5Q-2

質問応答システムにおける意味属性の利用

藤原 佑斗　浦谷 則好
東京工芸大学大学院工学研究科

1 はじめに

本研究では，factoid 型の質問応答システムにおいて，回答を絞るための情報として日本語語彙体系[1]によって得られる単語の意味属性や表層格の利用の有効性を示すものである．本稿では質問文から特徴語を取り出せない場合に対して，文の構成や属性を利用して回答候補を絞り込む手法を提案する．

2 関連研究

北條らの研究[2]のように factoid 型の質問応答システムにおいて質問文を解析し特徴語を用いて質問タイプを同定する研究は数多く存在しこれらは質問タイプを定めることにより高い精度を出している．

佐竹らの研究[3]では,新聞記事を対象に記事内の固有名詞の照応解析によって「ORGANIZATION」,「PERSON」,「LOCATION」,「DATE」,「ARTIFACT」,「TIME」,「MONEY」,「PERCENT」の８つの固有表現抽出を行なっている．

3 研究内容

本研究は，日本語語彙体系における意味属性や表層格の構文体系を利用することで回答となり得る属性を取得する．そして，Web 情報を用いて factoid 型の質問文に対して自動的に回答を求める．

3.1 質問文の制限

factoid 型の質問文はし，必ず回答が確定するものを限定する．また，「いつ」「どこ」といった質問文の型を示す特徴語が存在しないものを対象とする．

〈質問文の例〉
　ダイナマイトを発明したのは?
　徳川家康が征夷大将軍になったのは?
〈対象としない質問文の例〉
　電球を発明したのは誰?
　アメリカの首都はどこ?

Use of semantic attributes in question answering system
†Yuto Fujiwara, Graduate School of Engineering, Tokyo Polytechnic University
†Noriyoshi Uratani, Graduate School of Engineering, Tokyo Polytechnic University

3.2 意味属性の追加

日本語語彙体系から名詞に意味属性を付与する．しかし，日本語語彙体系のデータベースに存在しない名詞に対して属性を付与することができない．この問題を解決するために固有表現抽出を用いて名詞に意味属性を付与する必要がある．Cabocha で解析することで単語に対する固有表現を取得しそれらに意味属性を付与する．以下に固有表現ごとに付与する属性と例を記す．

表 1. 固有表現と意味属性の対応

固有表現	意味属性
ARTIFACT	具体
LOCATION	地域
ORGANIZATION	組織
PERSON	人間
DATE	年月日
TIME	時刻
MONEY	値・額
PERCENT	単位

例：1973 年　　DATE → 年月日
　東京都墨田区押上　　LOCATION → 地域
　アルフレッド・ノーベル PERSON → 人間

3.3 日本語語彙体系における含意関係の追加

日本語語彙体系には一般名詞と固有名詞がそれぞれの含意関係を構成している．さらに一般名詞と固有名詞で属性毎に対応を取っている．しかし含意は一部の属性同士間で行われており，今回のように様々な階層同士で含意関係を探る場合に上位や下位同士では含意の関係にならない．このため，対応している元の属性の上位と下位の 2 階層の含意関係を調べ含意関係を追加する必要がある．今回は独自に含意の関係になるであるものを追加した．

3.4 質問文のパターン

日本語語彙体系に存在する表層格の構文体系を利用するために質問文を二つのパターンに判別する．1 つは質問文の名詞を利用するパターンで意味属性のみを利用するものである．もう一つは動詞を含む質問文で，表層格を利用し回答候補となり得る属性を取得するものである．

4 システムの構成

本研究で構築したシステムを以下の図1に示す.

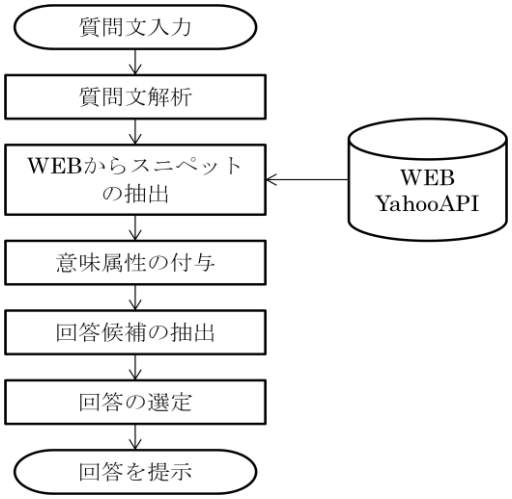

図 1. システムの概要

質問文の解析では形態素解析した結果に対して 3.2 で述べた方法で形態素に属性を付与する. さらに 3.4 の質問文のパターンに従って質問文を判別する.

質問文を形態素解析した結果から, 名詞・動詞・形容詞を検索キーワードとして Yahoo！API を利用して Web 検索を行い単語の出現頻度を取得する. 回答候補を抽出するためのスニペット(Summary)は 500 件とする. 検索キーワードの例を以下に示す.

質問文　：ダイナマイトを発明したのは?
↓
検索キーワード:「ダイナマイト」「発明した」

Web から取得したスニペット情報を形態素解析にかけ, 意味属性を付与する .

4.1 意味属性を利用

名詞の意味属性を利用した質問文の場合, 多くは「○○の××は」といった質問文の形態になる. このパターンの場合, 回答の意味属性が質問文と同じもしくはそれに近い意味属性を持つことが考えられる. 今回は質問文に含まれる名詞の意味属性が回答候補となる名詞の属性と同じもしくは含意の関係にあるものを回答候補として抽出する.

4.2 表層格の構文体系を利用

3.4 で述べたように質問文に動詞が含まれる場合, 表層格の構文体系を利用する. 構文体系には名詞・助詞・動詞が存在する. 構文体系と質問文との対応をとり質問文に沿う構文を取得し回答候補となり得る属性を抽出する.以下に取得した構文の例を記す.

例:〈質問文〉 ダイナマイトを発明したのは
〈取得した構文〉 N1 が N2 を 発明する

この時, N2 がすでに存在するので N1 になり得る意味属性が回答候補となる.ただし N1 が全ての属性になり得る場合は取得しない. しかしながら, 質問文の名詞の属性がマッチできない場合や, 名詞自体に属性が付与できない場合がある. この時は助詞・動詞から回答候補の属性を取得する. さらに WEB から取得したスニペットから助詞を用いて回答候補を絞り込む.上記の例では回答候補となる N1 の後に助詞「が」が存在するので助詞「が」の前に在る名詞のみを取得し属性を調べる. また, 質問文で取得した構文が全て埋まってしまった場合には属性「年月日」を回答候補とする. これは日本語語彙体系における表層格が基本的な構文になっているため, 年や日付といった任意書格が構文に存在し難いためである.

5 実験結果

3.1 で制限した factoid 型の質問文で意味属性 15 問, 表層格の構文体系 30 問を実験に用いた. それぞれの結果を表 2 に示す.

本手法で行った結果, 意味属性のみを利用した手法では精度, MRR ともに悪い結果になってしまった. これは, 今回質問文に含まれる属性を回答候補の属性とした結果, 一般名詞が上位に挙がったためである. 表層格を利用した構文体系では 5 位以上が 67%, MRR が 39%という結果になった.

表 2.実験結果

質問文	1 位	5 位以内	MRR
意味属性	6%(1/15)	13%(2/15)	0.10
構文体系	20%(6/30)	67%(20/30)	0.39

6 おわりに

本実験結果では意味属性の利用に対してあまり良い結果を得ることができなかった. 回答となり得る属性をさらに絞る必要がある. また, 一般名詞と固有名詞についても重みを変え差別化を図る必要がある. 表層格を利用した構文体系では回答となり得る属性を絞ることができ, また MRR の値から約 6 割の確率で回答が上位に出現することで属性を利用する有効性を示せた.

参考文献

[1]NTT コミュニケーション科学基礎研究所:日本語語彙体系, 岩波書店
[2] 北條奈緒美, 獅々堀正幹, 北研二:www 検索エンジンを用いた質問文内の用語特定手法, 情報処理学会研究報告, pp.97-102,2007
[3] 佐竹正臣, 白井清昭, 奥村学:照応関係を考慮した新聞記事の固有表現抽出,