

ウェブサービス検索システムにおける 自然言語検索文とオントロジに基づく動的表示方式

酒井 大[†] 多田 政美[†] 森 亜矢[†] 豊島 浩文[‡]
日本アイ・ピー・エム株式会社 ソフトウェア開発研究所[‡]

1. はじめに

ネット上のデータ量が爆発的に増える一方で、ユーザが欲している検索結果をユーザの望む形式で表示する方式は、いまだ確立されていない。セマンティックウェブ技術を用いた情報検索システムにおいても、ユーザの目的に合致する検索結果の効果的な表示方式は、重要な研究テーマになっている。

本稿では、ウェブサービスから取得した検索データの自然言語検索文とオントロジ[1]に基づく動的表示方式を述べる。オントロジとは、ある領域における概念間の関係の定義の集まりであり、概念のクラス定義や、階層構造による体系化された知識等を表現する。

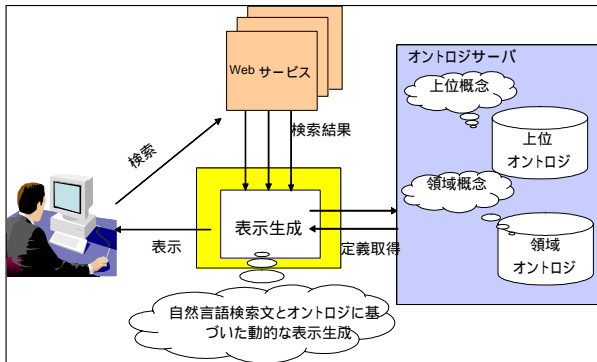


図1: オントロジに基づく動的表示方式の流れ

本方式のシステム構成を図1に示す。ユーザは、まず「安い腰掛便座が欲しい」のように自然言語で検索を行う。自然言語解析部で獲得された名詞（「腰掛便座」）を検索対象としてウェブサービスから検索データが収集される。表示生成部において、データはオントロジのクラスに対応したカテゴリとして分類され、検索対象を修飾する形容詞や形容動詞に応じてソートされた後、表としてまとめられる。このとき、オントロジサーバから概念間の関係の定義や概念を処理するための論理的なルールが取得される。このように検索結果に自然言語検索文の解析結果を適用することにより、ユーザの意図を反映したより見やすい表示を生成することができる。

2. 既存の表示方式

ユーザの意図を表示に反映する方法として、ユーザが行った操作の記録と設定に基づき表示を切り替える手法が存在する。例えば、オンラインショッピングなど様々な商業サイトで実現されている手法と

して、購買履歴をもとにユーザの興味のある新商品情報がトップ画面で表示される。また、ユーザがプロフィールに設定した嗜好をもとにデータの要約の有無を決定するサイトもある。しかし、この手法では、検索文に含まれる意図を表示に反映することが難しい。

3. 検索文に応じた動的表示方式

本方式は、自然言語検索文に応じた動的な分類表示方式と動的ソート表示方式によって構成される。

図2に、本研究で表示生成の対象とするデータの例を示す。対象とするデータはウェブサービスから返され、オントロジ記述言語の OWL[2]形式で構造化されているものとする。

```
<腰掛便座 rdf:ID="製品A">
  <定価 rdf:datatype="&xsd:positiveInteger">48000
</定価>
  <商品概要 rdf:datatype="&xsd:string">持ち運び可能。
  車輪付き</商品概要>
  <容量 rdf:datatype="&xsd:string">5リットル</容量>
</腰掛便座>
```

図2: OWL形式のウェブサービスのアウトプット例

3.1 検索文に応じた動的な分類表示方式

データは、検索文に応じてオントロジから探し出されたクラスをカテゴリとして分類される。例えば、介護オントロジにおいて「腰掛便座」の下位クラスに「ポータブルトイレ」、「補高便座」、「立ち上がり補助便座」と、「ポータブルトイレ」の下位クラスに「椅子型」、「組立型」があったとする（図3）。

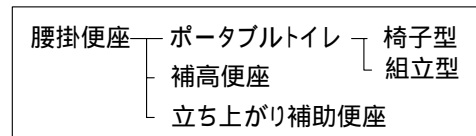


図3: 腰掛便座のクラス階層

ユーザが「腰掛便座が欲しい」と検索を行うと、「腰掛便座」を検索対象として獲得されたデータは、「腰掛便座」クラス（以下、分類クラス）の3つの下位クラスを候補として生成されたカテゴリに分類される。各データは、各クラス（図4）に含まれる語句とのキーワードマッチングによりカテゴリに振り分けられる（図5）。

検索対象を修飾する語句に応じてカテゴリは詳細化される。ユーザが「持ち運べる腰掛便座が欲しい」と検索したとき、検索データは、「ポータブルトイレ」クラスの下位クラスをカテゴリとして分類

Dynamic Display of the Web Service Search Results according to the Search Query

[†]Dai Sakai, Masami Tada, Aya Mori, Hirobumi Toyoshima

[‡]Yamato Software Development Laboratory, IBM Japan

される。名詞「腰掛便座」に応じて選ばれた3つのカテゴリ候補のうち「ポータブルトイレ」クラス(図4)が「持ち運べる」という語句と完全一致する特徴が定義されているので、分類クラスとなる。また、「持ち運べる」クラスに「持ち運び可能な」が類義語として定義されているので、「持ち運び可能な腰掛便座が欲しい」と検索されても同様の結果になる。

```
<owl:Class rdf:ID="腰掛便座" />
<owl:Class rdf:ID="ポータブルトイレ">
  <root:特徴 rdf:about="#持ち運べる" />
  <rdfs:subClassOf rdf:resource="#腰掛便座" />
</owl:Class>
<owl:Class rdf:ID="持ち運べる">
  <root:類義語>持ち運び可能な</root:類義語>
</owl:Class>
```

図4：「ポータブルトイレ」クラス

ポータブルトイレ

商品名称	定価	特徴	容量
製品A	48000	持ち運び可能 車輪付き	5リットル
製品B	58000	電池駆動	8.9L

立ち上がり補助便座

商品名称	定価	特徴	容量
製品D	98000	立ち上り時に自動的に上がり ます	---

補高便座

商品名称	定価	特徴	容量
製品E	28000	和式の上に設置可	---

図5：「腰掛便座が欲しい」の結果の例

このような仕組みにより、検索対象の詳細なカテゴリを表示することができるので、閲覧しやすくなる。また、「ポータブルトイレ」という名称を知らなくても「持ち運べる腰掛便座が欲しい」のように特徴を記述すれば該当データを見つけ出すことができる。

3.2 検索文に応じた動的ソート表示方式

データは、検索対象を修飾している形容詞や形容動詞に応じて動的にソートされ、表示される。例えば、ユーザが「赤い車が欲しい」と入力するとき、検索された車データは、「色」項目(表の列)において赤さ順にソートされる(図6)。また、「安価な車が欲しい」と検索されたときは、「価格」項目において低価格順にソートされる。

具体的には、表示生成においてまず形容詞と形容動詞が名詞に変換される。次に名詞に該当するクラス(あるいは、インスタンス)がオントロジから探し出される。それらから、データのソートを行うプロセス名とそのパラメータとなるソート対象項目(ソート項目)が取得される。例えば、「赤さ」インスタンスからは、RGBの値によってソートする「RGBソート」というプロセス名と、ソート項目である「色」が取得される。これにより検索データは、RGBソートプロセスによって「色」でソートされ表

示される。なお、ソート項目は表の最左列に表示される。

色	商品名称	価格	メーカー
赤	カーA	1,800,000	A社
ピンク	カーC	3,200,000	B社
オレンジ	カーD	698,000	C社

図6：「赤い車が欲しい」の表示の例

ソート項目が複数ある場合、ユーザにソート項目を選択してもらう。例えば「高い腰掛便座が欲しい」という検索文では、形容詞「高い」は「商品価格」なのか、「座高」なのか判断が付かない。このような場合は、検索文を入力し、結果を待っているユーザにソート項目の選択肢を表示する。

このような仕組みによって、ユーザの着眼点を反映した、データの比較をしやすい表示を生成する。

4. 結論

自然言語検索文とオントロジ技術に基づく検索結果の動的表示方式を研究開発した。図1で示したような検証システムを構築し、自然言語検索文に含まれるユーザの意図を反映した表示の生成が適切に行われることを検証した。

動的分類表示方式では、検索文に含まれる検索対象の名詞でカテゴリ分けするとともに、検索対象を修飾する語句に注目して、カテゴリ分けの詳細化を行った。また、動的ソート表示方式では、検索対象を修飾する形容詞または形容動詞に注目し、検索結果をユーザの意図に応じて並べ替えることを行った。

今回の研究では、検索対象を修飾する語句(特に形容詞)に注目したが、今後は、検索文に含まれる複雑な修飾節の取り扱いを課題とする。

謝辞

本研究は総務省の委託研究『ITを活用した企業間連携・起業支援プラットフォームに係わる要素技術に関する調査研究』の一部として実施したものである。研究の機会を与えて戴いたことに感謝する。

参考文献

- [1] 日高由布子, 榎谷一: オントロジを用いた情報配信システムの研究開発, FIT(情報科学技術フォーラム), 2002
- [2] OWL Web Ontology Language Overview, <http://www.w3.org/TR/owl-features/>