

I-9 土木用語データベースの構築と運用

Construction and Management of Civil Engineering Term Database System

今 尚之*, 高野伸栄**, 佐藤馨一***

Naoyuki KON, Shin-ei TAKANO, Keiichi SATOH

【抄録】 多岐にわたる専門分野を内包し、専門分化が著しい土木工学分野における専門用語を選定することは、非常に困難な作業となる。このため、データベースシステムを導入し、作業の合理化、効率化を図ることが必要である。本研究は、「質の確保」「網羅性」「大量性」「即時性」「協同性」が要求される土木用語データベースの構築にあたって、組織、運用されたシステムについて述べるものである。組織されたシステムの支援によって、データの質と網羅性を確保し、大量なデータを処理することが可能なデータベースを、短期間に作成することができ、専門用語の精度良い採語を行うことができた。

【Abstract】 It is very difficult to select technical terms in civil engineering, which is composed a wide special area. Therefore, a using of the database system is demanded for efficient work. It is required of adjust construction and management of the database system to 'quality', 'comprise', 'quantity', 'quicklime', 'cooperation'. This study describes construction and management of the database system which is for selection of civil engineering terms. We can select terms for the database system rationally and efficiently.

【キーワード】： 土木用語データベース, 採語作業, 採語支援, データベースの運用

【keywords】： civil engineering term database, selection of terms, support for selection of terms, operating of database system

1. はじめに

土木工学分野は、他の学問分野に比べてはるかに幅広い専門分野を内包し、かつ専門分化が著しい学問分野である。このため、専門用語の適確な理解や正しい使用がより強く望まれており、土木工学諸分野における専門用語の体系化およびその結果として辞典類が不可欠である。明治以来、現在までに土木工学の専門用語辞典の編纂が図1に示した経緯で取り組まれてきた。しかし、土木学会においては、土木工学諸分野の専門用語を網羅し、解説を付した専門用語大辞典の編纂はなされなかった。このことは、多岐にわたる専門用語を収集、選別する採語作業（辞典掲載用語を決定すること）が非常に困難であることが背景にある。

現在、土木学会創立80周年を記念した「土木用語大

辞典」（編集委員長：五十嵐日出夫 北海道大学教授）の編纂作業が行われている。そこでは多岐にわたる土木工学諸分野の用語を体系的に網羅し、精度の高い辞典を4年間という短期間に編纂することが要求されている。特に、専門用語の辞典編纂では採語の良否が辞典の完成度を左右するとまでいわれる。したがって、採語にあたってはデータベース管理システムを導入し、合理的、効率的に作業を進めることが必要である。

本研究は、採語作業の合理化、効率化のために導入した土木用語データベースの構築とその運用について述べるものである。

2. 採語作業とその問題点

*	正会員	小樽商科大学商学部社会情報学科	(〒047	小樽市緑3-5-21)
**	正会員	北海道大学工学部土木工学科	(〒060	札幌市北区北13条西8丁目)
***	正会員	北海道大学工学部土木工学科	(〒060	札幌市北区北13条西8丁目)

専門用語辞典を編纂するにあたって、まず、採語が行われる。その作業は一般に、ⅰ) 専門用語の収集→ⅱ) 用語の選別→ⅲ) 重要度のランク付け、の各段階を経て行われる。

また、採語作業においては、ⅰ) 収集の漏れ・偏り・重複等の網羅性に関する問題、ⅱ) 選別の基準、ランク付けの基準の設定など採語の質、に関する問題が発生する。従来の採語作業では、用語を書いたカードを一枚ずつ並べることによって対処していたが、この方法では採語作業に要する期間が非常に長くなり、仕事量も膨大なものとなる。このため、土木工学分野のように幅広い専門分野を持つ場合、全専門分野をカバーする用語辞典の編纂はほとんどなされなかった。

3. 土木用語データベース管理システムの設計²⁾

(1) システムの設計要件

土木工学は極めて専門性が高く、幅広い分野から成立している学問分野である。このため土木工学分野の専門用語データベース作成するためには、

- ⅰ) データの質を確保する (専門性)
- ⅱ) 幅広い専門分野の用語を網羅する (網羅性)
- ⅲ) 大量なデータを取り扱う (大量性)
- ⅳ) 時代の要求に合わせる (即時性)
- ⅴ) 数多くの専門家が協同作業を行う (協同性)

ことが必要である (図2)。

ⅰ) 専門性

従来のデータベース管理システムでは、データを効率良く利用するための整理、保管が重視され、データそのものの質は余り問題とされてこなかった。しかしながら、その内容や質によってそれを刊行した国、学会、出版社などの歴史や文化が評価されるといわれる辞書類の採語をデータベースの運用によって支援するならば、土木用語データベースに登録される用語は高い専門性や一般性を確保しなければならない。

ⅱ) 網羅性

土木工学の専門分野は幅広く、なおかつ専門性が高い。このため、ある特定分野の専門用語に偏ることなく、土木工学分野全体の用語をまんべんなく採語することが必要である。また、重要な用語を欠落させることがないように十分気を付ける必要がある。

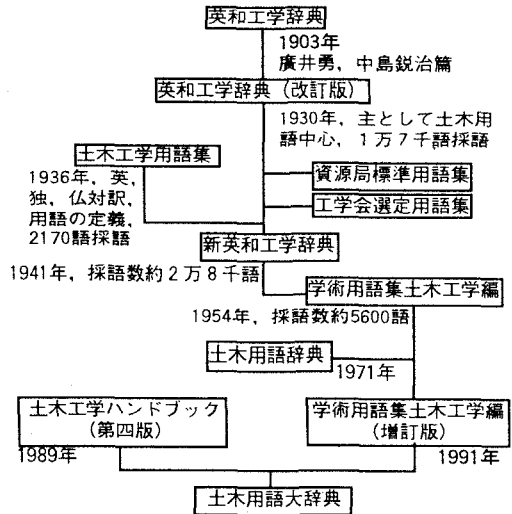


図1 土木用語辞典に係るこれまでの経緯

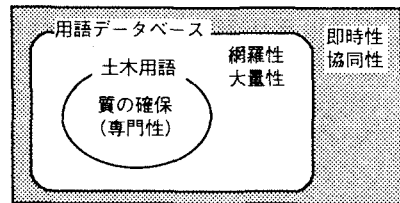


図2 土木用語データベースの要件

ⅲ) 大量性

土木工学分野の専門用語を網羅することから、データベースに登録されるデータ件数は膨大な数となる。また、採語の検討段階では、それ以上の大量な用語を管理し処理する必要もある。さらに、用語一つ一つに付加される諸情報も考慮すれば、土木用語データベースに登録される情報量は膨大な量となる。

ⅳ) 即時性

土木用語に付加される情報は、時代と共に変遷するものである。例えば、新技術の開発によって新しい用語が誕生し、用語の重要性が変化する。専門用語の採語は時代に則して常に改訂される必要がある。また編纂期間が4年と従来に比べて短期間なものとなっており、採語を速やかに終わらせることが必要である。

ⅴ) 協同性

土木工学の専門用語データを管理、運用し、専門用

語を採語するためには、日本全国を対象とした広範囲から人材を集め、各人が持つ能力を最大限に発揮できる協同作業体制が必要である。

(2) コンピュータによる作業支援と協同作業

土木用語の採語のためには、質の確保(専門性)、網羅性、大量性、即時性、協同性の5点を考慮した作業システムの組織と運用が必要であり、データベースを作成するだけでは不十分である。

例えば「質の確保(専門性)」、「網羅性」は、数多くの専門家が参集し、互いの知見を持ち寄り、情報の交換を行うことによって成されるものである。すなわち、個々の専門用語を持つ情報、例えば、用語の意味、用語の重要性等をグループによる協同作業によって一つ一つ吟味し、互いに調整しあうことによって、重要な用語を欠落させることが防げ、より専門性の高い採語が可能となり、土木用語の新しい情報が生成されることとなる。

このように協同作業によって新しい情報が生成されるが、それはまた、情報の交換という場面にフィードバックされ、再び、グループによる調整作業段階を経ることによって、より高い質と網羅性を確保できることになる。すなわち循環的な作業プロセスを経ることによって、土木用語の採語の質と網羅性が向上する。

以上のプロセスの「情報の交換」、「グループによる調整作業の準備段階」では、データベースや通信ネットワークなどコンピュータによる作業支援を導入し、作業の合理化、効率化を図る必要がある。(図3)

(3) データベースシステムのフレームワーク

i) 土木用語データベースシステムのフレーム

- 本研究では、
- a) パーソナルコンピュータによる、土木用語データベースの作成と利用
 - b) 各専門分野の専門家による小委員会組織よる採語作業
 - c) 編集委員会組織による用語データの全体調整と管理

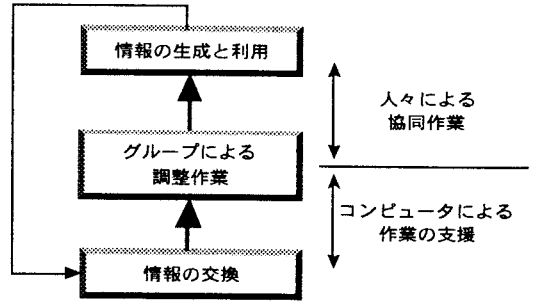


図3 採語作業でのコンピュータ支援と協同作業プロセス

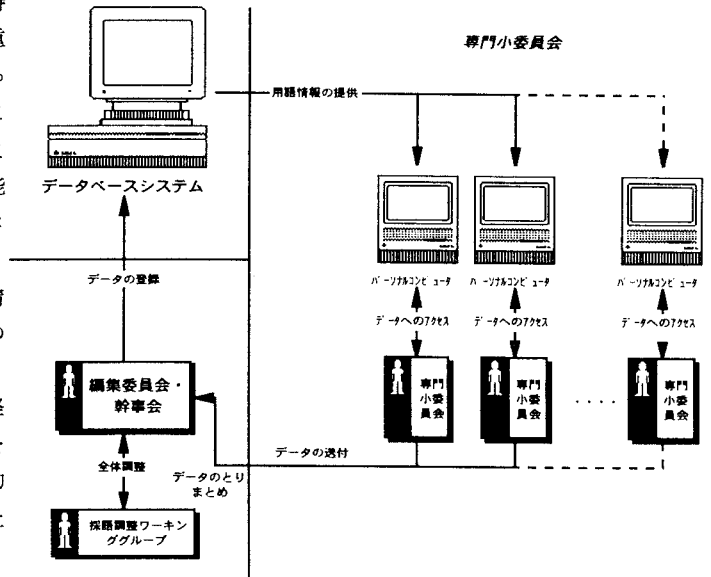


図4 土木用語データベース管理システムのフレームワーク

の3点をフレームとした全体システムを構築し、データベースを核とした情報交換を行い、採語作業を進めた(図4)。

ii) パーソナルコンピュータの導入と土木用語データベース

専門家各自において、データベースに直接アクセスでき、電子情報の形で情報交換が行われるならば、作業の精度、効率の向上を図ることができ、土木工学の専門用語を、高い質と網羅性を確保しながら採語することが可能となる。

近年、パーソナルコンピュータシステムの発展は目覚ましく、CPU処理速度の高速化、記憶容量の大容量

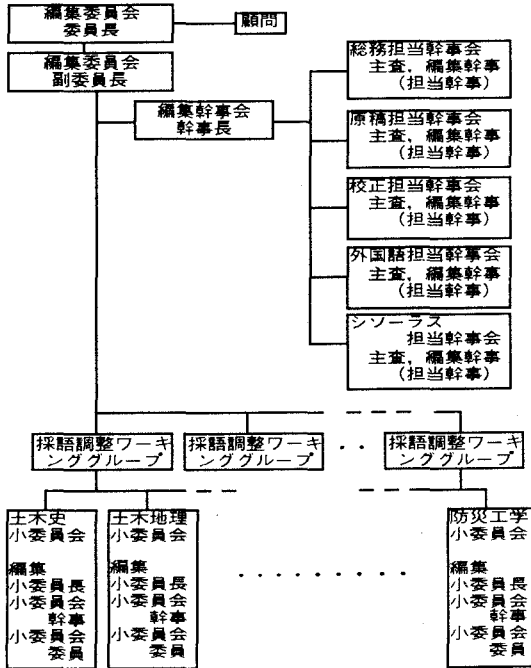


図5 土木用語大辞典編集委員会

化が進み、全体で20MByte以上に及ぶデータファイルを管理・運用できる性能を持つに至った。また、安価で高性能なデータベースアプリケーションソフトも市販されている。そのため、大型計算機を使用し、その上でアプリケーションソフトを自作して処理する必要がなくなった。さらに、パーソナルコンピュータのワープロやデータ処理利用も一般的となり、パーソナルコンピュータに対するリテラシーも向上している。

このような状況により、土木用語データベースは、パーソナルコンピュータベースで動作するものを構築し、各専門小委員会にデータベースソフト共に配布することとした。

iii) 土木用語大辞典の編集組織

採語の質（専門性）と網羅性を確保し、効率良く作業を進めるために、図5に示した編集委員会組織が編成された。小委員会組織は、各専門分野の用語を採語し、解説文の執筆を担当している。また、5つの担当幹事会より構成される編集幹事会は、編集作業のとりまとめを行い、質と網羅性を確保し、即時性、協同性などに対応している。さらに、担当幹事会、専門小委員会の枠を越えた採語調整ワーキンググループ（26部

門を関連する6部門に統合）が別途組織され、編集委員会内に設置された。

(4) 設計要件に対する対応

i) 採語の質と網羅性の向上

採語の質と網羅性の向上のために

- a) 専門小委員会の組織
- b) 用語のランク付け
- c) 学術用語集、土木工学ハンドブック掲載用語のデータベース作成
- d) 採語調整ワーキンググループの組織とそれによる重複採語等のチェック作業

などを行った。特に小委員会における採択用語選定作業結果が常にフィードバックされるような組織が編成された。

ii) 大量性と即時性

大量のデータを速やかに処理する必要から、

- a) パーソナルコンピュータベースのデータベースのシステム作成
- b) データの電子情報化とフロッピーディスクによるデータの交換

などを行った。その結果、データの再入力作業を省くことができ、作業期間の短縮、ケアレスミスの防止がなされ、効率化を図ることができた。

iii) 協同性

協同作業を円滑に進めるために

- a) パーソナルコンピュータベースのシステム作成
- b) FAXによる連絡、情報交換

などを行った。特に、FAXを利用したことによって記録の保持が可能となった。また、一部個別であるが試験的に商用ネットワークによるパソコン通信も利用した。

4. 土木用語データベースの構築³⁾⁴⁾⁵⁾

(1) 土木用語基礎データベースの作成

i) リレーショナルデータベース構造の採用

土木用語データベースの作成にあたっては、

- a) 編集作業において用語情報の更新作業が頻繁に行われること
- b) 小委員会毎にデータベースを分割・併合する作業が必要であること

などの理由からリレーショナルデータベース構造を採

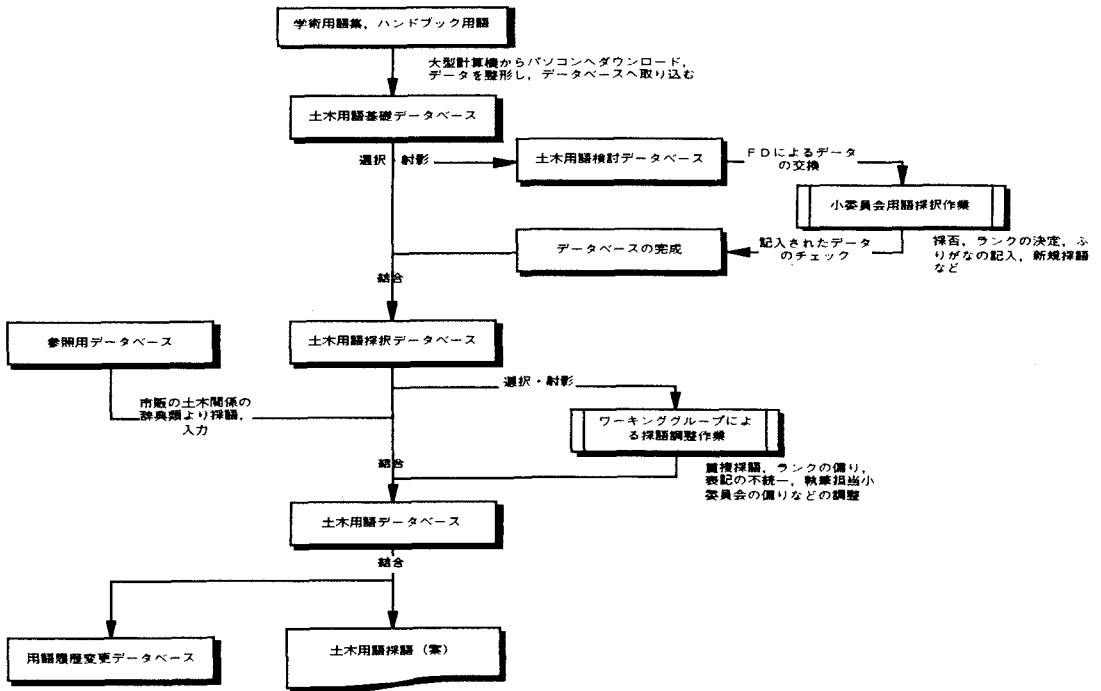


図 6 データベース作業の流れ

用することとした。図 6 は一連のデータベース処理である。

ii) 市販データベースソフトの採用

土木用語データベースのデータベースソフトには、その作業上以下の各点が要求される。

- a) パーソナルコンピュータ（日本国内で標準的に使われている NEC 製 PC-98 シリーズもしくはその互換機）上で動作し、ハードウェア能力を要求しないこと
- b) 各自が使い慣れた日本語入力フロントエンドプロセッサがそれぞれ使用できること
- c) ユーザーインターフェースが簡単であること
- d) 安価であること
- e) リレーショナルデータベース構造を採用し、編集作業の進展に伴いデータ構造を柔軟に変更できること
- f) 文字処理関数や統計処理関数さらに集計機能等のデータ処理機能を豊富に持つこと

しかしながら、時間的、コスト的な制約から、これらの要求を満たすデータベースソフトを独自に開発す

ることは不可能であった。そこで、市販の汎用性の高いデータベースソフトを利用することにした。

この場合、問題となるのがコストである。各専門小委員会がデータベースに直接アクセスするためには、専門小委員会分のアプリケーションソフトが必要となり、高価なソフトの導入は不可能である。しかし、安価であってもユーザーインターフェースが特殊であれば配布されても利用されないことが予想される。特に、データの追加、修正をデータベースに直接加えてもらうためには、使い慣れた日本語入力フロントエンドプロセッサが使用できることが重要である。

今回、上記項目を満たすデータベースソフトとして、アルゴ社が開発した「CALCBASE1」を採用した。このソフトは、完全なリレーショナルデータベースソフトではないが、基本的なリレーショナル演算が行え、豊富な文字処理関数、データ処理機能を持っている。その上非常に価格が安く、各小委員会に配布してもコストがかからないというメリットを持っている。

iii) 大型計算機からのデータダウンロード

各専門小委員会が土木用語を採録する際に、参考と

なるデータを電子情報の形で配布すれば、再入力の手間が省け、追加修正のみの作業となり、作業量が大幅に減少するなど作業の合理化が図れる。そこで「学術用語集（増訂版）」「土木工学ハンドブック（第四版）」の用語をまとめた「土木用語基礎データベース」を作成し、各専門小委員会に配布することとした。

このため、大型計算機にストックされているシソーラス構築用データをMS-DOSのテキストファイルに変換し、パーソナルコンピュータ上で「土木用語基礎データベース」を作成した。また、この時ダウンロードしたデータは正規化されていなかったためMS-DOSの汎用ツールであるSED、JGAWKを用いてデータの正規化、整形を行い、CSV形式のテキストデータファイルを作成し「CALCBASE1」に取り込んだ。

iv) 土木用語基礎データベースのデータ構造

大型計算機からダウンロードしたデータは約3万2千語である。また、土木用語基礎データベースのデータ構造は、用語番号、見出し語、フリガナ、採否カテゴリ、ランク、出典、英語等15項目から構成されるものである。

(2) 土木用語採択データベースの作成

i) 土木用語検討データベースと小委員会での検討

「土木用語基礎データベース」の作成後、専門小委員会における用語採択作業の参考となるように、小委員会が担当する分野をキー項目とした選択演算と射影演算によって「土木用語検討データベース」を作成し、データベースソフトと共にフロッピーディスクによって26の各小委員会に配布した。

小委員会では配布されたデータベースによって採否、用語の重要度のランク、修正等の検討を加え、配布された「土木用語検討データベース」上に直接記入し、データファイルを編集委員会に返送するものとし、この間約6ヶ月の作業期間を設けた。

返送されたデータファイルには、a) 記入漏れ、b) 重複採語、c) 用語ランクの偏り等のエラーがあることが予想される。そこで、編集委員会では、データベースソフトの集計、検索、並べ替え機能を用いて、それらのエラーの有無をチェックした。その結果、問題が存在する場合には該当小委員会と連絡を取り、編集委員会において修正を行い「土木用語検討

データベース」を完成させるものとした。

実際の運用では、a) ソフトの使用やデータについての問い合わせ、b) 返送されたデータファイルの不備などが見られた。さらに、軽微な修正で済んだ小委員会と大規模な修正を依頼した小委員会とが生じ、26の専門小委員会間でデータベースに対するリテラシーに差があることが見られた。

以上のように土木用語の採択結果を直接データベースに記入し、電子情報の形で交換する方法を採用した結果、データ再入力の手間が省け、チェック作業に専念でき、作業の能率化、効率化が図れた。

ii) 土木用語採択データベースの作成

全小委員会のデータベース完成後、採語された用語の全体像を知るために「土木用語採択データベース」を作成することとした。このため用語番号を結合属性として「土木用語基礎データベース」と26小委員会それぞれの「土木用語検討データベース」との自然結合演算を行った。

(3) 土木用語採択データベースの運用と土木用語データベースの構築

i) ワーキンググループによる専門分野別採語調整作業

「土木用語採択データベース」は各専門小委員会から提出されたデータベースを自然結合したのみである。このためa) 重複採語、b) ランクの偏り、c) 表記の不統一、d) 執筆担当小委員会の偏りなどが存在し、データの質や網羅性の確保がなされていない。このため、データベース全体をチェックし、その結果として土木用語データベースを作成する必要がある。

実際のチェック作業であるが、例えば、重複語のチェックは、JISコードによる並べ替え、検索で比較的簡単に行えるものと考えられるが、実際には同音意義語等が存在し、機械的に処理を行うことには限界がある。そのため、機械的に処理を行った上で、さらに一語一語の意味を考えながら調整する作業が必要となる。このことは、ランクの偏り、表記の不統一の修正、執筆担当小委員会の調整などでも同様である。そこで、26の小委員会を6の専門分野に大別して、専門家から成る採語調整ワーキンググループを組織し、調整作業を行うこととした。

このため、土木用語採択データベースの作成後、小委員会をキーとした選択演算と射影演算を行い、調整

作業用のデータベースをそれぞれの専門分野別に作成し、そのリストアウトを行った。各採語調整ワーキンググループでは、リストアウトされた用語の一語一語について、表記等の確認、重複採択用語、ランク、執筆担当小委員会等の調整作業を行った。

このように、調整作業では機械的に作業を進めることができず人手に頼ることが多く、この調整作業だけで約2ヶ月の日数が必要であった。

ii) 土木用語採択データベースによる採語全体調整作業

データの質と網羅性を確保するために、採択された用語を全体から調整する作業を行った。このためワーキンググループによる採語調整作業結果を基に、用語番号を結合属性とした自然結合演算によって、土木用語採択データベースを更新した。そして、更新されたデータベースを全体調整作業を行うための基礎データベースとした。

全体調整作業もワーキンググループによる調整作業と同様に、データベースのリストアウトを基に、専門家によって、採否、重複採択語、ランク、執筆担当小委員会の調整等の作業を行った。この調整作業では、多岐にわたる専門分野から構成される約3万語の用語を扱うことから、効率的に作業を行うためには、専門家によるFace to Faceの作業を必要とした。そこで専門家による合宿作業を行い、その結果を基にデータベースの更新作業を行った。その後、更新されたデータベースに対して、射影演算を行い「土木用語データベース」を作成した。

また、必要に応じてデータベースソフトの集計機能よりクロス集計を行い、専門分野による採択語数の偏りやランクの偏りが生じないようにチェックを行った。また、これら一連の作業において、小委員会やワーキンググループからの要請に応じて、選択演算と射影演算によりサブファイルを作成、送付を行った。

(4) 参照用データベースの作成によるデータの補完

専門小委員会によって採語作業を行う場合、いずれの専門分野にも属さないが重要な用語の採語漏れが生じる恐れがある。さらに、人為的ミスによる記載漏れなども考えられる。土木用語データベースは用語の質と同時に網羅性も要求されている。このため市販の土木関連辞典類7冊の見出し語から約2万語の「参照用

データベース」を作成した。そして、並べ替え機能を利用して複数の辞典類に採語されている用語チェックし、3冊以上の辞典類に採語されている用語を選択演算によって抽出した。さらに、用語名を結合属性とした演算による照合作業を「土木用語データベース」と行った。その結果、約200語の用語を追加採語した。これによって、複数の既存辞典に採語されている用語の採語漏れを防ぎ、網羅性を確保できた。

(5) 土木用語変更履歴データベースの作成

「土木用語データベース」の作成後、一連のデータベース更新作業を一覧できる「用語変更履歴データベース」を、用語番号を結合属性とした結合演算によって作成し、各小委員会からの問い合わせの対応や用語のデータベースからの欠落防止に用いることとした。また、執筆作業を進める上で修正されたランクなどの変更情報などを随時受け付けながら、データベースの更新作業を行った。

5. 土木用語データベースの運用結果

(1) 土木用語データベースを核とした情報管理システムの運用結果

i) 質の確保（専門性）と網羅性

学術用語集土木工学編（増訂版）、土木工学ハンドブック（第四版）より作成した土木用語基礎データベースに、専門家が追加、修正を加え、さらに編集組織においてデータのチェック機能を持たせることを行った。そして、作業手順が「情報の交換」→「協同作業によるデータのチェック」→「情報の生成」→「情報の交換」という循環系を構成するようにした。このことにより土木用語データベースの質および網羅性を確保することができた。

ii) 大量性と即時性

データベースを導入することによって、トータルで5万件に及ぶ用語情報を管理することができた。また、パーソナルコンピュータ上で扱えるデータベース管理システムとしたことにより、各専門小委員会毎に作業を進めることができた。さらに、電子情報としてデータを交換したことにより、情報の交換が比較的円滑に行われた。このため、作業日程の短縮、単純エラーの回避等が図られ作業が効率化した。

iii) 協同性

編集委員会、小委員会間の連絡には、主にファクシ

表1 演算所要時間

項目	選択	射影	結合 (1)	結合 (2)	並替え
時間	0m28s	2m30s	25m22s	4h3052s	3m08s

※ PC-H98modelU105(CPU486DX66Mhz,HDD100MByte,RAM8MByte,キャッシュメモリ-4MByte有効) 使用
 ※ データ件数23,279件
 ※ 結合 (1) は、23,279件のデータに対し、用語番号を結合属性として2,388件のデータを結合した結果である。
 ※ 結合 (2) は、23,279件のデータに対し、用語番号を結合属性として23,279件のデータを結合した結果である。

ミリを用いた。このため、速やかな情報交換と記録の保持が可能となった。また、パーソナルコンピュータでの使用を前提としたデータベース管理システムを導入したことにより、データの共有性が高まった。また、用語の意味などを考慮する必要がある場合には、専門家によるワーキンググループによって一語一語確認して対処した。

(2) データベースツールの性能評価

本研究では、一連のデータベース作業において、NEC社製PC-H98modelU105(CPU486DX66MHz,HDD100MByte,RAM8MByte)を用いた。データベース作業における演算の所要時間であるが、おおむね実用的な範囲であった。表1は、土木用語採語(案)データベース(データ件数23,279件)に対して、選択、射影、結合の各演算および並べ替えを実行した時の所要時間である。

(3) 本システムによるデータベースや手法の応用、今後の可能性、および課題

i) 新しい検索システムの導入

KWIC(Key Word In Context)索引等、データの電子情報化によって作成が容易となる検索システムを土木用語大辞典に導入することが容易となった。

ii) 土木用語の構造化・体系化への情報提供

本データベースにより、シソーラス策定への情報提供やキーワード索引による用語の類型化⁶⁾など、土木用語の構造化・体系化に対し情報提供が可能となった。

iii) 質・量を同時に確保することが必要なデータベースの構築

本研究で得られた方法は、データの質、量が共に要求されるデータベースの作成に対して有効なものと考えられる。具体的には、土木史的遺産の調査や土木事業のデータベース作成など、データが多様性に富み、個別領域の専門家の作業を集約、統一することが必要なデータベースの構築などが考えられる。

iv) ネットワークシステムの活用

連絡は主としてファクシミリを用い一部商用ネットワークによるパソコン通信も利用した。将来的には、インターネットなどによるメール交換を活用して、より一層コミュニケーションの円滑化を図る必要がある。そのためにはネットワークに関するリテラシーの向上が必要となろう。

6. おわりに

土木用語が持つ特徴を把握し、土木用語データベースを核とした情報管理システムを構築した。その結果、質を確保し、網羅性、大量性が必要な用語大辞典を短時間で編纂できるシステム運用を行えた。

【謝辞】

本研究の遂行に当たり、土木学会土木用語大事典編集委員会の幹事会、小委員会の諸先生方との議論によるところが多い。厚くお礼申し上げます。

【参考文献】

- 1) 土木学会土木用語大辞典編集委員会：『土木学会創立80周年記念出版 土木用語大辞典編集要綱』、土木学会、1993年
- 2) 高橋三雄：『情報管理学』、日本放送出版協会、1992年
- 3) 増永良文：『リレーショナルデータベース入門—データモデル・SQL・管理システム』、サイエンス社、1991年
- 4) 今 尚之、高野伸栄、佐藤馨一：『土木用語情報管理システムの構築と運用に関する研究』、土木学会北海道支部論文報告集平成5年度、1994年
- 5) K.Parsaye,M.Chignell,S.Khoshafian,H.Wong共著、近谷英昭訳：『知的データベース—オブジェクト指向・演繹・ハイパーメディア—』、オーム社、1992年
- 6) 細川寛、高野伸栄、佐藤馨一、五十嵐日出夫：キーワード索引による土木用語の類型化に関する研究、土木学会第49回年次学術講演概要集、1994年