

リレーショナル・データベースによる非計量データ処理について *

— 史的情報を例として —

Non-Numerical Data Processing based on Relational Database

— a case of Historical Data —

中岡良司^{**}, 森 弘^{***}, 五十嵐日出夫^{****}, 佐藤 馨一^{*****}

by Ryoji NAKAOKA, Hiroshi MORI, Hideo IGARASHI and Keiichi SATO

Relational Database is characterized by storing data in the form of a table. Processing of the data is done with relational operation. The relational operation remakes the table formation without changing, and irrespectively, of the contents of data. In this paper we will introduce a relational database taking up the data of Japanese sentences on history of civil engineering. The result enables us to get intended data freely from the enormous database and to study the history from various viewpoints.

1. はじめに

今日、パーソナル・コンピュータの実用化とリレーショナル・データベースの登場によって、我々はコンピュータによる情報処理に新たな局面を迎えようとしている。

その第1の特徴は、パーソナル・コンピュータがかな漢字を含む日本語処理に多大な能力を発揮し、本格的な情報処理時代の門戸を開いたという点である。従来、数値解析が主体の大型コンピュータにおいては、日本語処理はさほど重要な要素ではなかつ

た。しかしながら、わが国においては、表現力において（すなわち情報量において）日本語に勝るものはない。日本語処理を当初から考慮したパーソナル・コンピュータの出現によって、我々は漢字を含む多様な情報をコンピュータで容易に取り扱うことが可能となった。

第2の特徴は、情報の効率的な管理・運用に不可欠なデータベースが個人的なレベルで構築できる環境が整ってきたという点にある。従来、我々はデータベースを利用する者であり作成する者ではなかつた。データベースの作成には複雑な設計理論を必要としたからである。しかしながら、リレーショナル・データベースでは、表を作成するという単純なイメージで、しかもその規模を問わずデータベースの作成が可能である。データの内容は、数値に限らず何でもあっても良い。そして、強力なリレーショナル

* キーワーズ;リレーショナル・データベース,非計量

** 正会員 北見工業大学助手 工学部開発工学科 データ

*** 正会員 工修 北見工業大学助教授 工学部開発工学科

**** 正会員 工博 北海道大学教授 工学部土木工学科

***** 正会員 工博 北海道大学助教授 工学部土木工学科

演算機能によって、我々は必要とするデータだけを迅速に手に入れることができる。

さて、土木計画学の分野における計画情報は極めて広範囲に及ぶ。そして、その情報の形態は様々である。たとえば、意識情報において、問題設定において、計画目標において、意志決定プロセスにおいて、そしてあらゆる計画の基本となる歴史情報において、数値として現れないデータの形態や処理方法に十分な合意が得られている訳ではない。ただ、これら非計量データに、共通する情報形態として日本語が重要であることは明らかである。

そこで、本研究では、土木の歴史に関する日本語文章そのままをデータとする土木史データベースを構築し、非計量データに関するリレーショナル・データベースの有効性を検討することとした。その結果、我々は、土木史データベースから必要なデータを任意に選択し自由な史観で土木の歴史を眺めることができることを確認した。具体的な成果の一部は、既に他の機会に報告済みではあるが、本論文では改めて統一した論理の下でそれらも取り上げた。本論文で新たにその応用を展開するのは、主にリレーショナル演算における JOIN の機能についてである。

2. リレーショナル・データベース (Relational DataBase ; RDB)

リレーショナル・データベースは、数学における関係 (relation) の概念をデータ操作に応用したものである。その提唱は1970年に E. F. コッド (米、IBM) によってなされたが、当時のコンピュータの性能に限界があったため、実際にシステムが稼働したのは70年代後半になってからであった。その後、80年代にはパーソナル・コンピュータでの利用が可能となっている。

リレーショナル・データベースは、データベースが本来必要とする基本的な条件、

- ① 複数の利用者が共用できるデータ群であること
- ② データ群へのデータ入力・更新・削除などデータ管理機能を持つこと
- ③ データ群を処理するプログラムとデータ群とは独立していること

④ データベースの利用言語を持つこと

に加え、先行するネットワーク型データベースや階層型データベースに対し以下の特徴を持っている。

⑤ データベースの論理的構造と物理的構造が独立している

従来のデータベースはデータの論理的構造 (どのようなデータか) と物理的構造 (どのような形で記憶されているか) が一体化されていたために、目的プログラムの作成には両者を意識しなければならなかった。リレーショナル・データベースにおいては、物理的構造はコンピュータ側で管理され、利用者はどのようなデータを対象とするかだけを意識すれば良い。

⑥ データの関係は簡単な2次元の表として表現される

リレーショナル・データベースの論理的構造は、2次元の表で表現される。たとえば、ある学生の成績は表-1の形で表現される。この表には、氏名、数学、国語の間に3つの関係が含まれている。データベースとはこのような表の集合体と考えればよい。関係のこのような表現は、人間の知識の形態や論理的思考をそのまま反映した極めて分かり易い表現形式といえる。

表-1 リレーショナル・データベースの構造

項 目 →			
レ コ ー ド ↓	氏 名	数 学	国 語
	伊 藤 信 之	90点	65点
	岩 崎 有 男	78点	84点
	北 川 秀 樹	65点	60点
	鈴 木 秀	95点	87点
	田 中 英 治	74点	83点
	中 島 光 一	45点	62点

⑦ データの集合演算機能を持っている

リレーショナル・データベースでは、数学的な集合概念に基礎を置きリレーショナル演算が可能である。このリレーショナル演算では、SELECT、PROJECT、JOINの3つの機能が基本となる。SELECTとは、特定の表の中から1つ以上の制約条件をつけてレコードを抽出し新たな表を作成することである。PROJECTとは、特定の表からいくつかの項目を抜き出して新たな表を作成することである。JOINとは、複数の表から共通の項目をもとにして新たな表を作成することである。これら3つの機能に関しては、本研究のデータベースを用いて、後に具体的に示すこととする。

3. 土木史データベース

(1) 土木史データベースの内容

本研究では、非計量データの一例として土木史データベースを取り上げる。言うまでもなく、土木史の対象は広く史料は膨大である。したがって、それらを広範囲に収録したデータベースが既に在るわけではない。その必要性を痛感し本研究が独自に作成したものである。本研究における土木史データベースはそのデータ・ソースを一冊の本に拠っている。すなわち、総合研究開発機構が土木学会に委託して作成した「土木技術の発展と社会資本に関する研究」（昭和60年7月刊）である。同書は、「今日までの土木技術の発展とそれを支えた社会経済的背景、および土木技術の発展が社会資本に与えた影響を解明し、今後の土木技術の発展の方向性を見い出そうとする」目的で執筆された全六百余頁に及ぶ大部作である。そこには、明治期30年間と昭和期30年間の土木教育、土木事業、建設技術、施工技术などが幅広く取り上げられている。本研究では、その内、主に土木施設に関する事項を中心に2055件の文章を抜き出し土木史データベースとした。

(2) 土木史データベースの設計

リレーショナル・データベースの構築には、まず、

項目の設定とその内容の設計が必要である。今回のように、原データが日本語文章である場合、そこには2通りの方法が考えられる。

その一つは、共通する内容に関し可能な限り項目を設けて、作成者の判断でデータを埋めてゆく方法である（図-1）。この方法に拠るならば、後のデータ処理が容易になることは明らかであった。ただし、前述した広範囲なデータに共通の項目を設けることは困難であること、および入力データの作成に大変な時間を要するという問題もあった。

もう一つの方法は、要約した文章そのままをデータとして入力する方法である（図-2）。この方法では項目数は大幅に減少する。データの作成も容易である。しかし、日本語文章そのものをコンピュータでどのように取り扱えるかに関しては検討を要した。

ファイルの名称 B : KOU.DAT		
=====		
1. 名 称	;	逢坂山トンネル
2. 目 的	;	鉄道トンネル
3. 線 名	;	東海道本線
4. 駅名1	;	京都
5. 駅名2	;	大津
6. 延 長	;	665 m
7. 着工年	;	1878年
8. 完成年	;	1880年
9. 開業年	;	1880年

図-1 項目型入力データの例

ファイルの名称 B : BUN.DAT		
=====		
1. 見出し	;	逢坂山トンネル
2. 記 事	;	

2. 記 事		
京都から大津に向かって25/1000の下り片勾配のわが国最初の本格的な山岳鉄道トンネルで、明治13年6月に建設された。このトンネルは、外国人技師依存から脱して、わが国の技術者によって監督施工され、土木技術者自立の過程で画期的な役割を果たした。		

図-2 文章型入力データの例

結局、本研究においては、この後者の方法を採用したが、その背景には、日本語ワードプロセッサに見られるようにコンピュータへの日本語入力が大幅に改善されたという技術的側面もあった。また、あらゆる情報を内に取り込む日本語の柔軟性や、そのまま出力しても十分に意味が通る文章データの判明性は十分に魅力的であった。さらに、仮に、前者の方法において必要な項目が明確になった段階では、そのデータを後者のデータから作成することは可能であると考えた。

4. リレーショナル演算機能

リレーショナル・データベースでは、数学的な集合概念に基礎を置くリレーショナル演算が可能である。ここでは、作成した土木史データベースを対象に、その実際の適用例を示してゆく。

リレーショナル演算は、SELECT、PROJECTのように関係表内部の演算とJOINのように複数の関係表を操作する外部演算に分けられる。

(1) 内部演算 - SELECT, PROJECT, SORT

① レコードの検索

おそらく、データベースに対する最初の要求は、そのデータベースから必要なデータを取り出すことであろう。SELECTは、与えられた条件に該当するデータから新たな表を作り出す機能である。いま、土木史データベースにおいて、次のような条件を与えたとする。

* 【見出し】あるいは【記事】の項目データの中に、「工部大学校」という用語があるデータを全て抜き出せ。

その実行結果が、表-2である。注目を要するのは、「工部大学校」の用語が【記事】中のどこに在ってもよいという点である。このような文書中の部分的な検索は、同様な事を人力で行う場合と比較すると、その効果は明らかである。また、歴史データの場合、膨大な史料から任意の用語に関連する全てのデータを引き出し、自由な史観で歴史を構築できる意義は大きい。

② 項目の抽出

PROJECTは、関係表から必要な項目を抜き出し新たな表を作成する機能である。いま、次のような条件を与えたとする。

表-2 「工部大学校」による検索結果

No.	見出し	引用箇所	記 事
0022	工学寮	023-14	1871年工部省所管の下に東京虎の門内の門内旧延岡藩邸に設けられた。
0145	工部大学校	024-12	1877年には工部大学校と改められる。
0186	工部大学校	024-12	1877年に工学寮を改組拡充して開設された。初代校長には工作局長大島圭介が兼務し、教頭にはH.ダイヤを当てた。工部省が運営管理する官営企業（鉾山寮、鉄道寮、電信寮、制作寮）へ人材を供給することを、第一の目的として設立された。
0521	工部大学校 田辺朔郎	043-05	1885年に工部省が廃止されると同時に文部省に移管された。
0622	工手学校	055-01	1883年に工部大学校を卒業。卒業後に京都府御用掛となり、疎水工事の設計監督にあたった。1890年に帝国大学工科大学教授。1900年に理工科大学土木工学科教授。
1273	笹子トンネル	488-01	1886年設置。工部大学校の卒業生が中心となって設置したもの。土木、機械、電工、造家、造船、採鉱、冶金、製造倉庫の8学科を有した。また民間の寄附金によって設置されたわが国最初の私立工業大学でもあった。
			工部大学校1884年卒古川阪次郎が八王子鉄道局出張所長時代に計画実施。日本初の自家水力発電による工事。1931年8月に清水トンネルが竣工するまで日本最長のトンネルであった。

* 表-2から、[No]、[見出し]、[引用箇所]の項目から成る関係表を作成せよ

結果を表-3に示す。このような機能により、我々は、大きなデータベースを毎回取り扱う煩雑さから解放され、その時に必要なデータのみを操作する便宜を得ることができる。

表-3 項目の抽出 — PROJECT

No.	見出し	引用箇所
0022	工学寮	023-14
0145	工部大学校	024-12
0186	工部大学校	024-12
0521	田辺朝郎	043-05
0622	工手学校	055-01
1273	笹子トンネル	488-01

③ レコードの並び替え

レコードをある基準で並び替えると、内容の照合や関連事項の整理が容易になる。とりわけ歴史データの操作には欠かせぬ機能と言えよう。この機能、すなわちSORTは、リレーショナル・データベース特有の機能ではないが、いずれのソフトウェアにも備えられている基本的な機能なので、ここに例示しておく。表-4は、表-3の項目[見出し]をキー項目としてデータを並び替えたものである。この例が示すように、SORTは、数値に限らず文字列に対してもその文字コードから並び替えを行うことができる。

表-4 データの並び替え — SORT

No.	見出し	引用箇所
0022	工学寮	023-14
0622	工手学校	055-01
0145	工部大学校	024-12
0186	工部大学校	024-12
1273	笹子トンネル	488-01
0521	田辺朝郎	043-05

(2) 外部操作 — JOIN

④ 他のデータベースとの結合

リレーショナル・データベースの利点は、データ内部の強力な機能に留まらず、共通の項目がただ一つあれば他のリレーショナル・データベースと任意に結合できる点にある。この機能をJOINという。これは、リレーショナル・データベースの論理構造が単純な表形式だから可能なことである。

ここでは、土木史をより広い視野から眺めるために、日本史年表と結合させることを考えてみよう。そのために、簡単な日本史年表をリレーショナル・データベースで作成した。

* 土木史データベースに[西暦]の項目を追加し、それをキー項目として日本史データベースと結合せよ

この課題をそのまま実行すると膨大な作業となるため、ここでは土木史データベースのうち「佐久間ダム」の関連データに限って実行した。結果の一部を表-5に示す。土木史を研究する目的の一つは、このような総合年表を作成することにあるが、それが任意の事項に関して必要な項目からなる年表が作成できることを示した。

⑤ リレーションによる検索

JOINは、複数のリレーショナル・データベースを結合する機能であるが、検索機能を拡張する働きをさせることもできる。すなわち、用語そのものによる検索に対し、その用語の内容を含めた検索を可能とする。

例えば、前出の「工部大学校」による検索の例を考えてみよう。SELECTにおいては、「工部大学校」の文字列そのものによって検索を行っていた。しかし、我々が求める工部大学校の意味は単なる文字としての「工部大学校」にとどまるであろうか。仮に、工部大学校に関して、事前に知識を持っていたなら（たとえば、前身が工学寮で後に帝国大学工科大学に改組したこと、卒業生は二百十三名である

ことなど）、あるいは、表－２のような形でそれらの知識を得たならば、工部大学校の関係資料の検索はより幅広いものとなるはずである。その時、仮に必要な史料がリレーショナル・データベースの形で蓄積されていたなら、我々は、JOINによって関連するデータの一切を手にすることが可能となるのである。

残念ながら、現在の土木史データベースからこのような拡張検索の例を示すのは難しい。ここでは、あまり現実的ではないが、土木史データベースから作成したダム史データベースを用いてその例を示したい。なお、ダム史データベースは〔ダム名〕、〔河川名〕、〔竣工年〕の項目から作成されている。また、ダムに関して不足するデータは「ダム総覧」より追

加登録している。

ここで、次のような課題を設けてみよう。

＊ ダム史データベースから佐賀県にあるダムを検索せよ。

この場合、ダム史データベースには県名のデータが無いので、SELECTによる検索は不可能である。そこで、前述したリレーションによる検索を行う。そのために必要となるデータベースは別に用意した。すなわち、県と水系がまとめられた「県・水系データベース」と水系と河川がまとめられた「水系・河川データベース」である。

表－５ 「佐久間ダム」総合年表

西暦	年 号	科学技術	(施設名)	土 木 史 デ ー タ (記 事)	内 治	外 交	経 済
1953	昭和28	NHKテレビ放送開始 民間テレビ放送開始	佐久間ダム 佐久間ダム 佐久間ダム 建設工事 佐久間ダム 大原トンネル 佐久間ダム 仮水路トンネル 大規模貯水地式発電所改良	人力から大型機械により建設された初めてのダム アメリカ・アトキンス社の技術指導 大型機械はアメリカより輸入 この工事における外国からの大型施工機械の導入が我が国の機械メーカーを刺激し、その後の国産建設機械開発を促すきっかけとなった 1953年着工、米国アトキンス社の技術指導、徹底的な機械化施工、米国製大型機械の活躍 アメリカより大型施工機械および工事用プラント類を導入し、従来の人力主体の施工では不可能であった難工事を短期間に施工し、本格的な機械化施工への道を開いた 鋼製支保工の採用により全断面工法を採用、1953年佐久間ダムの建設のための飯田線付替工事に伴う作業 1953年、米国の機械化全面掘削工法の導入	町村合併促進法公布		経済自立三目標・四原則
1955	昭和30	第一回紫綬褒賞授与（発明改良）	佐久間ダム	1955年－1965年 佐久間ダム、黒部第四ダム	原子力基本法成立		ガット加入決定
1956	昭和31	原子力委員会発足 科学技術庁発足 東海道本線電化完成	佐久間ダム 佐久間ダム 佐久間ダム 佐久間ダム	1953年－1956年の間に天竜川中流に建設高さ155m 発電用重力式コンクリートダム 五十里ダムより40m余り高い世界的規模のダム 1956年4月完成 堤高155m コンクリート重力 全貯水量326848000立方m 有効貯水量205444000立方m 堤高155.5m 堤長293.5m 堤体積1120000立方m 1956年4月施工 形式C・G 堤高155.5m 堤頂長293.5m 堤体積1120000立方m 労働災害死亡数52人 死亡数／堤高＝0.33 佐久間発電所 パワーショベル59台 ダンプトラック90台 D-9およびD-8などのブルドーザ40台など 佐久間ダム以来急速に導入しあるいは開発した国産の建設機械を駆使する近代土木施工技術を発揮	日ソ共同宣言調印 国際連合加盟成る		百貨店法成立
1958	昭和33	バチスカーフ女川沖で潜水新特急こだま号運転開始	御母衣ダム		インドネシアと平和条約 ポーランドと通商条約		〇なべ底景気 一万円札発行

これらのデータベースに対し、図-3に示すようにJOINを重ねることによって、目的とするデータベースが得られた。ただし、今回の例では、同じ名称の河川や水系があるため重複した出力もあった。結果の一部を表-6に示す。

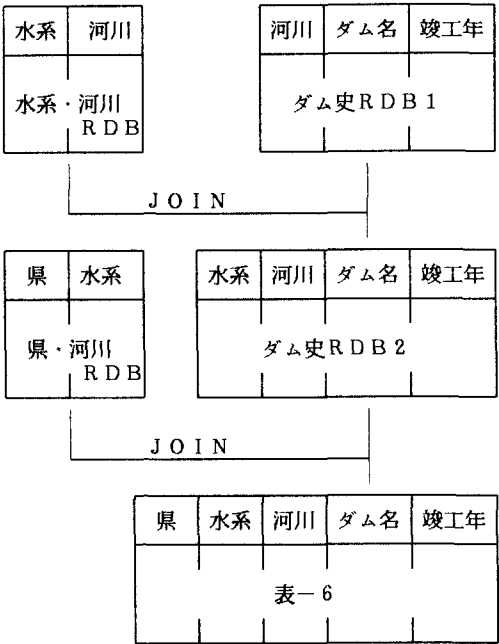


図-3 リレーションによる検索プロセス

表-6 佐賀県ダム一覧

県	水系	河川	ダム名	竣工年
佐賀	松浦川	鷺木川	鷺木	198X
佐賀	筑後川	城原川	城原	198X
佐賀	嘉瀬川	嘉瀬川	嘉瀬川	198X
佐賀	六角川	天山川	天山	1986
佐賀	鹿島川	中川	中木庭	198X
佐賀	塩田川	岩屋川内川	岩屋川内	1974
佐賀	嘉瀬川	嘉瀬川	北山	1956
佐賀	松浦川	伊岐佐川	伊岐佐	1979
佐賀	有浦川	藤の平川	藤の平	198X
佐賀	塩田川	吉田川	春日	198X

(全107ダム)

このように、JOINにおいては、データベース間に包含関係があるならば、その関係によって検索が可能なのである。

5. 応用プログラムの開発

リレーショナル・データベースでは応用プログラムの開発が容易である点を評価する者も多い。

そのプログラム言語として、大型コンピュータではSQL（構造化照会言語）に代表される「準言語」の利用が可能であるが、パーソナル・コンピュータのレベルにおいてはADL（アプリケーション開発言語）が提供されている場合が多い。このような言語によって、リレーショナル構造をとりながら、特定の利用目的に合わせた専用ソフトウェアを作成することが可能である。

一方、これらの専用言語を用いることなく、データを自由に操作する方法もある。パーソナル・コンピュータでは最も普及しているBASIC言語を用いる方法である。リレーショナル・データベースのデータ構造は極めて単純なので、そのプログラムの作成には若干のプログラム経験があれば足りる。ここでは、リレーショナル演算では出来ないデータの加工に関して2つの例を示す。

① データのローマ字変換

先に、データの並べ替えの例を示したが、漢字の読みは複雑なため正しく五十音順に並べられないことが多い。そのような場合には、一般に、[カタカナ]の項目を設けてその読みで並び替えるが、事典や用語集においては、読みとして[ローマ字]を付記することも多い。

そこで、[カタカナ]を[ローマ字]に変換するプログラムを作成した。そのプログラムの内容を紹介することは本論の目的ではないため割愛するが、データの読み込みは通常のファイル入力と同様である。そして、変換したデータを再び入力と同じ構造で書き込めば、新たに[ローマ字]項目を追加したリレーショナル・データベースを作成することができる。

(2) データ内容の項目分け

リレーショナル・データベースでは、項目内のデータ内容そのものは操作できない。そのため、たとえば、ダム技術の進歩を堤高と建設年の関係から調べたいと思っても、それが項目として独立していなければ、リレーショナル演算は使えない。

そこで、文章データの内容を項目別書き改める

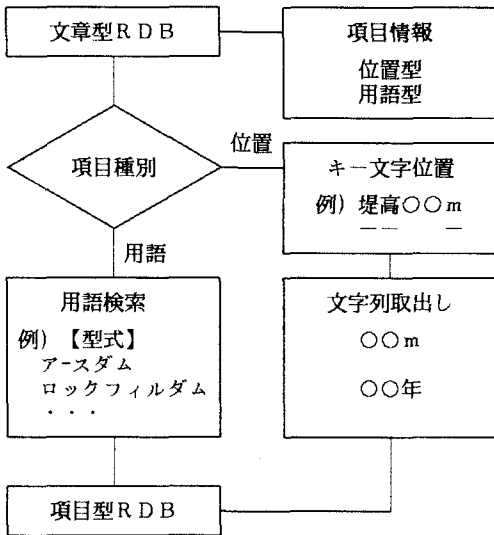


図-4 項目型RDB作成フロー

ファイルの名称 B : DAM. DAT	
1. 名 称	佐久間ダム
2. 河 川	天竜川
3. 型 式	コンクリート重力 C. G
4. 目 的	発電用
5. 堤 高	155.5 155
6. 堤頂長	293.5
7. 堤体積	1120000
8. 着工年	1953
9. 竣工年	1956

図-5 項目分けされたダム・データ

プログラムを開発することにした。現在、このプログラムで処理可能なデータは限られているが、ここでは、応用プログラムの一例として紹介する。概略的な作成フローを図-4に、そして表-5の佐久間ダム関連データから作成した項目分けの例を図-5に示す。項目3および項目5に見られるように、該当するデータが複数ある場合は、そのまま併記するようにしている。

6. おわりに

リレーショナル・データベースの歴史は浅い。とりわけ、わが国のパーソナル・コンピュータ上でそれが稼働を始めたのは、ここ1, 2年のことである。しかし、その普及は目ざましく、現在では、ソフトウェアの動向を注目する段階からその応用へ関心が向けられる段階に至っている。

リレーショナル・データベースで一体何ができるのか。この疑問こそ本研究の最大の関心である。本論文においては、土木史料を用いて、現時点で可能なことのほぼ全容を報告してきた。しかしながら、本研究で作成した土木史データベースは、まだ規模も小さく、リレーショナル・データベースの効果を十分に発揮しているわけではない。今後は、様々な非計量データへの応用を検討するとともに、分散的なデータベースの構築が可能なリレーショナル・データベースの利点を生かして、より多くの研究者と共同研究を進めて行きたい。

参 考 文 献

- 1) 佐藤馨一・五十嵐日出夫・堂柿栄輔・中岡良司、明治以前日本土木史年表の試作について、第4回日本土木史研究発表論文集、1984.6
- 2) 田中重夫・佐藤馨一・五十嵐日出夫・中岡良司、リレーショナル・データベースによる土木史情報の体系化について、土木学会北海道支部論文報告集第41号、1985.2
- 3) 中岡良司・佐藤馨一・五十嵐日出夫、リレーショナル・データベースによる土木史情報支援システムについて、第5回日本土木史研究発表論文集、1985.6
- 4) 佐藤馨一・田中重夫・五十嵐日出夫・中岡良司、パーソナル・コンピュータを用いた土木史データベースの構築について、第40回土木学会年次学術講演会講演概要集、1985.9