでは，「いつ」と質問されると年月日単位でしか解候補を提示できないため，対応が必要である．また，日本史では，「昭和 3 年」などの元号での表現が多数使われているが，考慮に入れておらず，そのために不正解となった問があった．このような表現にも対応が必要である．

選択肢の正誤を判定する問は一番多く，全体の正解率に最も大きな影響を与えたと考えられる．文の正誤の判定が成功した例を図 8 に示す．誤文の選択肢 1 では，エディソンを想定解とした「19 世紀後半，誰がダイナマイトを発明したか」という質問文に対して，解候補にエディソンが入らないなど狙い通りの結果となり，誤文と判定された．また，正文の選択肢 2 では，フォードを想定解とした「誰は，自動車の大量生産方式を生み出したか」という質問文に対して，1 位の解候補にフォードを出力するなどして，正しく正文と判定された．

正解できなかった問では，大きく分けて以下のようなパターンがあった．

(1) 質問文の生成はうまくいっているが，QA システムが出力した解候補と想定解の照合がうまくいかなかった
(2) 質問文の生成がうまくいっていなかった
(3) 今回用いた，用語を疑問詞に置き換える手法では，正誤の判定ができない文であった

(1) では，どの質問文でも解候補の中に想定解が全く現れず，正文らしさのスコアがどの選択肢でも 0 になってしまい選択肢が選べないことが起こっていた．想定解となるものが，QA システムで質問タイプが設定されていないものが多く，解候補の抽出がうまくいかなかったことが原因と考えられる．QA システム自体の精度を向上させることで，正解率の向上が期待できる．また，想定解と解候補の部分一致を許したために間違って判定されてしまったものもあった．部分一致を許したのは，「東ローマ帝国」と「東ローマ」の一致を許容するなどの目的であったが，「東ローマ帝国」と「帝国」が一致してしまうなどの問題もあるた

ⓒ 2014 Information Processing Society of Japan

め，許される部分一致のルールなどを考える必要がある．

(2) では，疑問詞に置き換える語の判定がうまくいっていない場合と，限定表現や参照表現の参照先の抽出がうまくいっていない場合とがあった．疑問詞に置き換える語の判定がうまくいっていない場合では，「国家」「全国」といったあまり正誤判定に重要そうでない語が疑問詞に置き換えられている場合があった．ただし，実際はどの部分に間違いがあるかわからないため，疑問詞に置き換える必要のない語の判定は困難である．疑問詞に置き換える必要があるにも関わらず，置き換えられていない場合もあった．こちらは疑問詞に置き換える語の抽出方法や，リストの補強などの対応を検討する必要がある．

限定表現や，参照表現の参照先については，指示文に「○○世紀に起こった出来事について…」と簡潔に書かれていたり，下線部の内容が単純な語だったりすると適切に抽出できる場合が多かった．逆に，下線部の内容が複雑だったため，限定表現の抽出に失敗している例もあった．例えば，下線部の内容が「水野忠邦もいくつかの政策を試みたが」とあり，限定表現としては「水野忠邦」のみが抽出できるのが理想であるが，そのような抽出は行っておらず，質問文の生成がうまくいかなかった例があった．また，選択肢の文中に「この条約」とあった場合には，指示文や下線部に条約名が書いてあることを想定していたが，実際には，試験問題中には条約の内容文のみが載っており，条約名は試験の解答者が導かなければならない場合もあった．

(3) については，用語を入れ替えただけでは誤文を正文に直せないような例である．例えば，「徳川氏の一族の大名は，関東地方のみに配置され…」という文の場合，関東地方『のみ』ではないので誤文であるが，用語を入れ替えただけでは正文には直せない．また，「ユスティニアヌスは，イタリア半島の領土を失った」という文では，実際には領土を獲得しており，述語の部分に間違いがある．さらに，全くでたらめの文であり，文の一部を直して正文にするのが不可能な誤文も存在した．このような文は用語の間違いを前提としている本手法では解けない．まったくでたらめの文の場合は，質問文を作っても解候補が取れずに誤文と判定される可能性はあるが，それ以外の者は判定が困難である．含意関係認識など，他の手法と組み合わせることで解答が可能だと考えられる．事前の調査で，このような誤文は世界史 B では約 25%以下であることが分かっているが，日本史 B では調査していない．日本史 B では，世界史 B に比べ，文の正誤の判定をする問の正解率が悪く，本手法が使えない誤文の比率が多い可能性もあり，さらに調査が必要である．

最後に，文の正誤の組み合わせを選ぶ問については，選択肢の正誤の判定と同様の問題がある．また，文の正誤の判定に単純な閾値を設けているが，実例を見ると，単純な閾値ではうまくいっていない場合が多かった．例えば，二つの文の正文らしさのスコアに大きな差がある場合，どちらも閾値を超えていても，スコアの小さい方は誤文である場合が多かった．このため，単純な閾値以外で正文，誤文の判定をすることで正解率の向上が期待できる．

6. 終わりに

本稿では，多岐選択式の歴史科目の試験問題の一つとして，センター試験の問を解くための，既存の QA システムを用いる手法を提案した．本手法では，QA システムそのものには変更を加えず，センター試験の問を解くのに適した質問文を生成し，QA システムに入力し，解候補を得たのちに，問の解答となる選択肢を選ぶ処理を行う手法を採用した．センター試験の問には，空欄に入る語を選ぶ問や，文の正誤の判定をする問など様々な形式があるが，本手法では，問の形式ごとに質問文の生成方法を変えることによって，形式の違いに対応した．センター試験の過去問題及び代々木ゼミナールが作成したセンター試験模試を用いた評価実験の結果，世界史で約 4 割，日本史で約 3 割 5 分の正解率が出せることが分かった．正解率の向上のためには，自動で質問文を作る部分の精度向上や，QA システム自体の精度向上が必要であることが分かった．

謝辞

本研究の実施にあたっては，センター試験過去問データと代々木ゼミナールのセンター模試を使用した．また，東京書籍及び山川出版の教科書データを使用した．独立行政法人大学入試センター様，株式会社ジェイシー教育研究所様，代々木ゼミナール様，東京書籍株式会社様，株式会社山川出版社様に感謝いたします．

参考文献

[1] Ferrucci, D. A.: Introduction to "This is Watson", *IBM Journal of Reserch and Cevelopment*, Vol. 56, No. 3.4, pp. 1:1–1:15 (2012).
[2] Kanayama, H., Miyao, Y. and Prager, J.: Answering Yes/No Questions via Question Inversion, *Proceedings of COLING 2012*, pp. 1377–1391 (2012).
[3] Mori, T., Kanai, A., Ishioroshi, M. and Sato, M.: Effect of combining different Web search engines on Web question-answering, *Proceedings of the 10th Conference of Pacific Association for Computational Linguistics (PACLING 2007)*, pp. 325–332 (2007).
[4] Mori, T., Ohta, T., Fujihata, K. and Kumon, R.: An A* Search in Sentential Matching for Question Answering, *IEICE Transactions on Information and Systems*, Vol. E86-D, No. 9, pp. 1658–1668 (2003). Special Issue on Text Processing for Information Access.
[5] Shima, H., Lao, N., Nyberg, E. and Mitamura, T.: Complex cross-lingual question answering as a sequential classification and multi-document summarization task, *Proceedings of the 7th NTCIR Workshop Meeting*, pp. 33–40 (2008).
[6] 石下円香，佐藤　充，森　辰則 : Web 文書を対象とした

ⓒ 2014 Information Processing Society of Japan

質問の型に寄らない質問応答手法，人工知能学会論文誌，Vol. 24, No. 4, pp. 339–350 (2009).
[7] 新井紀子，松崎拓也：ロボットは東大に入れるか？-国立情報学研究所「人工頭脳」プロジェクト，人工知能学会論文誌 (2012).
[8] 田　然，宮尾祐介：関係代数に基づく推論の含意関係認識への応用，人工知能学会全国大会 (第 27 回) 論文集，pp. 2A4–4 (2013).
[9] 狩野芳伸：統合研究基盤:質問応答システムの互換コンポーネント化による再利用性向上と開発自動化支援，人工知能学会論文誌 (2012).
[10] 狩野芳伸，典子神門：大学入試センター試験は教科書の肯定的表現密度のみで解けるか，情報知識学会誌，Vol. 23, No. 2, pp. 179–184 (2013).
[11] 宮尾祐介，川添　愛：「大学入試問題を解く」ことから見える言語，知識，世界理解に関する研究課題，人工知能学会論文誌 (2012).

ⓒ 2014 Information Processing Society of Japan