専門用語管理支援システムの実装

小山 照夫
(国立情報学研究所)
KOYAMA Teruo
(National Institute of Informatics)

Abstract

本研究の目的の一つは、実用的な形で利用可能な専門用語管理支援システムのプロトタイプを実際に実装し、その機能要件を確認することである。本稿ではシステムとして用語データベースを中心に、抄録文献管理機能、用語候補提案機能、形態素管理機能を統合し、Web ユーザインタフェースから操作可能な形で実際に構築したシステムに関して、その満たすべき要点と具体的な実装方法の詳細について述べる。

1 システムの要件

この研究の中で、研究代表者らは実際に稼働するプロトタイプシステムを実装して [1]、その評価と改善点に関する検討を進めてきた。以下では実際に作成したシステムに関して、その詳細を説明する。

既に全体的まとめの中でも述べているが、実装すべきシステムの機能的要件について再度まとめておく。

1.1 用語管理の分野依存性

用語の問題は分野ごとに検討されるべきであるという用語管理問題の特質から、構築するシステムはまず、複数の様々な分野について、分野ごとに独立に用語や用語間関係を管理できるものでなければならない。

このためには複数の、任意に設定できる分野の用語を蓄積・検索・修正するため、複数の独立したデータベースが定義できるように、システムが構築されている必要がある。すなわち、システムは同時に複数の用語データベースを並列して管理できるものでなければならない。

1.2 Web アプリケーションとしてのユーザインタフェース

また、実際の利用者の作業を考えるなら、システムは利用者の側で特別な環境を整備する必要なく、どこからでも利用可能なものであることが望まれる。この目標を達成する代表的な方法は、システムのユーザインタフェース（UI) を Web アプリケーションとして実装することである。

システム UI が Web アプリケーションとして実装されているなら、インターネットが利用可能な環境と PC があればどこでも用語管理作業が可能となる。

今回開発するシステムでは、基本的なデータベースを構築するとともに、このデータベースにアクセスするためのユーザインタフェースを Web アプリケーションとして実装し、この UI を通してデータベースを操作する枠組みを構築している。

1.3 共同作業の必要性

一般に一つの分野で管理すべき用語は膨大なものとなる。ここに見られるような広範なデータの管理には複数作業者の協力が不可欠であるから、一つのデータベースに対して、それぞれ複数の作業者が同時にアクセスして利用できることも重要である。

このためには、データベース管理システムが、複数ユーザの同時アクセスに対してデータ整合性を保証できるものでなければならない。

1.4 データベースへのアクセス権管理

複数データベースが管理されている状況ではまた、システム利用者はそれぞれ、アクセス権限を持つ適切なデータベースにのみ随時アクセスでき、かつ、権限のないデータベースにアクセスする事が許されない枠組みが必要である。

このことを可能にするためには、適切なユーザアカウント管理の枠組みを設け、アカウントごとに操作できるデータベースが制限できなければならない。それぞれのユーザはユーザごとの特定のアカウントを用い

てシステムにアクセスし、かつアカウントに許されたデータベースのみを操作できる形を採る必要がある。

1.5 用語収集支援

用語集編纂に当たっての問題点の一つは、用語候補をどこからどの程度、どのようにして収集するかである。この問題が本格的な用語集を編纂する上での大きな問題となっている。

この問題を解決するために、システムには用語収集の元となるデータを管理する機能および、用語定義支援機能を備えることが必要となる。

今回のシステムでは、用語を収集する大元として、ユーザがそれぞれの分野における研究抄録を登録し、全文検索可能な形でデータベースに蓄積して、用語候補の用例を検索・参照可能にすると同時に、抄録から用語と考えられる文字列を抽出する機能を組み込むことにより、抽出された用語候補を表示することによって、用語登録を支援する機能を用意することとした。

ここで用語候補の抽出に当たっては、研究代表者らがこれまでに開発してきた、用語抽出アルゴリズムを使用している [2],[3]。

1.6 データのアップロード機能

様々な研究分野では、特定の目的に従って部分的な用語集が作られていることがある。また、どの分野においても、基本的な用語については、ある程度まとめてオフラインで編集した方が効率的である場合も多い。

このことから用語集編纂に当たっては、用語をオンラインでデータベースに登録する機能に加えて、オフラインで編集したデータをデータベースにアップロードする機能も必要である。

この機能を備えることにより、既に確定しているデータについては、その電子化を外部に委託するなど、専門家の負担を軽減させることに役立つ。

今回のシステムでは、用語データ、用語間関係データ、抄録文献データを、ローカルに作成・編集した上で、データベースにアップロードする機能を提供する。

1.7 用語データのダウンロード機能

定義された用語データは、辞書の形で参照できる必要がある。また、このデータに基づいて各分野での用語辞書を作成する目的で、編集された用語データをローカルに編集可能な形でダウンロードできる必要がある。

今回のシステムでは、それぞれの分野について、蓄積された用語情報の参照と、用語データを辞書形式でダウンロードする機能を提供する。

1.8 形態素解析辞書の分野最適化

用語候補抽出のためには形態素解析と形態素辞書を必要とする。一般に専門分野では、一般的な日本語文書ではあまり使用されない分野固有の形態素が数多く用いられる傾向がある。また、一般に用いられる複合語の省略形が、解析精度を落とす要因となることもある。

このことから形態素辞書は、本来は対象分野の特異性に合わせて調整すべきものであり、ユーザがその内容を変更できる事が望ましい。

今回のシステムでは、初期の形態素辞書として既存のものを使用するが、必要に応じて形態素辞書に形態素を追加／削除する編集機能を備えることにより、用語候補抽出に用いる辞書を分野に対して最適化する機能を用意している。

以下では今回のシステムを構成する上での基本的な構成について紹介した後、それぞれの機能について概説する。

2 システムの構成

これまでに述べてきた要件を満たすために、システムのハードウェア・ソフトウェア両面からの構成を決定する必要がある。

2.1 システムのハードウェア構成

システムは、ハードウェア的には intel x86-64 アーキテクチャの 4 コアサーバ上に構築されており、外部から HTTP を通じてアクセス可能な形でネットワークに接続されている。

2.2 システムの基本ソフトウェア

このシステムの上で稼働させるソフトウェアはその基本的構成として、

- オペレーティングシステム
- データベース管理システム
- HTTP サーバ

- HTTP サーバとアプリケーションとのインタフェース
- アプリケーションプログラム群

から構成されている。これらの内、アプリケーションプログラムは主として python3(python3.2) によって開発されている。

より具体的には、オペレーティングシステムとしては linux(Debian7.0, Wheezy[4]) を採用している。Debian は、安定性が高く、長期にわたるメンテナンスが期待できる linux ディストリビューションであり、長期にわたって安定的にシステムを運営できると期待できる。

このオペレーティングシステムの下、可能なものについては、ディストリビューションで標準的に用意されているソフトウェアを利用することにした。

順序が前後するが、アプリケーション開発言語としては、python3 を主として利用する。この言語での開発を前提に、HTTP サーバとしては apache2.2(mpm-prefork) を採用し、アプリケーションインタフェースとしては libapache2-mod-wsgi-py3 を用いてステートフルな接続を保証している。

以下ではデータベース管理システム及び、データベース管理システム用 API について述べる。

2.2.1 データベース管理システム

先にも述べた通り、データベース管理システムは、複数分野のデータを並列して管理できるとともに、データアクセスに関するアクセス権限の設定が出来る必要がある。また、複数ユーザの同時アクセスに対して、データ整合性が保証されなければならない。

また、このシステムでデータベース管理システムの扱うデータの多くは日本語テキストであり、データ検索のためには日本語を含む全文検索機能を備えることが望ましい。

データベース管理システムはアプリケーションソフトウェアと連携して動作する必要のあることから、アプリケーションソフトウェアからのアクセスを許す適切な API を備えることも重要である。

以上の要件を考慮して、用語管理の中核となるデータベース管理システムは、mysql-5.5 を採用し、mysql に全文検索機能をプラグインとして追加する第三者提供ソフトウェア mroonga[5] を採用した。

mysql では、任意個数のデータベースを並列して管理することが可能であるからこの機能を利用して、一つの分野のデータを、一連のテーブルを有する一まとまりのデータベースとして定義することが可能となる。

2.2.2 データベース管理システム API

アプリケーション開発言語である python3.2 との API としてはいくつかのものが存在するが、今回のシステムでは mysql-connector-python を利用する。linux ディストリビューションの中には python3.2 に対応した mysql インタフェースを用意しているものも存在するが、Debian/Wheezy では用意されていないため、PyPI(python package index[6]) からインストール可能モジュールとして提供されているものをダウンロードして利用している。

2.3 HTTP サーバとユーザ認証

同一のデータベース管理システムによって、複数のデータベースを管理する環境では、ユーザごとにアクセスを許可するデータベースを制限する必要がある。

このためにはユーザアカウントを設定し、ユーザ認証を行う必要がある。ここで実際に認証を行う方法とタイミングとしてはいくつかの方法が考えられるが、ネットワーク経路上のパスワード暗号化を簡便に実現できる所から、今回は apcheHTTP サーバの提供するダイジェスト認証を利用することとした。

実際にシステム先頭ページにアクセスすると、ダイジェスト認証のためのダイアローグが表示されるので、ユーザはそれぞれに割り当てられたアカウントとパスワードでログインすることとなる。

apache サーバアカウントはまた、mysql のユーザとも対応する形で管理されており、mysql 内でアクセス可能なデータベースが決定されている。

この対応づけにより、ログイン後はアカウントに対応するデータベース名とともに、初期画面が表示される。また、一度特定のアカウントでログインした後は、関連付けられたデータベース以外のデータにはアクセスできない形になっている。

2.4 データベースの構成

データベース管理システムは複数の分野に対応する複数のデータベースを管理しているが、各分野に対応するすべてのデータベースにはそれぞれテーブルとして

- 用語
- 用語間関係
- 多義語
- 基礎文献

- 登録文献
- 用語候補
- 形態素

が用意されている。ここで用語データベースの本体は

用語／用語間関係／多義語

テーブルであり、その他のテーブルは用語管理支援のためのものである。用語関係のテーブルは、初期画面から定義と検索ができるようになっており、検索や検索結果を修正する機能も用意されている。

以下では用語管理支援のためのそれぞれのテーブルについて、その詳細を述べる。

2.4.1 基礎文献テーブル

基礎文献は、システムの利用開始当初から用語候補検索を有功にするためのもので、初期の用語候補を抽出するために利用される、研究抄録データの集合である。一度データが作成されると、ユーザはその内容について変更することはできない。また、このテーブルはオプションであり、必ずしも作成する必要はない。

基礎文献データが登録されると、文献に対して用語抽出が行われ、結果が用語候補テーブルに蓄積される。

2.4.2 登録文献テーブル

登録文献はユーザが用語管理に当たって必要とする抄録データを登録、管理するためのもので、随時登録／参照／変更が可能である。

登録文献にデータが入力されると、入力されたデータから用語抽出が行われ、抽出された用語候補が用語候補テーブルに追加登録されることになる。

2.4.3 用語候補テーブル

各分野のデータベースはそれぞれ用語候補テーブルを持つ。これまで述べた通り、用語候補テーブルのデータは、基礎文献および登録文献に登録されたデータから、用語抽出機能によって追加されるものである。

用語候補データは、頻度と用語状態という二つの状態を持っているが、頻度は基礎文献および登録文献を通じて抽出された頻度であり、用語状態は用語として登録済みかあるいは用語として不適切と判断されたかを記録するものである。

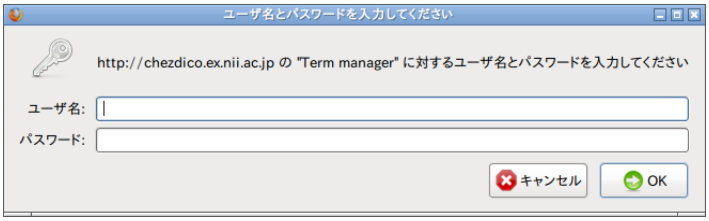

図 1: ログインダイアローグ

2.4.4 形態素テーブル

最後に形態素テーブルは、分野固有の形態素を追加したり、分野の特殊性を考慮して有害と思われる形態素を削除したりするためのものである。

なお、このテーブル上での変更は、そのままでは用語抽出のための形態素変更とはならず、明示的に現在のテーブル上の形態素集合で、システムの形態素辞書を置き換える指定をする必要がある。これは、形態素辞書の変更は辞書のコンパイルを伴い、ある程度計算時間を要するため、個々の形態素変更ごとではなく、ある程度まとまった変更が行われたタイミングで辞書入れ替えを行う方が適当であると判断したことによる。

3 システムの動作例

以下では実際のブラウザ画面を参照しながら、実際のシステムの動作について説明する。

3.1 初期画面

実際のシステムでは、システムにアクセスすると、ログインダイアローグ（図.1）が表示され、ユーザ ID とパスワードの入力が求められる。

ここで正しいユーザ ID とパスワードが与えられるとログインが行われ、図.2 に示す初期画面が表示される。

ログインに使用するアカウントは、データベースユーザ ID と関連付けられており編集が許されたデータベースにアクセスすることになる。初期画面左上には、現在のアカウントで編集可能なデータベース名が表示されている。

初期画面ではオンラインで、登録された用語の検索と用語の新規登録（同時に 5 語まで）、用語間関係の検索と新規登録（同時に 5 関係まで）、多義語の検索と新規登録を行うことができる。新規登録はいずれも当該入力エリアに必要なデータを記入し、それぞれに対応

する登録ボタンを押せば良い。用語間関係定義では、広義、狭義などの関係の分類はポップアップメニューから選択する事が出来る。

3.2 用語データ検索と変更処理

用語データである用語／用語間関係／多義語データは、いずれも初期画面から検索する事が出来る。検索に当たっては、文字列の完全一致に加え、指定文字列を部分列とする全てのデータの検索が可能である。ここでは mroonga の提供する全文検索機能が利用される。

図.3 に、「知識」を部分文字列として含む、登録済みの用語を検索した例を示す。

検索された用語等には右側に機能選択ボタンが配置されており、ボタンを選択することにより、検索されたデータを削除したり、データを編集する画面を呼び出したりすることができる。用語編集画面の例を図.4 に示す。

用語編集画面では現在の登録内容が表示されるから、見出し以外の項目を変更して登録することができる。見出しそのものは編集することはできず、一度レコードを削除した後、登録し直すこととなる。

同様に、用語間関係や多義語についても、検索や検索結果に対する削除処理または編集処理を指定できる。

3.3 用語データのアップロード機能

初期画面からはまた、用語データおよび用語間関係データに対するファイルアップロードを行うことができる。それぞれの新規登録フィールドの下部に、アップロードファイル選択と、アップロード実行のボタンがあるので、これらを用いてローカルのファイルを指定し、実行ボタンを押すことでアップロードを行うことができる。

現在アップロードするファイル形式としては、用語および用語間関係では CSV 形式をサポートしている。

3.4 用語候補検索と登録

本システムで重要な役割を果たすのが用語候補検索機能である。これは用語候補の部分文字列および、現在までの基礎文献および登録文献の中に出現する頻度を指定して、あらかじめ基礎文献および登録文献から自動的に抽出された用語候補を検索する機能である。

検索された用語候補は、一画面 20 づつ表示される。表示された候補の右には、ラジオボタンが配置されており、ボタンを用いて選択することにより、候補を用語として登録したり、用語候補から削除したりすることを指定できる。図.5 は「推論」をキーワードとし、頻度 10 以上を指定して検索を行った結果の最初の部分である。

実際にラジオボタンの各選択肢に対しては次の処理が行われる。

- 用語ボタンは当該候補を用語として登録する
- 非用語ボタンは当該候補を非用語とし、以降表示しない
- 保留ボタンは何も行わない

用語候補検索機能から用語登録を行うことにより、手作業で細かな用語データを入力するのとは比較にならない高い効率で用語の登録が行われる。

3.5 その他の機能

初期画面からはこの他に、用語出力機能、登録文書管理機能と、形態素辞書編集機能を呼び出すことができる。

3.5.1 用語出力機能

用語出力機能では、現在登録されている用語、用語間関係、多義語定義をファイルに出力する。出力された結果はブラウザ上で表示させるか、ファイルとしてダウンロードするかを選択できる。図.6 は出力された用語リストの一部を示すものである。

3.5.2 登録文書管理機能

文献登録機能では、入力済み文献の検索と編集、新規文献の入力、ローカルで作成した文献ファイルのアップロードが可能である。

図.7 に文書管理の先頭画面を示す。上部には新規文献入力のためのフォームが表示され、その下には既存登録文献検索のためのフォームと、ファイルアップロードの項目が表示される。現在アップロードファイルとしては、XML 形式のものを想定している。

この画面から新規文献を入力し、登録を行うと、図.8 に示す確認画面が表示される。この画面では、データの確認を行うとともに、手入力で新しい用語データを入力する事が出来る。

この画面の下には、図.9 に示すように、新規データから抽出された用語候補が表示される。ここでも各候補の取扱いをラジオボタンで選択することができ、用語登録、非用語判定、判定保留のいずれかを選択する事ができる。

文書管理先頭画面からは、登録済み文書の検索や変更も可能である。検索結果は図.10 のような形で表示されるが、この画面から文書の削除や編集が可能である。また、詳細画面は基本的に入力後に表示される確認画面と同様であり、内容の確認が出来る他、手入力による用語の追加や、提案候補の取扱いなども出来る。

3.5.3　形態素辞書編集機能

本システムは、各分野の特殊性に応じて形態素情報を管理する機能を備えており、形態素の追加や編集、解析辞書の変更などができる。図.11 は形態素辞書管理の先頭画面である。

この画面では上部のフォームから新規形態素の追加が可能である。また、形態素を検索し、検索結果から形態素の削除や編集が可能となっている。形態素検索結果の例を図.12 に示す。

ここで編集機能を選択すると図.13 に示す編集画面が表示され、形態素データを編集することができる。

形態素編集先頭画面からは、辞書データの入れ替えや、例文を用いて変更の効果を確認することができる。図.14 は例文管理画面であるが、ここでは例文として抽出するデータに参照文献を含めるかどうか、どのような文字列を含む例文をいくつ選択するかが指定でき、選択された例文について変更前と変更後とで解析結果がどのように変化するかを確認できる。

4　まとめ

以上、今回実際に稼働できる形で実装したシステムについて概説してきた。今回作成したシステムについて識者の意見を求めた結果、機能的には評価できるが、ユーザインタフェースには改善の余地があるとの意見が得られている。

今後は、ユーザインタフェースの見直しを行うとともに、システム機能や画面遷移の最適化などが必要となると考えている。また、システム保守性を向上させるために、Web フレームワークの本格的な適用や、データベース構造の見直しなども検討中である。

参考文献

[1] 小山照夫他: “用語管理システムの開発”, 情報処理学会自然言語処理研究報告, NL-212, 2013.

[2] 小山照夫他: “日本語専門分野テキストコーパスからの複合語用語の抽出”, 情報処理学会自然言語処理研究報告, 2006-NL-176, pp.55-60, 2006.

[3] 小山照夫他: “候補の接続関係を考慮した複合語用語抽出”, 情報処理学会自然言語処理研究報告, NL-193, 2009.

[4] https://www.debian.org/

[5] http://mroonga.org/ja/

[6] https://pypi.python.org/

情報処理学会

用語検索

検索語: 完全一致 部分一致

シソーラス検索

検索語: 完全一致 部分一致

用語情報登録

用語: ヨミ:
説明:

用語: ヨミ:
説明:

用語: ヨミ:
説明:

用語: ヨミ:
説明:

用語: ヨミ:
説明:

登録

シソーラス登録

用語1: 用語2: 関係: NT
説明:

用語1: 用語2: 関係: NT
説明:

用語1: 用語2: 関係: NT
説明:

用語1: 用語2: 関係: NT
説明:

用語1: 用語2: 関係: NT
説明:

登録

用語ファイルアップロード

参照... ファイルが選択されていません。
実行

シソーラスファイルアップロード

参照... ファイルが選択されていません。
実行

用語候補検索

検索語: 頻度:
検索

多義語検索

検索語: 完全一致 部分一致

多義語登録

用語:
説明: 登録

辞書出力
文書情報管理
形態素辞書管理

図 2: スタート画面

検索結果

知識表現	チシキヒョウゲン	編集	削除
知識ベースシステム	チシキベースシステム	編集	削除
語彙知識	ゴイチシキ	編集	削除
知識ベース	チシキベース	編集	削除
知識処理	チシキショリ	編集	削除
知識獲得	チシキカクトク	編集	削除
知識処理技術	チシキショリギジュツ	編集	削除
経験的知識	ケイケンテキチシキ	編集	削除
知識源	チシキゲン	編集	削除
知識処理システム	チシキショリシステム	編集	削除

戻る

図 3: 用語検索結果

用語編集

用語: 知識ベースシステム
ヨミ: チシキベースシステム
説明:
登録

戻る

図 4: 用語編集画面

検索結果

デフォルト推論	デフォルトスイロン	○用語	○非用語	◉保留
ファジィ推論	ファジィスイロン	○用語	○非用語	◉保留
推論部	スイロンブ	○用語	○非用語	◉保留
高速仮説推論システム	コウソクカセツスイロンシステム	○用語	○非用語	◉保留
非単調推論系	ヒタンチョウスイロンケイ	○用語	○非用語	◉保留
推論速度	スイロンソクド	○用語	○非用語	◉保留
前向き推論	マエムキスイロン	○用語	○非用語	◉保留
後向き推論	ウシロムキスイロン	○用語	○非用語	◉保留
推論系	スイロンケイ	○用語	○非用語	◉保留
高速仮説推論	コウソクカセツスイロン	○用語	○非用語	◉保留
推論機能	スイロンキノウ	○用語	○非用語	◉保留
並列推論マシン	ヘイレツスイロンマシン	○用語	○非用語	◉保留
並列推論マシンPIM	ヘイレツスイロンマシン	○用語	○非用語	◉保留
推論制御機構	スイロンセイギョキコウ	○用語	○非用語	◉保留
推論モデル	スイロンモデル	○用語	○非用語	◉保留
推論システム	スイロンシステム	○用語	○非用語	◉保留
非単調推論	ヒタンチョウスイロン	○用語	○非用語	◉保留
推論規則	スイロンキソク	○用語	○非用語	◉保留
推論方式	スイロンホウシキ	○用語	○非用語	◉保留
推論機構	スイロンキコウ	○用語	○非用語	◉保留

登録

次へ

戻る

図 5: 用語候補検索例

辞書データ

アイコン	アイコン
曖昧性	アイマイセイ
アセンブリ言語	アセンブリゲンゴ
アプリケーションプログラム	アプリケーションプログラム
仮説推論	カセツスイロン
仮想空間	カソウクウカン
関係データベース	カンケイデータベース
関数型言語	カンスウガタゲンゴ
R NT	C言語
R BT	プログラミング言語
画像処理	ガゾウショリ
画像データ	ガゾウデータ
機械翻訳	キカイホンヤク
計算機	ケイサンキ
計算機システム	ケイサンキシステム
計算時間	ケイサンジカン
計算モデル	ケイサンモデル
計算量	ケイサンリョウ
形式化	ケイシキカ
CASEツール	ケイスツール
形態素解析	ケイタイソカイセキ
検索システム	ケンサクシステム
高級言語	コウキュウゲンゴ

次へ

戻る

図 6: 用語データ出力例

文書情報登録

標題 : 日本語専門分野テキストコーパスからの複合語用語の抽出

著者 : 小山照夫、影浦峡、竹内孔一

キーワード: 用語抽出、形態素解析、複合語、語構成規則

抄録:

テキストコーパスからの用語抽出は自然言語処理技術の重要な応用である。従来テキストコーパスから用語候補を抽出する方法として、主として候補出現頻度に関わる統計的指標を用いて用語性を判定する方法が採用されて来たが、統計的手法では出現頻度の低い候補についての判定が困難であった。今回の発表では、複合語に注目し、用語性を損なう形態素出現パターンを排除する形での用語候補抽出を行うことにより、高い精度で複合語用語抽出が可能となる事を示す。

出版物: 情処技研NL-176-6

出版日: 2006/06/11

学会名: 情報処理学会

登録

文書検索

検索語:

登録日:

登録者: 検索

文書ファイルアップロード

参照... ファイルが選択されていません。

実行

戻る

図 7: 登録文書管理先頭画面

文書情報

表題: 日本語専門分野テキストコーパスからの複合語用語の抽出
著者: 小山照夫、影浦峡、竹内孔一
キーワード: 用語抽出、形態素解析、複合語、語構成規則
抄録:

テキストコーパスからの用語抽出は自然言語処理技術の重要な応用である。従来テキストコーパスから用語候補を抽出する方法として、主として候補出現頻度に関わる統計的指標を用いて用語性を判定する方法が採用されて来たが、統計的手法では出現頻度の低い候補についての判定が困難であった。今回の発表では、複合語に注目し、用語性を損なう形態素出現パターンを排除する形での用語候補抽出を行うことにより、高い精度で複合語用語抽出が可能となる事を示
す。

出版物: 情処技研NL-176-6
出版日: 2006-06-11
学会名: 情報処理学会

用語情報登録

用語: ヨミ:
説明:

用語: ヨミ:
説明:

用語: ヨミ:
説明:

用語: ヨミ:
説明:

登録

図 8: 文書入力確認画面

新規用語候補

テキストコーパス	テキストコーパス	○用語 ○非用語 ◉保留
用語抽出	ヨウゴチュウシュツ	○用語 ○非用語 ◉保留
自然言語処理技術	シゼンゲンゴショリギジュツ	○用語 ○非用語 ◉保留
用語候補	ヨウゴコウホ	○用語 ○非用語 ◉保留
候補出現頻度	コウホシュツゲンヒンド	○用語 ○非用語 ◉保留
統計的指標	トウケイテキシヒョウ	○用語 ○非用語 ◉保留
用語性	ヨウゴセイ	○用語 ○非用語 ◉保留
統計的手法	トウケイテキシュホウ	○用語 ○非用語 ◉保留
出現頻度	シュツゲンヒンド	○用語 ○非用語 ◉保留
形態素出現パターン	ケイタイソシュツゲンパターン	○用語 ○非用語 ◉保留
用語候補抽出	ヨウゴコウホチュウシュツ	○用語 ○非用語 ◉保留
複合語用語抽出	フクゴウゴヨウゴチュウシュツ	○用語 ○非用語 ◉保留

登録

戻る

図 9: 用語候補表示部分

文書検索

検索結果

Web情報を利用したQAデータの自動生成とベスト回／ 早川晃央,韓東力 編集 詳細 削除

固有名詞の所属国推定における表層情報の利用／ 延澤志保,佐野智久,菊地弘晶, 編集 詳細 削除

用語候補提案

戻る

図 10: 文書検索結果画面

形態素管理

新規登録

見出し:
品詞　: 名詞-サ変接続
重み　: 3999
読み　:
発音　:

新規登録

形態素検索

検索語:
部分一致
完全一致

辞書データ出力
辞書入替
例文管理

戻る

図 11: 形態素管理先頭画面

形態素検索

検索結果

知識人 名詞-一般 編集 削除

知識 名詞-一般 編集 削除

知識工学 名詞-一般 編集 削除

知識欲 名詞-一般 編集 削除

戻る

図 12: 形態素検索結果

形態素編集

見出し: 知識欲

品詞 : 名詞-一般

重み : 3999

読み : チシキヨク

発音 : チシキヨク

活用 :

置換

戻る

図 13: 形態素編集画面

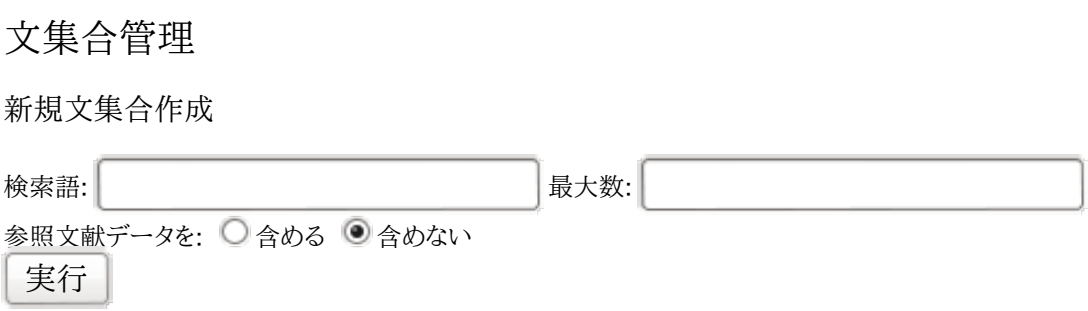

図 14: 試験文書選択画面