

質問応答システム Metis の回答精度向上

- 検索フェーズの改良を中心として -

久保田裕章, 平塚飛将, 吉川ひかる, 松田源立, 原田実
青山学院大学 理工学部 情報テクノロジー学科

1. はじめに

近年、自然文で書かれた質問文に対して回答を抽出する質問応答システムの研究が盛んに行われている。先行研究では、質問文と知識文をある基底表現に言い換えてから照合を行うシステム[1]、質問文と知識文の係り受け構造の類似性から回答候補を得るシステム[2]などが提案されている。原田研究室では、語の意味と語間の深層格を評価できる意味解析を用いれば現在提案されているシステム以上に高精度な質問応答システムが開発できることに着目し、意味解析システム Sage[3]を用いて意味グラフベースで質問文と知識文を照合することで回答抽出を行う質問応答システム Metis を開発している。Metis2006[4]では NTCIR[5]の評価型ワークショップに参加したが、プロトタイプであったため回答抽出精度は他システムと比べて中程度のものであった。本研究の目的は、質問文解析、知識文検索や回答抽出などシステム全体において改良を加え、更なる精度向上を目指すことである。

2. 質問応答システム Metis の概要

Metis は、図 1 に示すように、まず自然言語で入力された質問文と新聞記事や Web 中の文章（知識文）の両方を Sage で意味グラフに変換する。ここでは、文は文節を頂点、係り受けを辺にしたグラフで表現され、文節の主辞には EDR 辞書の語意が、辺には Sage が定めた 37 種の深層格のどれかが割り当てられる。この両グラフの共通部分グラフ（閾値以上の語意類似度を持つ頂点ペアとその間の閾値以上の格類似度を持つ辺ペア）の大きさで文間の類似度を判定する。例えば、図 1 に示すように Metis に質問文「ペスト菌を発見した細菌学者は誰ですか。」が入力されると意味解析を行った後、質問文解析を行う。この処理では疑問詞が特定され、それに与える意味制約（誰）を決定する。次に、質問文から検索エンジン呼び出すためのキーワード（ペスト菌、発見、細菌学者）を抽出し知識文検索を行う。そして得られた知識文と質問文の意味グラフを照合し回答（北里柴三郎）を抽出する。最後に抽出した回答に順位付けを行い表示する。

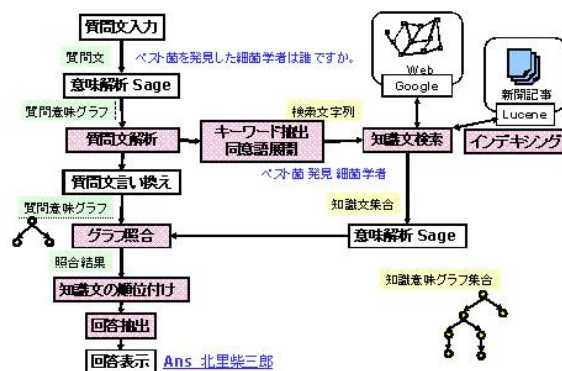


図 2-1 Metis システム構成図

2.1. 質問文解析の流れ

まず、Metis に入力された質問文は Sage により意味解析され質問グラフに変換される。質問グラフ中に疑問詞があれば質問箇所の特定に進む。質問箇所の特定処理は質問が何を回答として求めているのかを判別し、factoid 型の質問の場合には質問グラフ中において、疑問詞を持つノードに質問箇所であることを示すフラグを立てる。質問箇所を特定した後、factoid 型質問(人物、時間といった単語を回答とする質問)の場合に質問箇所がどのような概念であるべきかを表わす意味制約を与える。

2.2. 知識文検索の流れ

知識文検索の流れを図 2 に示す。まず、質問文解析が終了した質問グラフに対してキーワードを抽出し、さらにこれらのキーワードの組合せを生成する。生成したキーワードの組合せから検索エンジンの仕様にあった検索文字列 (AND 検索) を作成し、検索エンジン呼び出す。検索された知識文を Sage で意味解析して知識グラフを獲得し、その中から述語節が「疑問」、「推量」、「仮定」、「否定」のモードを持つグラフ以外を確実な事実を表す知識グラフとして選び出し、グラフ照合へと進む。

2.3. グラフ照合の流れ

Metis は質問グラフと検索された知識グラフを照合することによって質問に対する回答を抽出する。グラフ照合により計算されたグラフ類似度によって抽出した回答の信頼性を判定している。

2.4. 回答抽出の流れ

Metis はグラフ照合の後、質問に対する回答を抽出する。回答抽出とは質問の内容に沿った回答が抽出できているかを判断し、抽出する回答の整形を行う処理である。

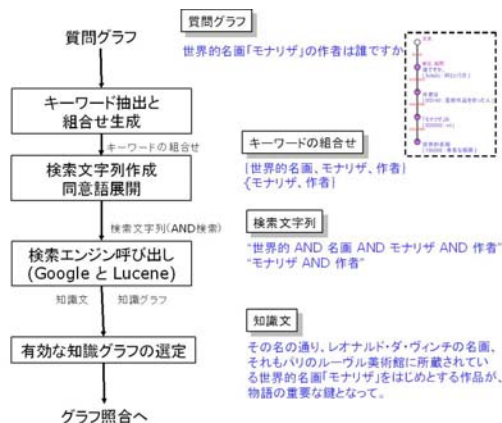


図 2.4-1 知識文検索の流れ

3. 従来の Metis の問題点

以下に、具体的な4つの工程における問題点を挙げる。

1. 質問文解析における問題点

- ・意味制約を広い範囲でしか与えていない。
- ・疑問詞のない質問文に対応していない。
- ・Yes/No 文に対応していない。

2. 知識文検索における問題点

- ・キーワードが形態素単位ではなく文節単位で抽出されている。
- ・従来のキーワードでは回答を含まない知識文も検索してしまう。
- ・カタカナ語の表記のブレのため知識文が得られない。

3. グラフ照合における問題点

- ・数詞や固有名詞の類似度が人の評価とズレている。

4. 回答抽出における問題点

- ・キーワードを多く含む知識文が優先されない。
- ・同格表現から回答が得られない。

4. 精度向上策

以下、各フェーズにおける改良点を述べる。

4.1. 質問文解析における改良

Metis2007 における質問文解析の処理の流れを図 3 に示す。質問文に疑問詞がない場合には、適切な疑問詞の判定を行う。さらに、適切な疑問詞が判定できなかった場合には Yes/No 文の判定を行う。

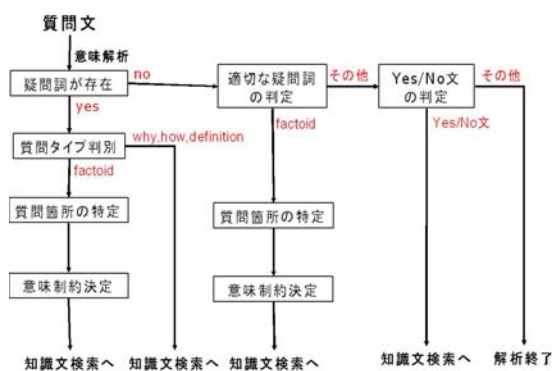


図 4.1-1 質問文解析の流れ

4.1.1. Factoid 型質問に対する意味制約

factoid 型質問において、適切な回答のみ抽出するため意味制約を詳細化する。例えば質問タイプが「何処」の場合、従来は「地名」と「組織」を意味制約として与えたが、地名を問う質問に対し組織名を回答としてしまう場合がある。主題(共起関係子「は」の文節)の概念が組織と地名のどちらに近いかを計算し意味制約を与え、質問内容に沿った照合と抽出を行う。

4.1.2. 疑問詞のない質問文への対応

質問グラフ中に Metis が認識する疑問詞が存在しない場合、適切な疑問詞の判定を行う。具体的には、図 4 に示すように、「疑問ノードの主辞の概念 ID」と「質問タイプの概念 ID」(表 1 参照)の類似度が最も高いものを疑問詞として追加する。例えば、「黒沢明氏が亡くなった月は？」を入力すると、疑問ノードの主辞(この場合は「月」)の概念 ID : 0fe60c と表 1 に示した全ての概念 ID と類似度を求める。類似度が最も高いのは「30f776:時」と「3f9882:時間点」の概念であり、類似度は共に 0.465 である。なお、その他の概念 ID との類似度は 0 であった。結果として最も類似度が高い概念に対応した質問タイプである「いつ」を適切な疑問詞として、疑問詞ノードを末尾に加える処理を行う。

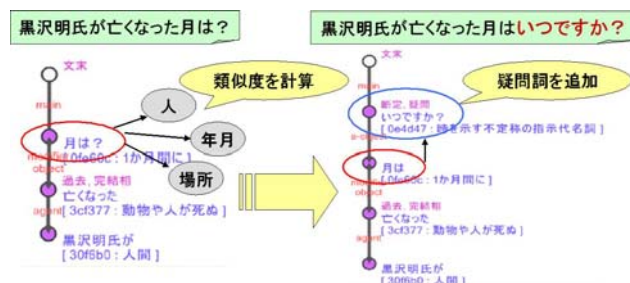


図 4.1.2-1 疑問詞のない質問文への対応

表 1 質問タイプ別の概念 ID

質問タイプ	概念 ID	概念 ID 説明
誰、だれ	3f961c	人間の具体的な名前
	444846	人間の支持的な呼称
	30f6b0	人間
	0f15aa	役割で捉えた人間
	0e3252	彼(照応関係があった場合に有効)
何処、どこ	3aa911	人間または人間と似た振る舞いをする主体
	4449b4	組織のいろいろ
	3f965a	地域
	444a40	国名で捉えた国家
	3f9659	行政単位に対応する地名
	3aa932	建造物
	3aa938	場所
いくら、いくつ	3aa912	自立活動体
	30f93d	計量の単位
	30f7aa	数量
	3f9875	具体物の大きさ
何時、いつ	30f776	時
	3f9882	時間点
何、なに	30f7e4	事象
	3d017c	物事
	30f6ae	具体物
	444d86	抽象物
	0f3177	作ったもの
	4445e8	情報が記されている物
	444db1	物事の属性
	3aa963	状態
	0f15aa	呼び名
その他	0f15aa	呼び名

4.1.3. Yes/No 文への対応

適切な疑問詞が判定できなかった場合には、Yes/No 疑問文の判定を行う。質問タイプが判断できなくても語尾に「か?」があれば質問タイプ YES/NO として判断する。図 5 に示すように、YES/NO に対する回答は基準となる閾値を超えるグラフ類似度を得られた知識文の割合で 3 種の中から適切なものを返す。

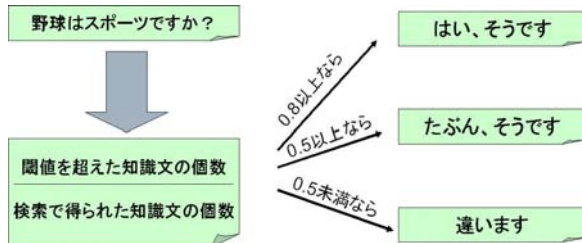


図 4.1.3-1 Yes/No 文への対応

4.2. 知識文検索における改良

知識文検索における改良を以下に述べる。

4.2.1. キーワード抽出法の改善

従来のキーワード抽出方法では、キーワードが形態素単位ではなく文節単位で抽出されていた。WEB 検索とローカル DB 検索ではインデックス方式が異なるため、キーワードの抽出方法を分ける。WEB では Google がキーワード形態素への分割を自動的に行うので文節で指定し、ローカル DB では Lucene[6]を用いて Index 自体を形態素で作成するので、形態素で指定する。例えば「ハーバード大学ケネディ・スクールの学長は誰ですか?」という質問文に対して、WEB 検索では「ハーバード大学ケネディスクール」「学長」が抽出され、ローカル DB 検索では「ハーバード大学」「ケネディ」「スクール」「学長」が抽出される。

4.2.2. ローカル DB におけるインデキシング

Index 対象となる語は、ノードのリファレント、表記より助詞を除いた語、形態素より助詞・括弧・句読点を除いた語、形態素の基本形、である。また、Index の作成においては図 6 に示すように、語と深層格（語の役割）をペアにした Index を登録する。深層格を含めることで「1979 年に、米中が国交を正常化した。」という文は「正常化」の「time」「agent」「object」格で表される知識を持つことがわかる。

知識文: 1979 年に、米中が国交を正常化した。

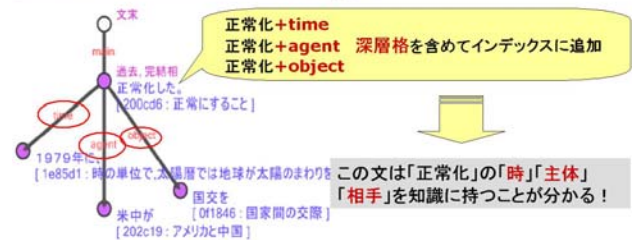


図 4.2.2-1 語と深層格をペアにした Index 作成

4.2.3. 答えを含まない知識文検索の阻止

「米中が国交を正常化したのはいつですか?」という質問文に対して、キーワードを「米中 国交 正常化」として文を検索しても、得られた知識文に国交が正常化したのが「いつ」なのかという答えが含まれているとは限らない。そこで、図 7 に示すように、例えば「いつ」の質問タイプであれば深層格 time を含めたキーワード「正常化+time」を指定することで、「正常化」が time 格を持つ知識文のみを検索できるようになる。言い換えれば、後続のグラフ照応による質問グラフに対する全体的な照合の前に、語の深層的な役割に基づく小規模な意味検索をインデックスを用いて高速に行うことができ、有効なスクリーニングになる。

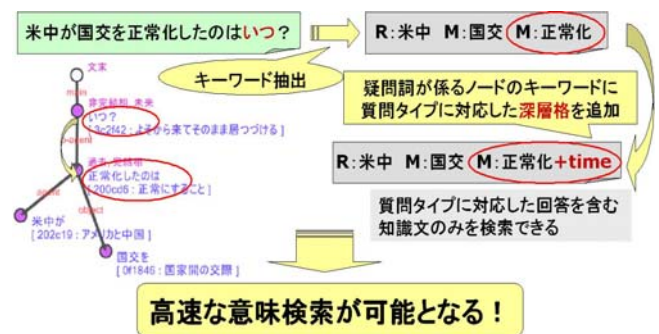


図 4.2.3-1 深層格を含めたキーワード検索

4.2.4. カタカナ表記のブレへの対応

カタカナには表記にブレがあるため、知識文が得られない場合がある。図 8 に示すように、カタカナ語に対してブレを吸収するため表記のバリエーションを増やす処理を行う。

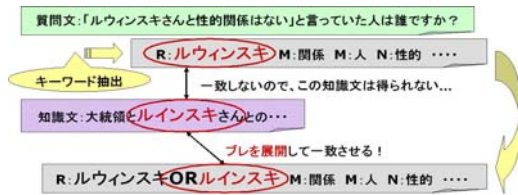


図 4.2.4-1 カタカナ表記のブレへの対応

4.3. グラフ照合における改良

グラフ照合における改良を以下に述べる。

4.3.1. ノード類似度計算の改善

年数や日付のような数値表現では「1999年」と「1998年」を違うものと認識し、完全一致の類似度を与える。主辞・副主辞の文字列を比較し一致した場合のみ類似度を1.0、一致しない場合は類似度0とする。ただし年数においては「1999年」を「99年」と略記することもあるため、このような場合は類似度を1.0とする。

4.4. 回答抽出における改良

回答抽出における改良を以下に述べる。

4.4.1. キーワード得点の追加

Metis2006では、キーワードを多く含む知識文が回答抽出の際に優先されないことがあった。そこで、図9に示すように、知識グラフがキーワードを含む量に応じて回答スコア(グラフ類似度)に経験値30点を基本にして点数を加算していく処理を行う。

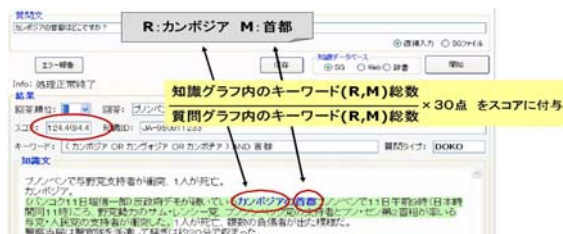


図 4.4.1-1 キーワード得点の追加

4.4.2. 同格表現の展開

Metis2006では「首都プノンペン」のような同格表現から回答を得ることができなかった。そこで図10に示すように、同格表現部分を展開することで回答を抽出するように改良した。

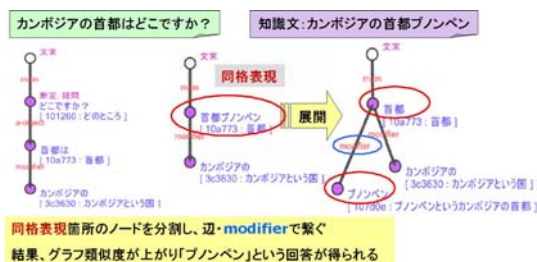


図 4.4.2-1 同格表現の展開

5. 評価実験結果

クイズミリオネアから factoid 型の質問を 100 問、WEB を知識源として実験を行った結果を図 11 に示す。この検索 100 問中 4 問はインターネットから有効な知識文を検索できなかった。

	検索	正解		正解抽出精度		回答精度
		1位	3位まで	1位	3位まで	
2006年度Metis	96/100	27	48	28.1%(27/96)	50.0%(48/96)	48.0%(48/100)
2007年度Metis	96/100	49	70	51.0%(49/96)	72.9%(70/96)	70.0%(70/100)

図 5-1 クイズミリオネア実験結果(知識はWEB)

また、NTCIRでのテストコレクションCLQA(factoidが200問、新聞記事2年分を知識源として実験を行った結果を図12に示す。

	検索	正解		正解抽出精度		回答精度
		1位	5位まで	1位	5位まで	
2006年度Metis	158/200	33	65	20.9%(33/158)	41.1%(65/158)	32.5%(65/200)
2007年度Metis	189/200	78	142	41.3%(78/189)	75.1%(142/189)	71.0%(142/200)

図 5-2 NTCIR の CLQA 実験結果(知識は新聞記事)

評価実験の結果、2006年度のMetisと比べて回答抽出の精度が32.5%から71.0%へ大幅に向上した。この精度は2007年NTCIRのCLQAコンテスト中最高位の59.5%を上回るものだった。

6. 結論

システム全体に渡って改良を行い、回答抽出精度が大幅に向上した。特に、疑問詞のない質問文への対応が再現率・回答精度向上に大きな影響を与えた。また、意味制約ルールを詳細化することで質問内容に沿った回答抽出が実現された。知識文検索においては深層格込み検索により、正答を含む知識文が上位に検索できるようになった。

7. 参考文献

- [1]Boris Katz : Using English for Indexing and Retrieving, MIT Press, Vol.2.(1990).
- [2]倉田岳人, 岡崎直観, 石塚満 : 係り受け関係に基づくグラフ構造を用いた質問応答システム 情報処理学会自然言語研究会報告 NL-158, pp.69-74(2003).
- [3]川口純一,青木洋,松田源立,原田実"意味解析システム SAGE の精度向上"情報処理学会第 69 回全国大会論文集,1C-04,第 2 分冊 pp. 77-78. (2007.3).
- [4]Minoru Harada, Yuhei Kato, Kazuaki Takehara, Masatsuna Kawamata, Kazunori Sugimura, and Junichi Kawaguchi: "QA System Metis Based on Semantic Graph Matching "Proc. of the 6th International Conference on NII Test Collection for IR Systems(NTCIR6), Tokyo, Japan, pp.448-459, (2007.5).
- [5]NTCIR : <http://research.nii.ac.jp/ntcir/>
- [6]Lucene.net:<http://incubator.apache.org/lucene.net/>