

## 詳細化された質問タイプによる質問応答システム

遠藤 哲哉† 福本 淳一 ††  
立命館大学大学院 理工学研究科†  
立命館大学 理工学部 情報学科 ††  
t\_endo@nlp.cs.ritsumei.ac.jp†  
fukumoto@cs.ritsumei.ac.jp††

### 概要

我々が QAC-1 に参考出展した質問応答システムは IREX 準拠の固有表現抽出を用いていた。そのため根本的に回答できるものが少なく、また分類の不足から質問の内容にそぐわない答を返すことが多かった。これら問題点を解決するために固有表現を階層的に詳細化し、それを用いて質問解析のタイプの詳細化を行った。また固有表現抽出だけでは解析できない質問文に対しても、詳細化した固有表現と動詞との共起を取ることで対応した。この手法を用いて構築した新たなシステムは QAC-1 時の結果と比較して精度の向上がみられ、良好な結果を得た。

## QA System using Classified Type of Named Entity

Tetsuya ENDO† Junichi FUKUMOTO††  
Graduate School of Science & Engineering, Ritsumeikan University†  
Department of Computer Science, Ritsumeikan University††  
t\_endo@nlp.cs.ritsumei.ac.jp†  
fukumoto@cs.ritsumei.ac.jp††

### Abstract

In this paper, we described a question and answering(QA) system using classified type of named entity. We submitted our QA system to Question and Answering Challenge 1(QAC-1) and evaluated it. We found insufficient and ambiguous question type in our system. We expanded named entity types to about 70 in order to handle wider question type. For ambiguous question type, collocations of named entity and verb are used. We evaluated our new QA system using question set of QAC-1. The new system had better performance than the previous one.

## 1 はじめに

近年の情報化の過程において、大量の文書情報の中からユーザーの質問に対して自動で回答を提示する質問応答（以下 QA）技術が注目されている。その QA システムにおいて用いられる技術としては情報検索や情報抽出などがあり、MUC[1]

や IREX[2] などで研究・評価がなされ、向上が図られてきた。また、QA システムの評価として Question and Answering Challenge -1（以下 QAC-1）[3] が NTCIR-3 Workshop（2002 年 10 月）のサブタスクとして行われており、我々の構築したシステム [4] も参考出展という形ではある

が参加し、評価を行った。その際に幾つかの問題点が発見されており、本稿ではその各問題に対して行った改良点について述べる。

まず、我々の QA システムは大まかに質問解析・テキスト検索・答候補抽出といった流れで構成され、これらの基幹部として固有表現 (Named Entity:NE) 抽出技術が用いられている。しかしながら QAC-1 へ出展したシステムでは、前半の質問解析・テキスト検索と後半の答候補抽出の部分が異なる設計で開発されたものであったため、全体を通して固有表現の統一が為されていなかった。このために質問から得られた情報を用いてテキスト中から目的の情報を探し出すことが難しく、正しい答を導き出せないことがあった。またさらに、固有表現抽出は IREX に準拠した人名、地名、組織名など 8 種類を識別する手法をとっていたため、QA システムで用いるには分類が非常に大まかであり、相応しくないものを回答とすることもあった。例えば「立命館大学のびわこ草津キャンパスはどこにありますか」という質問に対し、答候補の中に「日本」「滋賀県」「草津市」「野路東」とあった場合、これら全てを同列の「地名」として扱うことになっていた。このときに「日本」を答としてもそれはユーザーの望むものではない可能性が高く、かといってユーザーの前提知識が不明瞭な状況で「野路東」を回答とするのも相応しくないとと思われる。

次に、質問解析の精度の問題があげられる。QAC-1 に出展したシステムでは質問文の解析を 70 種ほどの単純なパターンマッチで行っていた。そのため簡単に質問タイプの判定できる「～は誰」などの表現であれば誤りは少なかったが、「～はどこ」「～は何」といった疑問表現を持つ質問に対しての判定精度は低かった。特に「～はどこ」という質問では地名と組織名どちらを答えるのかの判別が難しく、例えば「Java を開発した会社はどこですか」と「Java を開発したのはどこですか」という質問では、前者はパターンマッチで対処できるが、後者の場合は疑問表現に係る動詞へ着目しなければならず、我々のシステムでは解析が不可能であった。

そこで上記 2 点の問題に対処すべく、まず固有表現を階層的に分類・拡張し、質問タイプのレベルに応じて答となる情報の選定を目指した。また詳細化した固有表現の分類を利用し、QAC-1

時では困難であった質問タイプの解析も可能とさせた。その上で同一設計において新たに構築した QA システムの評価を行った。

## 2 固有表現の拡張

QAC-1 で作成された質問文とその答の情報、また新聞記事 (毎日新聞 98・99) 中に頻出するものなどを分類することで、固有表現を階層的に拡張・詳細化していった。基本的には IREX 定義の下に詳レベルの分類を作成したかったが、一部改訂を加えている。例えば「金額」と「割合」は共に「NUM(NUMERICAL)」分類の下位レベルの分類とし、「固有物名」は削除し、「人名」や「地名」「組織名」以外をまとめた分類「PROPER」を新たに設けている。現在は基本となる分類を 7 種、その下位レベルとして 60 種、他品詞情報 5 種を定義している。それを図 1 に示す。

QA システムへの実装にあたっては、パターンマッチや新たに作成した固有表現用辞書を用いた。この固有表現辞書は形態素解析器 ChaSen[5] の出力と新聞記事中の単語を分析した結果から作成しており、約 20,000 語を分類している。

## 3 質問解析

### 3.1 詳細なタイプ判定

前節で述べた固有表現の分類を用いて質問タイプも詳細化させることが出来る。例えば「火星が地球に最接近するのは何日になりますか」という質問の場合、以前のシステムでは「DATE」を尋ねている質問文であるとしか判定できず、誤った答を返す可能性が高かった。しかし本システムでは「DATE-DAY」と判定できるので、答候補として「26日」「2003年」「8月」「23時」等の情報がある場合においても、「DATE-DAY」表現である「26日」を答とすることができる。

### 3.2 共起を用いたタイプ判定

第一節の例問「Java を開発したのはどこですか」のように、以前のシステムで解析が不可能であった質問は疑問表現が「どこ」のものが多

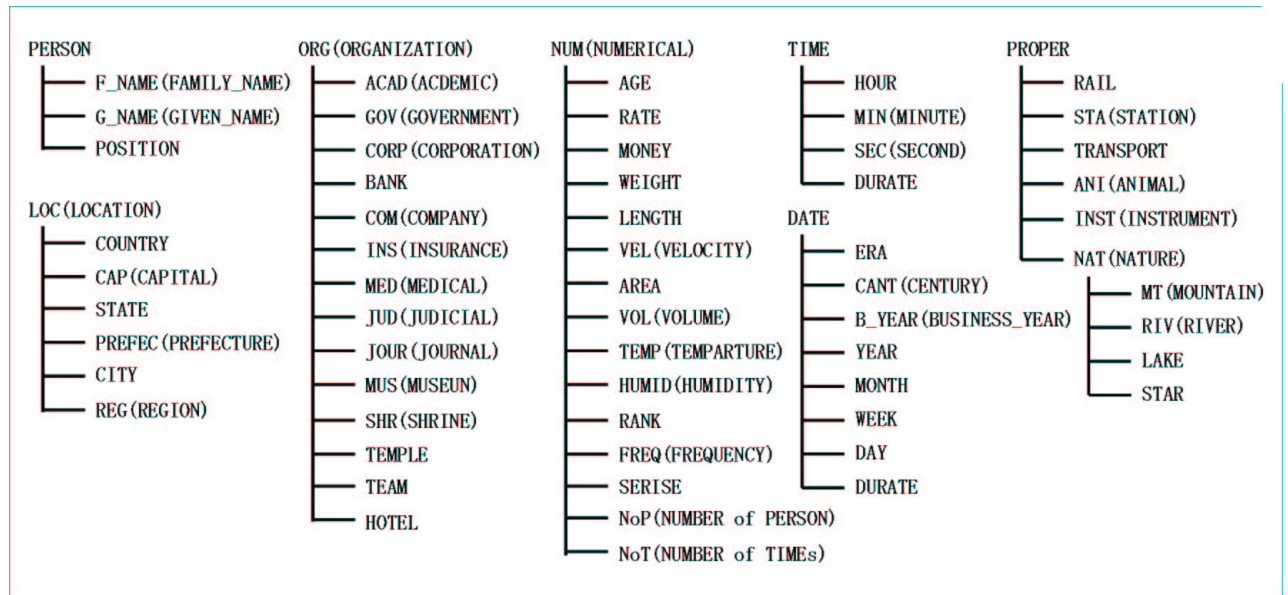


図 1: 固有表現の分類

かった。そこでこのタイプの質問の場合、答となる情報は「LOC」もしくは「ORG」であるという前提において、前節で作成した分類のうち、この2つを最上位に持つ各固有表現と、動詞またはサ変接続名詞の共起の統計を取り、その結果を利用して判定を行うことにした。対象データとして1994～2001年の毎日新聞記事と1998・1999年の読売新聞記事を用い、若干の係り受けを考慮して共起を計って約16,000語の動詞とサ変接続名詞の抽出を行った。その結果の例を表1に示す。

実際のシステムでは、各固有表現の共起頻度の降順において、隣り合う頻度の差が最大になる部分までを判定結果として返している。例えば「合併」の場合、頻度の差が最大になるのは「BANK」と「CITY」なので、判定結果は「COM」と「BANK」になる。またこの共起情報を用いると、「Javaを開発したのはどこですか」の質問の場合においては、疑問表現に係るサ変接続名詞「開発」と最も多く共起する「COM」が質問タイプの判定結果となる。

## 4 パッセージ抽出

検索で得られたテキストには回答に不必要な情報も多く含まれている。この不必要な部分を取

り除き答候補の検索効率を向上させるため、質問文から抽出されたキーワードを用いてパッセージを抽出を行って、答候補の検索範囲の絞り込みを行う。

### 4.1 テキスト検索

テキストの検索には全文検索システム Namazu[7]を利用している。検索キーワードは質問文から抽出された単語のみを使い、類似語などの言い換え表現は使用していない。まず各キーワード毎のDF値を得るために予備検索を行う。次にキーワードのDF値が低いものから用いていき絞り込み検索を行う手法を取ったが、数値や日時表現、用言のように異なる表現が可能なキーワードは例外的に優先度を低くした。

### 4.2 答候補検索範囲の絞り込み

パッセージの定義とその抽出範囲を決めるため、まず上位n位までの検索テキストが正解データをどの程度含んでいるのかを調べた。次に1パッセージを「検索キーワードを含む文とその前後m文」とし、mを順に増やしながら各テキストからパッセージ抽出を行って同様に含まれる正解

表 1: 共起頻度

開発		合併		こだわる		けん制		移り住む		下回る	
COM	457	COM	230	COUNTRY	178	COUNTRY	294	CITY	135	GOV	87
COUNTRY	246	BANK	197	GOV	125	GOV	86	LOC	111	COM	86
GOV	82	CITY	93	LOC	79	LOC	12	COUNTRY	53	COUNTRY	50
LOC	55	LOC	60	CITY	33	BANK	6	CAP	34	LOC	43
CITY	43	GOV	32	COM	10	CITY	2	PREF	22	CITY	35
ACAD	21	COUNTRY	23	BANK	7	CAP	1	REG	4	BANK	24

数を計った。計測には QAC-1 Task1 の問題セット 200 問を用い、検索対象は毎日新聞 98・99 年とした。その結果を表 2 に示す。

これよりシステムが回答可能な正解データはテキスト検索結果の上位 10 位までで全て出揃っており、正解質問数の伸びが顕著なのは 4~6 位までであることがわかった。またパッセージの範囲をキーワードを含む文のみにした場合と、その文と前後 1 文を含む場合とした場合とで差は殆どなく、今回の場合は前後 1 文を加えれば全ての正解データを含むことが分かった。よってこの結果から本システムにおいては、パッセージの定義を検索キーワードを含む文とその前後 1 文とし、パッセージを抽出する範囲は検索されたテキストの、上位 5 位までを上限とすることにする。

### 4.3 パッセージに対する重み付け

前項のパッセージの範囲を決めるための検証を進めていくと、1 テキストから複数のパッセージが抽出された場合、パッセージが重なり合う箇所に多くの正解データが含まれることが発見できた。これはこの箇所が検索キーワードが集まる部分であり、正解に関係する情報が多く含まれているためである。よってパッセージが重なり合う場合、その重なり合う箇所の文が重要であると見なし、答候補の抽出時に重みを与えることとした。

## 5 答候補の検索

前節の手法で抽出したパッセージから答候補を検索する。検索する答候補の固有表現分類は質問解析の結果から得られたものである。もし質問が解析不可能であった場合、疑問表現に掛かる名詞語句を含むパッセージから、「固有名詞的表現」や固有表現として分類されなかった名詞・未知語を検索する。各候補語の重みは TF/IDF 値で求め

表 5: 正解順位の統計

	1st	2nd	3rd	4th	5th	Total
System-A	58	26	9	5	5	103
System-B	64	27	12	6	5	114

るが、検索テキストの順位も考慮する。以下に答候補の重みを求める式 (1) を示す。

$$tfidf_{ac} = tf_{pac}(1 + \log(N/df_{ac}))$$

$$W_{ac} = tfidf_{ac} \cdot \frac{5 - (Rank_T - 1)}{Rank_T} \quad (1)$$

ここで  $tf_{pac}$  はパッセージ中における答候補  $ac$  の頻度であり、 $Rank_T$  はパッセージを抽出したテキストの検索時の順位である。またパッセージに重なり合う部分がある場合は以下の式 (2) を適用する。

$$W'_{ac} = W_{ac} \cdot \frac{1 + P_{freq}}{1 + \frac{P_{freq}}{2}} \quad (2)$$

$P_{freq}$  はその箇所におけるパッセージの重なり回数であり、この式によって 2 倍を超えない範囲で  $W_{ac}$  に重みを与える。

## 6 実験と評価

前節までの手法を用いて新たに構築した QA システムの評価を行った。実験には QAC-1 Task1 の問題セット 200 問を用いた。まず、今回拡張した固有表現分類の精度を確かめるため、質問解析で判定された分類のみを答候補として検索するシステムを「System-A」として採点を行った。その結果を表 3 に示す。

次に質問解析で判定された固有表現とその上位レベルの分類を答候補として検索するシステ

表 2: 正解が含まれる質問数

上位 n 位	1	2	3	4	5	6	7	8	9	10	...	20
テキスト全体	138	147	152	154	155	156	157	157	157	158	...	158
パッセージ (m=0)	137	146	151	153	154	155	156	156	156	157	...	158
パッセージ (m=1)	138	147	152	154	155	156	157	157	157	158	...	158
パッセージ (m=2)	138	147	152	154	155	156	157	157	157	158	...	158

表 3: System-A 実験結果

	Output	Correct(A)	Correct(Q)	MRR
QAC1 FormalRun	551	43	31	0.099
System-A	857	134	103	0.381

ムを「System-B」として採点を行った。ただし「PERSON」に関しては下位レベル「PERSON-F\_NAME」を同時に検索した。結果を表4に示す。また、それぞれのシステムが返した答が第何位で正解であったのかの統計を表5に示す。

表3と表4から、どちらも精度は前回のQAC1の平均値は上回ることができ、大幅にスコアの向上がみられた。また表5から、システムの返した正解データの約56%が第1位の回答であるという結果を得た。

また表2とシステムの精度から、回答可能な質問の約72%にシステムは正解データを返していることがわかる。

## 7 結果の考察

System-A と System-B の精度の差の原因となった質問の例として「2003年10月1日に合併することが決まった通信三社はどこですか」などがあった。これは質問解析の結果は「ORG-COM」であったが、正解である「DDI」「ID O」「KDD」がどれも固有表現辞書に「ORG-COM」として定義されてなく、固有表現抽出の結果が上位分類である「ORG」となっていた。そのため判定された分類のみを答候補として検索する System-A では正解を返すことができなかった。また「国際デジタル通信はどこの会社を買収されましたか」とその正解「ケーブル&ワイヤレス」も同様の原因であった。次に、逆に System-

Bの方が精度が落ちた質問もあったが、それらは全て人名に関するものであった。例えば「速水優の前の日銀総裁は誰ですか」という質問では、System-Aでは「PERSON」のみを検索しており第2位で正解「松下康雄」を回答していた。しかし System-Bでは「PERSON-F\_NAME」も同時に検索したため、これが雑音となって正解を第3位として返していた。

また、質問解析が不可能であった質問が53問存在した。これらは例えば「宇多田ヒカルの初アルバムは何ですか」や「世界で最も高いビルは何ですか」のように疑問表現が「～は何」であるものが殆どであった。これら解析できなかった質問の正解率は System-Bにおいて47%であり、MRRも0.336と全体の精度を下げている要因となっている。それ以外の質問に関しては、概ね今回作成した固有表現の分類に則した解析が出来ており、良好な結果を示している。固有表現の最上位分類での結果を表6に示す。

## 8 今後の課題

今回解析できなかった質問や、固有表現の分類の誤りは固有表現辞書の拡充で対処できるが、全ての表現を手動で分類するのは不可能であり、効率の良い学習方法等を検討していかねばならない。また System-Bのように、質問解析の結果だけでなくその上位レベル(一部下位レベル)の固有表現も同時に検索した方が精度が良かったが、

表 4: System-B 実験結果

	Output	Correct(A)	Correct(Q)	MRR
System-A	857	134	103	0.381
System-B	895	158	114	0.420

表 6: 固有表現毎の精度

分類	質問数	正解数	MRR
PERSON	42	22(52%)	0.389
ORG	14	10(71%)	0.530
LOC	33	22(67%)	0.441
TIME	4	2(50%)	0.208
DATE	17	10(59%)	0.451
NUM	29	18(62%)	0.435
PROP	8	5(62%)	0.417
misc	53	25(47%)	0.336

逆に邪魔になってしまう情報をどうふるいにかけるかも今後の課題である。さらに、今回は固有表現を階層的に拡張していったが、単純に1つの分類だけに当てはまらないものもある。質問の内容によっては「ORG」が「LOC」にもなり得るし、「PROPER」にもなり得る。このような場合に対応するためには、表層的な各固有表現分類や質問解析、答候補抽出を用いるのではなく、より意味的な解析が必要になってくると考えられる。

すでに本システム System-B を使用して QAC-2 Task1 の質問文は回答済みであるので、今後はその結果を考察してシステムに改良を加えていく予定である。また、丹羽 [6] らが数値表現に関する分類の詳細化と答候補の選定手法について提案しているが、本項で述べた固有表現分類はそれとは別途に調査・研究されたものであった。しかし数値表現に関しては似通った分類になっていることから、彼らの提案手法を組み込むことも検討している。さらに、今回拡張した固有表現の分類を利用したユーザとのインターフェイスを構築する事も可能であると考えられる。例えば質問文から得られる情報が不明瞭な場合、そのまま無理に答を探すのではなく、ユーザへ逆に質問を提示するなどし、対話的な解決方法で正解を導き出す事も実現できる。

## 参考文献

- [1] Message Understanding Conference (MUC)  
<http://www.muc.saic.com/>
- [2] Information Retrieval and Extraction Exercise (IREX)  
<http://cs.nyu.edu/cs/projects/proteus/irex.html>
- [3] J. Fukumoto, and T. Kato, An Overview of Question and Answering Challenge (QAC) of the next NTCIR Workshop, Proceedings of the Second NTCIR Workshop Meeting, National Institute of Informatics, 2001.
- [4] J. Fukumoto, T. Endo, T. Niwa, RitsQA : Ritumeikan question answering system used for QAC-1. Working Note of the Third NTCIR Workshop Meeting, National Institute of Informatics, 2002
- [5] Free Japanese Morphological analyser  
Computational Linguistics Laboratory,  
Graduate School of Information Science,  
Nara Institute of Science and Technology  
<http://chasen.aist-nara.ac.jp/>
- [6] 丹羽達洋, 福本淳一. "質問応答システムにおける回答候補の選定手法", 言語処理研究会 第9回年次大会 発表論文集, pp625-628, 2003
- [7] Namazu: a Full-Text Search Engine  
Namazu Project <http://www.namazu.org>
- [8] J. Fukumoto, T. Kato, and F. Masui, NTCIR-4 Question Answering Challenge (QAC-2)  
<http://www.nlp.cs.ritsumei.ac.jp/qac/index-e.htm>