

6S-7

大規模電子化辞書開発における高機能辞書エディタ
— ブラウザ —

中沢正幸, 鈴木美穂, 安達久博*, 三輪和弘
株式会社 日本電子化辞書研究所
株式会社 東芝

1. はじめに

これまでの電子化された辞書についての情報は、いわば、本のページをめくるといった、それぞれに独立した情報として提示されるというアプローチによって行われてきた。しかし、このようなアプローチでは、本を読むのと根本的に違いはない。日本電子化辞書研究所(以下EDR)における電子化された辞書の大きな特徴はそれぞれの情報は互いにリンクしており、相互参照を行えるという点にある。このような特徴のもとに現在は、単語辞書、概念辞書を核とする大規模な自然言語処理用の電子化辞書を構築中である。

本論文では、この複数の辞書にまたがるリンクに注目し、EDR辞書を大規模なハイパーテキストとして捉え開発された高機能辞書エディタのブラウザについて述べる。

2. ハイパーテキスト

大規模な辞書情報を1つの情報としてユーザに提示する場合、欲しい情報のみならず、関係のない情報まで提示されることが多い。このような情報の提示の仕方では、ユーザ自身、どのような情報が欲しかったのか、何を目的として情報を検索したのか、この情報は他の情報とどのような関係をもつのかわからなくなる。

本研究では、このような問題を解決するため、また、大規模な辞書情報を効率的に処理するためにハイパーテキストを採用している。大規模な辞書をハイパーテキストとして捉えることにより、ユーザに対して要求された情報を要求された形で提示できるからである。EDR辞書を大規模なハイパーテキストとして実現するには、次のような問題があると考えられる。

1) 現在、自分がどの位置にいるのか

モニタリング機能を組み入れることにより、自分がどのような経路をたどり現在に至っているのかを把握できるようにする。

2) 目的の情報にどのように到達したらよいのか

オブジェクト指向の概念を取り込み、複数の辞書にまたがる情報を効率的に処理する。目的の情報への経路を明らかにし、容易に情報を参照できるようにする。

3. ブラウザとして要求される機能

1) ネットワーク状に構成される辞書の情報検索機能

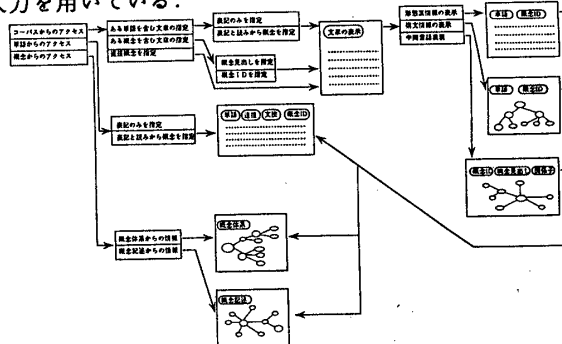
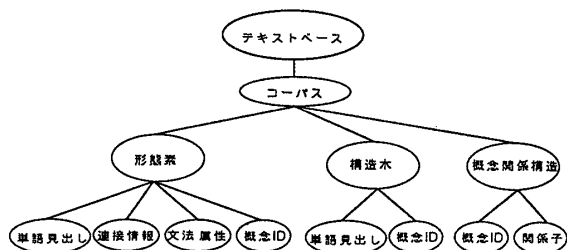
検索のキーとなるものを単語見出し、概念ID、概念見出しとし、複数の辞書を結び付ける。

2) 辞書情報を同時に表示することのできるマルチウィンドウ機能

ハイパーテキストとして情報検索するために必要不可欠な機能である。本エディタでは、Xウィンドウを用いて実現している。

3) 辞書情報の参照の操作を容易にするポインティング機能

情報の検索の指定が複雑、操作回数が多いなどの問題があれば、インタフェースの点から望ましい状態ではない。操作が簡単であればあるほど、情報の内容に注目することができる。本エディタでは、この検索の指定にマウス、キーボード入力を用いている。



High Quality Editor for Developing Electronic Dictionary
Hisahiro ADACHI, Kazuhiro MIWA, Masayuki NAKAZAWA and Miho SUZUKI
Japan Electronic Dictionary Research Institute, Ltd.

4. ブラウザの実現

4.1 オブジェクト指向

大規模な辞書情報を効率的に処理するために、オブジェクト指向を導入している。用いるクラスは次の通りである。

- 1) テキストベース
- 2) コーパス
- 3) 形態素
- 4) 構文木
- 5) 概念関係表現
- 6) 単語辞書 (単語見出し, 接続属性, 文法属性の情報をもつ。)
- 7) 概念辞書 (概念体系, 概念記述上の情報をもつ。)

これらのクラスは、モニタリングにおいては、現在どのオブジェクトを操作しているのかという状態を用いて知ることができ、ハイパーテキストとしてアクセスするときは、1つのノードとなる。したがって、複数の辞書にまたがる情報を効率的にアクセスできるといえる。また、これらのオブジェクトのアクセスの状態により、次に、どの状態に移ることができるかを決定することも可能である。

検索はそれぞれ相互に密接した情報を利用し、再帰的に処理を進めることができる。これら各クラスの階層関係は図ノのようになる。

メッセージ送信の起動は、マウスによるポインティングにより行われ、概念見出し、概念ID、テキストIDが各オブジェクトに送信される。このときの各オブジェクトに対するメソッドは次のように定義されている。

オブジェクト	メソッド
コーパス	テキストID, 単語, 概念IDをもとにテキストベースから文章を取り出す。
形態素	単語見出し, 接続属性, 文法属性をもとに形態素解析の結果を表示する。
構文木	単語見出し, 係り受け情報をもとに構文解析の結果を表示する。
概念関係表現	係り受け情報, 概念関係情報をもとに概念関係の結果を表示する。
単語辞書	単語見出し, 文法属性, 接続情報, 概念IDを表示する。
概念辞書	概念体系, 概念記述を表示する。

4.2 ブラウザの動作

本研究で開発した、高機能辞書エディタのブラウザの動作例を図3に示す。ここでの例は、“その”を含む文章をコーパスより検索し、その単語はどのような解析結果により得られたのかを示している。

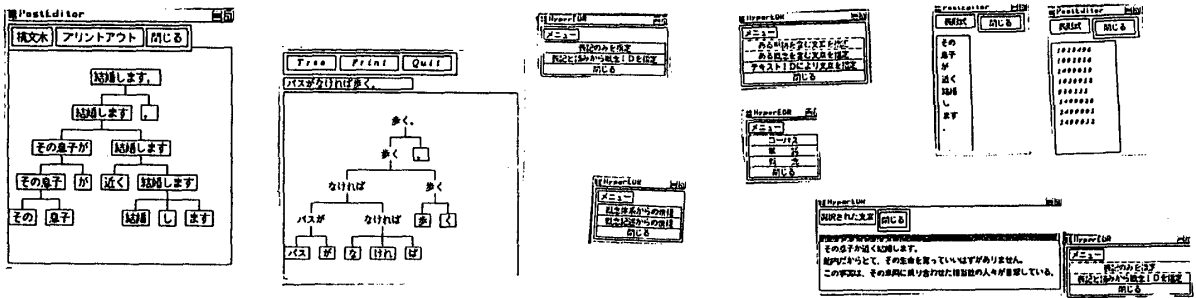


図3

5. おわりに

EDR辞書を大規模なハイパーテキストと考え、オブジェクト指向を採用した、高機能辞書エディタのブラウザについて述べた。本研究によって提案したようなブラウザによってユーザに対して有用な情報、容易な操作を提供できると考えられるが、ユーザの要求は、多種多様であり、また、大規模なハイパーテキストの検索に関しては、研究の余地があると思われる。今後は、この点に関して考察を加えていきたい。

謝辞

最後に、この研究の機会を与えて下さったEDRの横井所長、第四研究室吉田室長、及び終始貴重なご意見を頂いた第一研究室内田室長並びに研究員の皆様に感謝致します。

参考文献

- [1] "EDR電子化辞書(第2版) TR-016", 日本電子化辞書研究所, November 1989
- [2] "Hypertext: An Introduction and Survey", Conklin, J., COMPUTER, vol. 20: pp17-41, September 1987