

電子文書参照支援システム BENLI

— システムの構成と機能の概要 —

3 F - 7

小川 洋一, 和田 浩一, 石橋 由希夫
セコム (株) IS 研究所

1 はじめに

我々は、電子文書参照支援システム BENLI (Browsing System for Electronic Document with Natural Language Interface) の研究/作成を進めている。

BENLI は、ハイパーテキスト [1] をベースとした電子文書のためのブラウザである。ハイパーテキストのリンクに基づいた参照機能に加えて、ユーザの入力した自然言語文を検索キーとして、その検索キーと内容的に一致する部分を検索する内容検索の機能を持っている。

システムの対象文書としては社内向けの技術文書を想定している。各種警備システムや個々のセンサー類のマニュアル等の社員による参照を支援することが目的であり、現在、銀行向けの警備システムの技術マニュアルを対象としてプロトタイプシステムを構築中である。

以下では、BENLI のプロトタイプシステムの構成および内容検索の処理内容について述べる。

2 対象文書の構成

ブラウジングの対象となる電子文書はハイパーテキストとして格納される。文書は構造的なまとまりごとにノード (以後、このノードを“文書ノード”と呼ぶ) とされ、それらの間の参照関係や親子関係に従ってリンクが張られる。ユーザはこのリンクを辿ることで、対象文書の“拾い読み”をすることができる。

例えば、プロトタイプシステムが対象としている警備システムのマニュアルの場合、文書が、章や節、小節といった階層的な論理構造を持っているので、図 1 に示すようなツリー状のハイパーテキストになる。

なお、各々の文書ノードにはそれぞれ ID 番号が与えられており、これによって文書ノードを一意に特定することができる。

3 システムの構成

BENLI のシステムの構成を図 2 に示す。

BENLI は、キーセンテンス解析部、文書解析部、キーワード検索部、内容検索部といったサブシステムと GUI から構成されている。それらが全体としてハイパーテ

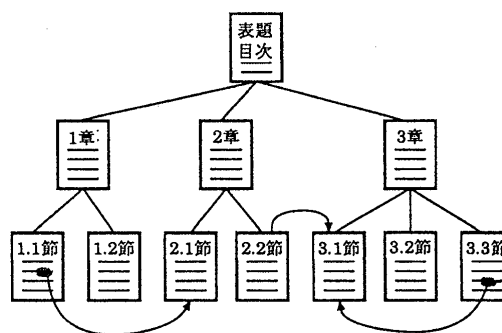


図 1: 対象文書の構成

キストに対するインターフェースとして機能することになる。

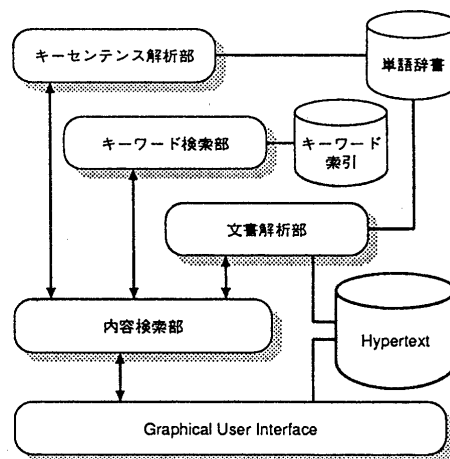


図 2: システムの構成

4 内容検索機能

BENLI における内容検索機能とは、ユーザが検索キーとして入力した日本語文 (以後“キーセンテンス”と呼ぶ) と内容的に一致する文書ノードをハイパーテキスト中から検索する機能のことである。

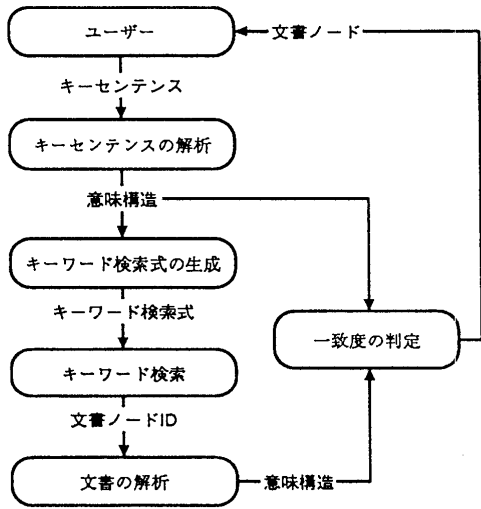


図 3: 内容検索の処理の流れ

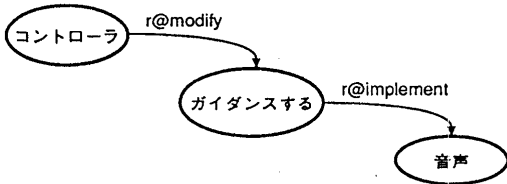


図 4: 「音声でガイドするコントローラ」に対する意味構造

内容的に一致するかどうかの判断は意味構造間のマッチングにより行なわれるが、対象文書の全ての文書ノードについてその処理を行なったのでは、システム全体の負荷が大きくなり過ぎると考えられる。そこで、意味構造を比較すべき文書ノードの数を、まず最初にキーワード検索を行なうことにより絞り込むという方法をとることとした。

処理の流れは以下のようになる (図 3 参照)。

- i) キーセンテンス解析部によりキーセンテンスを解析し、意味構造に変換する。
 キーセンテンス解析部は、形態素解析部と構文/意味解析部とから構成される。形態素解析部で単語ごとの構造体とされたキーセンテンスは、構文/意味解析部で単語間の修飾関係を解析され、意味構造に変換される。
 意味構造は文中の自立語 (名詞、動詞、形容詞等) をノード、その自立語間の修飾関係をアークとする意味ネットワーク (図 4) である。この意味構造が内容検索の際の基本的なデータ構造となる。
- ii) 意味構造中のノード (自立語) をキーワードとみなしてキーワード検索を行ない [2]、文書ノード ID を得る。

図 4 の例の場合だと、「コントローラ」、「ガイダンスする」、「音声」がキーワードとなる。

キーワード検索の際用いられるキーワード索引は、キーとしてキーワード、値としてそのキーワードが出現する文書ノードの ID のリストを持つハッシュテーブルである。この索引は、単語辞書を利用して対象文書から自動的に生成される。

- iii) ii) で得られた文書ノード ID を用いて、文書ノード中のテキストを得、それぞれについて文書解析部で解析を行ない、意味構造を得る。
 文書解析部の処理内容はキーセンテンス解析部のそれとほとんど同じである (プロトタイプシステムではキーセンテンス解析部を流用している)。
- iv) iii) で得られた意味構造と i) で得られた意味構造との間の一致度の判定を行ない [3]、各文書ノードをその一致度に基づいて順位付けする。
- v) iv) の順位が上位の文書ノードを検索結果としてユーザーに提示する。

5 おわりに

電子文書参照支援システム BENLI のプロトタイプシステムの構成および内容検索機能の概要について報告した。

このプロトタイプシステムは、GUI を含めたハイパーテキストの部分が HyperCard で、内容検索機能に関する各モジュール (キーセンテンス解析部、キーワード検索部、内容検索部) が Common Lisp でそれぞれ記述されており、MacIvory 上で稼働している。現在、内容検索の検索効率 (呼出率、適合率) についての評価を進めている段階である。

今後は、その評価結果に基づいて各サブシステムの改良を進めるとともに、社内文書を効率的にハイパーテキストとして構築するためのツールや環境についても研究を進めていく予定である。

参考文献

- [1] Conklin, J. "Hypertext: An Introduction and Survey", *IEEE Computer*, Vol.20, No.9, pp.17-41 (1987-9)
- [2] 石橋他 "電子文書参照支援システム BENLI における内容検索機能 — 前処理としてのキーワード検索 —", 第 45 回情報学大会 3F-08 (1992-10)
- [3] 和田他 "電子文書参照支援システム BENLI における内容検索機能 — 意味構造の一致度の判定手法 —", 第 45 回情報学大会 3F-09 (1992-10)