

土木工学ハンドブックシソーラスの構築支援システム

北海道大学工学部 上島 壯

1 はしがき

土木工学ハンドブック第四版（平成元年11月刊行）には活用を助けるためのツールとして資料編に「シソーラス」が付属した。これによってハンドブックの「事典」のような利用の仕方にも便利になったことと思う。例えば、「列車衝突事故」のニュースに接して鉄道の運転管理技術について関心を持ったとしよう。第36編「鉄道・軌道」の2ページ程のシソーラスを眺めることにより、第5章第2節がそれに関連する記事であることが直に分かり、記事の概要も把握できる。シソーラスを用いないとすると、「鉄道・軌道」編の目次を探し、表現の抽象度が高くそれだけ具体的イメージが浮かび難い見出し項目から見当を付けるか、密度の高い約70ページの本文を順にめくって探すことになろうし、それを考えるとささやかな知識欲であれば減退してしまう。実際に、各ページにこれだけ記事が詰め込まれると索引で引いたキーワードでさえ本文中に見いだすのもけっこう大変である。

このシソーラスを利用しての記事検索の流れとしては、着想語⇒総索引の検索⇒本文頁の数値⇒シソーラス編名の検索（シソーラス本文中の参照ページ範囲より）⇒内容の概要の判断⇒本文参照、の順が一般的であろう。

ここで、総索引の役割であるが、“鉄道の運転（管理／制御）～”として索引化されていればすぐに本文にアクセスできたかも知れないが、実際の索引項目は“運転制御システム”であり、最初の着想語はこのように詳細で明確でないのが普通である。この例の最初の着想語が“鉄道”か、“列車”か、“運転”か、あるいは“制御”かは対象とする図書の内容などによっても異なってくるが、内容が土木工学関連の図書であるから“鉄道”であろうか。ただし、この総索引には目次の見出しを索引語にしないという慣行もまだ残っていて、索引の“鉄道”から第36編「鉄道・軌道」にアクセスできないなどの不便さは残している。

この土木工学ハンドブックシソーラスのシソーラスとしての問題点は、概念上の上位下位関係・関連語などの親類関係が編章別に分断され、それらを結び付けるきずなが何もないこともあり得る。このように完成度においては歴史のある「J I C S T 科学技術用語シソーラス」などとは比較にならないが、見るシソーラスとしての楽しさの一つの特色といつてよからう。

このシソーラスと総索引は作業開始後短期間に完成したものであるが、そのために用いた計算機による支援システムSAURUSについて報告する。

第36編 鉄道・軌道(pp.1519~1588)

(省略)

[36] 5 オペレーションシステム

[36] 5.1 電力供給システム

電力 [11]3.7 [64]7.5

- 直流
- 単相交流
- 三相交流

電車線路

- パンタグラフ
- 架空単線式
 - 直接ちよう架式
 - 剛体ちよう架式
- シングルカタナリ式
- ツインシングルカタナリ式
- コンバウンドカタナリ式

- 第3レール・剛体複線式
- トロリ線

電気鉄道

- サイリスタ整流器
- シリコン整流器
- AT方式
- BT方式

[36] 5.2 運転制御システム

閉そく式

- 乗務閉そく式
- タブレット閉そく式
- 連発閉そく式
- 連動閉そく式
- 自動閉そく式

運転制御

- ATS
- ATC
- CTC
- ゴムトラック

図-1 資料編「シソーラス」内容例

2 シソーラス作成の計画

シソーラス小委員会（委員長 三浦裕二日本大学教授）は土木学会出版委員会の下部組織として昭和62年5月に活動を始めた。委員会の記録をもとにこの活動の当初の目的、方針などの概要を記す。

改定中の土木工学ハンドブックの索引代替となるシソーラスを作成することを当面の目標とした。この目的から完成するシソーラスは、ハンドブックの五十音順索引に上位語・下位語・関連語を併記する形式がもっとも自然と思われる。期待される効果として、i)索引用語のレベルをそろえる、ii)索引用語間のリンケージができる、iii)用語の標準化、iv)将来整備されるシソーラスの第1段階となる、などの項目が挙げられている。

土木学会で作成中の「土木工学学術用語(12,000語強)」の信頼性が高いのでこれをベースとし、ハンドブックの索引用語などで補強する。その他「新体系土木工学(全105巻)」の索引用語（延べ約40,000語）、講演・論文のキーワード（約2,000件/年）などの利用が検討された。

「土木工学学術用語」の16部門形式、「土木工学ハンドブック」の65編形式、土木学会の6部門と全国大会のセッション分けなどが検討の材料となり、最終的には「土木工学ハンドブック」の編単位が用いられた。

準備作業は概略以下のスケジュールで行われた。

- i) 「土木工学学術用語」の分類を単位とし、関係のハンドブック編に振り分け、各編の担当委員が取捨選択の作業を行った。
- ii) 選択された用語リストをハンドブック編執筆者に参考資料として提供した。標準用語として利用されることを期待したものである。
- iii) ハンドブック編執筆者に、配布用語にない追加用語リストの作成を依頼した。
- iv) ハンドブック編執筆者に、章・節単位で索引用語のリスト作成を依頼した。総索引の規模は約10,000語規模を想定し、これより逆算して4語/頁を目安とした。索引用語の記入リストは用語を階層化して入力する形式とし、シソーラス作成の資料としても利用できるようにした。（図-3左参照）

3 システムの概要

シソーラス構築支援システムSAURUSの主要部は土木工学ハンドブックシソーラスの最終的なプロファイルの確定前に作成された。以下、用いられなかった機能も含めて記す。

① シソーラスの要素と用語データベースの構造

主なデータ項目を示す。

- i) 用語属性：見出し、限定句、訳語（複数）、読み（複数）、出典、注釈。
- ii) 関係の種類：上位・下位関係、同義語、関連語。
- iii) 分類コード（用語単位）。
- iv) 部門：編章節。

なお、用語番号はデータベース構築のつど与えられる仮番号である。

データの管理は統計解析システムとして知られるSASを用い、図-2に示すように基本用語データベース部とそれを参照して作成する機能データデータベース部の二つのブロックより構成する。その他、出典ラベル、修正情報などに関するファイルが作成される。

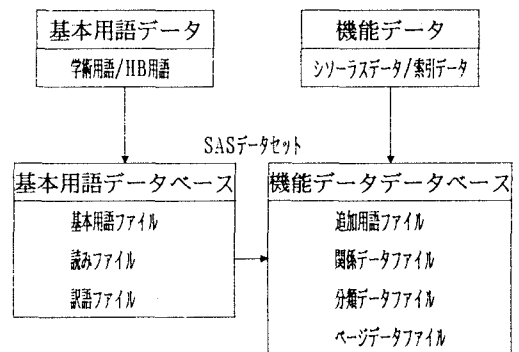


図-2 SAURUSデータベースの構成

② データの方式

レコードに見出し、読み、訳語、階層などを与えるデータの種類は、標準形式データ（シソーラスデータ入力向き）と引用符付き区切りデータ（索引データ管理向き）の2形式とした。以下の説明は標準形式データに関するものである。

シソーラスデータの作成は用語ごとに上位語、同義語、関連語などを入力する作業である。これを図-3の記入例に示すように、テキスト形式で用語群をコンマなどの記号を用いて階層的に表記する方式とした。上位語は階層の解析により求められる。

階層表記方式はパソコンのカード型データベース類似方式に較べて、i) 関連語の全体が視覚的に把握できてわかりよい、ii) 作業者はパソコンワープロなど手持ちのツールで入力できる、iii) 各用語は一度入力すればよい、などの利点がある。

用語部への読み、訳語の記入は下記の形式とし、シソーラスデータファイルからの用語情報の採取が可能である。《用語部形式》 見出し語：ヨミ₁、ヨミ₂、...、ヨミ_n:::term₁;term₂;...;term_n

関係の種類の定義は階層設定子（コンマ）の前に一字の文字で指定する方式とした。ただし、ハンドブックシソーラスの本作業のデータについては、同義語は見出し語の後にカッコ付きで表記することとし、その他の上位語・下位語関係については関係の種類を区別していない。

その他、制御行、削除行、階層内の独立領域の定義、同一レベルを区切る“パーティション”の設定、ブラケットによる上位語の切り出し、などの機能を設けている。

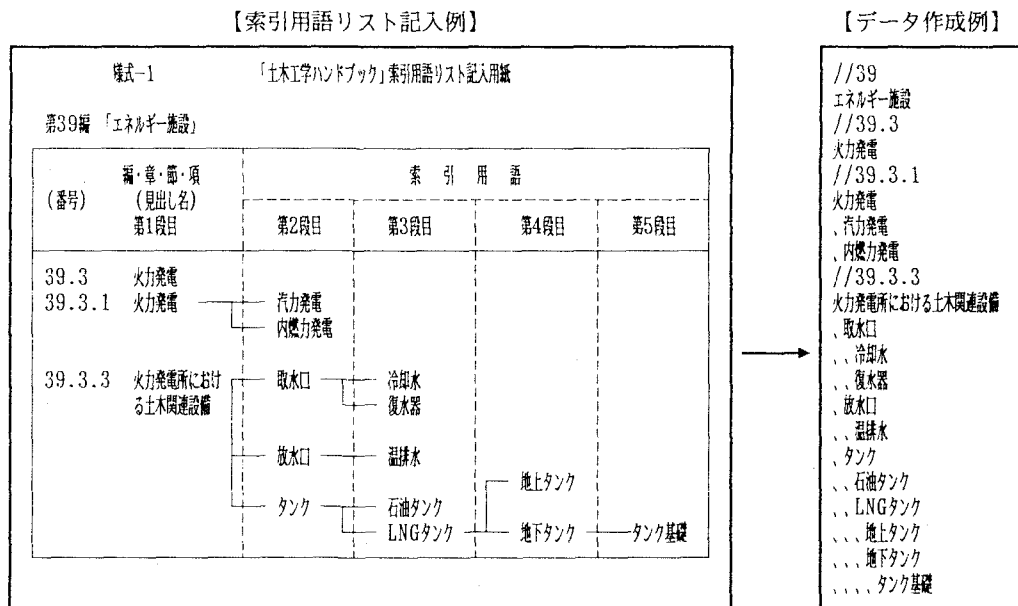


図-3 ハンドブック執筆で作成索引用語記入リストとそのコーディング例

③ 構造化とチェック機能

SASシステムは表形式でデータ管理を行うが、階層構造を構築するために用語関係の経路リストを生成する手法を用いた（図-4）。データの検査とその結果のプリントが以下の項目について得られる。i) 関係の多重指定チェックリスト：下位語で関連語の

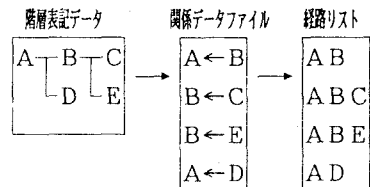


図-4 経路リストの模式図

場合など、ii)関連語（RT）グループ化リスト：(A,B)(B,C)→(A,B,C)、iii)サーキットチェックリスト：上位に自分自身がある場合、iv)バイパスチェックリスト：2語間に複数の経路がある場合、v)用語引き一覧表、vi)階層表示型一覧表、など。

④ データの修正・更新

修正・変更は機能データベースに対して修正ファイルを用いて行う。機能の種類は、i)新しい階層の設定と古い階層の強制的修正、ii)用語と関係の削除、iii)新しい用語の挿入などである。

階層構造の構築結果を出力することにより整理されたシソーラス更新データが得られる。

⑤ 索引作成サブシステム

索引の作成は以下の手順で行なわれる。i)初期作業として目次データ（章節とページ）を作成し、シソーラスデータを仮の索引データとして利用する。ii)索引データを機能データベース化し、目次データを用いて作業用索引語一覧表とパソコン可読データを出力する。iii)このデータをカード型データベースに取り込み採択する用語の相対ページ（所属の章節の始まりページを1とする）を入力する。iv)テキスト形式に再変換した索引データを得てii→ivの作業を修正が終了するまで繰り返す。v)索引を出力する。索引語は基礎用語データベースに登録されていないといけない。また、読みが複数ある場合は重複表示される。

4 ハンドブックシソーラス作成作業

① 様式の決定

シソーラス小委員会において最終的な様式は体系表示シソーラスと通常の五十音順総索引の組み合わせとなった。体系の編集は行わず章節単位に作成した体系をそのまま使い、他のブロックに重複する用語の参照を表示する形式とした。索引語は「学術用語」か否かの区別を行なうこととした。

② ハンドブックシソーラス入力作業支援プログラム

担当者へのパソコン入力作業支援プログラムとして階層化記号の罫線化印刷プログラムが配布された。また、入力用語をチェックする登録用語照合プログラムがマネージャーレベルで用いられた。

③ 使用された用語と土木工学学術用語の関連性

図-5に示すように土木工学ハンドブックに関連する用語は多くは編執筆から追加用語として提出され土木工学学術用語に匹敵する規模となったが、共通する用語は学術用語の3割程度であった。

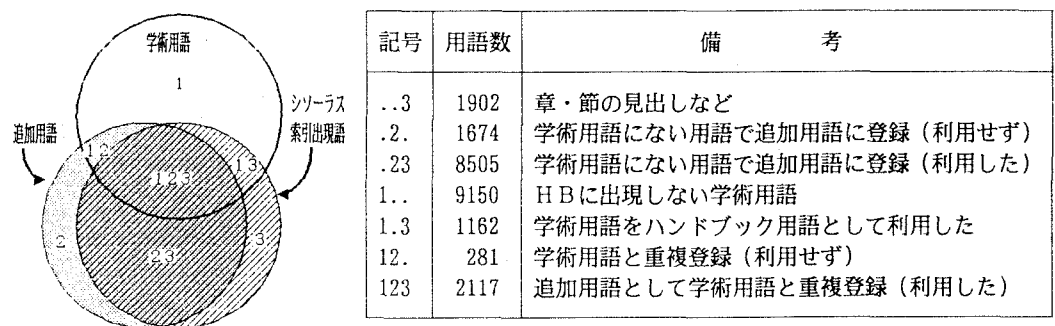


図-5 土木工学学術用語とハンドブック用語の関連性

5 おわりに

土木工学ハンドブックシソーラス構築支援のため開発されたこのシステムの多くの機能については小規模のデータでテストされたのみである。今後大規模なデータを長期的に管理することに利用するためには用語管理システムとしての改良が必要と考えている。