# ウェブを用いて４択クイズを解く

外 池 昌 嗣　　佐 藤 理 史

京都大学大学院 情報学研究科 知能情報学専攻
tonoike@pine.kuee.kyoto-u.ac.jp, sato@i.kyoto-u.ac.jp

## 1. はじめに

本研究では、ウェブの知識源としての有用性、および、ウェブを知識源とした計算機による問題解決の実現可能性を明らかにするために、ウェブを用いて４択クイズを解く方法について検討する。

一般に、文として与えられた質問に対する答をテキストに基づいて導くタスクは質問応答と呼ばれ、近年盛んに研究されている[1)~4)]。質問応答では、通常、解候補は与えられない。このため、例えば、「自由の女神はどこにありますか？」という問題に対して、「アメリカ」、「ニューヨーク」、「ニューヨーク港」、「リバティー島」などの答が考えられ、何を正解とするかを決定することは、それほど容易ではない。この問題を避けるために、我々は、４つの選択肢（解候補）の中から正解を選ぶという４択クイズを研究対象に選んだ。

本論文では、まず、2 章において、本研究で対象とする４択クイズについて説明する。3 章では、問題のタイプに考慮しない、汎用的な解法について検討する。4 章では、特定の問題のタイプに有効な解法について検討し、5 章では総合実験とその結果について述べる。

## 2. ４択クイズ

### 2.1 クイズ$ミリオネア

本研究では、４択クイズとして、トミー社から発売されているカードゲーム版の「クイズ$ミリオネア」の問題を用いる。ゲーム版の問題は、全部で 1960 問ある。それらの問題は、賞金額別に A(￥10,000) ~ O(￥10,000,000) の 15 クラスに分類されており、高額の問題ほど人間にとっては難しい問題となっている。ここでは、それぞれのクラスを 2 つに分割し、半分 (980 問) を分析用の問題セットとして、残り半分を評価用の問題セットとして使用する。以下に問題例を示す。

**A59** 日本最古の物語といわれる「竹取物語」に登場するお姫様は誰？
A. 親指姫, B. 白雪姫, C. かぐや姫, D. うりこ姫
[正解: かぐや姫]

なお、解候補は、ほとんどの場合は一単語、長い場合でも「A の B」程度の短い名詞句である。

### 2.2 ４択クイズに固有の問題

４択クイズには、解候補の存在が問題成立の前提となっている問題が存在する。典型的には、「次の」を含む問題である。このタイプの問題は、さらに、以下のような 3 種類に分類できる。

比較問題　「次の … うち、最も ~ なのはどれ？」
否定語を含む問題　「次の…うち、~ でないものはどれ?」
その他　「次の … うち、~ なのはどれ？」

分析用セットに含まれるこれらのタイプの問題数を表 1 に示す。

表 1　４択クイズに固有な問題の数

| 分類 | 個数 |
|---|---|
| 比較問題 | 19 |
| 否定語 | 24 |
| その他 | 53 |
| 合計 | 96 |

これらの問題は、先に示した A59 のような問題よりも、解くのが難しいと考えられる。たとえば、比較問題に分類される次の問題を解くことを考えよう。

**L25** 次の愛知県の都市のうち、人口が最も多い都市はどこ？
A. 豊明市, B. 豊川市, C. 豊田市, D. 豊橋市
[正解: 豊橋市]

この問題は、この 4 つの選択肢の中で人口が最も多い都市を選ぶことを求めており、愛知県で人口が最大の都市を聞いているわけではない。このような問題は、これまでの質問応答では扱われてこなかったタイプの問題である。

４択クイズは、解候補が与えられているので、通常の質問応答の（解候補が与えられない）問題より解くのが容易であるとみなされることが多い。しかし、上記のような４択問題に固有の問題が存在するため、異なる難しさが存在する。

## 3. キーワードに基づく汎用的解法

ここでは、まず、問題のタイプを考慮しない、汎用的な解法について検討する。この解法の概要を図 1 に示す。この解法は、キーワード抽出、ページ収集、照合、多数決の 4 つのステップから構成される。まず、キーワード抽出において、問題文からキーワードを抽出する。次に、このキーワードを用いて、ウェブページを収集する。その後、照合において、収集したそれぞれのページに対して、そのページに基づく解を決定する。最後に、多数決で、最終的な解を決定する。以下では、これらの処理に

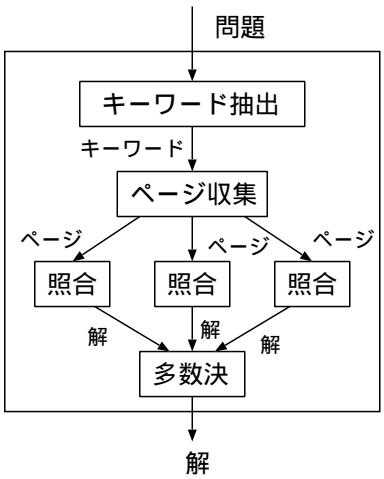


図 1　汎用的解法の概要

表 2　抽出すべきキーワードと重要度

| 重要度 | 基準 | 例 |
|---|---|---|
| 5 | 「」中の語句 | 「プラビ」「ビーチ」 |
| 4 | 人物名 (和・英) | 森光子、ボブ・サップ |
| 4 | 人物の関係を表す語 | 妻・夫・父親 |
| 3 | 固有名詞 (地名・組織名) を含む複合語 | 日本、天台宗、京都駅 |
| 2 | その他の名詞 | 都市、トンビ |
| 1 | サ変名詞 | 設計、考案 |
| 0 | ストップワード | 何、誰、現在、往年 |

ついて詳しく述べる。

### 3.1　キーワード抽出

キーワード抽出では、ページ収集および照合処理で使用する重要度付きキーワードを問題文から抽出する。キーワード抽出の流れを以下に示す。

(1)　問題文を形態素解析システム Juman[5] で形態素解析する
(2)　隣接する形態素が複合語を構成するならば結合する
(3)　キーワードにふさわしい語句を抽出する
(4)　キーワードに重要度を付与する
(5)　重要度が高いものを上位 $N$ 個出力する

ここで抽出するキーワードは、ページ収集 ( サーチエンジンによる検索 ) と照合の両方で用いるため、それらの処理に有用なものを抽出する必要がある。サーチエンジン検索では、正解を含むウェブページに含まれ、かつ、検索結果を絞り込むことのできるキーワードであることが求められる。一方、照合においては、正解を他の解候補と区別することができるキーワードが求められる。たとえば、人物名などの固有表現は重要であり、差別化に貢献しない一般的な語は不要である。

以上の点を考慮して、抽出すべきキーワードとその重要度を表 **2** に示すように定めた。抽出すべきキーワードはすべて名詞であり、その個数は、最大 4 個と定めた。

### 3.2　ページ収集

前節で抽出したキーワードをサーチエンジンに与えて、解を決定するのに用いるウェブページを収集する。収集方法として、次の 2 つの方法を実装した。

収集法 **A**　キーワードをクエリーとして $n$ ページ収集する。

収集法 **B**　キーワードに解候補を一つ加えたものをクエリーとして $n$ ページ収集する。これを、すべての解候補に対して行った後、それらをマージし、重複しているページを削除する。

### 3.3　照合と多数決

ここでは、収集したウェブページから、4 つの解候補のうちのどれを最終解とすべきかを決定する。ここでも、2 つの方法を実装した。

ページ単位の照合　ある解候補が、すべてのキーワードが存在するページに出現した場合に、1 点を与える。得点が最大の解候補を、最終解とする。この得点は、実際にウェブページをチェックして計算するのではなく、解候補とすべてのキーワードを AND 検索した時のヒット数をそのまま得点として用いる。すなわち、ウェブページの収集は実際には行なわない。

パッセージ単位の照合　ページ中にキーワードを含む文があれば、その前後 1 文を含めてパッセージとして抽出する。そのパッセージに解候補が存在した場合、その解候補にスコアを加算する。加算するスコアは、キーワードを含む文に解候補がある場合はそのキーワードの重要度を 2 倍した値、前後 1 文にある場合はそのキーワードの重要度とする。すべてのパッセージに対してこの処理を行い、スコア最大のものをそのウェブページに基づく解とする。最後に、すべてのウェブページに対して多数決をとり、最終的な解を決定する。

パッセージの例として、キーワード「竹取物語」を含む文とその前後 1 文を以下に示す。

> その他、かぐや姫 の里と言う事で、富士市の表玄関、新幹線の新富士駅なんかにもどか～んと壁画が描かれていたりします。
> 平安時代に書かれた 竹取物語 (いわゆる かぐや姫 の話 ) の発祥の地というふれこみなわけですが‥。確実な史実は、江戸時代の白隠さんというお坊さんが、この地に残る竹取り物語に似た伝承を記録しているという事らしいです。
>
> (http://www.netvision.co.jp/hist/kaguya.html より)

キーワード「竹取物語」の重要度を 5 とすると、中心行の解候補「かぐや姫」には 10 点が、第 1 行の解候補「かぐや姫」には 5 点が加算され、このパッセージにおける解候補「かぐや姫」のスコアは 15 点となる。

### 3.4　実験と検討

上記の提案手法を用いて、実際に 4 択問題を解く実験を行なった。実験では、分析用セット (980 問) を用い、次の 3 つの場合に対して、正解率を調べた。

(1)　ページ単位の照合 (AND 検索のヒット数)

表 3　ページ収集法ごとの正解率の比較

| ページ単位の照合 | 65.0% | | | |
|---|---|---|---|---|
| パッセージ単位照合 | n = 10 | n = 25 | n = 50 | n = 100 |
| + 収集法 A | 62.9% | 66.6% | 69.6% | 70.8% |
| + 収集法 B | 73.4% | 74.7% | 74.4% | |

表 4　4択クイズに固有な問題の正解率

| 分類 | 個数 | 正解数 | 正解率 |
|---|---|---|---|
| 比較問題 | 19 | 13 | 68% |
| 否定語 | 24 | 1 | 4% |
| その他 | 53 | 40 | 75% |
| 合計 | 96 | 54 | 56% |

(2)　収集法 A+ パッセージ単位の照合
(3)　収集法 B+ パッセージ単位の照合
その結果を表 **3** に示す。

この表より、まず、ページ単位の照合より、パッセージ単位の照合の方が優れていることがわかる。ウェブページのサイズは千差万別であり、同一ページにすべてのキーワードと解候補が存在したとしても、内容的には、それらの間に全く関係がない場合もある。より小さなパッセージ単位で照合する方がよい結果が得られることは、我々の直観にも合う。

一方、ページ収集では、収集法 A よりも収集法 B、すなわち、解候補も利用してページ収集を行なう方がよい結果が得られた。ページ収集に解候補を利用した場合、収集されるページには、すくなくとも一つの解候補が存在する。つまり、収集したページは、かならず照合で使われることになる。一方、キーワードだけでページを収集した場合は、得られたページに解候補がひとつも存在しない場合も多く、これらのページは、照合では全く使われない。つまり、解候補を利用した場合の方が、照合で実際に使用されるページを効率よく集めることができ、これが正解率の向上に寄与していると考えられる。

本章で提案したキーワードに基づく汎用的な方法は、比較的単純な方法にも関わらず、74.4%というかなり高い正解率を示した。これは、ウェブに各種の知識が記述されており、比較的単純な方法でも、これらの知識を利用できることを示している。しかしながら、4 問に 1 問は解けないことも事実である。

どのようなタイプの問題が解けていないかを調べるために、4 択クイズに固有な問題の正解率を調べた。パッセージ単位の照合と収集法 B を組み合わせて $n = 25$ を用いた場合の結果を表 **4** に示す。この表から、特に否定語「ない」を含む問題がはほとんど解けていないことがわかる。このような問題を解くためには、問題のタイプに応じた特別な解法が必要である。

## 4. 問題タイプを考慮した解法

本研究では、問題タイプを考慮した解法として、略称ルーチンと「ない」ルーチンの 2 つの方法を実装した。以下では、これらをタイプ別ルーチンと呼ぶ。

表 5　略称ルーチンの実行例

| 4 択 (原語) | クエリ | ヒット数 |
|---|---|---|
| ファシズム | ファシズム (ファックス) | 0 |
| ファッション | ファッション (ファックス) | 0 |
| ファクシミリ | ファクシミリ (ファックス) | 27 |
| ファクトリー | ファクトリー (ファックス) | 0 |

### 4.1　略称ルーチン

4 択クイズに含まれる略称問題 ( 略称に関する問題 ) には、次の 3 種類がある。
(1)　略称から原語を問う問題
(2)　原語から略称を問う問題
(3)　略称の一部に対応する原語を問う問題
これらの問題を解くためには、略称と原語の対応がわかればよい。このような対応は、テキスト中には次のような形式で現れることが多い。

- 原語 (略称) … 米国連邦準備制度理事会 (FRB)

以上に基づき、「原語 (略称)」のようなクエリーを生成してフレーズ検索を行い、そのヒット数が最も多いものを解とする方法が考えられる。たとえば、次に示す問題をこの方法で解く過程を表 **5** に示す。

**A58**　ビジネスに欠かせない「ファックス」を、略さずにいうと何？
A. ファシズム, B. ファッション, C. ファクシミリ, D. ファクトリー　[正解: ファクシミリ]

この場合、「ファックス」が略称で、4 択が原語である。ここでは、「ファクシミリ」に対応するヒット数が最も多いので、「ファクシミリ」を解とする。

この方法を実装して、実験を行った。ただし、問題文中の略称または原語の抽出には、簡単な構文ルールを用いた。対象とした問題は、分析用の問題に含まれる略称問題 (13 問) である。その結果は、汎用的解法で解いた場合は 10 問 (77%) 正解であったのに対し、略称ルーチンでは 12 問 (92%) が正解となった。

### 4.2　「ない」ルーチン

否定語「ない」が問題文中に含まれている問題は、否定語を考慮すべき問題と、否定語を考慮しなくてもよい問題に分けられる。否定語を考慮すべき問題として、例えば次のような問題がある。

**B41**　フランスの国旗に使われていない色はどれ？
A. 赤,　B. 緑,　C. 白,　D. 青　[正解: 緑]

汎用的な解法では、否定語「ない」をまったく考慮していないため、キーワードと最も関係が深い解候補が解として選ばれてしまう。実際、汎用的解法ではこのような問題は 24 問中 1 問しか解くことができなかった。

このタイプの問題は、仲間外れを探す問題と言える。仲間同士はウェブページ上で共起することが多いが、仲間外れはあまり共起しないと考えられる。すなわち、仲間外れ以外の 3 つの選択肢とキーワードをクエリとして検

表 6 「ない」ルーチンの実行例

| 4 択 | クエリ | ヒット数 |
|---|---|---|
| 赤 | フランス、国旗、緑、白、青 | 539 |
| 緑 | フランス、国旗、赤、白、青 | 962 |
| 白 | フランス、国旗、赤、緑、青 | 621 |
| 青 | フランス、国旗、赤、緑、白 | 746 |

表 7 総合実験の結果

| | 分析用セット | 評価用セット |
|---|---|---|
| 汎用的解法 | 74.7% | 71.7% |
| 組み合わせ | 75.6% | 71.0% |

索した場合のヒット数は多いが、仲間外れをクエリに含めた場合はヒット数が少ないことが期待される。事実、上記の B41 の問題に対してヒット数を求めると、表 **6** となる。すなわち、このようなクエリを生成して、ヒット数の最も多いものを解とする方法が有効と考えられる。

この方法を実装して実験を行った。問題としては、分析用セットに含まれる、否定語「ない」を考慮すべき問題 (24 問) を用いた。その結果、汎用的解法では 1 問 (4%) しか解けなかったのに対し、この方法では 10 問 (42%) 解くことができた。

## 5. 総合実験と考察

総合実験として、汎用的解法と 2 つのタイプ別ルーチンを組み合わせた場合の正解率を調べた。これを、分析用セットと評価用セットの両方の問題セットに対して行なった。その結果を表 **7** に示す。

この表からわかるように、分析用セットにおいては、タイプ別ルーチンを追加することによって、正解率を向上させることに成功したが、評価用セットにおいては、逆に正解率を下げる結果となった。この原因を調査したところ、タイプ別ルーチンにおける問題タイプの判定パターンの不備や、必要な情報を抽出するための規則の不備が、正解率を低下させる原因となっていることがわかった。これらの点を改善すれば、正解率は向上すると考えられる。

次に、分析用セットに対して、汎用的解法とタイプ別ルーチンを組み合わせた場合の正解率を、解のタイプ別に求めた。この結果を表 **8** に示す。この表から、解のタイプが固有表現の問題は、それ以外の問題より 10%程度正解率がよいことが分かる。すなわち、解が固有名詞の問題は比較的優しく、他の場合はそれに比べて解くのが難しいということである。

最後に、上記と同じ条件で、問題に設定されている賞金クラス (A ~ O) 別の正解率を求めた。この結果をグラフ化したものを図 **2** に示す。この図からわかるように、全体としては、人間にとって難しい問題は、計算機にとっても難しいという傾向が見られる。

表 8 解のタイプ別比較

| タイプ | 個数 | 正解数 | 正解率 |
|---|---|---|---|
| 組織名 | 89 | 74 | 83.1% |
| 名前 | 167 | 133 | 79.6% |
| 地名 | 182 | 142 | 78.0% |
| 固有物名 | 97 | 77 | 79.4% |
| 日付 | 22 | 18 | 81.8% |
| 数値 | 40 | 29 | 72.5% |
| 固有表現以外 | 383 | 268 | 70.0% |
| 合計 | 980 | 741 | 75.6% |

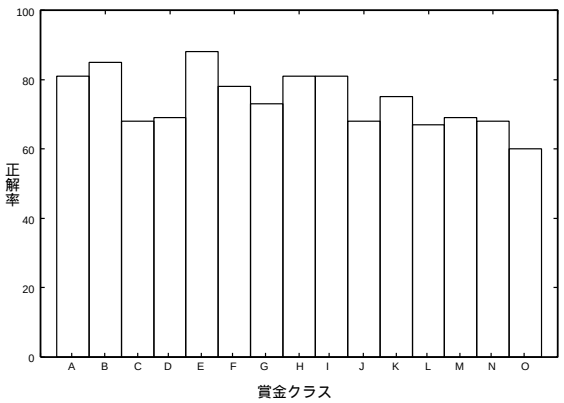


図 **2** 賞金クラス別の正解率

## 参 考 文 献


1) Kwok, C. C. T., Etzioni, O. and Weld, D. S.: Scaling Question Answering to the Web, *WWW10* (2001).
2) Clarke, C. L. A., Cormack, G. V. and Lynam, T.R.: Exploiting Redundancy in Question Answering, *SIGIR'01* (2001).
3) 村田真樹, 内山将夫, 井佐原均: 類似度に基づく推論を用いた質問応答システム, 情報処理学会研究報告 NL–135–12, pp. 181–188 (2000).
4) 佐々木裕, 磯崎秀樹, 平博順, 廣田啓一, 賀沢秀人, 平尾努, 中島浩之, 加藤恒昭: 質問応答システムの比較と評価, 信学技報 NLC–2000–10, pp. 17–24 (2000).
5) 黒橋禎夫, 長尾真: 日本語形態素解析システム JUMAN version 3.62 使用説明書 (1999).