

シソーラス管理システムにおけるカスタマイズ機能について

7L-7

西川 記史*、徳田 圭世**、細川 貴史*、間瀬 久雄*、辻 洋*

* (株) 日立製作所 システム開発研究所、** 日立西部ソフトウェア (株)

1. はじめに

近年のWWWの普及により、ネットワーク経由で多様な利用者が多分野で情報検索を行うようになってきている。この支援のため、我々は検索に用いる用語を体系づけて管理したシソーラス管理システムを利用することを提案してきた^[2]。

今回、各種検索要求に対応するための用語管理について特に考慮し、仮想的に用語と用語間の関連とを更新するカスタマイズ手段を開発した。本稿ではカスタマイズ機能の実装について述べる。

2. 情報検索支援用語管理の必要性

ネットワーク時代を迎えWWW上の情報検索サービスでは、インデックスロボットによる情報収集、指定された用語を含む情報を提示する全文検索が行われている。ネットワーク情報検索は、従来の検索システムと相対的に検索対象分野が広範であり、利用者が検索要求として入力するキーワードによっては、検索結果に多数の文書がヒットする確率が高い。検索に不慣れな個人利用者の場合、多数のヒット文書からさらに目的の文書を探し出すため、2次、3次プロセスに労力を費やすことが少なくない。

こうした背景から、情報検索時の利用者の発想を支援するため、検索目的に応じた関連用語を格納する用語管理技術が必要と考える。情報検索支援にシソーラスを利用する場合、利用形態の多様性を考慮した上で、従来のシソーラス管理システム^{[3]-[5]}が備えていた用語管理の特徴を再考する必要がある。

3. 従来のシソーラス管理システム

従来のシソーラスは特定分野向けに構築され、そ

Customization function of thesaurus management system for network information retrieval

Norifumi Nishikawa* Tamayo Tokuda**

Takafumi Hosokawa* Hisao Mase*

Hiroshi Tsuji*

* Systems Development Laboratory, Hitachi, Ltd.

** Hitachi Seibu Software, Ltd.

の質の高さを保持するため、利用者が容易に更新することが出来ない仕組みとなっていた^[1]。実際、複数で共有利用するシソーラスを各利用者が思い思いに更新することはできない。しかし、多様な利用者が多分野の情報検索を行うようになってきた今日、利用者が個別用途にシソーラスデータベースの内容を更新できる手段が必要であると考え、以下ではこの要求を満たすためのシソーラスの個人カスタマイズ機能について述べる。

4. 個人向けカスタマイズ機能

4.1. カスタマイズ機能の必要性

従来のシソーラスは、プロバイダによって長周期的にしか更新されることのない完全な体系を保持している。しかし、これを検索支援にそのまま利用するのは必ずしも有効とは言いがたい。各検索要求に適した用語体系は自ずと異なるため、利用形態に柔軟に対応できるシステムが必要である。

個人利用者向けのカスタマイズ機能はこの要求を満たすと考える。各利用者に与えられるデータ仮想更新手段は、オリジナルのシソーラスをベースとして個人の認識に沿った知識体系を構築する手段となる。逆に、検索時に利用者の検索傾向・用途に沿った用語体系を用いるためには、個人の認識に沿った用語体系を構築できる枠組みが必要である。

4.2. カスタマイズ機能の実装

上述の検討により、我々はシソーラス管理システムにカスタマイズ機能を実装した。

(1) カスタマイズ方式

図1にカスタマイズ方式の手順を示す。

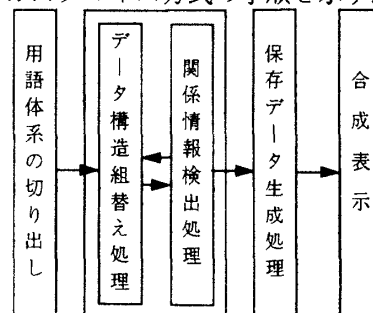


図1. カスタマイズ方式

利用者は、GUIを用いて個人用に用語体系をカスタマイズする。オリジナルシソーラスから利用したい用語体系を部分的に切り出し、これをベースにデータ構造を組み替えたり、より詳細な情報を追加して個人の用語体系を構築する。カスタマイズ操作は一時的に記憶され、利用者の保存信号発信によりカスタマイズデータフォーマットに生成される。

(2) データ構成要素

生成されたカスタマイズデータは、オリジナル辞書との差分データであり、データベースに格納する。データの構成要素を以下に挙げる。

- (2-1) 識別子情報
- (2-2) 利用する用語体系のルートノード
- (2-3) 生成された用語
- (2-4) 削除された用語
- (2-5) 生成された用語間関連
- (2-6) 削除された用語間関連

このうち、(2-1)(2-2)の要素は必ず存在するが、(2-3)~(2-6)の要素はカスタマイズ内容によって存在しないものがある。

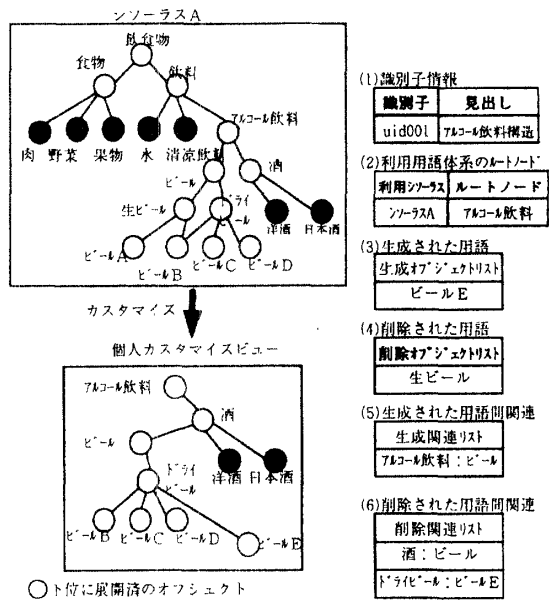


図2. カスタマイズデータとビュー

図2にシソーラスオリジナルの構造とカスタマイズ後の構造のビュー、およびそのとき保存されるカスタマイズデータの例を示す。利用する用語体系のルートノードを指定し、差分データのみを保存することでディスク容量を削減する。識別子情報に利用者情報を含むことで、利用者に対応するカスタマイズデータを認識する。

(3) データ保存場所

データ保存場所に対するシステム上の制約はなく、連携するアプリケーションのシステム構成に応じた保存場所に格納することが可能である。例えば、イントラネットで利用する場合はユーザ管理を行っているサーバと同一サーバに、個人利用者が単独でシソーラス管理システムを利用する場合はシソーラス管理サーバに、端末にプロキシ機能を有するソフトが組込まれている場合は利用者端末に保存する、などの形態が考えられる。

(4) カスタマイズ構造の参照

識別子情報から利用者に対応するカスタマイズデータを認識し、オリジナルシソーラスと合成して表示する。カスタマイズデータの保存場所がシソーラス管理サーバの場合は、データを合成した後クライアントに一括して転送・表示し、保存場所がクライアントの場合はオリジナルシソーラスの必要データのみを転送し、クライアント側で合成表示する。

5. まとめ

我々はシソーラス管理システムに個人向けのカスタマイズ機能を実装し、仮想的にシソーラスの用語および用語間関連を更新する手段を利用者に提供した。WWWブラウザでカスタマイズGUIを用いて個人用の用語体系を構築する。オリジナルシソーラスとの差分のみカスタマイズデータとして保存し、利用者情報を含んだ識別子を付与することにより利用者との対応を管理する。

これにより、利用者の用途に応じて柔軟に関連用語を提示するための用語管理を行い、シソーラス管理システムにネットワーク情報検索支援の上で重要な要素を備えることができると考える。

参考文献

- [1] 西川、他：オブジェクト・データベースを用いたシソーラス保守システムの試作、情報処理学会第50回全国大会講演論文集(4), PP33-34, 1995
- [2] 徳田、他：情報検索支援のためのシソーラス管理システムの提案、情報処理学会第53回全国大会講演論文集(3), PP163-164, 1996
- [3] 内藤衛亮他訳、シソーラス構築法、丸善株式会社, 1989
- [4] JICST科学技術用語シソーラス1993年版、日本科学技術情報センター, 1993
- [5] 日経シソーラス、日本経済新聞社, 1994