

検索クエリの理解に基づいた数値回答文の生成

Generation of numerical answer sentences based on an understanding of search queries

上原 良太^{*1}, 當間愛晃^{*2}
 Ryota UEHARA^{*1}, Naruaki TOMA^{*2}
 琉球大学 工学部 情報工学科

Faculty of Engineering Department of information Engineering, University of the Ryukyus^{*1}
 Email: e165716@ie.u-ryukyu.ac.jp

あらまし：web情報から金額情報を検索する場合、検索クエリに対して数値情報を付与する事がある。しかし、この数値情報が扱われる事は少なく、大抵は無視された結果が出力される。この出力結果は、ユーザが求めた本来の結果ではない。この事を解決するために検索クエリの数値情報と金額情報の関係性を理解し、算出演算による結果を出力する手法の提案について報告する。

キーワード：オントロジー、MeCab、CaboCha、数値回答

1. 研究背景・研究目的

近年、スマートフォンの普及によりインターネットを利用した情報検索が一般的になってきた。情報検索では数多ある情報からの的確な情報のみを引き出す能力が要求されるが、こと金額情報の検索においては検索結果に対して数値情報の加工を行う必要がある。例えば「ある商品を複数個買う値段」を検索した場合、結果としてweb上に掲載されている「商品の値段」が出力され「複数個」の情報が欠落されるだろう。この結果は、検索クエリの変数的な数値情報を扱うことが出来てないためだと考えられる。

本研究では、この問題を解決するためのアプローチとして、検索クエリ中の数値情報と検索結果である金額情報の関係性を理解し、算出演算を行うことで正しい結果として出力することが目的である。

2. 関連知識

2.1 オントロジー

人工知能分野では“概念化の明示的記述”⁽¹⁾と定義され、オントロジー工学の立場では“人間が対象世界をどのように見ているかという根源的な問題意識を持って物事をその成り立ちから解き明かし、それをコンピュータと人間が理解を共有できるように書き記したもの”⁽²⁾と定義される。

2.2 MeCab / CaboCha

MeCab⁽³⁾は形態素解析エンジンの1つである。形態素解析とは、文法や辞書の情報に基づき、意味を持つ表現要素の最小単位である形態素ごとに文を分割し、品詞情報などを出力する事である。

CaboCha⁽⁴⁾は係り受け解析器の1つである。係り受け解析とは、文章を形態素に分割し、文節ごとの修飾関係を解析する事である。

3. 研究設計・研究結果

本研究の目的は1節でも述べている通り、検索クエリと検索結果の関係性を理解し、算出演算を行う事であるが、ここでの検索結果とは実際のホーム

ページに掲載されている情報を受け取れるものとして、自前で用意した。データ形式としては、「文字列：数値」の辞書型による出力とする。この検索結果を出力する検索アルゴリズムの提案は、本稿では行わない。

web情報の検索範囲として、本研究では「琉球大学」の情報を取り扱う事にした。また、琉球大学内の金額情報である「学費情報」に対して、想定される検索クエリを受け取った場合に、クエリと検索結果の理解に基づいた数値回答文の生成を行う手法の提案を試みる。

3.1 専用オントロジーの作成

検索クエリの理解に対し、オントロジーを利用する。2節のオントロジーの定義より、語彙の概念を明示し、コンピュータでの理解に効果的であると考えたからである。そこで、オントロジーを作成する上での事前準備として、琉球大学公式ホームページ⁽⁵⁾での「学費」情報を図1に示す。

| 区 分 | 授業料(年額) | 授業料(前期) | 授業料(後期) | 入 学 料 | 検 定 料 |
|--------------------|---------|---------|---------|---------|--------|
| 学部(昼) | 535,800 | 267,900 | 267,900 | 282,000 | 17,000 |
| 学部(夜間主コース) | 267,900 | 133,950 | 133,950 | 141,000 | 10,000 |
| 特別支援教育特別課程 | 273,900 | 136,950 | 136,950 | 58,400 | 16,500 |
| 大学院研究科 | 535,800 | 267,900 | 267,900 | 282,000 | 30,000 |
| 法務研究科(*) | 804,000 | 402,000 | 402,000 | 282,000 | 30,000 |
| 転学・編入学・再入学(昼) | 535,800 | 267,900 | 267,900 | 282,000 | 30,000 |
| 転学・編入学・再入学(夜間主コース) | 267,900 | 133,950 | 133,950 | 141,000 | 18,000 |

図1 琉球大学の学費情報

図1より、学費には「授業料、入学料、検定料」が含まれる。特に授業料には「年額、前期、後期」と情報が複数あり、結果として受け取った場合に語彙のばらつきが起きる。また、一般的な理解として「入学料、検定料」は1度だけの支払い、授業料は年間ごとに支払う必要がある。これらの関係性を考慮し、学費情報に対して特化した琉球大学オントロジーの作成を行った。作成したオントロジーを図2で示す。

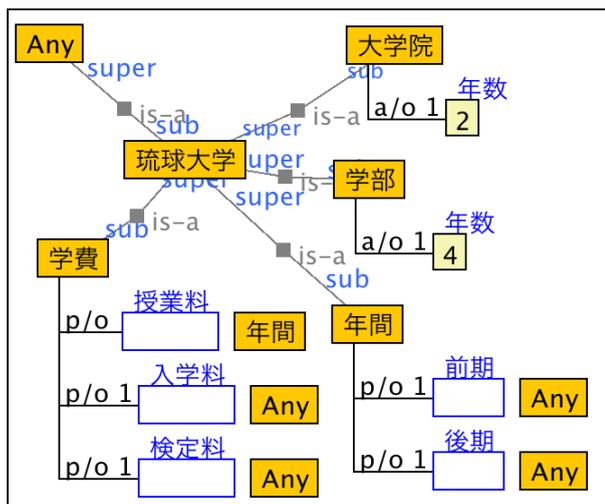


図2 琉球大学オントロジー

図2でのオントロジーは、オントロジーエディタである法造⁶⁾を用いて作成したものである。オントロジーの要素として「概念クラス」とそれらをつなぐ「意味リンク」で構成されている。「意味リンク」が表す関係や、その概念を表1に示す。

表1 意味リンクによって示される関係・概念

| 関係 | 概念 | 図中表現 |
|----------------|---------|-------------|
| is - a | 上位 / 下位 | super / sub |
| part - of | 部位 | p / o |
| attribute - of | 属性 | a / o |

「意味リンク」の隣にあたる数字は個数表現を表している。事前準備によって、「入学料、検定料」の支払いは一度のみ、「授業料」は「年間」により変化する変数属性を持った数値情報であると想定したため、明確な個数を示すのではなく「年間」概念を付与する形でオントロジーを形成した。また「学部、大学院」が持つ「年数属性」には予め標準値になり得る値を入力しておくことで、学費情報に特化した琉球大学オントロジーの完成とした。

3.2 検索クエリから必要情報の抽出

検索クエリで「琉球大学の4年間の学費はいくらか」と受け取った場合に、クエリ中の必要情報として「4年間」の数値情報がどの形態素に係り受けしているかを抽出する必要がある。そこでCaboChaによる係り受け解析を用いて「4年間」に係る情報として「学費」を抽出した。また、「4年間」の情報に対して品詞情報である「数詞」と結びつく「助数詞」をそれぞれ取得する必要がある。そこでMeCabによる形態素解析で、クエリ中から品詞情報の抽出を行った。

3.3 オントロジーの適応

3節の冒頭でも記載したが、今回の実験では予め

web検索による想定された結果が取得できるものと考えている。よってここでは、作成した琉球大学オントロジーの「概念クラス」に対して「検索結果」を適応させていく。適応というのは、図1での学費情報からオントロジーに適した金額情報を取得し、オントロジーの概念に対して値を入力することである。今回は「授業料」の情報が「年額、前期、後期」と記載がぶれているので、「授業料」に付与してある「年間」概念の持つ「前期、後期」を参照することで、「前期、後期」の値を含め、「年間」の値として記載のぶれに対応した。

3.4 検索結果を適した値に算出

3.2節で抽出した係り受け情報と品詞情報、3.3節での「検索結果」を適応させたオントロジーをそれぞれ用いて算出を行う。「学費」概念の中で、個数表現が「1」を持つ「入学料、検定料」はその値のまま加算を行う。個数表現を持たない「授業料」は代わりに付与されている「年間」を参照し、「年間」と結びつく数詞である「4」を利用しての乗算を行った上での加算を行い、「4年間の学費」を算出することとした。

4. まとめと今後の課題

本稿では、検索クエリの理解に基づいた数値回答文の生成に対して、オントロジーを用いた実験による解決手法の提案を行い、想定した条件下での結果を示した。結果より限られた条件下ではあるが、実際に検索結果に対して算出を行い、クエリ通りの結果を出力する事が可能であるといえる。

今後の課題としては、検索クエリが形態素の羅列である場合に対応できていないため、適した自然文に変化させる事を考えている。またオントロジーの生成には知識と時間を有する為、web情報を取得してのオントロジーの半自動生成法についても研究していく。

参考文献

- (1) Thomas R. Gruber: "A Translation Approach to Portable Ontology Specifications", ["https://tomgruber.org/writing/ontologia-kaj-1993.pdf"](https://tomgruber.org/writing/ontologia-kaj-1993.pdf) (2020/1/15)
- (2) 溝口理一郎: "オントロジー工学", オーム社 (2005)
- (3) MeCab: "Yet Another Part-of-Speech and Morphological Analyzer", ["https://taku910.github.io/mecab/"](https://taku910.github.io/mecab/) (2020/1/15)
- (4) CaboCha: "Yet Another Japanese Dependency Structure Analyzer", ["https://taku910.github.io/cabocha/"](https://taku910.github.io/cabocha/) (2020/1/15)
- (5) 琉球大学公式ホームページ, ["https://www.u-ryukyu.ac.jp/"](https://www.u-ryukyu.ac.jp/) (2020/1/15)
- (6) 法造 - オントロジーエディタ, http://www.hozo.jp/index_ip.html (2020/1/15)