

Web 検索エンジンを用いた Why 型質問応答システムに関する研究

田村元秀[†]・村上仁一[†]・徳久雅人[†]・池原悟[†]

膨大な情報の中からユーザの必要とする情報を見つけ出す技術として質問応答システムがある。本稿ではインターネット上の Web 検索エンジンを利用することにより、Why 型質問に回答する質問応答システムを試作し、出力された回答候補を人手で評価した。また明らかになった問題点に対して改善手法を示し、評価を行った。その結果、 F 値 0.200 から 0.352 に向上し、本システムにおける改善手法が有効であることを示した。

キーワード：質問応答システム，Why 型質問，non-factoid 型質問

A Study for the Why-type Question-Answering System using Web Search Engine

YUKIHIIDE TAMURA[†], JIN'ICHI MURAKAMI[†], MASATO TOKUHISA[†] and SATORU IKEHARA[†]

Question-Answering systems are used for finding needed information from an overload of information. A Question-Answering system that answers why-type questions by using existing Web search engines is proposed, and candidate answers are evaluated. In addition, we developed an advanced system by improving the search functions. Our results, 0.352 F-measure and 54% search success rate, showed that these improvement techniques were effective for this system.

Key Words: Question-Answering system , why-type question , non-factoid question

1 はじめに

Googleをはじめとする Web 検索エンジンの普及により、情報検索はより身近になっている。しかし大量の検索結果から必要な情報を取り出すことはユーザにとって容易ではない。そこで自然言語で記述された質問に対し、情報検索や情報抽出の技術を組合せ、最適な回答を得る質問応答技術が注目されている。

質問応答技術を評価するワークショップとしては米国の NIST(National Institute of Standards and Technology) などが主催する TREC¹において Question Answering Track(Dang, Lin, and Kelly 2007) があり、名称や日付・数値など事実に基づく回答を求める factoid 型質問や、複数の回答を求める list 型質問、定義や説明を求める definition 型質問などが扱われてきた。

また日本国内でも、国立情報学研究所主催の NTCIR²の QAC タスクがあり、TREC と同様に factoid 型質問(佐々木裕, 磯崎秀樹 2001) や、list 型質問、定義や説明を求める definition 型質

[†] 〒 680-8552 鳥取市湖山町 4-101, 鳥取大学工学部知能情報工学科, Faculty of Engineering, Tottori University, Minami 4-101, Koyama, Tottori, 680-8552, Japan

¹Text REtrieval Conference

<http://trec.nist.gov/>

²NII Test Collection for Information Retrieval and Text Processing
(<http://research.nii.ac.jp/ntcir/index-ja.html>)

問などが扱われてきた。しかし更に発展した課題に取り組むために、2006年4月から行われたQAC-4では、WhyやHowなどを問うnon-factoid型の質問も扱われることになった(J.Fukumoto and T.Mori 2007)。

non-factoid型質問応答技術に関する過去の研究として、諸岡らは、Why型質問と回答間に成立する意味的な関係とその手がかりとなる語に注目し、Why型質問の回答を抽出する手法を提案した(K.Morooka and J.Fukumoto 2006)。また東中らは、Why型質問に対して、原因表現のデータを得るためにEDRコーパスを使った学習アプローチを提案した(R.Higashinaka and H.Isozaki 2007)。これらの研究では共に新聞記事から回答候補の抽出を行っている。また英語におけるnon-factoid型質問に回答する研究として、Soricutらは、Web上のFAQにより学習した質問応答技術を提案した(Soricut and Brill 2004)。

ところで近年、情報爆発によりインターネット上に存在する文書はますます増加している。情報源としてWeb上の文書を利用する場合、新聞記事を利用する場合に比べ、関連する大量の文書が得られるが、各文書の信頼性は低いという問題がある。しかし、得られる大量の文書から抽出される単語の出現頻度を用いて、信頼性の高い文書を得ることができると考えている。また日本語において、Web上の文書を用いてWhy型質問に回答する議論はあまりされていない。

そこで、本研究ではWeb検索エンジンから得られる検索結果を利用し、Why型に回答するために検索結果に含まれる単語の出現頻度に着目した質問応答システムを試作する。また出力された回答候補を手で評価し、得られた結果より本システムの問題点を明らかにすると同時に、改善手法を提案し評価する。実験の結果、提案した改善手法はWhy型質問に対する有効性を示した。

2 Why型質問応答システム

2.1 システム概要

本研究では、検索エンジンにより得られる結果を対象に、Why型質問に回答する質問応答システムを試作する。例えば質問文「地震が起きるのはなぜ(どうして)?」に対して、Web検索結果をもとに「地震が起きるのは地球の表面を覆っているプレートが動くためです。」と回答するシステムである。

本研究におけるWhy型質問応答システムは、質問応答技術として一般的な4つのモジュールにより構成される。本システムの構成を図1に示す。

本システムは、(1)質問文解析、(2)情報検索、(3)回答候補抽出、(4)回答候補ランキングより構成されている。以下に各モジュールの詳細を示す。

2.2 質問文解析

入力された質問文を解析し、情報検索用キーワード(以下検索KW)と回答候補抽出用パターン(以下回答候補抽出PT)を作成するために、質問文の言い換えを行う。次に、言い換え後の質問文から検索KWと回答候補抽出PTを作成する。

質問文の言い換え

疑問語を用いた質問文はその位置(文頭、文中、文末)により大きく3つに分類できる。以

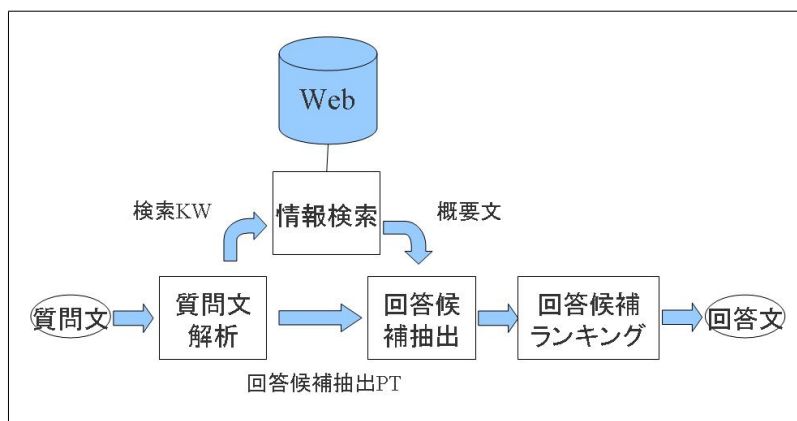


図 1 Why 型質問応答システム

下に例を示す．

- (1) 疑問語が 文頭 にある
(例 1) どうして地球は自転しているのですか。
- (2) 疑問語が 文中 にある
(例 2) 空はどうして青いのですか。
- (3) 疑問語が 文末 にある
(例 3) 雷が鳴るのはどうしてですか。

本研究における質問文の言い換えとは、疑問語の位置を“文末”にすることである．以下に質問文の言い換え手順を示す．

言い換えの手順

- step1. 疑問語 (なぜ、どうして) を削除
- step2. 文末表現 (の、のですか。) を削除
- step3. 助詞「は」を助詞「が」に置換
- step4. 文末に「のはなぜ」を追加

上記の(例 1),(例 2),(例 3)について言い換えを行った結果を図 2 に示す．

- (例 1) 地球が自転しているのはなぜ
 (例 2) 空が青いのはなぜ
 (例 3) 雷が鳴るのはなぜ

図 2 言い換え結果

検索 KW の作成

質問文「空が青いのはなぜ」に回答する最も単純な文として“空が青いのは～”が考えられる．本研究では、このような文を検索するために、Web 検索エンジンを利用する．Web 検索エンジンに与えるキーワードを作成するために、言い換えを行った質問文から疑問語部分を削除

する．図 3 に質問文「空が青いのはなぜ」における検索 KW の作成例を示す．

<質問文> 空が青いのはなぜ
疑問語部分を削除
<検索 KW > 空が青いのは

図 3 検索 KW を作成する例

回答候補抽出 PT の作成

検索結果から回答候補を抽出するために，検索 KW をもとに回答候補抽出 PT を作成する．
図 4 に質問文「空が青いのはなぜ」における回答候補抽出 PT³の作成例を示す．

<検索 KW > 空が青いのは
<回答候補抽出 PT > 空が青いのは*

図 4 回答候補抽出 PT を作成する例

2.3 情報検索

インターネット上にある膨大な文書から必要な情報を得るために情報検索を行う．まず，質問文解析によって得られた検索 KW をクエリとし，Web 検索エンジンに与える．本研究では Web 検索エンジンとして Google(グーグル株式会社)を使用する．図 5 は検索 KW「空が青いのは」における Google の検索結果の一部である．

³ただし回答候補抽出 PT に使用する “*” は任意の文字列である

空は何故青いの？

空が青いのは、主に空気の分子による光の散乱です。空気の分子のように、光の波長よりも長さが短い粒子による光の散乱を「レイリー散乱」と言います。一方、夕焼けは、この「レイリー散乱」に加えて、「ミー散乱」という物も関係しています。 ...

www.tensaiji.net/answer06.htm - 12k - キャッシュ - 関連ページ

NIDEK 目のおはなし 空が青く見えるのは？

そうなる空が青いのは太陽が原因なののでしょうか。結論から言えばその通りで太陽の光も関係しています。それではまず太陽の光のお話です。太陽と地球との距離は約 1 億 5000 万キロメートルあり、太陽の光が地球に届くには約 8 分 19 秒かかるそうです。 ...

www.nidek.co.jp/bluesky.html - 13k - キャッシュ - 関連ページ

パーティサファイア サファイアの魅力

また、古代ペルシャの伝説では、サファイアは大地を支える石と呼ばれ、空が青いのは、サファイアの輝きが空に映し出されているからだと言われてきました。実際、サファイアは空気を敏感に感じる石で、曇りの日と晴れの日では、輝き方が違い、「最も神に ...

home.iprimus.com.au/austsapp/miryoku.html - 11k - キャッシュ - 関連ページ

図 5 検索結果の一部

図 5 に示すように、Google の検索結果はページのタイトル、概要文 (snippet), URL 等で構成されている。本研究では概要文に対して「。」などを手がかりに文単位で抽出する。

2.4 回答候補抽出

得られた概要文と、質問文解析で作成した回答候補抽出 PT とのマッチングにより回答候補を抽出する。図 6 は、回答候補抽出 PT 「空が青いのは*」で抽出される回答候補の例である。

< 適合文 1 > 空が青いのは、主に空気の分子による光の散乱です。

< 回答候補 1 > 空が青いのは、主に空気の分子による光の散乱です。

< 適合文 2 > 空が青いのは太陽が原因なののでしょうか。

< 回答候補 2 > 空が青いのは太陽が原因なののでしょうか。

< 適合文 3 > また、古代ペルシャの伝説では、サファイアは大地を支える石と呼ばれ、空が青いのは、サファイアの輝きが空に映し出されているからだと言われてきました。

< 回答候補 3 > 空が青いのは、サファイアの輝きが空に映し出されているからだと言われてきました。

図 6 回答候補を作成する例

2.5 回答候補ランキング

回答候補抽出において、通常複数の回答候補が得られるため、各回答候補をスコアリングする必要がある。情報源として Web 上の文書を利用する場合、新聞記事を利用する場合と比べ信頼性が低いと考えているが、本研究では「検索結果の複数の概要文に多く含まれる単語は、質問の正解文にも含まれる」と仮定をし、信頼性の高い文書を得る。そのためスコアリングには、概要文の単語の出現頻度を用いた単語頻度テーブルを利用する。

単語頻度テーブル

単語頻度テーブルとは、Web 検索結果の概要文を収集した際に名詞（ただし非自立語を除く）を抽出し、その出現頻度を示したテーブルである。表 1 に検索 KW「空が青いのは」における単語頻度テーブルの高頻度名詞上位 5 件、表 2 に低頻度名詞上位 5 件を示す。

表 1 高頻度名詞上位 5 件

名詞	スコア
空	724.0
光	300.0
色	121.0
海	117.0
波長	110.0

表 2 低頻度名詞上位 5 件

名詞	スコア
世	1.0
細工	1.0
霜	1.0
範囲	1.0
視界	1.0

スコアリング

作成した単語頻度テーブル T に基づき、各回答候補 a に対するスコア S_a は以下の式で求める。 M_a は回答候補 a における名詞の数であり、 ω_i は回答候補 a における i 番目の名詞である。

$$S_a = \sum_{i=1}^{M_a} T(\omega_i) \quad (1)$$

作成した単語頻度テーブルに基づき、各回答候補をスコアリングする。ただし、“？”や“なぜ”は回答候補に悪影響を与えるため、それらを含む回答候補のスコアに対して、基準値⁴(score500)を減算する。最後にスコアを降順でソートすることでランキングを行う。図 7 は回答候補ランキングの例である。

3 実験

3.1 実験条件

実験に使用する質問文は、「問題と回答がペアになっていること」、また「一般に科学の世界には“なぜ”や“どうして”を問う質問が多いこと」から学研サイエンスキッズ(学習研究社)の「科学なぜなぜ 110 番」の自然(地球・気象)ジャンルを参考に 50 問作成する。また正解文は、同サイトから本論文の著者が、正解を含む部分を引用し作成する。これらを用いて検索実験を

⁴本研究における基準値は、最大検索件数 500 件としたことに基づいている(3.1 節参照)

<回答候補 1 > 空が青いのは、主に空気の分子による光の散乱です。(score 1104)
<回答候補 2 > 空が青いのは太陽が原因なのでしょうか。(score 831)
<回答候補 3 >、空が青いのは、サファイアの輝きが空に映し出されているからだと言われてきました。(score 772)

図 7 回答候補ランキング例

行う。本研究では Web 検索エンジンとして Google を使用し、検索結果最大 500 件から概要文を抽出する。また形態素解析器として MeCab(工藤拓)を用いる。

3.2 評価

本研究では、システムの性能を評価するために 4 種類の評価指標を用いる。以下に詳細を示す。

人手による評価

本システムの精度を確認するために、人手による評価を行う。評価は QAC-4(J.Fukumoto and T.Mori 2007) で用いられた評価基準に基づき、ランキング 1 位の回答候補に対して“A”，“B”，“C”，“D”，“なし”で評価する。以下に評価基準，図 8 に評価例を示す。

評価基準

評価 A 出力された回答は正解とほぼ等しい内容である

評価 B 出力された回答は正解に加え、他の情報を含む

評価 C 出力された回答は正解の一部を含む

評価 D 出力された回答は正解の内容を含まない

評価 なし 出力なし

評価 “A”

<質問文>虹はどうしてできるのですか。

<正解文>プリズムにた形の雨のつぶに光が当たると、何色にも分かれて見えるようになるのです。

<回答候補>虹ができるのは、雨のあとなど空気中にうかぶ水蒸気の粒に太陽の光があたると、水滴がプリズムの働きをするため光が反射・屈折して光の帯となって見える自然現象なのです。

評価 “A”

<質問文>海はどうしてしょっぱいのですか。

<正解文>海の水がしょっぱいのは、中に塩がとけているからです。

<回答候補>海の水がしょっぱいのは、水や二酸化炭素、塩のもとになるナトリウムや塩素が含まれているから。

評価 “B”

<質問文>虹はどうして7色なのですか。

<正解文>虹が7色なのは、虹のもとになる太陽の光が、およそ7色の光からできているからなのです。

<回答候補>虹が7色なのは太陽の光が水蒸気などでそれぞれの波長に分解され見えるのですが、人間には見えない波長も存在します。

評価 “C”

<質問文>雷はどうして光るのですか。

<正解文>雷の電気が流れたところの空気の温度はかなり高くなります。すると、熱くなった空気は光ります。

<回答候補>雷が光るのは、もちろんこれによって発生した光によるものです。

評価 “D”

<質問文>雪はどうして冷たいのですか。

<正解文>雪が溶けるとき融解熱を奪っていくために、雪は冷たいものだと感じるのです。

<回答候補>雪が冷たいのは、とても当たり前なことなのに、今までこんな風に俳句に作った人はいなかったみたいです。

評価 “なし”

<質問文>山に登るとどうして空気が薄くなるのですか。

<正解文>高い所では、地球の引力が弱いためにたくさんの空気を引きつけておくことができないこととなります。

<回答候補>出力なし

図 8 評価例

F 値による評価

システムの出力評価を行うために情報検索でよく利用される F 値を求める．ただし本研究では，評価 “A” または “B” を正答とし， $Question$ は質問数， $Correct$ は正答を出力した数，また $Output$ は本システムが回答を出力した質問数である．以下に $Recall$ ， $Precision$ ， F 値の式を示す．

$$Recall = \frac{Correct}{Question} \quad (2)$$

$$Precision = \frac{Correct}{Output} \quad (3)$$

$$F\text{-measure} = \frac{2 * Recall * Precision}{Recall + Precision} \quad (4)$$

MRR による評価

回答候補ランキングの評価のため，ある質問に対して最上位正答順位の逆数を平均した値である MRR (*Mean Reciprocal Rank*) (Dang et al. 2007) を用いる．例えば，ある質問の最上位正答が回答候補ランキング 2 位に存在した場合， MRR は 0.5 であり，正答が上位で絞られるほど MRR は高くなる．以下に MRR の式を示す．

$$MRR = \frac{1}{N} \sum_{k=1}^N \frac{1}{rank_k} \quad (5)$$

ただし N は問題数， $rank_k$ は問題 k における正答の最高順位である．

累積検索成功率による評価

システムの実質的な出力評価のために，ランキング上位 10 位の回答候補に正答を含む割合として累積検索成功率を用いる．

3.3 実験結果

表 3，表 4 に Why 型質問文 50 問における各評価の結果を示す．ただし，表 3 において分母は質問数，分子は評価の数である．

表 3 人手による評価の結果

A	B	C	D	なし
10%	8%	4%	56%	22%
(5/50)	(4/50)	(2/50)	(28/50)	(11/50)

表 4 F 値， MRR ，累積検索成功率の結果

F 値	MRR	累積検索成功率
0.20	0.285	40%

この表から分かるように，回答候補ランキング 1 位の回答候補の約 8 割は不正解または出力なしであった．また累積検索成功率も 4 割にとどまっていた．

3.4 システム出力の分析

評価の結果、ランキング 1 位が正答にならない理由として大きく 2 つあることが分かった。

(1) 回答候補抽出の失敗

評価“なし”が 2 割以上であることから、「Web 検索結果が存在しない」、「Web 検索結果は存在するが、回答候補が存在しない」ことが考えられる。例えば「雨が降っても海がいっぱいにならないのはなぜ」という質問から「雨が降っても海がいっぱいにならないのは」で検索した結果、回答候補は出力されなかった。これは検索 KW が長いから、Web 検索結果が存在しなかったと考えられる。

(2) 回答候補ランキングの失敗

回答候補の中に正答は存在するが、誤答がより上位に存在してしまう場合がある。図 9 は誤答（回答候補 1）が正答（回答候補 2）のスコアを上回った例である。

<質問文> 雲はどうして動くのですか。
<正解文> 雲が動くのは、風が吹くからというのが答えです。
<回答候補 1> 雲が動くのは風ではなく、自力で動いてるんだと思っていた。(score705)
<回答候補 2> 雲が動くのは風があるからだ。(score704)

図 9 回答候補ランキングの失敗例

4 改善手法

システム出力の分析により 2 つの問題点があることが確認された。本研究では“回答候補抽出の失敗”に対して“検索 KW のスリム化”を、また“回答候補ランキングの失敗”に対して“単語頻度テーブルの最適化”を行うことによりシステムを改善する。

4.1 検索 KW のスリム化

回答候補抽出の失敗の原因として、「Web 検索結果が存在しない」、「Web 検索結果は存在するが、回答候補が存在しない」が挙げられる。

これらの問題に対して質問文を解析し、検索 KW を“従属節の名詞+主節”とすることで、Web 検索結果が多く出力されるようにして、より多くの回答候補を抽出する。主節の判定には助詞の“は”または“が”を利用する。日本語において主節は文の後半に置かれることが多いことから、検索 KW から“のは”を除いた文に対して、最も後ろにある“は”、“が”を元に判定を行う。また主節部分を回答候補抽出 PT とする。

以上を本研究では、検索 KW のスリム化と呼ぶ。図 10 に「雨が降っても海がいっぱいにならないのはなぜ」における検索 KW のスリム化の例を示す。

4.2 単語頻度テーブルの最適化

「回答候補ランキングの失敗」の原因として、単語頻度テーブルの不備が考えられる。例えば、単語頻度テーブルに余計な語が含まれていたり、原因・理由を表す語（“から”など）を評価していないことが挙げられる。以下に単語頻度テーブル最適化の手法を示す。

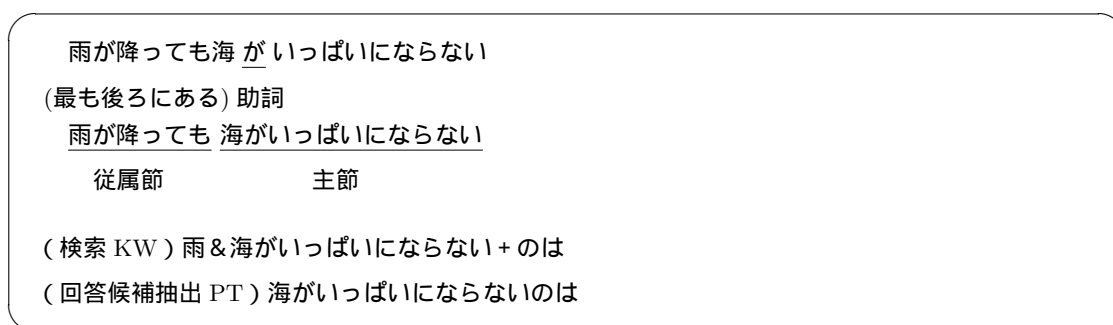


図 10 検索 KW のスリム化の例

低頻度削除

Web 検索エンジンの結果には必ず一定の不必要な情報(ノイズ)が存在する。例えば、検索 KW 「雪が白いのは」で検索した結果の「雪が白いのは神がそうしたから。」は明らかに質問の意図に反する場合である。ノイズの影響を減らすために、検索結果中に存在する出現頻度 1 以下の単語を削除する。

スコア加算

本システムの評価過程において、正解には“から”をはじめとする原因・理由を表す語が含まれることが分かった。これらの語を含む文は正解である可能性が他よりも高くなるため、回答候補のスコアに対して、基準値 (score500) を加算する。図 11 はスコア加算の対象文例である。

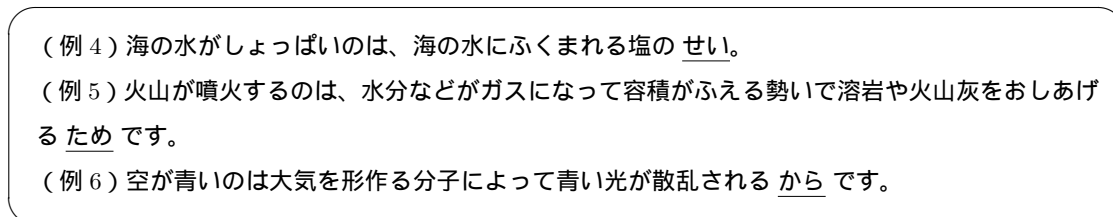


図 11 スコア加算の対象文例

スコア減算

本システムの評価過程で、“当たり前”や“当然”など回答候補ランキングに悪影響を及ぼす単語があることが分かった。回答候補ランキングへの悪影響を減らすため“当たり前”，“当然”，“常識”を禁止語とみなし、回答候補のスコアに対して、基準値 (score500) を減算する。図 12 はスコア減算の対象文例である。

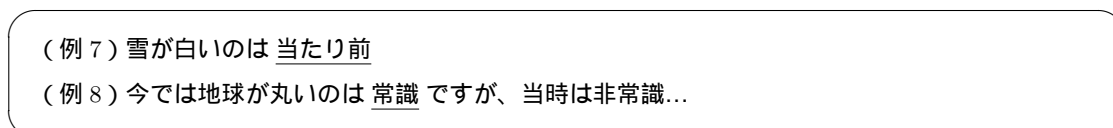


図 12 スコア減算の対象文例

名詞限定

単語頻度テーブルは回答候補をランキングするのに大きくかかわっている．本研究では登録する単語として名詞を用いたが，質問の回答として代名詞“あれ”はあまり重要ではない．表5に，本研究に使用したIPA品詞体系の名詞の一部を示す．

表5 IPA品詞体系における名詞の一部

type1	type2	例
名詞	一般	「大根」「加速度」
	数	「ゼロ」「億」
	接続副詞的	「対」「兼」
	固有名詞	「北海道」「ニューヨーク」
	接尾	「化」「州」
	代名詞	「そこ」「俺」

そこで，名詞の中でもあまり重要でない単語による回答候補ランキングの影響を防ぐために，単語頻度テーブルの登録名詞は“一般名詞”，“固有名詞”のみにする．

4.3 結果

表6，表7に改善手法別の評価結果を示す．ただし，スリム化や低頻度削除など各手法の値は改善前のシステムに対して個別に得ている．また表中の(all)は改善手法すべてを適用した後のシステムの結果である．

表6 人手による評価の結果

手法\評価	A	B	C	D	なし
改善前	10% (5/50)	8% (4/50)	4% (2/50)	56% (28/50)	22% (11/50)
スリム化	12% (6/50)	8% (4/50)	4% (2/50)	58% (29/50)	18% (9/50)
低頻度削除	12% (6/50)	8% (4/50)	2% (1/50)	56% (28/50)	22% (11/50)
スコア加算	18% (9/50)	4% (2/50)	4% (2/50)	52% (26/50)	22% (11/50)
スコア減算	10% (5/50)	8% (4/50)	6% (3/50)	54% (27/50)	22% (11/50)
名詞限定	16% (8/50)	8% (4/50)	4% (2/50)	50% (25/50)	22% (11/50)
(all)	20% (10/50)	12% (6/50)	2% (1/50)	48% (24/50)	18% (9/50)

表7 F値，MRR，累積検索成功率の結果

手法	F値	MRR	累積検索成功率
改善前	0.202	0.285	40%
スリム化	0.220	0.308	44%
低頻度削除	0.225	0.283	42%
スコア加算	0.247	0.333	50%
スコア減算	0.202	0.299	42%
名詞限定	0.270	0.325	44%
(all)	0.352	0.409	54%

表6より，検索KWのスリム化によって評価“なし”の数が改善前に比べ減っていることを

確認した。また表 7 より、単語頻度テーブルの最適化の各手法のうち最も MRR の増加が大きかったのは“スコア加算”であり、回答候補ランキングに効果が高いことが分かった。

改善手法すべてを適用した結果、 F 値、 MRR 、累積検索成功率すべての値が大きく増加していることから、システムの実質的な性能が向上したことが確認できる。

4.4 改善手法に関する考察

検索 KW のスリム化の影響

表 6 より、評価“なし”の割合が減少し、検索 KW のスリム化は有効であることが確認できた。図 13 に、評価“なし”から評価“A”になった例、また評価“A”から評価“D”になった例を示す。

(例 9): 評価が良くなった例

<質問文> 山に登るとどうして空気が薄くなるのですか。

<スリム化後 KW > 「山」, 「空気が薄くなるのは」

<正解文> 高いところでは、地球の引力が弱いためにたくさんの空気を引きつけておくことができないこととなります。したがって、山の上では空気の量がうすくなるのです。

<回答候補> 空気が薄くなるのは、そこより上にある空気の重さが少なくなるからです。

(例 10): 評価が悪くなった例

<質問文> 海にどうして波があるのですか。

<スリム化後 KW > 「海」, 「波があるのは」

<正解文> 海の上では、必ず風がふいています。そこでできた波が岸まで伝わってくるのです。

<回答候補> 波があるのは南の海で、そこに行くだけでも大変だからである。

図 13 スリム化の影響

例 9 のように、検索 KW のスリム化により検索結果が出現し、評価が良くなった例が 2 件あった、また例 10 のように、逆にスリム化により評価が悪くなった例が 1 件あった。以上より、すでに検索結果が存在する場合、必ずしも検索 KW のスリム化は有効ではないことが分かった。

単語頻度テーブルの最適化の影響

表 6, 7 より、単語頻度テーブルの最適化におけるいずれの手法も効果があることを確認した。以下に手法別の詳細を述べる。

(1) 低頻度削除

MRR にほとんど変化が無く、回答候補ランキングに与える影響は少なかった。今後、頻度 2 以下、頻度 3 以下など削除対象となる名詞の頻度を変えて調べる必要がある。

(2) スコア加算

回答候補ランキングに最も効果が高かった。これは、Why 型質問応答において「から」などの原因・理由の表現は、回答を抽出する上で大きな手がかりとなることを意味している。

(3) スコア減算

低頻度削除と同様に MRR にほとんど変化が無く、回答候補ランキングに与える影響は少なかっ

た．しかし，“当たり前”や“当然”など回答候補ランキングに悪影響を与える禁止語を増やすことで“スコア減算”の効果が上がると考えている．

(4) 名詞限定

スコア加算には及ばないものの，同程度の効果が期待できるという結果になった．この結果，単語頻度テーブルに加える品詞を限定することは有効であるが，今後，“一般名詞”や“固有名詞”以外の意味のある語についても検討していきたい．

単語頻度テーブルと改善手法の影響

本研究では，得られた複数の回答候補をランキングするために，「検索結果の概要文に多く含まれる単語は，質問の正解文にも含まれる」と仮定し，スコアリングに名詞頻度テーブルを利用した．図 14 は，正答を含む回答候補を得た質問をランダムで 5 問選択し，横軸に回答候補ランキング，縦軸に各質問における平均正答出現数を示したものである．

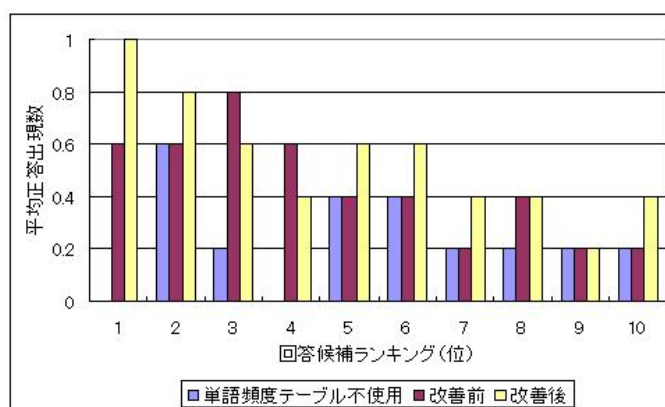


図 14 回答候補ランキング別平均正答出現数

図 14 の結果より，単語頻度テーブルの有効性を示し，本研究における仮説“検索結果の概要文には，質問の正答に含まれる単語が多く存在している”はほぼ正しいと考えている．また，特に回答候補ランキング 1 位において改善策が有効であることも確認できた．

4.5 回答候補に関する考察

回答候補のスコア分布について

図 15 は回答候補ランキング上位 5 件の回答候補において，横軸にスコア，縦軸にその出現頻度を示したものである．

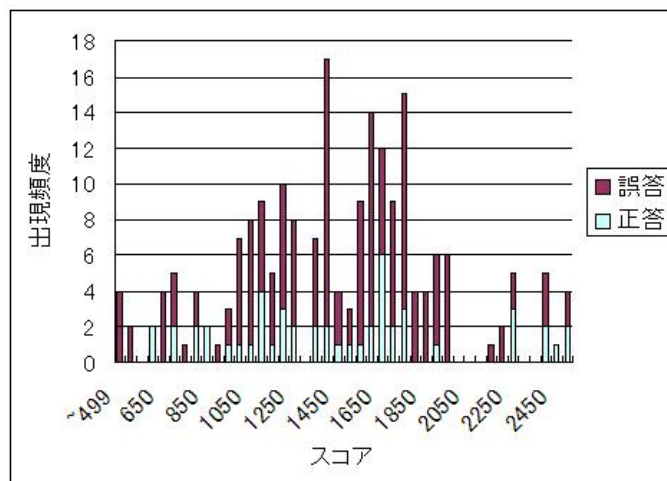


図 15 回答候補のスコア分布

図 15 より，高いスコアに誤答が存在し，また低いスコアに正答が存在することから，スコアと正答，誤答間に関連性が低いことが確認できた．なお，ランキング上位 5 件に占める割合は正答 24.6%(50/203)，誤答 75.4%(153/203)であった．

回答候補の抽出元サイトについて

評価“A”と“B”の 16 件の回答候補について，抽出元サイトを調べた．その結果，質問作成元サイトからの回答候補抽出は無かった．表 8 は抽出元サイトを種類別に示したものである．

表 8 抽出元サイトの割合

種類	割合	(例)
ブログ	31%(5/16)	個人のブログ等
掲示板	25%(4/16)	初心者の天文学質問箱，Yahoo 知恵袋
その他	44%(7/16)	こどもニュース，東北で起きた大きな地震の仕組み

表 8 から，本システムにおける回答候補の抽出元サイトは様々であることが確認できた．

評価における別解の扱いについて

本研究では正解文をもとに人手評価を行い，別解を誤答としたため厳しい評価となった．そこで，著者の基準で評価を行ったところ累積検索成功率 20%増加し 64%になった．つまり，本システムでは実質的には表 6，表 7 よりも良い結果が得られる．図 16 に別解の例を示す．

<質問文>雪はどのように降るのですか。

<正解文>雲の中で水のつぶは，はじめ小さな氷のつぶになります。そして、その氷のつぶのまわりに空気中にある水蒸気がくっついて、だんだん大きな氷の結晶となっていくのです。

<回答候補>雪が降るのは、シベリアってトコロから冷たい空気がやってきて北海道の高い山にぶつかるからなんだって。

図 16 別解の例

5 おわりに

本研究では、Web 文書利用時の信頼性の問題に対し、得られる大量の文書から抽出される単語の出現頻度を用いて、信頼性の高い文書を得ることができると考え、Why 型質問に回答するために検索結果に含まれる単語の出現頻度に着目した質問応答システムを試作した。そして得られた結果より、本システムの問題点を明らかにすると同時に、改善手法を提案し、評価した。その結果、本研究において単語頻度テーブルの有効性を示し、「検索結果の複数の概要文に多く含まれる単語は、質問の正解文にも含まれる」ことが確認できた。また本システムに対して各最適化は有効であることを示し、 F 値 0.352、累積累積検索成功率 54%を得た。

ところで本研究では、回答候補は引用した正解文をもとに評価を行ったため、正解文に記述されている内容以外は誤答とした。そこで、著者基準の評価では累積検索成功率で約 20%の増加が見られ、最終的に累積検索成功率 64%を得た。

今後は、特に評価“なし”の対策として、回答候補抽出パターンのバリエーションを増やしたり、正答または誤答の回答候補について、特有の表現を抽出することによる回答候補の絞り込みによって、精度の高い Why 型質問応答システムの構築を目指したい。

謝 辞

謝辞はこちらに記述して下さい。

参考文献

- Dang, H. T., Lin, J., and Kelly, D. (2007). “Overview of the TREC 2007 Question Answering Track.” In *Proceedings of the Sixteenth Text REtrieval Conference (TREC 2007)* pp. 99-116.
- J.Fukumoto, T.Kato, F. and T.Mori (2007). “An overview of NTCIR-6 QAC-4.” In *In Proc. of the 6th NTCIR Workshop Meeting*.
- K.Morooka and J.Fukumoto (2006). “Answer extraction method for why-type question answering system.” In *IEICE Technical Report, volume 105, pp.7-12*.
- R.Higashinaka and H.Isozaki (2007). “NTT’s Question Answering System for NTCIR-6 QAC-4.” In *In Proc. of the 6th NTCIR Workshop Meeting*.
- Soricut, R. and Brill, E. (2004). “Automatic Question Answering: Beyond the Factoid.” In *HLT-NAACL 2004: Main Proceedings pp. 57-64*.
- グーグル株式会社. “Google Search Engine.”, <http://www.google.co.jp>.
- 学習研究社. “学研サイエンスキッズ.”, <http://kids.gakken.co.jp/kagaku/index.html>.
- 工藤拓. “MeCab.”, <http://mecab.sourceforge.net/>.
- 佐々木裕, 磯崎秀樹 (2001). “SAIQA : 大量文書に基づく質問応答システム.” 自然言語処理研究会 / 情報学基礎研究会, FI64/NL145-12, pp.77-82.

略歴

著者 1 氏名：著者 1 の紹介をこちらに記述して下さい。

著者 2 氏名：著者 2 の紹介をこちらに記述して下さい。