

## 技術革新に関する企業内資料の調査研究(4)

### —全文検索を利用した資料分析法—

## A survey of corporate documents concerning technological innovation (4)

### Documents analysis method using full text retrieval

下川 雄一<sup>1</sup>・山崎 幹泰<sup>2</sup>・中森 勉<sup>3</sup>・竺 覚暁<sup>4</sup>・丹治 明<sup>5</sup>・松野 建一<sup>6</sup>

SHIMOKAWA Yuichi<sup>1</sup>・YAMAZAKI Mikihiko<sup>2</sup>・NAKAMORI Tsutomu<sup>3</sup>・CHIKU Kakugyo<sup>4</sup>・TANJI Akira<sup>5</sup>・MATSUNO Kenichi<sup>6</sup>

デジタルアーカイブ、工作機械、画像ファイル、データベース  
Digital Archive, Machine Tool, Pictorial Data, Database

### はじめに

本研究は、社会や文化のニーズと技術革新との相関関係を知るための基礎資料を蒐集、組織化し、知識基盤として公開することを目指している。ここでの基礎資料とは、製品カタログなどの販促宣伝資料、仕様書・規格・技報・設計図などの技術資料、企画書、開発資料などの社内資料、などを指している。我が国の基幹産業技術企業における、主として昭和戦後期のこうした資料を蒐集して、デジタル画像化して記録し、さらにこの画像資料に適切な資料組織法、検索システムを開発して、画像データベースに組織し、研究用に公開することを目的としている。

平成18年度は、前年度に構築したシステムに基づき、工作機械分野の企業から借用した資料のデジタルアーカイブ化を進めた。また、今年度からは日本工業大学工業技術博物館と協同し、同博物館が構築を進めている同館収蔵の工作機械文献資料のデータベースとデータを共有することとなった。一方、工作機械産業と関連して、自動車産業の画像データベース作成も開始した。現在はトヨタ博物館の協力により、同館が所蔵する自動車カタログのデジタル画像化を進行中である。

本稿では、現在までのデータベース構築作業の進行状況と、データベースの試験利用（キーワード検索等）を通して抽出された問題点やその対応方法に関して報告を行う。

### データベース構築作業の状況と課題

これまでに、工作機械分野の企業14社から借用した資料（約5,000点）のスキニングを終え、現在はその内4社分のデータベース登録を終えた段階である。トヨタ自動車の資料に関しては43車種分のカ

タログ資料のデータベース登録を終えている。前回までの報告で述べたように、登録された資料はPDFファイル（以下、ファイル）となっており、企業名>資料群名>資料名というように、蒐集した資料の包含関係を階層的に保存することができるようになっている（図1）。



図1 データベース登録された資料の階層構成

### 全文検索による資料内容の統計的な把握

登録された資料は自動OCR処理されているため、キーワードによる横断的な全文検索が可能であり、任意のフォルダ内（企業内、資料群内など）に絞った検索にも対応できる。

データベース登録を終えた工作機械企業4社の資料について、簡単なキーワード検索を含め、資料の内訳を整理したものを表1に示す。表1上部の資料形態の内訳から、資料の大半が製品カタログやパンフレット等であること分かる。一方で、開発・設計

1.2. 金沢工業大学 博士(工学)

3. 金沢工業大学 工修

4. 金沢工業大学 工博

5. 日本工業大学工業技術博物館

6. 日本工業大学工業技術博物館 工博

1.2.4 Kanazawa Institute of Technology, Dr. Eng

3. Kanazawa Institute of Technology, M. Eng

5. Nippon Institute of Technology, Museum of Industrial Technology

6. Nippon Institute of Technology, Museum of Industrial Technology, Dr.Eng

関係文書や実物写真等もまとめて入手されるケースも見られる。

表1 工作機械企業4社の資料摘要

	マシナリー シチズン	テクノワシノ	村田機械	東京精機 工作所	計
資料 PDF	280	82	34	8	404
ブフォルダ	10	0	2	0	12
資料形態の内訳 <sup>1</sup>					
写真(実物)			1		1
写真( )					
製品	223	81	1	8	32
告・					
開発・設計関係文書	56				56
・図					
図書		1			1
その	1				1
キーワード検索 (PDF ) <sup>1</sup>					
1 盤	8	14	13		116
2 研 盤	41	11		6	58
3 フ イス	23			1	24
4 中 り	14				14
5 ール	43			1	44
6 マシニング ンタ	8	1			
計	218	26	13	8	

<sup>1</sup> 記入は0とする  
 PDF 1つにつき、いれかを するようにしている  
 企業 の全資料を対して全文検索を行った

製品カタログやパンフレットについては、ないもので ージ、いものでは ージとなる。これらや開発・設計関係文書が いど、テキスト 報が え、全文検索による分 がより 的となる。 え、表1下部は、工作機械に関するキーワード検索を行った (16のそれ れで ットしたファイル) を示している。これにより、資料内 や (どのような種 の工作機械の資料がどの 度存在するか) を大まかに することができる。このような資料内 に関する 計データは、より な分 に り かる前の 作業として有 であると えられる。

#### データベース検索の機能性の改善

全文検索を試験的に り 途中で、その利点や今後の 題等が明らかになった。利点としては、上記した通り、キーワード に ットしたファイル や、それらファイル群の階層内での分 を見ることにより、登録された資料の全 像を し なるという点である。これに対し、 入のシステムでは、検索方法や検索 の表示方法がま 要 のものしかない、表1下部のような単 な集計 を作成する けでも、(キーワード 企業) の回の検索作業が 要である。その、下記のような問題が られる。

- ・ のキーワードを一 して検索できない。
- ・ の企業( のフォルダ)を一 して検索できない。
- ・ 1つのPDFファイル内でのキーワードの ットが抽出されない。
- ・ 一 表示や検索 においてPDFファイルの が表示されない。
- ・ 表やグ フ等の ジアルな検索 表示に対応していない。

以上のように、現 点での検索システムは資料内の分 という点では 分であり、今後の大 な が められる。

#### まとめ

これまで資料の蓄積という 点から、データベースシステムの とデータの蓄積を 行して進めてきた。データベース登録の機能についてはその仕様がある 度 定したので、今後は 登録のスキニング画像を データベースに登録してい と共に、全文検索機能を応用した資料内 の分 機能を さ てい 要がある。

一方、分野における資料分 のためのキーワードの抽出・整理やシ ースの利用も今後の 題である。現在は工作機械分野に いで、自動車産業のデータベース構築に 手した段階であるが、自動車製品の内部には工作機械で 作された部品も 組 まれており、実態としては、産業 の がりがある。分 の際には、そのような産業 の関係構 や技術革新の連関を か 上がら るような形で分 することも 定する 要がある。