

6S-8

大規模電子化辞書開発における高機能辞書エディタ  
— ユーザモデル —

鈴木美穂, 安達久博†, 三輪和弘, 中沢正幸  
(株)日本電子化辞書研究所

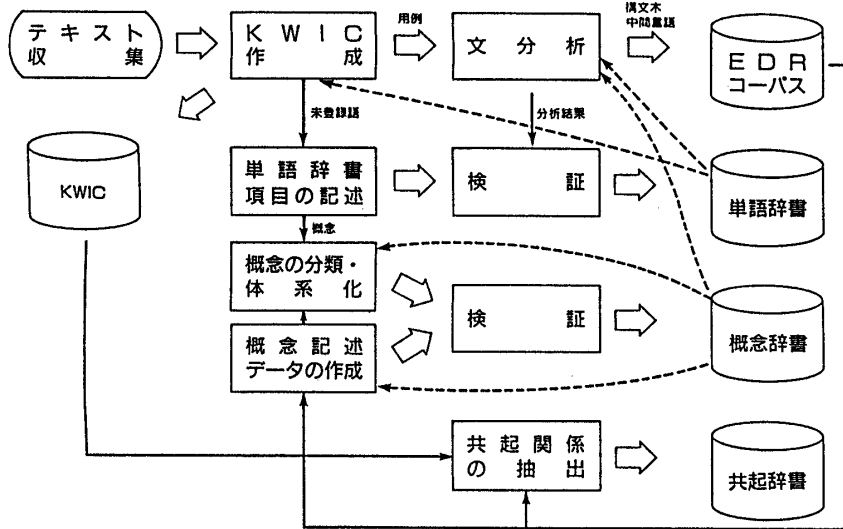
1. はじめに

(株)電子化辞書研究所(以下EDRと略す)は、単語辞書・概念辞書をはじめ、コーパスやKWICデータを構築中である。  
現在は、実際の文章データを分析して得た様々な言語情報を、EDRコーパスとして蓄積するとともに、単語辞書・概念辞書を核とする既存の各辞書に反映しながら、辞書データの開発を進めている。  
EDR辞書は現段階において約400メガバイト、また、今後コーパスデータ・KWICデータは30ギガバイトにも及ぶ予定である。  
このような大規模辞書を開発するには、辞書間に張られたリンクを辿り、各辞書情報への影響を考慮しながら、様々な観点からの開発項目が存在する。  
本論文では、それらの辞書開発を行うために利用される高機能辞書エディタのユーザモデルについて述べる。

2. EDR辞書の各開発における辞書エディットについて

2.1 EDR辞書開発について

実際の文章データ、及び各EDR辞書を用いてのEDRにおける大規模辞書開発が、どのような流れで行われているかを下図に示す。



EDR辞書の開発手順

2.2 各開発項目における辞書エディットについて

(1) 単語辞書項目の記述

既存の単語辞書を利用して実際の文章データを分析し、形態素情報(単語表記・文法属性・接続属性・概念見出し)を得る。  
得られた形態素情報に誤りがあった場合、概念体系・概念分類・概念関係表示への影響を考慮しつつ単語辞書へ反映させる。

(2) 概念分類, 体系化

何十万という膨大な概念を扱う上では、分類項目を設定し、概念を分類することにより概念の数を圧縮することが必要である。  
概念の体系化は、概念の意味的な階層構造(上位-下位)を表わし概念を抽象化したものである。  
この体系化を行うにあたって、ある概念が他の概念とどのような関係にあるかを表示し、また、EDRコーパスデータより概念分類項目の検証を行う。  
さらに単語辞書レベルでの概念見出しも参照する。

\* 現在, (株)東芝 総合研究所

## (3) 概念統合化

実際の文章データを分析し、得られた形態素情報（概念見出し）の中で、同義であると判断された概念見出しを統合化し、概念の数の発散を抑える。  
この際、概念分類・概念体系の情報を参照し、さらに単語辞書レベルでの概念見出しを参照する。

## (4) 概念記述データ作成

実際の文章データを分析し、得られた概念関係表現（概念の係り受け情報）に基づき、概念関係をあらかじめ定義された関係子を用いて記述する。  
この際、概念分類・概念体系の情報を参照し、さらに単語辞書レベルでの概念見出しを参照する。

## 3. ユーザモデルについて

2. 2にあげたように、EDR辞書開発では開発項目ごとに、各辞書への様々な観点からの、様々な辞書情報の参照が要求される。

このような大規模なEDR辞書への参照要求に答えるために、辞書間に存在する概念見出しを利用した概念リンクをはじめ、各種リンクを利用する。

ユーザは、開発において必要な情報、また情報間の様々な関係を、辞書間にはられたこれらのリンクを辿ることによって入手することが可能である。

どの情報もこれらのリンクを辿ることにより参照可能であるが、ある参照の経過において、参照された辞書データのエディットでは他の開発項目への影響を考慮し、参照不要とされた部分情報は保障されていなければならない。

この高機能エディタでは、ハイパーテキストの概念を導入し、EDR辞書情報相互間とEDRコーパスデータを、先に述べたリンクに双方向性を持たせて接続し、各開発項目の参照、エディット要求に対し個別のエディタを構成する。

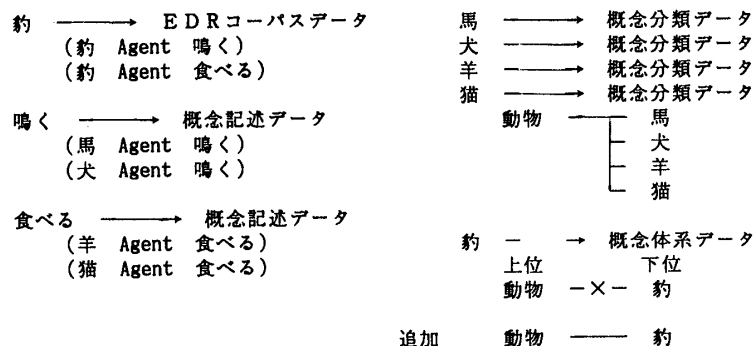
その結果、様々な観点から、必要なだけのデータの参照・エディットを実現する。

概念分類・体系化の開発を例にとる。

[豹という概念が動物という概念の下位関係に位置する]という情報をエディットする際に、EDRコーパスデータにおいて豹が他概念とどのような関係をとっているかを参照し、豹と関係子で結ばれたそれぞれの概念が概念記述データで、どのような概念と多く結び付くかを調査する。

そこで絞り込まれた概念が、概念分類データで動物に属するかを参照する。

さらに、概念体系データにおいて豹が動物の下位に属するか、単語辞書レベルでの概念見出しも参照する。すべてにおいて、整合性が取れている場合は[豹という概念が動物という概念の下位関係に位置する]という情報をエディットすることが可能である。



## 4. おわりに

大規模なEDR辞書を開発するにあたり、種々の開発項目により、様々な観点からのEDR辞書のマルチプルの参照、辞書情報のエディット機能を中心としたユーザモデルを作成した。

今後は、一つの情報の変更によるEDR辞書全体の動きをシミュレーションする場合を考慮した、辞書アクセス、各種リンク情報を含むユーザモデルを検討中である。

## 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、第七研究室太細室長、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：

- 日本電子化辞書研究所, EDR電子化辞書(第2版) TR-016,  
日本電子化辞書研究所, 1990. 11.  
小野寺夏生, ハイパーテキスト コンピュータによる知識の構造化,  
情報管理 3(11), 1989.  
Negroponte, N. P., ヒューマン インタフェース(吉成真由美 訳),  
日本経済新聞社, 1985.  
日本電子化辞書研究所, 概念辞書(第2版) TR-012,  
日本電子化辞書研究所, 1990. 11.