

## プログラミングヘルプシステムにおける 複数ヘルプコンテンツ間の横断的検索機構について

藤巻伸洋 永田雄大 大園忠親 新谷虎松

名古屋工業大学 知能情報システム学科

e-mail: {fujimaki, yuta, ozono, tora}@ics.nitech.ac.jp

### 1 はじめに

本論文では、プログラミングを支援に関連して、プログラミングヘルプシステムにおける、複数ヘルプコンテンツ間の横断的検索機構を提案する。プログラマーは、プログラミング中に、わからないことがあると、使用している言語のマニュアル(仕様書)や、電子掲示板などに蓄積されている Tips など参考にするために、ヘルプシステムなどを利用する。マニュアルも、Tips も、自然言語で記述されたテキストであるので、本研究で実装するプログラミングヘルプシステムはテキスト検索システムをベースに実装する。ここでは、システムのユーザは、知りたい事柄を検索質問として表現し、システムに入力することで、ヘルプのためのコンテンツを検索結果として手に入れる。

一般的には与えられた検索質問を、すべての検索対象データベース(マニュアル、電子掲示板など)に対して入力し、その結果を統合するという手段がとられている。しかし、我々が対象とするヘルプコンテンツにはその種類に応じた特徴があり、その特徴に適した検索質問が必要となる。また、ユーザは足りない知識を補うためにシステムを利用しているので、目的のコンテンツが適切に選択できる検索質問を入力することは難しい。

本研究では、与えられた検索質問を、対応するデータベースに適する検索質問に修正することを目的とする。その修正に、テキストの言い換え技術を利用する。これにより、単純な横断検索に比べ、よりユーザの検索質問の意図を反映したヘルプ結果を得ることを可能にする。

### 2 複数コンテンツ間の横断的検索機構

本研究で検索対象とするヘルプコンテンツの種類とその特徴を以下に示す。

- プログラミング言語のマニュアル：プログラミング言語の開発者によって言語の仕様や利用可能な

API などリファレンスが記述されたテキスト。テキストの記述者が開発者に限られているため、同一概念が異なる表現で記述されていることが少ない。

- Tips：マニュアルには記述されていないプログラミングするのに便利な技法やコツが記述されたテキスト。本システムではユーザや開発者から電子掲示板を通して収集することを想定しているため、テキストの記述者が限定されず、表記の揺れが発生しやすい。
- サンプルコード：なんらかの処理をするためのプログラムの見本として記述されたソースコード。ソースコード中に自然言語でコメント文が記述されている可能性がある。

ここで、我々は、ヘルプコンテンツにおけるテキストの記述者の違いから生じる表記の揺れの問題に着目した。表記の揺れの問題とは、同一概念を表す複数の表現が存在するため、テキストが一致するかどうかを、表現の一致のみで判定してしまうと、漏れが発生するという問題である。この問題を解決するために、我々は「文の意味を保存したまま別の文に言い換える」言い換えの技術を利用する。また、ユーザの意図をより正確に表現するためには、検索質問を語の出現の有無のみで表すことは不十分であると考えた。そこで検索質問に含まれる語(文節)の係り受け関係を利用することで、よりユーザの意図を検索に反映する。言い換えについては2.1節で、係り受け検索については2.2節で詳細に述べる。

#### 2.1 言い換えに基づく検索質問の修正

表記の揺れの問題に対して、我々は検索質問を言い換えることで、与えられた検索質問と同義の表現を含むコンテンツの検索漏れを防ぐ。検索質問の言い換えには言い換えの中でも比較的实现が容易とされている構文的言い換えを用いる。構文的言い換えとは、構造のマッピングに基づいて、単語の同義語、類義語への置換や構造の変形を行う。本研究では、互いに言い換え可能な表現の集合を辞書として管理し、それによって言い換え処理を行う。辞書の作成は人手で行っている。

On Cross Retrieval Mechanism from Multiple Databases for Programming Help

Nobuhiro FUJIMAKI, Yuta NAGATA, Tadachika OZONO, Toramatsu SHINTANI

Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya 466-8555 JAPAN

言い換えは、検索対象に応じて処理を行う。マニュアルがテキストとして一貫性のあるものに対し、Tips や (サンプルコードの) コメントが表記の揺れを多く含むものであることから、

- マニュアル検索では検索質問の表記の揺れをマニュアルで使われている表現に縮退するように言い換える
- Tips (サンプルコード) 検索では可能な限り表現を拡張するように言い換える

ことで、表記の揺れの問題を解決する。マニュアルで使われている表現は、言い換え辞書の中で指定する。

## 2.2 係り受け検索

係り受け関係とは、日本語における文節間の依存関係を表したものである。そもそも日本語では、述語に対する動作主体や動作対象などの役割を「が」「を」「に」などの格助詞によって指定している。この役割を格という。そして、それぞれの格と述語との関係を格関係といい、その格関係に基づいて、文節間の依存関係を表したものが係り受け関係である。二つの文節が修飾、被修飾の係り受け関係にあるとき、文頭側の文節が文末側の文節に「係る」といい、文末側の文節は文頭側の文節を「受ける」という。

係り受け関係を利用する場合、語の出現情報のみを利用する場合と比較して、文中での語の関係をより正確に表現することができる。例えば、「プログラムは短い、実行時間は長い」という文を語の出現のみで考える場合、〈プログラム、短い、実行時間、長い〉の4つの単語が共起するという情報が得られる。しかし、この文に対して「実行時間が短い」〈実行時間、短い〉という検索質問が与えられた場合、明らかに逆の意味にもかかわらず、検索結果として選択してしまう。係り受け関係を利用することで、元の文を〈プログラム、短い〉、〈実行時間、長い〉、〈短い、長い〉といった対(本論文では係り受け対と呼ぶ)の集合として表現できるので、先の検索質問の対象から外すことができる。

本機構では、係り受け対を索引として利用することで、テキストの係り受け検索を実現する。

## 3 プログラミングヘルプシステムへの適用

我々は、論理型言語 MiLog[Fukuta 00] の為の知的ヘルプシステム [永田 03] に提案する横断検索機構を適用した。本システムは、言語開発者によって用意されたマニュアルと、MiLog に関するディスカッションが行われている電子掲示板に記述されたテキストをヘルプコンテンツとする。図 1 では、本システムに、提案する

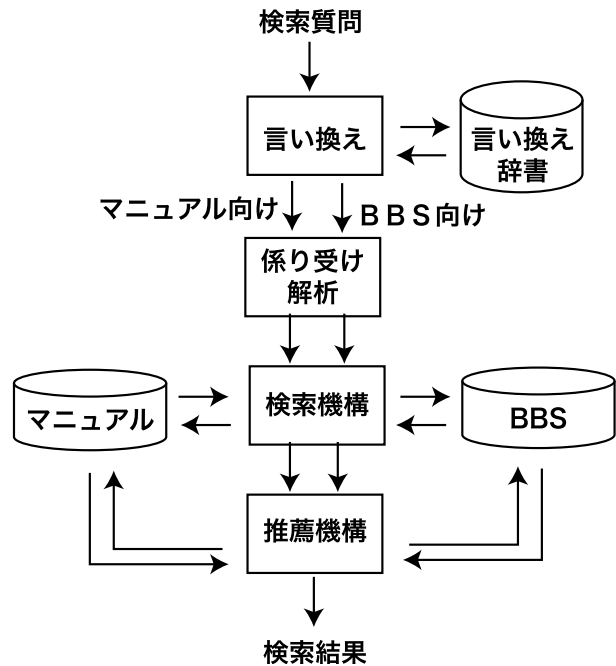


図 1: 知的ヘルプシステムにおける横断検索機構  
横断検索機構を適用した構成とその処理の流れを示している。与えられた検索質問は、言い換えモジュールによって各データベースに合わせて拡張(縮退)される。修正された検索質問は係り受け解析され、あらかじめ係り受け対で索引付けされた各種コンテンツデータベースを検索する。係り受け解析には Cabocha<sup>1</sup> を利用している。最後に、検索の結果得られたコンテンツには、本システムの推薦機構によって関連するコンテンツが付加され、ユーザに提示される。

## 4 おわりに

本論文では、プログラミングヘルプシステムにおける複数ヘルプコンテンツ間の横断検索機構について提案し、その実装を示した。ヘルプコンテンツに適した索引付けの手法と検索質問の拡張方法を示した。

## 参考文献

- [Fukuta 00] Naoki Fukuta, Takayuki Ito, and Toramatsu Shintani: “MiLog: A Mobile Agent Framework for Implementing Intelligent Information Agents with Logic Programming”, PRIIA2000, pp.113-123, (2000).
- [永田 03] 永田雄大, 藤巻伸洋, 大園忠親, 新谷虎松: “論理型言語 MiLog における知的ヘルプシステムの試作について”, 第 65 回情報処理学会全国大会論文集, (2003). (講演予定)

<sup>1</sup> <http://cl.aist-nara.ac.jp/taku-ku/software/cabocha/>