

Web上の辞書構造を利用したメタ辞書の構築

青田匠[†] 松山学[‡] 大園忠親[‡] 伊藤孝行[‡] 新谷虎松[‡]

[†]名古屋工業大学 知能情報システム学科 [‡]名古屋工業大学 大学院工学研究科 情報工学専攻

e-mail: {aota, manabu, ozono, itota, tora}@ics.nitech.ac.jp

1 はじめに

近年、WWWの普及により、様々な新しい情報がWWW上から入手できるようになっている。しかし、ユーザがWebページを閲覧中に知らない語句に直面することも多々ある。知らない語句の語意をWeb上の辞書を用いて調べる際に、知りたい語句が登録されている辞書を知らなければ、ユーザは語意を調べることができない。

Web上の辞書を用いて語意を調べる際のユーザの作業負担を軽減するシステムとして、Web上の辞書のメタ検索システム(メタ辞書)が存在する[1]。現在では、WWW上で有料サービスとして様々なサイトでメタ辞書が公開されている[2][3]。しかし、既存のメタ辞書では、検索語句がWeb上の辞書に存在するかを調べる段階に留まり、Web上の辞書の閲覧性までは考慮されていない。英和など文字数の多い語意では、1つの辞書の検索結果に目を通すだけでもユーザの負担は大きい。また、辞書毎に記述方法に違いがあるため、ユーザは辞書に対する知識も必要である。ユーザに辞書知識がなくとも簡単に検索結果を閲覧でき、辞書毎の記述の違いを一覧できる表示方法は、ユーザにとって有用だと考えられる。

そこで、本論文では、Webブラウザ上で選択した語句に対して、Web上の辞書構造を利用し統合した語意を、即座にポップアップ表示するメタ辞書を提案する。本システムでは、Webブラウザ上で選択した語句を検索語句とし、Web上の辞書を用いたメタ検索を自動で行う。そして、Web上の辞書から抽出した検索結果を統合してユーザに提示する。

2 Web上の辞書構造を利用したメタ辞書の概要

Web上の辞書は、要素(英和辞書:品詞,意味,用例,成句)が決まったタグ構造で表記されているという特徴を持つ。本論文における辞書構造とは、Web上

Construction of a meta-dictionary using the dictionary structure on Web

Takumi AOTA, Manabu MATUYAMA, Tadachika OZONO, Takayuki ITO, Toramatsu SHINTANI

Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Gokiso, Showa-ku, Nagoya 466-8555 JAPAN



図1: システムの実行例(太枠:ポップアップ)

の辞書のHTML文書におけるタグ構造を指す。本システムでは、辞書毎に辞書構造を利用したラッパーを生成し、要素単位で語意を取得することにより、語意の統合を図る。

本システムは、既存のWebブラウザ上で動作する。図1では、ユーザがWebブラウザ上で選択した語句「smile」に対して、語意がポップアップ表示されている。ポップアップでは、辞書構造を利用し統合した語意が表示される。ポップアップを用いる利点は、JavaScriptを組み込んだHTML文書を利用できる点、およびユーザの好みの大きさで表示できる点である。本システムでは、フレーム表示、畳み込み表示、およびツールチップ表示を用いることにより、コンパクトで閲覧効率を向上した表示を実現している。

本システムにおけるユーザの操作を以下に示す。

1. 本システムを起動させ、Webブラウザに対して本システムのプロキシを用いる設定をする
2. Webブラウザ上で閲覧中のWebページにおいて、調べたい語句を選択しCtrl+クリック(右クリック)する
3. 自動的にWeb上の辞書から語意が検索され、検索語句が存在した辞書の結果を統合した語意がポップアップ表示される

3 エージェントに基づく実装

図2に本システムの構成図を示す。本システムは、Proxyエージェント、Wrapperエージェント、および

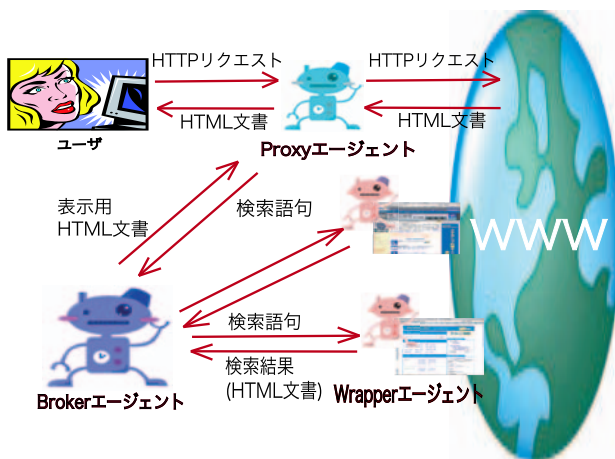


図 2: システム構成図

Broker エージェントの3つのエージェントで構成される。本システムの実装について、エージェントの働きと合わせて以下で述べる。

3.1 Proxy エージェントによる HTML 文書編集

Proxy エージェントでは、ユーザがリクエストした URL の Web ページの HTML 文書に JavaScript を組み込み、新たな HTML 文書を作成しユーザに提示する。ユーザに提示する HTML 文書を編集することで、Web ブラウザ上からの語句抽出、および語意のポップアップ表示を可能とする。Web ブラウザ上で選択された語句は、検索語句として Broker エージェントに送信され、Web 上の辞書を用いて語意検索される。そして、検索結果を Broker エージェントから受信した時点で、語意のポップアップ表示を行う。

3.2 Wrapper エージェントによる語意抽出

Wrapper エージェントでは、Web 上の辞書から語意となる要素抽出を行い、抽出された要素を用いて HTML 文書を作成する。まず、Broker エージェントから受信した検索語句が存在するかを判断する。検索語句が存在するかの判断は、検索型の辞書では実際にクエリを送信し、他の辞書では、検索語句が存在するであろうページを推定して行う。検索語句が存在しない場合、「Not Found」を検索結果とし、検索語句が存在する場合、Web 上の辞書のタグ構造を利用し、検索結果の表示がされている Web ページの HTML 文書から要素記述部分を抽出する。そして、予め作成されているテンプレートに要素記述を挿入し、ユーザへ提示するための HTML 文書を作成し検索結果とする。検索結果は Broker エージェントに送信される。

3.3 Broker エージェントによる検索結果統合

Broker エージェントでは、検索結果の統合を行う。まず、Proxy エージェントから送信された検索語句を

基に、Wrapper エージェントを統括して、Web 上の辞書から検索結果を収集する。検索結果をすべて受信した時点で、検索語句が存在しなかった辞書の検索結果の削除を行う。そして、検索結果を辞書分野毎に統合する。統合された検索結果を参照するリンクリストを作成し、表示用 HTML 文書に記述する。また、参照先のファイルとして、検索結果である HTML 文書をファイルに保存する。作成された表示用 HTML 文書は、Proxy エージェントに送信され、ユーザへポップアップ表示される。

4 本システムの特徴

本システムでは、i)Web ブラウザ上でメタ辞書が利用できる、ii)Web 上の辞書構造を利用し統合された語意が表示される、の2点が主な特徴である。既存のメタ辞書では、Web 上の辞書へのアクセスは不要となったが、辞書引きをするたびにメタ辞書システムへアクセスする必要があるため、ユーザの作業軽減として十分でないと考える。また、Web 上の辞書に検索語句が存在するかの判断に留まり、辞書の記述をそのまま引用するため、ユーザは登録されているすべての辞書に対する知識が必要になる。本システムでは、Web ブラウザ上で選択した語句を検索語句としてメタ検索を行うため、辞書引きにおけるユーザの作業は、調べたい語句を選択するだけである。また、辞書構造を利用して記述方法を統一することで、ユーザに必要な辞書知識を軽減した。そして、Javascript を用いて統合した語意をポップアップ表示することにより、コンパクトな語意表示を可能とした。

以上の点から、本システムは、既存のメタ辞書に比べ閲覧効率が向上し、ユーザの辞書引きにおける作業負担を軽減したシステムであると考えられる。

5 おわりに

本論文では、Web 上の辞書構造を利用したメタ辞書を提案し、本システムの仕様と特徴を示し、有用性を示した。今後の課題として、Web ブラウザや OS の違いによる JavaScript の挙動の違いへの対応が必要である。また、ユーザによって必要な辞書は異なるため、容易に辞書を追加できる実装を行う必要がある。

参考文献

- [1] 南野朋之, 奥村学: Web 上の辞書を利用したメタ辞書の構築, 第 65 回情報処理学会全国大会論文集, 2003 年
- [2] KOD: <http://kod.kenkyusha.co.jp>
- [3] 三省堂 Web Dictionary: <http://www.sanseido.net>