

単語解答を求める複雑な質問文を対象とした知的 Web 検索方式

源明 和也[†] 渡部 広一[‡] 河岡 司[‡]

[†] [‡] 同志社大学大学院 工学研究科 〒610-0394 京都府京田辺市多々羅都谷 1-3

E-mail: [†] dth0707@mail4.doshisha.ac.jp, [‡] {hwatabe, tkawaoka}@mail.doshisha.ac.jp

あらまし 本稿では、質問文に対する単語解答を Web から自動的に獲得する知的 Web 検索方式を提案する。Web 文書から単語解答を選択するために、語が有する意味特徴を語と重みで表現する概念ベースと語と語の間にある意味的な関連性を数値として算出する関連度計算方式を用いる。また、ノード得点計算を利用し、柔軟に質問の解答を選択できる構成をとっている。

キーワード WWW, 質問応答, 概念ベース, 関連度

Intelligent Web Retrieval Method Targeting Complex Question Sentence to Obtain the Answer

Kazuya GENMEI[†] Hirokazu WATABE[‡] and Tsukasa KAWAOKA[‡]

[†] [‡] Doshisha University 1-3 Miyakodani, Tatara, Kyotanabe-shi, Kyoto, 610-0394 Japan

E-mail: [†] dth0707@mail4.doshisha.ac.jp, [‡] {hwatabe, tkawaoka}@mail.doshisha.ac.jp

Abstract This paper proposes the Intelligent Web Retrieval Method, the technique obtaining the word answer for the question sentence from Web. Moreover, the technique in which the Concept Base, the Degree of Association Algorithm, the Node Score Calculation Algorithm are used to obtain the word answer from Web document. The Concept Base generates semantics from a certain word, and the Degree of Association Algorithm uses the results of the semantics expansion to express the relationship between one word and another as a numeric value.

Keyword WWW, Question Answering, Concept Base, Degree of Association

1. はじめに

現在、Web の急速な普及と利用者の劇的な増加に伴い、膨大な電子化された文書がオンライン上に蓄積されている。既存の検索システムでは、膨大な量の情報を提示されるため、ユーザが要求する情報が埋もれてしまい、必要な情報だけを手に入れることは困難である。そこで、的確にユーザが必要としている情報のみを獲得するための技術が求められている。この期待に応えられる技術として、質問応答技術が考えられる。一般に質問応答技術は、段階的な解析が必要であり、解析には知識をデータベースとして登録することが不可欠である。さらに、質問に用いられる表現および Web 文書の表現は自由度が非常に高い。

本稿では、質問文に対する単語解答（固有名詞を対象とする）を Web から自動的に獲得する手法（知的 Web 検索方式）を提案する。この手法を実現する上で、質問文が求めている解答の種類を特定し、さらに、解答の適切さを判断する必要がある。特に、解答の適切さを判断する上で、固有名詞が解答の種類に適するかを判断することは困難である。本稿では、単語の表記のみに依存しない柔軟な連想機能を可能とする概念ベ

ース^[1]と関連度計算方式^[2]（連想メカニズムとする）を利用する。連想メカニズムをベースとして、質問文意味理解システム^[3]、未定義語の属性獲得手法^[4]と最適ノード決定手法・ノード得点計算^[5]（Web 情報を利用する技術）を用いることで、柔軟に解答を選択できる方式を実現する。

2. 知的 Web 検索方式の概要

知的 Web 検索方式の構成を図 1 に示す。知的 Web 検索方式では、ユーザが入力した質問文に対して、解答を Web から選択する。例えば、質問文「2008 年のオリンピックはどこで開催されたか？」を入力した場合、正解「北京」を出力することが目標となる。

知的 Web 検索方式は、質問文解析、解答候補の獲得、解答スコア付けから構成される。質問文解析には、質問文から質問対象語（質問文が求めている対象）を取得できる質問文意味理解システムおよび最適ノード決定手法を利用する。解答スコア付けには、未定義語の属性獲得手法および最適ノード決定手法で用いられているノード得点計算を利用する。手法全体として、複数の国語辞書や新聞などから機械的に構築した大規模な知識ベースである概念ベースと、概念と概念の関連

の強さを定量的に評価する関連度計算方式（連想メカニズム）を利用する。また、多様な表現を許している Web 文書を利用することで、質問の解答のための情報を柔軟に獲得できると考えられるため、既存の検索エンジン^[6]を利用する。なお、検索エンジンを一朝一夕で構築することが困難で現実的ではないため、既存の検索エンジンを利用している。

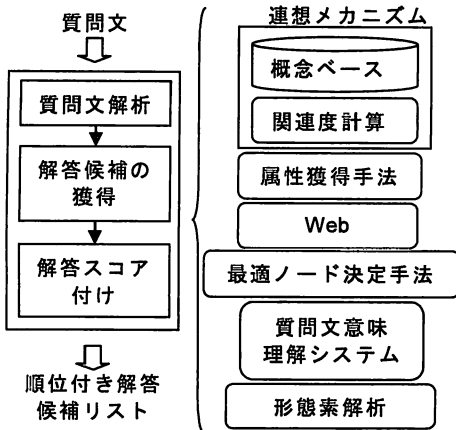


図 1 知的 Web 検索方式の概要

3. 連想メカニズムと質問文意味理解

連想メカニズムは概念ベースと関連度計算方式で構成されている。概念ベースは、ある単語から語意の展開を行い、関連度計算は、語意の展開結果を利用し、単語の間にある関連性の強さを数値として表す手法である。質問文意味理解システムは、質問文が求めている対象の語を質問対象語として獲得でき、本稿では、質問対象語を質問文が解答として求めている単語の種類（解答タイプ）を判断するために利用する。

3.1. 概念ベース

概念ベース^[1]は、電子化辞書や電子化新聞、シソーラス等から機械的に構築された大規模知識ベースである。見出し語である概念 A を、概念 A を特徴付ける語である属性 a_i と特徴付ける強さである重み w_i の対の集合として次式のように定義する。

$$A = \{(a_1, w_1), \dots, (a_i, w_i), \dots, (a_n, w_n)\} \quad (1)$$

図 2 に概念ベースの一部を示す。任意の属性は、必ず概念ベース内に含まれる語で定義している。従って、概念の属性を 1 次属性とした場合、1 次属性に対してそれぞれの 1 次属性を導くことができる。この集合を 2 次属性と呼ぶ。同様に、概念は任意の次元までの属性 (n 次属性) を導出することが可能である。以降の節では、単に「属性」と表記する場合、各概念の 1 次属性を表すものとする。なお、概念自身は必ず属性として定義している。

語	属性(1次属性)
雪	(雪, 0.61), (白い, 0.30), (下る, 0.27), ...
白い	(雪, 0.16), (白地, 0.14), (色, 0.14), ...

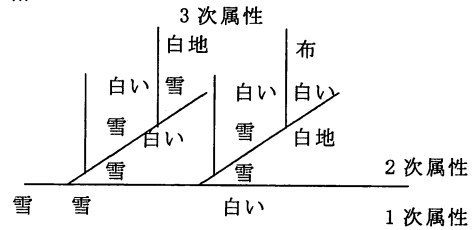


図 2 概念ベース(一部)

3.2. 関連度計算方式

関連度計算方式^[2]とは、概念ベースを利用して、概念と概念の関連の強さを定量化する手法である。

3.2.1. 重み比率付き一致度

関連度は、1 次属性の一致度から計算される。よって、関連度計算を説明する前に一致度計算について説明する。

以下のような概念 A, B があるとする。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_M, w_M)\} \quad (2)$$

$$B = \{(b_1, v_1), (b_2, v_2), \dots, (b_N, v_N)\} \quad (3)$$

M, N は、それぞれ概念 A, B の属性数である。また、 $a_i, w_i (1 \leq i \leq M)$ は、概念 A の属性とその重みである。同様に概念 B も $b_j, v_j (1 \leq j \leq N)$ で表される。2 次属性についても、以下のように定義される。

$$a_i = \{(a_{i1}, w_{i1}), (a_{i2}, w_{i2}), \dots, (a_{im}, w_{im})\} \quad (4)$$

$$b_j = \{(b_{j1}, v_{j1}), (b_{j2}, v_{j2}), \dots, (b_{jn}, v_{jn})\} \quad (5)$$

このとき、1 次属性 a_i と b_j の重み比率付き一致度 $DoM(a_i, b_j)$ は以下のように定義される。

$$DoM(a_i, b_j) = \sum_{a_u=b_v} \min(w_{iu}, v_{ju}) \quad (6)$$

3.2.2. 関連度

前節で求めた一致度を 1 次属性全ての組合せに対して行い、一致度が大きいものから順に対応を決めていく。式 (2) の概念 A に対して、一致度が最大となる組合せになるように、概念の属性を並べ替えたものを以下に示す。

$$B = \{(b_{x1}, v_{x1}), (b_{x2}, v_{x2}), \dots, (b_{xN}, v_{xN})\} \quad (7)$$

よって、これらの概念 A, B の関連度は次のようになる。

$$DoA(A, B_x) = \sum_{i=1}^l DoM(a_i, b_{xi}) \times \frac{(u_i + v_{xi})}{2} \times \frac{\min(u_i, v_{xi})}{\max(u_i, v_{xi})} \quad (8)$$

関連度の値は概念間の関連の強さを 0~1 の間の連続値で表す。概念 A と B に対して関連度を算出した例を表 1 に示す。

表 1 関連度計算方式の例

概念 A	概念 B	関連度の値
自動車	車	0.912
	飛行機	0.130
	学校	0.012

3.3. 質問文意味理解システム

質問文意味理解システム^[3]は、構文解析ツール^[7]を利用して、質問文から質問対象語（質問文が求めている対象）を取得するシステムである。質問文意味理解システムでは、質問文中に疑問詞「誰」や「場所」の表現があった場合、質問対象語として「人物」、「場所」を獲得できる。また、疑問詞が「何」の場合や疑問詞が省略される場合でも質問対象語を獲得できる。

質問文意味理解システムは、約 90%の成功率で質問対象語を正しく獲得できると報告されており、優れたシステムといえる。

4. Web 情報利用技術

本稿では、概念ベースに登録されていない概念を未定義語と呼ぶ。未定義語は語意の展開を行うことができないため、関連度を計算することができない。しかし、未定義語の属性獲得手法を適用することで属性とその重みを定義することができる。本稿では、ある単語に属性と重みの集合を与えることを概念化と呼ぶ。最適ノード決定手法は、5 章で述べる解答タイプ候補の中で解答タイプを決定できる手法である。

4.1. TF・IDF

TF・IDF 法^[8]とは、語の頻度と網羅性に基づいた重み付け手法である。TF はある文書中 d に出現する索引語 t （文書の内容を構成する要素）の頻度を表す尺度である。IDF はある索引語が全文書中のどれくらいの文書に出現するか（特定性）を表す尺度であり、式 9 で定義される。なお、 N が検索対象となる文書集合中の全文書数、 $df(t)$ が索引語 t が出現する文書数である。

$$idf(t) = \log \frac{N}{df(t)} + 1 \quad (9)$$

4.2. Web-IDF

4.1 で説明した IDF は一般的な文書（新聞や書籍など）を用いて索引語の特定性を考慮する手法である。一方、Web-IDF^[4]は Web にある文書のみを用いて索引語の出現頻度を考慮する手法である。Web-IDF では式 9 の N を Google^[6]が保有している日本語のページ数、 $df(t)$ を索引語 t を Google で検索を行ったときのヒット件数とする。なお、Google は全言語において保有しているページ数は公開されているが、日本語のページとして保有している数は公開されていないため、日本語の文書として最も使われている「は」で検索を行ったヒット件数（1,740,000,000-2009 年 1 月 31 日現在）を Google が保有している日本語の全ページ数としてい

る。

4.3. 未定義語の属性獲得手法

未定義語の属性獲得手法^[4]とは、未定義語 X の意味的特徴を表す属性（単語）とその重要性を表す重みの組を Web を用いて獲得する手法である。

- 1) 未定義語 X をロボット型検索エンジンに入力し、検索上位 100 件の検索結果ページを獲得する。
- 2) HTML タグなど不要な情報を取り除いた文書群に対して、形態素解析^[9]を行い、自立語を獲得する。
- 3) 獲得した検索結果ページに含まれる自立語の出現頻度と Web-IDF の算出を行い、TF・Web-IDF 重み付けを行う。
- 4) 自立語を重み順に並び替え、なおかつ、概念ベースに存在する自立語とその重みを X の属性として抽出する。

この手法を用いて未定義語 X の属性とその重みの組を構成する。未定義語 X の属性は式 10 のように構成される。

$$X = \{(x_1, w_1), (x_2, w_2), \dots, (x_n, w_n)\} \quad (10)$$

この手順により、未定義語に属性が与えられるため、未定義語に対しても関連度を算出することが可能となる。

4.4. 最適ノード決定手法

最適ノード決定手法^[5]とは、入力キーワードが大局的のどのような意味を持つのかを、意味分類体系として定義されたノード（所属候補ノード）の中で、その語が所属するべきノードを提示する手法である。キーワードと所属候補ノードの比較に、関連度計算およびノード特有の情報を利用することで、キーワードと所属候補ノードをより詳しく関係付けられる点が特徴として挙げられる。

本稿では、意味分類体系として 5 章で述べる解答タイプ候補を用いる。本稿では、解答タイプ候補を用いた最適ノード決定手法を解答タイプ決定処理と呼ぶ。解答タイプ決定処理により、質問が解答として求めている単語の種類（解答タイプ）を決定できる。

5. 解答タイプ候補

質問文が何を解答として求めているかを判断することは、質問の解答決定に大きな影響を与える。例えば「～はどこか？」という質問文であれば、解答として「場所」を期待していることがわかり、「場所」に適するかどうかを考慮する必要がある。本稿では質問文が解答として求めている単語の種類を解答タイプと定義する。解答タイプを決定する上で、候補となるものが必要となる（表 2）。

表 2 解答タイプ候補（一部）

人	場所	製品	行事
---	----	----	----

解答タイプ候補は、質問に頻繁に用いられる疑問詞「誰」や「どこ」から想定される「人」、「場所」、「組織」、さらに Web 検索において検索対象となりやすいと考えられる単語の意味分類体系「製品」や「作品」などを考慮して 12 種類を定義している。本稿では、解答タイプ候補を対象とする質問文の種類とする。

6. 知的 Web 検索方式

知的 Web 検索方式の流れを図 3 に示す。

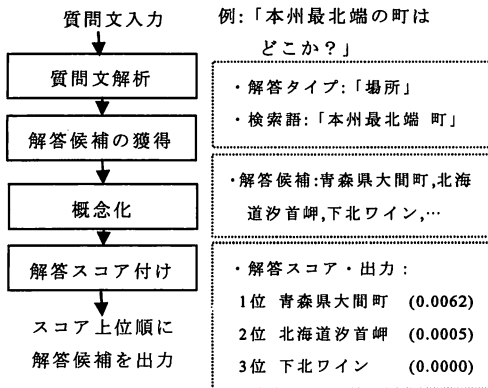


図 3 知的 Web 検索方式の流れ

まず、質問文を入力した後に、質問文解析を行い、質問解答のための検索条件(解答タイプおよび検索語)を決定する。次に、検索語を入力として Web から解答候補を獲得する。ここで、次処理の解答スコア付けを行うために検索語の概念化と解答候補の概念化を行う。最後に、解答候補に対して解答スコアを与える。スコアの上位順に順位付き解答候補リストとして出力する。

6.1. 質問文解析

6.1.1. 質問の解答タイプの決定

質問文を質問文意味理解システムに入力し、質問対象語を獲得する。解答タイプは質問対象語をもとに決定する。基本として質問対象語が、どの解答タイプに所属するかを示した知識ベースをあらかじめ作成することで解答タイプを決定する(表 3)。なお、質問対象語の表記が知識ベース上に存在しない場合は以下の処理を行う。

表 3 解答タイプの決定例

質問文	解答タイプ
同志社大学の前身である同志社英学校を創立したのは誰か?	人
本州最北端の町はどこか?	場所
毎年 5 月にフランスで開かれている国際映画祭は何か?	行事

質問対象語が「人物」であれば、解答タイプ「人」とする。質問対象語が「場所」であれば、質問対象語と解答タイプ候補「場所」および「組織」の関連度をそれぞれ計算し、関連度が大きい解答タイプ候補を解

答タイプとする。「人物」、「場所」以外の場合、質問対象語を入力として解答タイプ決定処理を行い、提示された解答タイプ候補を解答タイプとする。

6.1.2. 質問の解答タイプの決定

質問文に対して形態素解析ソフト「茶筌」^[9]を用いて形態素解析を行い、自立語や複合語(数字やアルファベットの連続)を、質問文のキーワード(動詞・形容詞を除く)として抽出し、スペース区切りで繋げた語を検索語とする(表 4)。

表 4 質問の解答タイプの決定例

質問文	検索語
本州最北端の町はどこか?	「本州最北端 町」

以上の手順によって、質問の内容をキーワードとして抽出し、有効に利用する。そこで、質問文が単文、複文に関係なく、あらゆる質問に対応することが可能であると考えられる。

6.2. 解答候補の獲得

質問の解答候補を Web から獲得する。解答候補の獲得の流れは次の通りである。

- 1) 検索語を入力として検索エンジン^[6]を用いて検索を行い、検索上位 100 件の検索結果ページの内容を獲得する。
- 2) HTML タグなど不要な情報を取り除いた文書群に対して、形態素解析^[9]を行い、自立語を獲得する。
- 3) 形態素に対して「名詞・数字・アルファベットの連続は複合」の条件を用いて、複合語を獲得する。
- 4) 抽出した自立語および複合語に対して、出現頻度と Web-IDF の算出を行い、TF・Web-IDF 重み付けを行う。
- 5) 自立語および複合語を重み順に並び替え、重み上位 50 件を解答候補として獲得する。

解答候補の獲得例を表 5 に示す。

表 5 解答候補の獲得例

検索語「本州最北端 町」	
解答候補	重み
青森県大間町	699.298
北海道汐首岬	39.582
下北ワイン	13.194
...	...

6.3. 概念化

4.3 節の未定義語の属性獲得手法を利用して、検索語および解答候補の概念化を行う(表 6)。これにより、6.4 節の解答スコア付けにおいて、関連度を算出できるようにしている。

検索語の概念化に利用する Web 文書は、質問文のキーワードから構成される検索語を入力として Web 検

索を行い得られたものである。これにより、質問の解答のために必要な文書を網羅的に扱い、その文書の意味特徴を属性として獲得できる。すなわち、概念化した検索語とは、質問の適合性を判断するための対象と捉えることができる。

表 6 概念化の例 (属性は一部)

検索語「本州最北端 町」		解答候補「青森県大間町」	
属性	重み	属性	重み
大間	1859.522	大間	4035.559
本州	1522.526	地図	188.496
町	276.172	天気	151.084

6.4. 解答スコア付け

解答候補に対して、質問に適するか、および、解答タイプに適するかの観点をもとにスコアを与える。質問の解答として適するかを考慮したスコア計算を 6.4.1 に、解答タイプに適するかを考慮したスコア計算を 6.4.2 に示す。6.4.3 にスコア付け手法を示す。

6.4.1. スコア計算 (質問の解答として適するか)

解答候補 ac_i の検索語 QW に対する関連度 $DoA(ac_i, QW)$ (検索語関連度とする) を計算する。検索語関連度を解答候補が質問の解答として適するかのスコアとする。

6.4.2. スコア計算 (解答タイプに適するか)

解答候補 ac_i の解答タイプ $AnswerType$ に対するノード得点 $NodeValue(ac_i, AnswerType)$ を以下の式 11 で計算する。ノード得点⁵⁾とは、解答候補と解答タイプの関連度に加え、解答タイプの持つ特有の情報を利用して、より詳しく関係性を定量化する得点である。

$$NodeValue(ac_i, AnswerType) = DoA(ac_i, AnswerType) \times \log(VerbHit(ac_i, AnswerType)) \quad (11)$$

$DoA(ac_i, AnswerType)$ は解答候補と解答タイプの関連度、 $VerbHit(ac_i, AnswerType)$ は解答候補にノード動詞を連結したキーワードの検索を検索エンジンで行ったときの HIT 数を表す。

ノード動詞とは、「ある解答タイプに所属する具体語は、その具体語の直後に現れる助詞を伴う動詞が同じである」という関係を利用して定義されたキーワードである。例えば、「行事」に所属すると考えられる「祭」と「オリンピック」はどちらも直後に「に参加」という表現を取ることが可能である。この関係性を利用して Web の HIT 数を求めることで、解答候補が解答タイプに適する語かを判断できる。本稿では、12 種類の解答タイプ候補すべてに対して、「製品」ならば「を販売」のようにノード動詞を与えている。

ノード得点を用いることで、固有名詞を含むあらゆる語に対して、柔軟に解答タイプへの適合性を値として算出できる。

6.4.3. スコア付け手法

解答スコア $Score(ac_i)$ の計算式を式 12 に示す。

$$Score(ac_i) = DoA(ac_i, QW) \times NodeValue(ac_i, AnswerType) \quad (12)$$

解答スコアは、検索語関連度 $DoA(ac_i, QW)$ およびノード得点 $NodeValue(ac_i, AnswerType)$ の積で求めている。検索語関連度は解答候補が質問の解答として適するかをスコア化しており、ノード得点は解答候補が解答タイプに適するかをスコア化している。この 2 つの観点から、解答スコアを求めることで、より高得点の解答候補ほど質問の解答に相応しいと判断している。スコア付け手法の例を表 7 に示す。

表 7 スコア付けの例

解答候補	検索語関連度	ノード得点	解答スコア
青森県大間町	0.107	0.0575	0.0062
北海道沙首岬	0.070	0.0071	0.0005
下北ワイン	0.008	0.0004	0.0000

7. 評価

本稿で提案している手法の評価を行うために、テストセットとして、アンケートにより収集した固有名詞が解答となる質問文 120 問を用いる (表 8)。なお、テストセットの質問文には、正解となる解答 (正答とする) および正解となる解答タイプ (正解解答タイプとする) を与えている。正解解答タイプは 12 種類の解答タイプ候補の中から一つを与えている。

表 8 テストセットの質問文の例

質問文	本州最北端の町はどこか？
正解解答タイプ	場所
正答	青森県大間町

本稿では、質問の解答タイプの決定の評価 (7.1)、解答候補の獲得の評価 (7.2) を行った上で、解答スコア付けを含めた提案手法全体の評価 (7.3) を行う。

7.1. 質問の解答タイプの決定の評価

本稿では、テストセットの各質問に対して、与えた正解解答タイプをもとに解答タイプ決定の知識ベースを作成している。よって、質問の解答タイプ決定は必ず成功するものとしている。しかし、これはすべての質問に対して適用できるものではなく、知識ベースに存在しない質問文の入力に対しての質問の解答タイプの決定の成功率を調査する必要がある。そこで、テストセットの質問文に対して正解の知識を与えない場合の質問の解答タイプの決定の精度を評価した。想定した正解解答タイプと一致した場合は○、質問文に対して正解解答タイプとは一致しないが、人間が正しいと判断できる場合は△、正解解答タイプと一致しなかった場合は×とする。

評価結果は、○ (96)、△ (14)、× (10) となり、

表 9 評価結果

精度	MRR
0.758	0.562

○および△を合わせて 110 問が正しく決定された。なお、△や×のように想定と異なる解答タイプ候補に分類される場合も想定し、解答タイプに対する適合性をスコアとして与える方法を用いているので、柔軟に判断できる。

7.2. 解答候補の獲得の評価

質問文 120 問に対して解答候補の獲得を行うことで、解答候補中に正答が獲得できるかを評価した。正答が獲得できた場合は○、獲得できなかった場合は×とする。なお、正答として、正答の表記の一部が存在する場合や表記揺れも含めている。

評価結果として、質問文 120 問中 113 問 (94.1%) に対して正答が獲得できた。また、各質問に対して平均 4 個の正答が獲得していることがわかった。

7.3. 提案手法全体の評価

7.3.1. 評価方法・尺度

提案手法の評価として、質問文に対して正答を獲得できるか、質問文の正答が上位に獲得できるかの観点で評価することで提案手法の性能が測れると考えられる。よって、評価尺度として以下を用いる。本評価では、評価尺度 1) を精度と呼ぶことにする。正答として、正答の表記の一部が存在する場合や表記揺れも含めている。

- 1) 正答がシステム出力の上位 5 位以内に含まれる割合 (精度)
- 2) Mean Reciprocal Rank (平均逆順位, MRR)

なお、MRR はシステムの求解精度を測る尺度である。Reciprocal Rank (逆順位, RR) は、システムが出力した順位付き解答候補リストにおいて、実際の正解が最初に現れた順位の逆数であり、MRR はその全質問平均である。RR ならびに MRR は 0 以上、1 以下の値をとり、値が大きいシステムほど求解精度が高いことになる。なお、本稿では質問文に対して、正答が 5 位以内に現れない時には RR の値は 0 とし失敗としている。

7.3.2. 評価結果・考察

図 4 に提案手法が出力したスコア付き解答候補リストに対して、正答が出現した順位の分布を示す。なお、図 4 における失敗とは、正答が 5 位以内に現れなかった場合である。出力 1 位に正答が出現した質問が 51 問となり、上位 5 位にかけて 91 問 (75.8%) に対して正答が獲得できることがわかった。

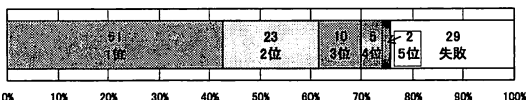


図 4 提案手法の出力における正答出現順位の分布

図 4 から、精度および MRR は表 9 となった。

なお、表 9 の結果に対して、解等スコアとして 6.4.1 (質問の解答として適するか) のみを利用した場合、精度は 0.667, MRR は 0.470 となった。また、解答スコアとして 6.4.2 (解答タイプに適するか) のみを利用した場合、精度は 0.5, MRR は 0.284 となった。提案手法では、スコア (質問の解答として適するか) がベースとなり、さらにスコア (解答タイプに適するか) を積として解答スコアを求めることで、より高い精度を得られたと考えられる。

8. おわりに

本稿では、質問文に対して Web から解答を選択する知的 Web 検索方式を提案した。手法として、語が有する意味特徴を語と重みで表現する概念ベースと語と語の間にある意味的な関連性を数値として算出する関連度計算方式、多様な表現を許している Web 情報を利用した未定義語の属性獲得手法および最適ノード決定手法、質問文意味理解システムを利用した。質問文の解答タイプに適するかという観点と質問に適するかという観点において、柔軟に解答スコアを求めることで、質問文の求める単語を解答できる手法となったと考えられる。

文 献

- [1] 奥村紀之, 土屋誠司, 渡部広一, 河岡司: “概念間の関連度計算のための大規模概念ベースの構築”, 自然言語処理, Vol.14, No.5, pp.41-64, Oct.2007.
- [2] 渡部広一, 奥村紀之, 河岡司: “概念の意味属性と共起情報を用いた関連度計算方式”, 自然言語処理, Vol.13, No.1, pp.53-74, Jan.2006.
- [3] 古川成道, 渡部広一, 河岡司: “概念ベースを用いた知的検索における曖昧な質問文の意味理解”, 第 18 回人工知能学会全国大会論文集, 2D1-10, Jun.2004.
- [4] 辻泰希, 渡部広一, 河岡司: “www を用いた概念ベースにない新概念およびその属性獲得手法”, 第 18 回人工知能学会全国大会論文集, 2D1-01, Jun.2004.
- [5] 後藤和人, 土屋誠司, 渡部広一, 河岡司: “Web を用いた未知語検索キーワードのシソーラスノードへの割付け手法”, 自然言語処理, Vol.15, No.3, pp.91-113, Jul.2008.
- [6] Google : <http://www.google.co.jp/>
- [7] 奈良先端科学技術大学院大学 構文解析システム「CaboCha」
- [8] 徳永健伸: “言語処理と計算 5 情報検索と言語処理”, 東京大学出版会, 1999.
- [9] 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明: “日本語形態素解析システム『茶筌』version1.0 使用説明書”, NAIST Technical Report, NAIST-IS-TR97007, 1997.