

マニュアル文書の自然な検索の実現

5G-2

余田直之、茂木健、湯村武、野崎康夫、西田行輝

三洋電機(株) 情報通信システム研究所

1. はじめに

現在、我々は、テキストデータベースの検索に関する研究を進めており、その一環としてマニュアル文書の検索手法についての考察を行ったので報告する。

近年、電子機器・システムの大規模化・高機能化に伴い、わかりやすいマニュアル(取扱説明書)が重要視されてきており、ソフトウェアシステムのマニュアルについては、その多くが電子化され、キーワード検索やハイパーテキスト的な機能を持つものも多くなっている。しかし、マニュアル文書およびその検索手段がユーザのあらゆる要求を満足させることは非常に難しい。例えば、ワープロなど情報処理機器のマニュアルはコマンドの機能を調べるのには好都合であるが、実際の作業においては、しばしば、特定の目的を実現するためには複数の機能をどのような手順で用いれば良いかが要求される。

このような要求を満たすためには、従来のように各機能の定義文を出力するだけではなく、目的を達成するための一連の作業の中でこれらの機能をどのように用いればよいかを回答として出力する必要がある。すなわち、ユーザの持つ検索要求をいかに直接的かつ正確に実現するかということが問題となる。

そこで、本稿では、上述したような作業を行うための一般的な検索要求に対して、マニュアル文書から該当する箇所を抽出し、抽出された単位をまとめあげて回答を出力するための一考察について述べる。

2. マニュアル検索システムの基本的な考え方

通常、検索システムにおける検索の流れは図1のように表すことができる。

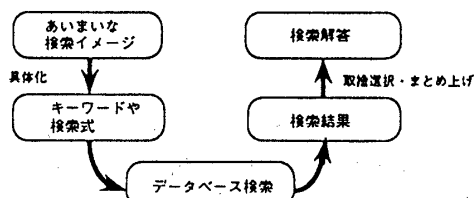


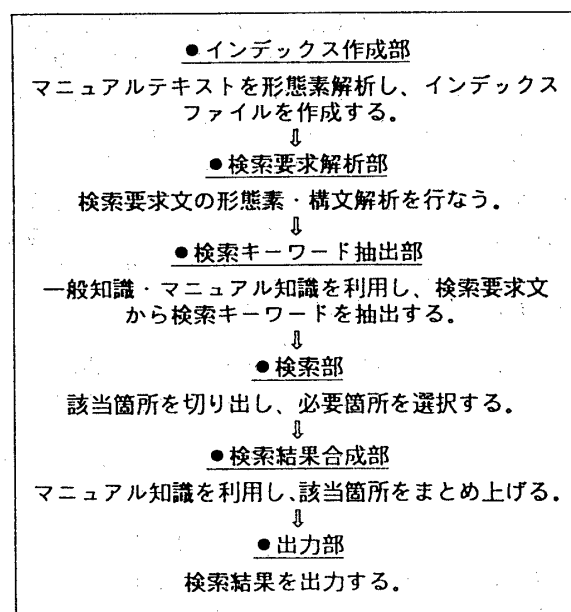
図1 検索の流れ

図のようなシステムにおいて、ここで目的とする「目的を達成するための一連の作業の中で、どのような機能を用いればよいか」といったようなユーザの検索要求を満足させるためには、次に挙げるような特徴を備えたマニュアル検索システムが構築されることが望ましいと考える。

- ユーザの検索イメージと検索キーワードの間、並びに、検索結果と検索回答の間に存在するギャップをマニュアル特有の特別な知識を用いて埋めることができる。
- 通常、マニュアルはある一定の観点から書かれるが、リファレンスマニュアルから手続き的な情報を、あるいは、手続きマニュアルからリファレンス的な情報を得ることが可能である。
- 検索要求に関連し、複数箇所に書かれている記述を順序だててまとめて提示できる。
- どのような状況のもとで検索要求が発せられたのかを考慮して検索要求を解釈したり、検索回答を提示できる。

3. 検索システムの概要

以下に、マニュアル検索システムの概要について述べる。本マニュアル検索システムは、日本語マニュアルテキストを対象とし、自然言語による検索要求の入力を想定しており、次に挙げるモジュールから成り立つ。



4. 検索システムの動作例

次に、実際に機械翻訳システムのマニュアル文書を用いて行なった検索例を示す。

検索要求文：「この文書を翻訳して、日本語と英語が対になった形で出力したい。」

まず、上記検索要求文の形態素・構文解析を行ない、キー要素（図中の円で囲まれた部分）を抽出する。

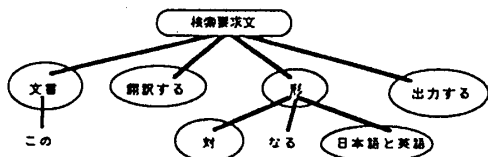


図2 検索要求の解析結果

そして、まず一般知識、さらに、マニュアル知識を利用して、検索キーワードへの展開を行なう。ここで一般知識は、キー概念辞書並びに類義語辞書により実現されている。また、マニュアル知識は、図3のように、一般知識と関連付けられた形で登録されている、対象とするマニュアルにのみ特有の知識である。

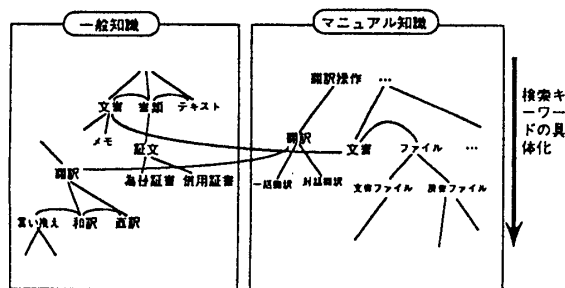


図3 知識データ

さて、上述の検索要求文においては、「文書」「翻訳」「出力」のように、マニュアルではかなり広い意味において使われている語彙が出現する。そこで、図に示されるような一般知識やマニュアル知識を参照して、これらの語彙の下位の階層の語彙へと展開することにより具体化を計る。また、もしも、検索要求が具体化のための十分な情報を備えていない場合には、マニュアル知識に基づいて、以下に示すように、ユーザへ選択枝を添えて再度要求内容の問い直しがなされる。

- 翻訳 ⇒ 一括翻訳（バッチ翻訳）／対話翻訳
- 出力 ⇒ 端末に出力／ファイルに保存／印刷

このようにして検索要求文から以下に挙げるような検索キーワードが求められる。

- 文書、ファイル、テキスト、読み込む、開く、選択
- 選択する、指定する、指定、翻訳、一括翻訳、実行
- 開始、対訳、形、形式、フォーマット、ファイル、
- 文書、テキスト、出力、出力する、書き込む、保存

上記検索キーワードを用いて、マニュアルテキストの検索がなされる。検索結果のテキストは、さらに、パラグラフや、マニュアルの小見出しにより限定される範囲にまで絞り込まれる。各見出しレベルでの範囲決定には前述したマニュアル知識を利用する。また、検索結果において重複する部分はここで除去され、最後に、次のような規則にしたがって、検索結果のまとめあげが行われる。

- ・基本的に検索要求文でのキーワードの出現順に並べる
- ・検索キーワードにより検索された箇所がマニュアルテキストの同一の見出しレベル内にある場合は、見出しレベル内での出現順位を優先する

5. まとめ

以上に述べた様なマニュアル検索システムを構築し、リファレンス・マニュアル的な正確を持つ情報機器のマニュアルから、実際の作業に必要なとなる手続き的な情報を抽出するための考察を行なった。

特に、対象を特定のマニュアルテキストに絞り、そのマニュアルに特有の知識を設定して、検索要求の展開や検索結果のまとめあげに利用し、より自然な形の検索要求への対応を試みた。

しかしながら、ユーザが意図している手続きと、実際の作業において生じる一連の手続きとではしばしばその順序が異なり、上述したような検索結果の合成方法では正しい作業手順にそったマニュアルを生成できないことも多い。また、検索の結果に基づいて抽出されるテキストの範囲が広がることが多く、全体として読み辛いものになってしまうことが多い。

今後は、これらのような問題点の解消を図ると共に、様々な種類のマニュアルによる実験も行なっていく予定である。

参考文献

[1] 石川徹也、「日本語テキストを対象とした自動索引システムの課題：総論」、情報の科学と技術、Vol. 42 1992 No. 11 pp. 994-1002