

# 土木データベースに大いなる期待をよせて

関西大学 総合情報学部 田中成典

E-mail: tanaka@civil.kansai-u.ac.jp

## 1. まえがき

情報処理産業が著しい発展を遂げるとともに、それに纏わるハードウェア/ソフトウェアの周辺技術もめざましい進歩を遂げた。その中でも、ソフトウェアのデータベース技術が最も躍進し、関連する多くの要素技術が確立した。土木分野においても、インハウスデータベースから商用データベースまで様々なデータベースシステムと密接な関係にある。また最近では、情報処理分野のインフラ整備にともない、広域ネットワークが普及してきたため、商用の土木データベースが注目されてきた。

そこで本稿では、データベースの概念を認識する上で原点に遡り、データベースとそれに関連する用語を定義し、データベースの課題を洗い出す。次に、土木データベースに着目して、大学の利用者側から眺めた商用の土木データベースのニーズと利用状況、および理想像と将来展望について述べる。そして最後に、「官」と土木学会に大いなる期待をよせて、土木データベースの構想を提案する。

## 2. データベースの定義

データベース(database)は、情報資源(テキスト、図形、画像、音など)を効率良く収集・整理、登録・蓄積、生産・供給することを目的とする科学技術である、と定義されることが多いが、それに関連する用語の定義は曖昧である。そこで、データ、情報、知識、データ処理、情報処理、知識処理、データ工学、情報工学と、知識工学についての関連用語を定義する。

わが国では、「データ(data)」と「情報(information)」の用語が、ほとんど明確な区別のないままに使用されていることが多いが、「データ」と「情報」は全く別物である。「情報」は、「データ」を受ける相手(receiver)の存在を前提として初めて成立する概念であって、受け手は人とは限らず、機械でも何らかのシステムでも構わない。その「情報」は受け手にとって意味のある場合もあり、そうでない場合もある。「情報」に対する用語として「知識(knowledge)」があるが、認識論によれば、知識とは、正当化された真となる命題であるとも、信念(belief)のうちの真なる信念であるとも説明されている。言い換えれば、実世界に関して既に明らかになっている真なる内容を指している。「情報」の中には、「知識」が含まれていることも、「知識」の要素が含まれていることもある。情報の受け手が、「情報」の意味を解釈して意味(meaning)を付与したものが「知識」となる。よって、「データ」は、「情報」となり、最終的に「知識」へと成長する。ただし、情報処理分野では、コンピュータなどの立場から見て「データ」を処理することは、人間の立場から見ると「情報」を処理することを意味するため、データ処理(data processing)と情報処理(information processing)は、同義に使用されることが多い。

データ処理に対応する「データ工学」は、「データ」を収集・整理し、それらをモデル化して登録・蓄積し、そこから得られる「情報」を生成して、最後に「情報」を受け手に供給する一連の方法を論じた学問、まさしくデータベースについての学問である。また、情報処理に対応する「情報工学」は、「データ工学」

のみならず、ハードウェア工学やソフトウェア工学などの情報処理全般の学問を指す。最近では、対象とする世界の分析方法や把握する方法、およびそれらを記述する方法が重要視され、情報学として確立されつつある。

また、システムが、「データ」や「情報」のみならず「知識」を取り扱って問題解決を効率良く行うために、「知識」を整理し体系化した「知識情報」が注目されている。「知識情報」をシステムに組み込み、その知的財産を利用すれば、誰もが高度な「知識」を共有できる。この目的のために考案された情報処理技術が「知識処理(knowledge processing)」で、「知識情報処理」とも呼ばれている。このように人間の知的な働きと同じ働きを機械によって実現した知識情報処理システムを開発するための学問が「知識工学(knowledge engineering)」である。

### 3. データベースの課題

データベースの一般的な課題をデータモデル設計、データベース設計とメンテナンスから眺める。また、データベースの登録データの課題として信頼性について触れ、また、データベースの二次的な課題として知的権利についても触れる。

- 1) データモデル設計：伝統的なデータモデルの枠に適合しないデータを取り扱うために、将来に予想される新型のデータモデルを準備しておくことが必要である。マルチメディアのデータ（文字・数字、文章、図形、画像、映像、音）のモデリングについては、異なるデータ型、時間概念、抽象化階層などの多次元性の取り扱いや、データモデル間の関連を維持するためのインターフェイスの支援が要請される。しかし、マルチメディアの促進による多様な要求は、限り無く広がる可能性があるため、システムの機能を適切な規模に纏め上げることは難しい。したがって、マルチメディアデータベースの実現を大きく妨げる。
- 2) データベース設計：大量の多種多様なデータを蓄積するために、データベースの設計は非常に難しい。利用者からの多様な要求を満たし、社会環境の変化に適応するデータベースを設計するためには、高度な工学的アプローチが必要とされ、データ工学の重要課題に上げられる。
- 3) メンテナンス：データベースのメンテナンスには、データの追加、修正、削除の作業項目が上げられ、これらの作業には多大な労力が必要とされる。特に、新たなデータを追加するためには、データの収集・整理する作業から始めなければならない。また、データを登録・蓄積する前にデータを分析して、そのデータを加工するための前処理作業を行わなければならない。
- 4) 登録データの信頼性：データベースの登録データの信頼性は、重要な課題である。データベースの信頼性が低下するのは、データを登録する時に生じる不手際から間違ったデータが登録・蓄積される場合や、生データの不備から生じる意味的に異常なデータが登録される場合などが考えられる。よって、生データを分析し、加工しながらキーワードなどの意味付与を行ってデータを登録・蓄積するには、細心の注意を払わなければならない。また、時が経つに連れて蓄積されたデータの陳腐化による問題もある。
- 5) 知的権利：データベースに纏わる権利は、知的所有権と著作権から議論されることが多いが、そもそも、知的所有権は、特許、実用新案、意匠、商標などの工業所有権と著作権との総称である。そこには、生データの権利、加工された登録データの権利、データベースのシステムの権利など様々な知的権利が複雑に絡み合っているため、そこから生まれる利害関係がデータベースの飛躍的進歩を妨げる要因になっている。

#### 4. 大学側から考える土木データベース

土木の特性を考慮して、大学の利用者側から眺めた商用の土木データベースのニーズと利用状況について考え、またそのイメージから捉えた理想像について述べる。そして、情報処理分野のデータベースの動向と、土木データベースの将来展望についても言及する。

##### 4.1 ニーズと利用状況

大学は、商用の土木データベースから研究材料となるデータや情報、もしくは知識を入手したいと考えているが、要求にあった研究材料は、土木データベースからほとんど入手できないのが現状である。なぜなら、商用の土木データベースは実験データ、解析データや統計データなどを扱っておらず、また、公共性に富んだデータも知的権利の問題から扱うことを避けている。文献の情報に関しても、電子化された文献の一次情報を入手することができず、同様に、設計図などの図面情報や施工現場などの写真情報も入手できない。よって、大学は、研究材料を商用の土木データベースから入手できないため、ほとんど利用していないのが現状であろう。

それではどのように研究材料となるデータや情報などを入手しているかと言うと、昔ながらの文献から読み取って獲得するか、または、情報源から直接入手するかのいずれかである。

##### 4.2 理想像

土木事業は、社会における貢献度が非常に高く公共な特性を持っている。土木データベースも同様に公共性がなくてはならない。それには、社会の共通財産である土木データベースを「官」が一元管理し、社会への還元を目的とした「何時でも何処でも誰でも」がその財産を共有でき環境を作り上げなければならない。よって、土木データベースは商用データベースとして開発されるのではなく、フリーデータベースとして開発され、情報の公開を無償で行われるべきである。実際に、情報処理分野においては、フリーソフトウェアという概念が確立されつつある。土木分野においては、フリーデータという概念が確立されるべきである。

##### 4.3 将来展望

データベースに格納されているデータ、情報や知識は、伝統的な資源（人、物、金）と同じように価値があるという認識から、新しい資源（情報資源）の一つであるという考え方が定着しつつある。情報処理分野においては、この情報資源を管理するための情報資源辞書(information resource dictionary)についての構想が練られている。これは、データやプログラムなどのシステム全体を構成する情報資源の総括的な管理簿である。現在、国際標準化機構(ISO: international standard organization)では、情報資源辞書システム(IRDS: information resource dictionary system)の規格案作りを推進している。このシステムでは、システムの下部構造はプロセスではなくデータであるという考えに立ち、情報資源としてのデータは、様々な視点から収集・整理されている。また、データは、情報を生産するための最重要部品として標準化し管理される。IRDSに近い将来に開発されたとすると、データ管理の基本的な考え方が組織を越えて共通化され、そのデータの管理のみならず、データを情報や知識として利用するためのスキルの向上に大きく役立つものであると期待されている。その他の標準化動向として、データベース言語、遠隔データベースアクセス(RDA: remote database access)などが検討されている。前者は、ネットワーク型データベース言語(NDL: network database language)と関係型データベース言語(SQL: structured query language)の検討であり、後者は、OSI(open system interconnection)の環境下におけるデータベースの遠隔操作を行うためのサービ

スプロトコルについての検討である。また最近では、これらの規格案に準拠した電子図書館の実験システムの研究開発も行われている。

以上のように、情報処理分野では、様々な将来展望が検討されている。土木分野の土木データベースについても、社会動向に応じた早急な対応が望まれる。そこで、土木データベースの将来展望として以下の問題点を提起する。

- 1) 標準化設計：データの互換性を高めるために、データ形式を統一するための標準化設計の規格制定が急がれる。実現すれば、土木データベースの共有化が加速的に促進される。
- 2) インフラ整備：標準化設計を補佐するためのインフラ整備として、データ処理のためのソフトウェア開発が必要である。それは、様々なデータを標準化設計に沿ったデータ形式に整えるためのツールの充実を意味する。
- 3) 知的検索機構：従来のデータベースで主として用いられてきたキーワード検索を廃止し、全文検索手法を採用すれば、インテリジェントな情報検索(information retrieval)が可能になる。それによって、土木データベースの登録データの信頼性が高まる。
- 4) 自己組織化技術：データベースの価値は、データ量とその質との二つの側面から問われることが多いが、まず、データ量が充実することが先決である。ただし、データベースに膨大なデータを登録・蓄積させるには、データ量に比例して経費がかかる。そこで、土木データベースのデータを自己組織的に増加させるための仕掛けが必要である。自己組織化技術が躍進すれば、土木データベースは自ら成長して行く。
- 5) ネットワーク対応：インターネットのような広域ネットワークの普及とともに、データベースとネットワークは、密接な関係になりつつある。土木データベースもネットワーク対応の早急な検討が必要である。それは、電子図書館の構想を念頭に据えながら検討しなければならない。
- 6) 権利制限：権利は主張されて然るべきものであるが、権利に重きを置くと土木データベースの飛躍的な発展は望めない。権利の制限が検討されるべきである。現行制度のもとにおいて、権利の行使と社会公共との調和について検討されるべきである。また、土木データベースにおいては、フリーデータ概念を確立させるべきである。

## 5. あとがき

土木データベースの特性から、「官」の役割は非常に重要である。なぜなら、前章でも述べたように土木データベースを開発するには膨大なデータを収集・整理し、それをデータベースに登録・蓄積して構築しなければならない。これらの作業は、言うまでもないが莫大な予算が必要である。しかし、土木データベースは、社会に還元されることを主目的としているため、情報の公開は無償で行われることが望ましい。そのため、利益を追求する「産」にそれを押しつけるのは不条理である。また、「学」は、マンパワーも予算もない。このような状況下で「官」が、積極的に土木データベースについて取り組まなければならない。「官」の旗揚げ如何によっては、「産」と「学」は、それぞれの範疇で協力を惜しまないであろう。

また、土木情報システム委員会の下部組織である土木データベース小委員会に対して、土木データベースに関する規約を早期に制定することを提案する。最後に、今後の土木データベースの動向に大いなる期待をよせる。