

指示詞を考慮したテキスト検索方法の検討 Text retrieval in consideration of the function of demonstratives

鈴木 晋[†]
Suzuki Susumu[†]

Abstract A new method is proposed to retrieve Japanese texts. For the method, a text manager makes, from a Japanese text, a formal text in the form that a computer can process easily, considering the function of demonstratives in the Japanese text, beforehand. The method retrieves Japanese texts by using each formal text. Since a formal text gives, to a Japanese text, a closer approximation than key words and tags, the new method can answer queries that can not be done by text retrieval methods using key words and tags.

1. はじめに

テキストを検索する方法として、キーワードやタグを用いた検索方法が普及している。筆者は文献[1]において、より高度な検索を目指して[2,3,4,5,6]、日本語テキストの新しい検索方法（形式化テキスト検索法と呼ぶ）を提案した。この検索方法では、予め、テキスト中の各日本語文に対して、その文法構造を手で付加した文（形式化文と呼ぶ）を作成し、これらの文の集まり（形式化テキストと呼ぶ）を用いて検索を行う。形式化テキストは、計算機で処理しやすい形に日本語テキストを近似したものであり、キーワード集合やタグ集合より正確に日本語テキストを近似しているので、キーワードやタグを用いた検索では答えられないような質問に答えることができる。形式化テキスト検索法の回答の正確さを向上させるためには、形式化テキストが日本語テキストを近似する精度をあげることが重要である。本稿では、文献[1]の形式化テキストに指示詞（これ、それ、この、その）の表現を導入し、その形式化テキストを用いた検索方法を提案する。この新しい形式化テキスト検索法は文献[1]の形式化テキスト検索法より高い精度で質問に答えることができる。

本稿は以下、次のように構成される。2節において形式化テキスト検索法の全体の流れを説明する。3節で、例として用いる日本語テキストと日本語質問文を与える。本稿の形式化テキスト検索法はこれらの例をうまく扱うことができる。4節以降は、形式化テキスト検索法において特

に重要である、形式化テキストの作成法と、形式化テキストに対する質問処理について述べる。4節では指示詞の表現の仕方を、5節では日本語文から形式化文を作成する仕方を説明する。3節の日本語テキストの例に対してこのようにして作られた形式化テキストを6節にまとめて示す。7節で形式化テキストに対する質問処理を説明し、最後に8節でまとめる。

2. 形式化テキスト検索法の全体の流れ

形式化テキスト検索法の全体の流れを図1に示す。図1の①から⑨を説明する。

- ① 予め、テキスト管理者は日本語テキストから形式化テキストを作成する。
- ②, ③ 検索者は日本語質問文を形式化質問文に変換して、計算機に質問を行う。
- ④, ⑤ 計算機は形式化質問文に対して形式化テキストを検索して回答（形式化回答）を作成し、検索者に返す。
- ①から⑤ 検索者は、計算機から返された形式化回答が十分でないかと判断した場合、形式化質問文を修正して質問を繰り返す。
- ⑥ 形式化回答は、もし人が日本語テキストと日本語質問文を読んだら返すであろう正しい回答の近似であるので、検索者は最終的に、日本語テキストの中で形式化回答の作成に関連した部分を読んで、回答が正しいか否かを判断する必要がある。そのために、検索者は日本語テ

[†] 愛知工業大学 情報科学部 情報科学科（豊田市）

キストの中の関連した部分を計算機に要求する.

⑦, ⑧ 計算機は関連する部分を検索者に返す.

⑨ 検索者をそれより回答の真偽を判断し, 必要ならば日本語回答を作成する.

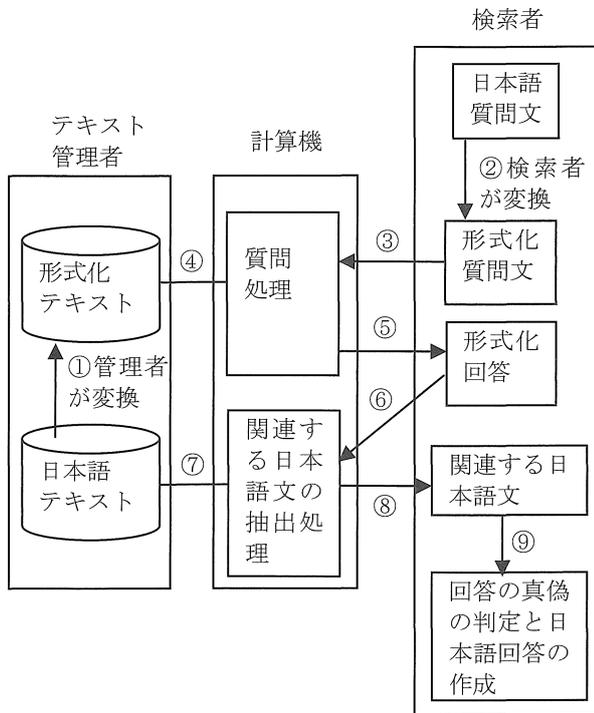


図 1. 形式化テキスト検索法の全体の流れ
Fig.1 The overall flow of the formal text retrieval method

3. 日本語テキストと日本語質問文の例

3.1 日本語テキスト

日本語テキストの例を次に示す. 下線は指示詞である.

S1: 太郎は毎年, 正月に, 自分でとった写真を友達に送っている.

S2: 一昨年のその写真は電車の中で漫画を読んでいるサラリーマンの写真であった.

S3: 昨年のその写真は予備校生についてであった.

S4: それらの写真のどちらも面白かった.

・・・(*1)

3.2 質問文 (原文)

(*1)の日本語テキストに対する質問文の例を次に示す.

Q1: 太郎は正月に友達に写真を送りますか?

答: はい

Q2: 太郎は写真を誰に送りますか?

答: 友達

Q3: 太郎は漫画を読んでいるサラリーマンの写真を友達に送りましたか?

答: はい

Q4: サラリーマンの予備校生の写真はありますか?

答: いいえ

Q5: 何が面白いのですか?

答 1: 写真

答 2: 太郎が一昨年, 正月に, 自分でとって友達に送った写真で, かつ, 電車の中で漫画を読んでいるサラリーマンの写真

答 3: 太郎が昨年, 正月に, 自分でとって友達に送った写真で, かつ, 予備校生についての写真

・・・(*2)

4. 指示詞の表現

4.1 普通名詞が表すもの

3.1 節の文 S1,...,S4 (*1) には普通名詞 (固有名詞でないもの. 例えば「写真」や「友達」) が使われている. これらの普通名詞, 例えば「写真」は, 写真の集合 (集合「写真」と記す) のある要素を表していると考えることができる. たとえば, 「太郎が一昨年に, 正月に, 自分でとって友達に送った写真」や「太郎が昨年に, 正月に, 自分でとって友達に送った写真」などを表していると考えることができる.

4.2 指示詞の機能

同じ普通名詞が複数の文に現れる場合, 各文の普通名詞は集合の同じ要素を表している場合もあるし, 表していない場合もある. ある文の中の普通名詞が, 前にあるいずれかの文の普通名詞と要素として同一であることは, 通常, 指示詞 (この, その) によって表される. 人は指示詞付きの普通名詞をみて, それがどの文の中の普通名詞と要素として同一であるか判断する. なお, 指示詞は省略されることもあるので, そのような場合には指示詞を補完した後に, この判断を行う必要がある. 日本語テキストを正しく理解するためには, 普通名詞の要素としての同一性を正しく理解する必要がある.

3.1 節の(*1)の中の普通名詞「写真」を使って具体的に説明する (複数の文に現れる普通名詞は他にはない). 文 S2 の中の指示詞 (下線で示されている) 付きの名詞「その写真」は文 S1 で説明された写真 (文 S1 を満たす写真) と同じ要素を表すと考える. すなわち, 「太郎が一昨年に, 正月に, 自分でとって友達に送った写真」と「太郎が昨年に, 正月に, 自分でとって友達に送った写真」と…を表す

と考える。この結果、S2 中の「一昨年のその写真」は、それらの写真の中で一昨年のも、すなわち、「太郎が一昨年に、正月に、自分でとって友達に送った写真」を表すと考える。S3 の「その写真」は、一昨年の写真と昨年の写真が同じ要素であることはないので、S1 を満たす写真（集合「写真」のある要素）と同一であると考え。また、S4 中の「それらの写真」は S2 を満たす写真（集合「写真」のある要素）と S3 を満たす写真（集合「写真」のある要素）と同一であると考え。

4.3 普通名詞と指示詞の機能の形式的表現

上記に説明した普通名詞と指示詞の機能を計算機でも正しく処理できるように、(*1)の日本語テキストを次のように書き換える。

$\exists a, \exists b, \exists c, \exists d, \exists e \in \text{集合「写真」},$
 $S1'(a) \wedge S2'(b) \wedge S3'(c) \wedge S4'(d,e)$

ここで

$S1'(X) = A1(X)$

$S2'(X) = S1'(X) \wedge A2(X)$

$S3'(X) = S1'(X) \wedge A3(X)$

$S4'(X,Y) = S2'(X) \wedge S3'(Y) \wedge A4(X,Y)$

A1(X): 太郎は正月に、自分でとった X を友達に送っている。

A2(X): 一昨年の X は電車の中で漫画を読んでいるサラリーマンの写真であった。

A3(X): 昨年の X は予備校生についてであった。

A4(X,Y): X と Y のどちらも面白かった。

・・・(*3)

(*3)において、

- ・変数 X,Y は集合「写真」の要素を表す変数である。
- ・ $Ai(X)$ は日本語文 Si 中の「写真」を変数 X で置き換え、指示詞を削除したものである。
- ・ $Si'(X)$ は、X が日本語文 Si を満たす要素であることを表す。
- ・ $Si'(X)$ の式、たとえば $S2'(X) = S1'(X) \wedge A2(X)$ は、日本語文 S2 を満たす写真の要素 X が、S1 中の写真が表す要素であり（すなわち、X が文 S1 を満たしており）、かつ、S2 の指示詞を除く部分の説明を満たす写真の要素であることを表す。
- ・式：

$\exists a, \exists b, \exists c, \exists d, \exists e \in \text{集合「写真」},$
 $S1'(a) \wedge S2'(b) \wedge S3'(c) \wedge S4'(d,e)$

は、各文の中の「写真」が集合「写真」の中のある要素であることを表す。

4.4 存在限量子の書き換え

4.3 節の(*3)では、変数 X が集合「写真」のある要素であることが、一目見ただけでは分かりにくい。そのため、変数 X を写真[I]と記す。ここで、N を普通名詞として、 $N[I]$ は集合「N」のある要素を表す。i=j のとき、 $N[i] = N[j]$ である。この記法を用いてテキスト(*3)を書き換えたものを次に示す。

$S1'(\text{写真}[I]) \wedge S2'(\text{写真}[2]) \wedge S3'(\text{写真}[3])$
 $\wedge S4'(\text{写真}[4], \text{写真}[5])$

ここで

$S1'(\text{写真}[I]) = A1(\text{写真}[I])$

$S2'(\text{写真}[I]) = S1'(\text{写真}[I]) \wedge A2(\text{写真}[I])$

$S3'(\text{写真}[I]) = S1'(\text{写真}[I]) \wedge A3(\text{写真}[I])$

$S4'(\text{写真}[I], \text{写真}[J]) = S2'(\text{写真}[I]) \wedge S3'(\text{写真}[J])$
 $\wedge A4(\text{写真}[I], \text{写真}[J])$

A1(写真[I]): 太郎は正月に、自分でとった写真[I]を友達に送っている。

A2(写真[I]): 一昨年の写真[I]は電車の中で漫画を読んでいるサラリーマンの写真であった。

A3(写真[I]): 昨年の写真[I]は予備校生についてであった。

A4(写真[I], 写真[J]): 写真[I]と写真[J]のどちらも面白かった。

・・・(*4)

5. 簡単な形式的な文への分解

5.1 単文への分解

複雑な形をした日本語文を計算機で処理するのは容易ではない。そこで、複雑な日本語文を複数の単文（動詞を1つ含む文）に分解する。このとき合わせて、名詞を修飾する句や節は、処理が容易になるように、非修飾の形に書き直す。時制等の情報は、処理が難しいので、削除する。この単文への分解では元の日本語文のもつ情報の一部が失われる。故に、得られる単文の集合は日本語文の近似（緩和）になる。

4.4 節で得られたテキスト(*4)の中の日本語文 A1, A2, A3, A4 を単文へ分解すると次が得られる。

A1-1(写真[I]): 太郎は写真[I]を撮る。

A1-2(写真[I]): 太郎は正月に写真[I]を友達に送る。

A2-1: サラリーマン[2]が電車の中で漫画を読む。

A2-2(写真[I]): 写真[I]は一昨年のものである。

A2-3(写真[I]): 写真[I]はサラリーマン[2]についてである。

- A3-1(写真[I]): 写真[I]は昨年のものである.
 A3-2(写真[I]): 写真[I]は予備校生についてである.
 A4-1(写真[I], 写真[J]) (= A4(写真[I], 写真[J])) :
 写真[I]と写真[J]のどちらも面白い.

・・・(*5)

ここで,

- ・ A1-1 と A1-2 は A1 から, A2-1 と A2-2 と A2-3 は A2 から, A3-1 と A3-2 は A3 から得られる単文である.
- ・ A2 の分解により, 普通名詞「サラリーマン」が 2 つの単文 A2-1 と A2-3 に現れるようになった. このため, 両者のサラリーマンが同じ要素であること表すために, サラリーマン[2]と記した.

5.2 単文への文法構造の付加

日本文の文法構造 (主語, 動詞, 目的語等) を利用すると, 文を柔軟に検索することができる. 文法構造は計算機による形態素解析, 構文解析により求めることができるが, 正しい解析が得られない場合もある. そこで, 本稿の形式化テキスト検索法では, 5.1 節のやり方で得られた各単文に対して, その文法構造を人手で付加する. 品詞を次のように略記する.

主:主語, 動:動詞, 目:目的語, 直目:直接目的語,
 間目:間接目的語, 補:補語, 他:その他

例えば, (*5)の単文 A1-2(写真[I]) :

太郎は正月に写真[I]を友達に送る.

に対して, その文法構造を付加して, 文 B1-2(写真[I]) :

{(主,太郎,は),(動,送る),(直目,写真[I],を),(間目,友達,に),(他,正月,に)}

を作る. このように作られた文は, 単文を組(品詞, 主要語, 付属語)の集合{(品詞, 主要語, 付属語),...}として表す. 5.1 節の単文(*5)を文法構造付きの文に書き換えたものを次に示す.

- B1-1(写真[I]): {(主,太郎,は),(動,撮る),(目,写真[I],を)}
 B1-2(写真[I]): {(主,太郎,は),(動,送る),(直目, 写真[I],を),(間目,友達,に),(他,正月,に)}
 B2-1: {(主,サラリーマン[2],が),(動,読む),(目,漫画,を),(他,電車,の中で)}
 B2-2(写真[I]): {(主, 写真[i],は),(動,である),(補,一昨年,の)}
 B2-3(写真[I]): {(主, 写真[i],は),(動,である),(補,サラリーマン[2],についての)}を表す.
 B3-1(写真[I]): {(主,写真[i],は),(動,である),(補,昨年,の)}
 B3-2(写真[I]): {(主,写真[i],は),(動,である),(補,予備校生,についての)}

- B4-1(写真[I], 写真[J]): {(主, 写真[I] and 写真[J],は),(動,である),(補,面白い)}

・・・(*6)

ここで, 各 Bi-j は(*5)の Ai-j を書き換えたものである.

6. 形式化テキスト

日本語テキストに対して 4 節と 5 節の書き換えを行って最終的に得られる各式を形式化文と呼び, その集合を形式化テキストと呼ぶ. 3 節の日本語テキスト(*1)から得られる形式化テキストを次に示す.

T1(写真[I]) ∧ T2(写真[2]) ∧ T3(写真[3])

∧ T4(写真[4], 写真[5])

T1(写真[I]) = B1-1(写真[I]) ∧ B1-2(写真[I])

T2(写真[I]) = T1(写真[I]) ∧ B2-1 ∧ B2-2(写真[I]) ∧ B2-3(写真[I])

T3(写真[I]) = T1(写真[I]) ∧ B3-1 ∧ B3-2(写真[I]) ∧ B3-3(写真[I])

T4(写真[I], 写真[J]) = T2(写真[I]) ∧ T3(写真[J]) ∧ B4-1(写真[I], 写真[J])

B1-1(写真[I]):

...

B4-1(写真[I], 写真[J]):

} (*6)と同じ

・・・(*7)

ここで, T1, T2, T3, T4 は, 各々, 4.4 節の(*4)の S1', S2', S3', S4'の右辺の Ai を Bi-1 ∧ Bi-2 ∧ ... に置き換え, 名前 Si' を Ti に変更したものである.

7. 形式化テキストに対する質問処理

7.1 形式化テキストの基礎的意味

形式化テキストに対する質問処理を考えるには, 形式化テキストが表していることを, より簡単な形で表せると便利である. 6 節の形式化テキスト(*7)について考える. 例えば, 日本語文 S2 から作られた T2(写真[2])は, 式

$$T2(写真[I]) = T1(写真[I]) \wedge B2-1 \wedge B2-2(写真[I]) \wedge B2-3(写真[I])$$

を使って,

$$T1(写真[2]) \wedge B2-1 \wedge B2-2(写真[2]) \wedge B2-3(写真[2])$$

に展開することができ, この中の T1(写真[2])は, 式

$$T1(写真[I]) = B1-1(写真[I]) \wedge B1-2(写真[I])$$

を使って,

$$B1-1(写真[2]) \wedge B1-2(写真[2])$$

に展開することができる。B1-1(写真[2])と B1-2 (写真[2])は展開することはできない。T2(写真[2])は、このように、

$$B1-1(\text{写真}[2]) \wedge B1-2(\text{写真}[2]) \wedge B2-1 \wedge B2-2(\text{写真}[2]) \\ \wedge B2-3(\text{写真}[2])$$

に展開することができる。各 $T_i(\text{写真}[i])$ からこのようにして得られる各式 B_{i-j} を基礎的形式化文と呼ぶ。基礎的形式化文は変数を含まない。全ての基礎的形式化文の集合を形式化テキスト(*7)の基礎的意味と呼ぶ。

形式化テキスト(*7)の基礎的意味を次に示す。

●T1(写真[1]) から得られる基礎的形式化文：

B1-1(写真[1]): {(主,太郎,は),(動,撮る),(目,写真[1],を)}

B1-2(写真[1]): {(主,太郎,は),(動,送る),(直目,写真[1],を),(間目,友達,に),(他,正月,に)}

●T2(写真[2]) から得られる基礎的形式化文：

B1-1(写真[2]): {(主,太郎,は),(動,撮る),(目,写真[2],を)}

B1-2(写真[2]): {(主,太郎,は),(動,送る),(直目,写真[2],を),(間目,友達,に),(他,正月,に)}

B2-1: {(主,サラリーマン[2],が),(動,読む),(目,漫画,を),(他,電車,の中で)}

B2-2(写真[2]): {(主,写真[2],は),(動,である),(補,一昨年,の)}

B2-3(写真[2]): {(主,写真[2],は),(動,である),(補,サラリーマン[2],についての)}

●T3(写真[3]) から得られる基礎的形式化文：

B1-1(写真[3]): {(主,太郎,は),(動,撮る),(目,写真[3],を)}

B1-2(写真[3]): {(主,太郎,は),(動,送る),(直目,写真[3],を),(間目,友達,に),(他,正月,に)}

B3-1(写真[3]): {(主,写真[3],は),(動,である),(補,昨年,の)}

B3-2(写真[3]): {(主,写真[3],は),(動,である),(補,予備校生,についての)}

●T4(写真[4], 写真[5])から得られる基礎的形式化文：

B1-1(写真[4]): {(主,太郎,は),(動,撮る),(目,写真[4],を)}

B1-2(写真[4]): {(主,太郎,は),(動,送る),(直目,写真[4],を),(間目,友達,に),(他,正月,に)}

B2-1: {(主,サラリーマン[2],が),(動,読む),(目,漫画,を),(他,電車,の中で)}

B2-2(写真[4]): {(主,写真[4],は),(動,ものである),(補,一昨年,の)}

B2-3(写真[4]): {(主,写真[4],は),(動,である),(補,サラリーマン[2],についての)}

B1-1(写真[5]): {(主,太郎,は),(動,撮る),(目,写真[5],を)}

B1-2(写真[5]): {(主,太郎,は),(動,送る),(直目,写真[5],を),(間目,友達,に),(他,正月,に)}

B3-1(写真[5]): {(主,写真[5],は),(動,である),(補,昨年,の)}

B3-2(写真[5]): {(主,写真[5],は),(動,である),(補,予備校生,についての)}

B4-1(写真[4], 写真[5]): {(主,写真[4] and 写真[5],は),(動,である),(補,面白い)}

・・・(*8)

7.2 形式化質問文の質問処理

検索者は日本語の質問を形式化質問文に書き換えて、計算機に質問を行う。形式化質問文は、これまで説明した、日本語文から形式化文を作る仕方とほぼ同じやり方で作ることができる。一般に、1つの日本語質問文は複数の形式化質問文として表される。計算機は複数の形式化質問文が与えられると、その全てが形式化テキストの基礎的意味の中にある基礎的形式化文と一致するかどうかを調べ、一致するときは「はい」を、1つでも一致しないときは「いいえ」を検索者に返す、以下に、3節にあげた日本語質問文 Q1, Q2, Q3, Q4, Q5 を使って詳しく説明する。

(1) Q1: 太郎は正月に友達に写真を送りますか?

・形式化質問文

R1-1 (太郎は正月に友達に写真を送る)：

{(主,太郎,は),(動,送る),(直目,写真,を),(間目,友達,に),(他,正月,に)}?

・質問処理

R1-1 は形式化テキストの基礎的意味(*8)の中の例えば B1-2(写真[1]):

{(主,太郎,は),(動,送る),(直目,写真[2],を),(間目,友達,に),(他,正月,に)}

に一致する。ここで、B1-2 の中の写真[2]の[2]は無視する。R1-1 が B1-2 に一致するので、計算機は「はい」を検索者に返す。

(2) Q2: 太郎は写真を誰に送りますか?

・形式化質問文

R2-1 (太郎は写真を X? に送る)：

{(主,太郎,は),(動,送る),(直目,写真,を),(間目,X?,に)}?

ここで、X?は変数への代入を尋ねている。

・質問処理

R2-1 は、たとえば、X=友達として、(*8)の中の B1-2(写真[1]):

{(主,太郎,は),(動,送る),(直目,写真[1],を),(間目,友達,に),(他,正月,に)}

に一致する。ここで、B1-2 の中には R2-1 にない項(他,正月,に)があるが、R2-1 のすべての項(主,太郎,は),…(間目,X?,に) が B1-2(写真[1])の中に含まれているので、R2-1 は B1-2(写真[1])に一致すると考える。計算機は「X=友達」を検索者に返す。なお、R2-1 は B1-2(写真[2]), B1-2(写真[3]), B1-2(写真[4]), B1-2(写真[5])とも一致するが、「X=友達」

以外の答はない。

(3) Q3: 太郎は漫画を読んでいるサラリーマンの写真を友達に送りましたか?

・形式化質問文

R3-1 (サラリーマン [I]が漫画を読む) :

{{(主, サラリーマン[I],が),(動,読む),(目,漫画,を)}}?

R3-2 (写真[J]はサラリーマン[I]についてである) :

{{(主, 写真[J],は),(動,である),(補,サラリーマン[I],についての)}}?

R3-3 (太郎は写真[J]を友達に送る) :

{{(主,太郎,は),(動,送る),(直目,写真[J],を),(間目,友達,に)}}?

ここで, R3-1 のサラリーマン[I]と R3-2 のサラリーマン[I]はIの値が同じであることを, R3-2 の写真[J]と R3-3 の写真[J]はJの値が同じであることを要求している。

・質問処理

例えば, サラリーマン[I]=サラリーマン[2], 写真[J]=写真[2]とすると, R3-1 が(*8)の中の B2-1:

{{(主,サラリーマン[2],が),(動,読む),(目,漫画,を),(他,電車,の中で)}}?

に, R3-2 が B2-3(写真[2]):

{{(主, 写真[2],は),(動,である),(補,サラリーマン[2],についての)}}?

に, R3-3 が B1-2(写真[2]) :

{{(主,太郎,は),(動,送る),(直目, 写真[2],を),(間目,友達,に),(他,正月,に)}}?

に一致する。故に「はい」を返す。この回答は, 指示詞の機能を利用して2つの日本語文 S1 と S2 の情報から作られている。なお, R3-1, R3-2, R3-3 は他の基礎的形式化文とも一致する。

(4) Q4: サラリーマンの予備校生の写真はありますか?

・形式化質問文

R4-1 (写真[I]はサラリーマンについてである) :

{{(主, 写真[I] ,は),(動,である),(補,サラリーマン,についての)}}?

R4-2 (写真[I]は予備校生についてである) :

{{(主,写真[I],は),(動,である),(補, 予備校生,についての)}}?

・質問処理

(*8)の中に R4-1 と R4-2 の両方に一致する基礎的形式化文はない。故に, 「いいえ」を返す。

(5) Q5: 何が面白いのですか?

・形式化質問文

R5-1 (X?は面白い) :

{{(主, X? ,は),(動,である),(補,面白い)}}?

・質問処理

X=写真[4] and 写真[5]のとき, R5-1 は(*8)の中の B4-1(写真[4], 写真[5]):

{{(主, 写真[4] and 写真[5],は),(動,である),(補,面白い)}}?

に一致する。ゆえに「X=写真」を返す。もし「X=写真」についてより詳しい説明が必要ならば, 「X=写真[4] and 写真[5]」として, 写真[4]の補足説明として(*8)の中の B4-1(写真[4], 写真[5]), B1-1(写真[4]), B1-2(写真[4]), B2-1, B2-2(写真[4]), B2-3(写真[4])を, 写真[5]の補足説明として(*8)の中の B4-1(写真[4], 写真[5]), B1-1(写真[5]), B1-2(写真[5]), B3-1(写真[5]), B3-2(写真[5])を検索者に返す。前者は3節のQ5の答2に, 後者は答3に対応している。

8. 終わりに

日本語テキストから人手で形式化テキストを作成し, これを用いて日本語テキストを検索する方法(形式化テキスト検索法)を提案した。この検索方法は指示詞の機能を考慮しており, キーワードやタグを用いた従来の検索方法では答えられないような質問に答えることができる。今後, 効率的な実装方法を検討したい。また, より高い精度で質問に答えることができるように, 検索方法を改良したい。

参考文献

- 1) 鈴木晋, 文法構造を付加したテキストに対する情報検索方法の検討, 情報処理学会第74回全国大会講演論文集, pp.1-535--1-536, 2012年3月。
- 2) オールドウド, アンデルソン, ダール 著, 公平珠躬, 野家啓一 訳, 日常言語の論理学, 産業図書, 1979年9月。
- 3) 長尾真, 佐藤理史, 黒橋禎夫, 角田達彦, 自然言語処理, 岩波書店, 1996年4月。
- 4) 奥村学, 難波秀嗣, テキスト自動要約, オーム社, 2005年3月。
- 5) 磯崎秀樹, 東中竜一郎, 永田昌明, 加藤恒昭, 質問応答システム, コロナ社, 2009年8月。
- 6) トビー・セガラン, コリン・エバンス, ジェイミー・テイラー 著, 玉川竜司 訳, セマンティック Web プログラミング, オライリー・ジャパン, 2010年6月。