

柔軟な辞書検索の機能について

5R-5

大熊智子 石崎俊

慶応義塾大学 政策・メディア研究科

はじめに

キーワードのマッチングによる従来の辞書検索では、あくまで辞書に記述された情報のみを取り出すに過ぎなかったため、発展性や応用性に乏しかった。また、シソーラスなどを用いた電子化辞書の検索ではそのシステムの構築に必要なデータの作成や管理といった作業が人手によるものであるため大きなコストを要する。

本システム KMA (Keyword in Multi Ambience) では談話理解の一貫として、意味解析に用いる概念辞書 [1] などのリソースによって国語辞典に知識を補い、柔軟な辞書検索を行う。これによって、独立した個々の単語の情報に構造を持たせ、対象語に関連するすべての語をひろうことが可能である。また、その軌跡を辿ることによって語と語の相互間の関係が明らかになり、メタファーを含む構文にもロバストに対応できる。

さらに、検索対象の単語があいまいな場合でも、属性や関連する語などから概念フィルタリングを用いて、そこから特定の語を挙げる。KMA のインターフェイスは X11 上の GUI として作成されているため、使用者は常にインタラクティブな動作で柔軟に検索を行うことが出来る。

1 システムの構成

1.1 データの概要

辞書は以下のような国語辞典電子化版を用いる。

見出し語：いし

漢字表記：石

語釈文：1) 岩の小さなかけらが風化したり

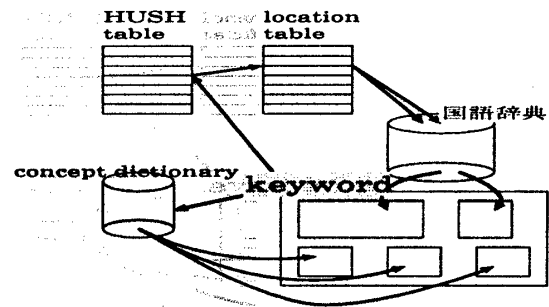


図 1: データの流れ

流水で角 (カド) が取れたりして丸くなったもの。(以下省略)

検索の際には、ハッシュコードで辞書引きを行い、単語の位置情報を持ったテーブルを参照する。さらに、概念辞書から属性や上位下位概念、動作環境を引き出す。

1.2 使用法

検索語入力欄に検索したい語を入力すると、国語辞書の該当する項目が結果表示欄に、その隣の関連語欄に、関連する語が表示される。下段の概念の欄には属性や上位/下位概念、動作概念が表示される。各欄に表示された語はすべてハイパーテキストによって、検索対象語に切り替えることが可能である。(図2参照)

対象語があいまいで、属性などで検索したい場合はそれぞれの欄に入力すると、候補欄 (位置は関連語の欄と同じ) に、該当する語が表示される。これらの語もクリックして、検索語とすることができる。

2 システムの機能

辞書検索: 検索語をかな・漢字入力して、辞書から該当する見出し語/漢字表記/語釈文を得ることができる。また、同音異義語が存在する場合、該当す

File

検索語: 石

検索結果

関連語

いし「石」
1) 岩の小さなかけらが風化したり流水で角(カド)が取れたりして
丸くなったもの。(砂より大きく、土と共に、至る所にあり、広義
では角のある砕石をも指す。堅い物、価値の無い物や、かんこ・外
産・無精などにたとえられる) 2) 建材・盛装用の岩石や砕物の総称
。「一をみがく」③特定の砕物加工品・宝石の称。「時計の一・ラ
イターの一(=発火合金)・一を打つ(=毒石を持ち、勝負する)」
「4) (じゃんけんて、はさみ・紙に對し) にぎりこぶしの称。ぐう
。「一が流れて木の葉が沈む」物事が進むことのこと。一
にかじりついて」どんな苦しい事があったも、一に就くマグラ
する」山野に暮る。「一に矢が立つ」信念を持てれば、出来
ないことはない。「一の上にも三年」つらい事でも長い間辛抱すれ
ば、いいことがある。

合成
財産
さす
砂漠
ざらざら
三角
地
磁石
自然
死ぬ
砂利
白

フィルタリング

属性: 上位概念 下位概念

属性	概念	動作環境
固い	上位	動作概念
小さい	砕物	拾う
	下位	投げる
	砂利	
	砂	動作環境
	石灰石	川原
	鉄鋼石	田舎道
	宝石	道路
	石碑	その他
	毒石	川

最初に戻る 進む 現在地に戻る

図 2: 入出力インターフェイス

る語すべてを候補として挙げ、そこから任意の語を検索対象語として確定することが可能である。

概念辞書の検索: 検索語が確定した場合、それと連動して行われる概念辞書の検索から属性/上位下位概念/同義概念/動作概念を得られる。

関連語の抽出: location table(図1参照)には、語釈文中にその語が現れる項目も関連語として格納されているため辞書検索時に関連語も抽出される。この機能によって、項目と項目の間にリンクが張られ、関連語から関連語への情報散策[2]が出来るようになる。また、直接関係を持たないように見える単語間のつながりが、リンクをたどることによって、顕在するようになるため、メタファーなど柔軟な解析を必要とする構文の理解を助けることが期待できる。

軌跡の記憶: ハイパーテキストによって、検索対象語が切り替わっていく過程は逐一記憶される。従って、その軌跡から単語と単語の相互関係が明らかになり、辞書全体を構造としてとらえることができる。概念フィルタリング: 使用者にとって検索語が明

確でない場合、概念の特徴から検索語を明らかにすることが出来る。例えば、「甘い食物」という属性が「甘い」で上位概念が「食物」という概念構造を持つ語を検索すると「菓子」「アップルパイ」「果物」「りんご」などが候補として挙げられる。さらにそこから「赤い果物」という条件の下で検索を行えば「りんご」が検索語として絞り込まれる。

3 まとめと今後の課題

本稿では自然言語理解の一貫としての辞書検索システムについて述べた。辞書そのものには手を加えず、他の複数のデータによって知識を補うことによって、様々なコストをおさえることが可能になると思われる。

KMAでは関連語の抽出や、概念フィルタリングの機能によって、柔軟に辞書検索を行うことが可能である。

今回は、概念辞書のみを背景知識として用いたが、より柔軟な検索を行うためにはさらなる知識も必要になると思われる。このデータ作成のための実験システムの構築も今後の課題である。

しかし、概念辞書を始めとする背景知識のデータを拡張する場合、概念フィルタリングによる検索時間の増加に伴うと予想されるため、処理の実時間性が問題になる。

本研究は文部省科研費重点領域「音声対話」および、(財)人工知能研究振興財団の助成を受けた。

参考文献

- [1] S.Ishizaki, T.Ikeda, Y.Sakamoto, H.Isahara
:Conceptual Dictionary as Interlingua
-Relation, Attribute and Modality-
:BUI.Electrotech.Lab Vol.54 No.4(1990)
- [2] 仁木和久 田中克己:ニューラルネットワーク技術の
情報検索への対応:人工知能学会誌 Vol.10
No.1(1995)
- [3] Kamran Parsaye, Mark Chignell Sretrag
Khoshafian Harry wong/近谷英昭訳:知的データ
ベース:オーム社(1992)