

保存本

01-開-03

データベース構築促進及び技術開発に関する報告書

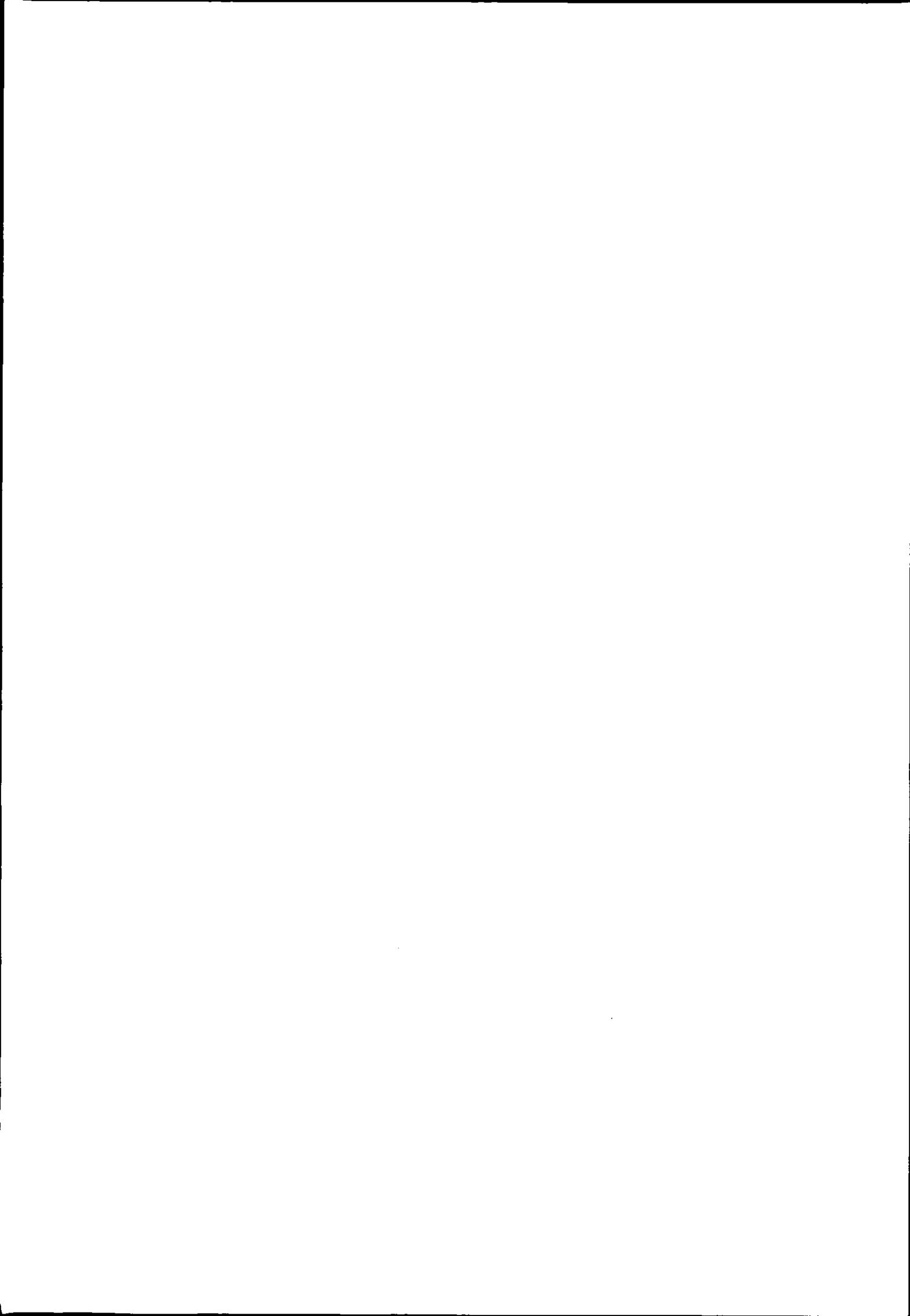
# 災害情報システムの構築

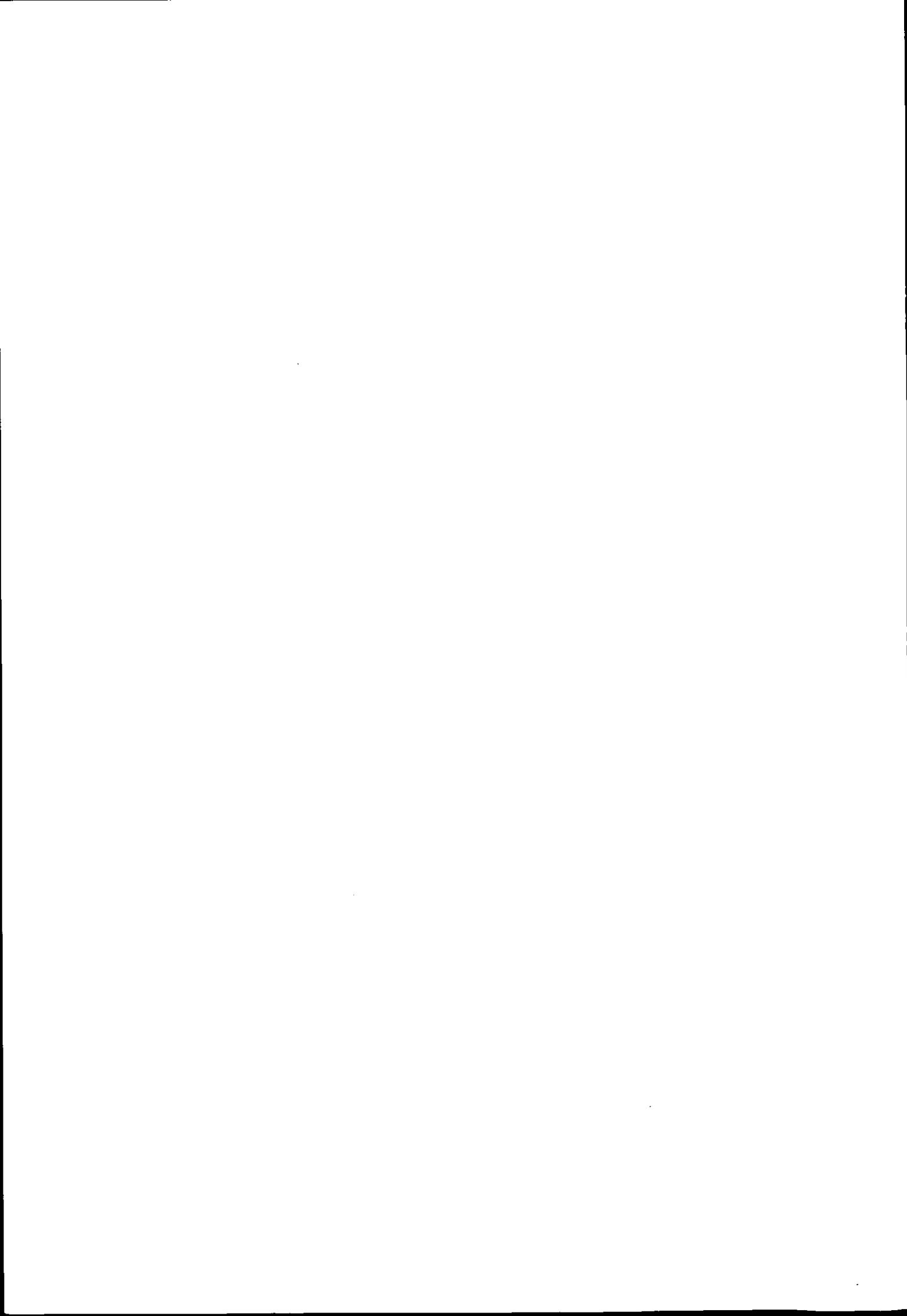
平成2年3月

財団法人データベース振興センター

委託先 株式会社防災都市計画研究所

本報告書は、日本自転車振興会から競輪収益の一部である機械工業振興資金の補助を受けて作成したものである。





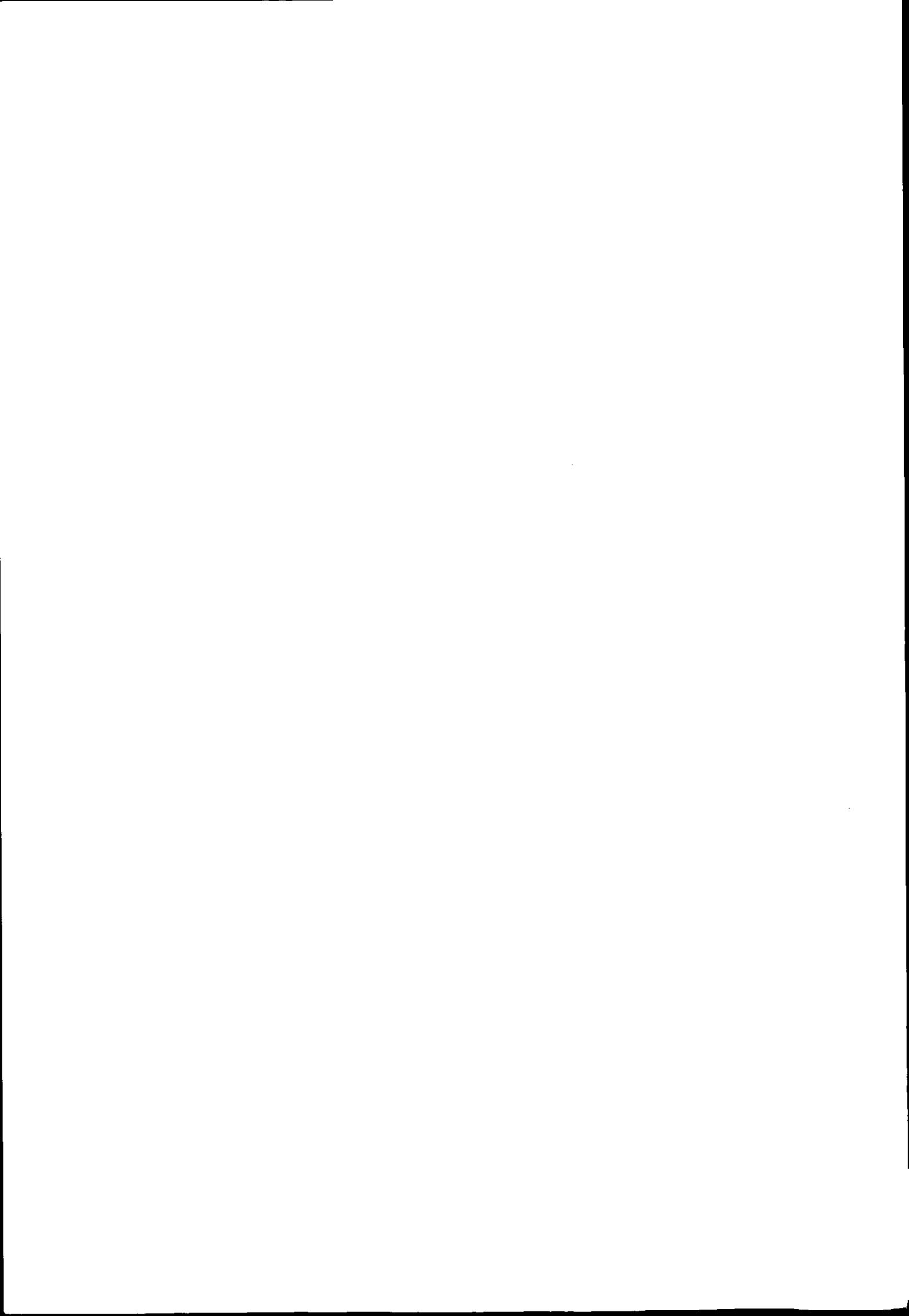
データベース構築促進及び技術開発に関する報告書

# 災害情報シソーラスの構築

平成2年3月

財団法人 データベース振興センター

委託先 (株)防災都市計画研究所



## 序

データベースは、わが国の情報化の進展上、重要な役割を果たすものと期待されている。今後、データベースの普及により、わが国において健全な高度情報化社会の形成が期待される。さらに海外に対して提供可能なデータベースの整備は、国際的な情報化への貢献および自由な情報流通の確保の観点からも必要である。しかしながら、現在わが国で流通しているデータベースの中でわが国独自のものは1/4にすぎないのが現状であり、わが国データベースサービスひいてはバランスある情報産業の健全な発展を図るためには、わが国独自のデータベースの構築およびデータベース関連技術の研究開発を強力に促進し、データベースの拡充を図る必要がある。

このような要請に応えるため、(財)データベース振興センターでは日本自転車振興会から機械工業振興資金の交付を受けて、データベースの構築および技術開発について民間企業、団体等に対して委託事業を実施している。委託事業の内容は、社会的、経済的、国際的に重要で、また地域および産業の発展の促進に寄与すると考えられているデータベースの構築とデータベース作成の効率化、流通の促進、利用の円滑化・容易化などに関係したソフトウェア技術・ハードウェア技術である。

本事業の推進に当って、当財団に学識経験者の方々に構成されるデータベース構築・技術開発促進委員会(委員長 東京工科大学教授 西野博二氏)を設置している。

この「災害情報シソーラスの構築」は平成元年度のデータベースの構築促進および技術開発促進事業として、当財団が(株)防災都市計画研究所に対して委託実施した課題の一つである。この成果が、データベースに興味をお持ちの方や諸分野の皆様方のお役に立てば幸いである。

なお、平成元年度データベースの構築促進および技術開発促進事業で実施した課題は次表のとおりである。

平成2年3月

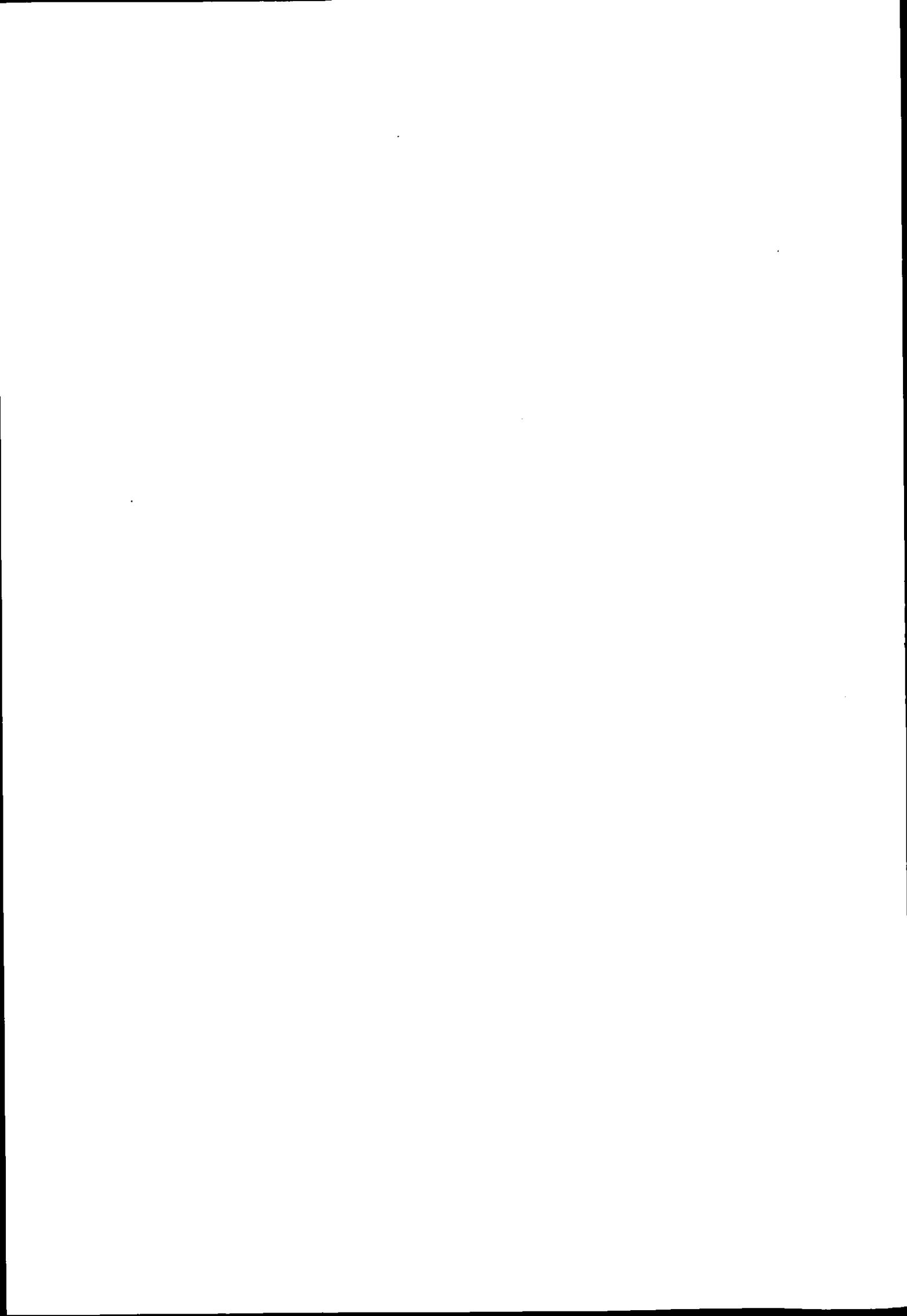
財団法人 データベース振興センター

平成元年度データベース構築・技術開発促進委託課題

分野	課題名
社 会	1 気候情報データベースの構築
	2 電磁波環境障害に関するデータベースの構築
	3 災害情報センサーの構築
	4 意味情報を中核とした医療評価データベースとコミュニケーションシステムの構築
	5 ハンディキャップパーソンの情報ニーズに即したライフサポートデータベースの構築
	6 博物館情報データベースシステムの構築
	7 中央省庁での電子計算機利用に関する報告書のデータベース化
地 域 活 性 化	8 沖縄地域における文化情報データベースの構築
	9 九州地域の人材情報データベース構築に関する調査研究
中 小 企 業 振 興	10 高岡市商圈データベースの構築
	11 地域の物産・人材・文化情報のデータベース構築と新しい地域間交流推進に関する調査研究
地 図	12 マルチメディア型地図データベース構築のための調査研究
エ ネ ル ギ ー ・ 資 源	13 燃焼技術と燃焼装置設計のデータベース作成
部 品 ・ 材 料	14 技術支援システムにおける産業機械部品データベースの構築
	15 マイクロコンピュータのプログラマブル周辺デバイスのデータベース構築
標 準 化	16 イオンクロマトグラフィー・データベースの構築
	17 CD-ROMマルチメディアデータフォーマットの調査
海 外	18 データベース構築のためのターミノロジーの調査研究
技 術	19 異種データから構成されるデータベースの総合的処理技術に関する調査研究
	20 バイナリモデルに基づく先端的文書検索システムの開発

# 目 次

1. 調査の目的 .....	1
2. データベースの概要 .....	2
2. 1 対象とする事故・災害の分野 .....	2
2. 2 カルテの様式 .....	3
2. 3 キーワード検索システム概要 .....	4
3. キーワードシソーラスの考え方 .....	9
3. 1 災害情報データベースの特徴 .....	9
3. 2 統制キーワードとフリーキーワード .....	12
4. 統制キーワード .....	14
4. 1 統制キーワードのシソーラス .....	14
4. 2 シソーラスの全体構成 .....	18
4. 3 シソーラス例 .....	23
5. フリーキーワード .....	28
5. 1 自動抽出用辞書の考え方 .....	28
5. 2 自動抽出例 .....	32
6. 成果と今後の課題 .....	38
6. 1 成果 .....	38
6. 2 今後の課題 .....	39



## 1. 調査の目的

災害情報データベースは、日々発生している事故・災害のデータを総合的にストックし、複雑化・多様化する現代社会における事故・災害の防止および安全な技術開発のために活用することを目的としたデータベースである。昭和57年以来、さまざまな分野の安全・防災の専門家が集まって、構築を進めてきた。現在では、事故・災害事例数にして4万件以上が登録され、平成2年度にはオンラインによる公開を控えている。

このデータベース公開に向けて、昨年は、検索システムの検討および既往の重要災害事例の登録を実施した。その結果、今後に残された重要な課題として、キーワード体系の構築が取り上げられることとなった。

これまで、当データベースの構築にあたっては、それぞれの事故・災害事例について、その特徴や内容を表現する単語を任意に抽出してキーワードとして付与する方法を取ってきた。しかし、一般への公開に当たっては、ユーザーにとって重要な検索手段となるキーワードの体系を確立し、検索ニーズに十分応えられるようなシソーラスを構成する必要がある。

本調査は、こうした観点から、災害情報の効率的な検索に供するキーワードシソーラスの構築を行うものである。

## 2. データベースの概要

### 2.1 対象とする事故・災害の分野

当データベースでは、表2-1に示すように、さまざまな分野・領域の事故・災害を網羅的に対象としている。これらは、災害分類としてキーワードとは別途に登録され、検索のインデックスとしても利用される。

表 2-1 対象とする災害

1. 風水害	1. 洪水 2. 強風 3. 豪雨 4. 高潮 5. 台風 6. 前線 7. 低気圧 8. 龍巻 9. 高波 10. 浸水 11. 湛水
2. 気象災害	1. 長雨 2. 干ばつ 3. 異常乾燥 4. 濃霧 5. 冷害 6. 寒波 7. 熱波
3. 雪害	1. 豪雪 2. 雪崩 3. 吹雪
4. 氷害	1. 凍結
5. 雷	1. 落雷 2. 雹
6. 地盤変動	1. 山崩れ 2. 崖崩れ 3. 土石流 4. 地滑り 5. 陥没 6. 隆起 7. 落石 8. 落盤
7. 地震	1. 前震 2. 余震 3. 群発地震 4. 断層 5. 液状化現象 6. 津波 7. その他
8. 火山爆発	1. 火砕流 2. 熔岩流 3. 火山泥流 4. 火山弾 5. 火山灰 6. 噴気 7. 海底火山
9. 火災	1. 大火 2. 放火
10. 爆発	1. ガス 2. 蒸気 3. 粉塵 4. 破裂 5. 爆破
11. 漏洩	1. ガス 2. 流出 3. 流失 4. 放出
12. 中毒	1. ガス 2. 酸欠 3. 食中毒 4. Alcohol 5. 被曝
13. 崩壊・破壊	1. 崩壊 2. 破壊 3. 転倒 4. 飛来・落下
14. 故障	1. 停電 2. システム
15. 交通災害	1. 自動車 2. 鉄道 3. 船舶 4. 航空機 5. 二輪車
16. その他災害	1. 群衆災害 2. 動物災害 3. 医療災害 4. 飢饉 5. 疫病
17. 公害	1. 汚染 2. 地盤沈下
18. Human Error	
19. 人身災害	1. 火傷 2. 薬傷 3. 感電 4. 山岳遭難 5. おぼれ 6. 傷害
20. その他	1. 犯罪 2. 戦争

## 2. 2 カルテの様式

個々の事故・災害事例は、1件ごとに定められたフォーマットで整理される。このフォーマットは、カルテと呼ばれている。

表 2-2 カルテ表示例

【 事例 No. 15436 】							
[ 日 時 ]	開始	1986年	11月	1日 (土)	早朝		
	経過	年	月	日			
	終了	年	月	日			
[ 場 所 ]	1. Europe	Swiss					
[ 死傷者 ]	死 者	行方不明	小 計	重 傷	軽 傷	小 計	死傷者計
	0 人	0 人	0 人	0 人	0 人	0 人	0 人
[ 分 類 ]	1. Fire						
	2. Public Nuisance	汚染					
[ 産業分類 ]	1. 製造	化学					
[ 記 事 ]	Swiss, Basel 郊外 化学肥料工場倉庫の染料から出火炎上。有害物質が消火用水等とともにライン川に流入。広域汚染。						
[ 抄 録 ]	倉庫に.....						
[ 特記事項 ]	この事故を契機として.....						
[ キーワード ]	1. 化学	2. 肥料	3. 工場	4. 倉庫	5. ライン川	6. 水銀	7. 避難
	8. 水道	9. 染料					
[No.]	[光ディスク番号]	[頁]	[著者：表題：誌名：巻号頁：コメント]	[扱い]			
1199	144.	1989-04---	029101 3 西川光一：化学工場の事故対策に関する国際動向について：保安月報：No. 291 P. 2-4；	¥500			
376	152.	1989-01-11-	0120 1 K S 日経産業新聞（朝）染料。欧米勢、公害対策に悩む	公開			
74	102.	1987-10-19-	0203 1 M 毎日新聞（夕）スイスでも緑の党躍進	公開			
66	082.	1987-06-06-	0109 1 A 朝日新聞 ライン川汚染は染料失火が原因	公開			
35	065.	1986-11-17-	0002 1 L LLOY'S LIST	公開			
			⋮				
			⋮				
41	065.	1986-11-15-	0001 1 E International Herald Tribune	公開			
42	065.	1986-11-11-	0002 1 E Financial Times	公開			
46	065.	1986-11---	000011 20 Frankfurter Allgemeine	非公開			
6	A01.	1986-11---	000001 120 △○夫：災害事例集：○○出版：P. 151-271	公開			

## 2.3 キーワード検索システム概要

当データベースの検索システムでは、前述のカルテ・フォーマットに設けられた各検索フィールドのうち、キーワード・フィールド（フィールド名：KEY）及び記事フィールド（フィールド名：AD）についてキーワード検索が可能となっている。また、将来的には、文献著者名や文献表題などのフィールドについてもキーワード検索を可能とする予定である。

キーワードに関する検索には、以下のような機能がある。

### (1) 完全一致検索

完全一致型検索は、SEARCHコマンドおよびXSコマンドを用い、検索の対象とするフィールドとキーワードを指定して行う。

SEARCHコマンドでは、複数の条件を指定した場合にも、検索結果のみの集合セット（Snn：nnは番号）が作成される。一方、XSコマンドで複数の条件を指定すると、それぞれの条件毎の集合セットおよびそれらを組み合わせた検索結果の集合セットの両方が作成される。

[コマンド例] 下線部分を入力する。

```
> SEARCH 自転車/KEY
```

```
S1 : 500 自転車/KEY
```

\*キーワード・フィールドに『自転車』という単語が登録されている事例 500件が検索されたことを示す

[コマンド例]

```
> SEARCH (自転車*バイク)/KEY
```

```
S2 : 100 (自転車*バイク)/KEY
```

```
> XS (自転車*バイク)/KEY
```

```
S3 : 500 自転車/KEY
```

```
S4 : 700 バイク/KEY
```

S 5 : 100 (自転車\*バイク) /KEY

\*この2例では、キーワード・フィールドに『自転車』と『バイク』の両方の単語が登録されている事例100件が検索されている。このうちXSコマンドでは、最終結果以外に『自転車』『バイク』それぞれの検索集合セットも作成されている。

## (2) トランケーション検索

トランケーション検索機能としては、前方一致、後方一致、中間一致の3種類があり、SEARCHコマンド、XSコマンドとも利用可能である。

[コマンド例]

> SEARCH 運転? /KEY

S 6 : 100 運転? /KEY

> SEARCH ?運転 /KEY

S 7 : 50 運転? /KEY

> SEARCH ?運転? /KEY

S 8 : 200 ?運転? /KEY

## (3) シソーラス検索

キーワード辞書の持つシソーラスを活用した検索方法であり、同義語検索・上位語検索・下位語検索・関連語検索（それぞれの詳細は後述）などが可能である。

シソーラス検索を行う場合には、SEARCHコマンド、XSコマンドの修飾子にそれぞれ次のような指定を行う。（レベル：何階層まで遡るかを数字で指定）

同義語検索 /UT\_THESAURUS

上位語検索 /BT\_THESAURUS=レベル

下位語検索 /NT\_THESAURUS=レベル

関連語検索 /RT\_THESAURUS

[コマンド例]

```
> SEARCH/UT_THESAURUS B747/KEY  
S9 : 80 B747/KEY  
> SEARCH/BT_THESAURUS=2 飛行機/KEY  
S10 : 600 飛行機7/KEY
```

なお、デフォルトでシソーラス検索を設定する場合には、以下のコマンドを用いる。

```
同義語検索 SET/THESAURUS=UT  
上位語検索 SET/THESAURUS=BT  
下位語検索 SET/THESAURUS=NT  
関連語検索 SET/THESAURUS=RT
```

#### (4) EXPANDコマンド

EXPANDコマンドはキーワードの有無を調べるコマンドであり、前方一致及び後方一致のトランケーション機能が利用できる（中間一致は使用できない）。このコマンドを実行した結果、対象とする単語の前後9つのEXPAND集合セットが作成される。

[コマンド例]

```
> EXPAND 医療/KEY  
E1 : 10 医師/KEY  
E2 : 20 医者/KEY  
E3 : 5 医薬/KEY  
E4 : 50 医薬品/KEY  
*E5 : 100 医療/KEY  
E6 : 50 医療器具/KEY  
E7 : 50 医療機器/KEY  
E8 : 200 医療行為/KEY  
E9 : 300 医療施設/KEY
```

このようにして作成されたEXPAND集合セットは、次のように検索に利用することができる。

[コマンド例]

```
>SEARCH (E3+E5) /KEY
S11 : 200 (E3+E5) /KEY
>XS E1 /KEY *誤り /AD
S12 : 200 E1 /KEY
S13 : 300 誤り /AD
S14 : 50 S4 * S5
```

#### (5) その他のキーワード関連処理機能

上記のほか、キーワードに関連する処理機能としては、以下のようなものがある。

##### ①同形異義語処理

キーワードには、同じ表現をする文字でも意味が異なるものがある。キャッシュディスクペンサー、コンパクトディスク等の略称である「CD」がその一例である。その場合どちらの意味を選択するかは、SEARCHコマンド及びXSコマンドでの検索時に明示的に指示する方法と検索時に選択させる方法の2種類での対処が可能である。

[コマンド例]

```
>SEARCH CD (キャッシュディスクペンサー) /KEY
S9 : 100 CD (キャッシュディスクペンサー) /KEY

>SEARCH CD /KEY
1 CD (キャッシュディスクペンサー)
2 CD (コンパクトディスク)
次ページ:N 選択:[1-9] 終了:Exit >1
S10 : 100 CD (キャッシュディスクペンサー) /KEY
```

## ②同音異義語処理

同音異義語とは、例えば『火気』と『牡蛎』のように異なる文字で読みが同じものをいう。現在は、キーワードの読みを登録していないため特に問題とはならない。しかしながら、将来的にはキーワードの読みを登録することにより『読み仮名』による検索を可能とするなど、さらに検索機能のユーザーインターフェースを向上させる必要があり、その際に同音異義語の処理が必要となる。同音異義語については、どちらの意味を選択するかをSEARCHコマンド及びXSコマンドでの検索時に選択させる方法での対処が可能である。

[コマンド例]

```
> SEARCH カキ/KEY
```

```
1 火気
```

```
2 牡蛎
```

```
3 夏期
```

```
次ページ：N 選択：[1-9] 終了：Exit > 1
```

```
S11 : 100 火気/KEY
```

## ③片仮名キーワード処理

片仮名のキーワードの場合、小文字、長音、中点などの付け方が検索ユーザーにとって大きな問題となる。このシステムではそうした問題を内部的に処理し解釈するような機能をもっている。

[処理例]

```
『ソフトウェア』と『ソフトウエア』  
『コンピュータ』と『コンピューター』 } ⇒それぞれ同一とみなされる。  
『データベース』と『データ・ベース』
```

ただしこの処理では、『アンカ』と『アンカー』というキーワードがある場合などに問題となる。

### 3. キーワードシソーラスの考え方

#### 3. 1 災害情報データベースの特徴

当データベースは、あらゆる分野の事故・災害を総合的に蓄積するという、他に類を見ないデータベースである。このため、入力されるデータにはいくつかの特徴があり、それを考慮しなければ効率的・合理的なキーワードシソーラスを構築することは不可能と言える。ここでは、これらの特徴について述べる。

##### (1) さまざまな分野のデータを対象とする。

すでに述べてきたように、当データベースでは、あらゆる分野の事故・災害を対象としている。事故・災害の情報を収集するデータベースはほかにもいくつかの例があるが、このように各分野を網羅的に収集するものはほとんど見られない。

これまで、事故・災害に関するデータの多くは、各分野の専門家が独自に収集・分析し、その分野の安全対策や防災対策に役立ててきた。しかし、現代社会が複雑化・多様化するにつれて、事故・災害の現象もある特定の分野のみにとどまらず、社会の様々な側面に広く影響を及ぼすようになってきている。複雑化する社会の中では、事故・災害のメカニズムさえも単純なものではなくなり、もはや単一の分野・領域における安全研究や防災対策だけでは十分とは言えなくなってきた。

これまで分野ごとに行なわれてなかなか交流しにくかった安全・防災の知識や技術を、分野を越えて共通のものとし、より総合的な安全対策を推し進めるために、さまざまな分野のデータを対象としているのが、当データベースの最大の特徴である。

##### (2) 利用者は専門外の分野へのアクセスも行う。

さまざまな分野のデータを共通のものとして利用するため、ある特定の分野の事例に関しても、利用者は必ずしもその分野の専門知識を十分に持っているとは限らない。むしろ、当データベースの目的とするところから考えると、利用者は自分の専門以外の分野へも積極的にアクセスすることが望まれている。

こうした利用者の状況を考慮すると、求められた事故・災害の分野に関して専門知識が

あることを前提とした検索システムを構築することはできない。その一方で、これまで各分野で安全・防災の研究を行ってきた専門家が自分の専門分野に関する情報にアクセスすることも多いことも予想される。したがって、こうした専門家の利用に対しても耐えられる精度を持ちつつ、しかも専門知識を持っていない利用者でも情報検索などができるよう、システム側が柔軟性を持って対応できる必要がある。

### (3) キーワードの種類が多岐にわたる

データのある特定の分野に限定したデータベースであれば、その分野に必要な項目を羅列して、それに応じたキーワード体系を作成することも容易であろう。現に、航空機事故を対象とする I C A O (国際民間航空機関) の事故情報データベースの例では、航空機の機種、装置の名称や発生した事象の名称がほぼ限定されるため、そうした情報のほとんどをコード化して整理している例がある。

しかし、当データベースのようにさまざまな分野を総合的に扱うデータベースの場合には、データの種類は多岐にわたり、それに従って付与されるキーワードの種類も非常に多種多様なものとなる。

多くの分野を扱うことから、たとえば次のようなキーワードが発生する。

#### ①各分野に共通して用いられる単語

例) 避難

裁判

#### ②ある特定の分野にだけ使用される単語

例) 集中豪雨

E C C S (原子炉の緊急炉心冷却装置)

ターンアラウンド (着陸復行)

#### ③ほとんど類似の単語であるが、分野によって通常使用される単語が異なるため、結果としてニュアンスの異なる単語

例) 運転手 (トラックなど)

運転士 (鉄道)

運転員 (原子力発電所)

このように複雑で多岐にわたるキーワードを全体としてどのように体系づけるかは、非常に難しい問題となる。

#### (4) 十分な情報を得られない場合もある

一般に事故・災害に関する情報は、なかなか詳細なデータまでが公開されず、十分な情報を得られにくい種類の情報である。特に、企業機密が関係したり、刑事・民事の裁判などで問題となったりしている場合などでは、他に対する影響の大きさなどもあって事故・災害の状況に関する詳細は一切公開されない場合もある。

また、他の一般のデータベースでは、直接の当事者（たとえば、文献の著者、事故の当事者など）が、定められたフォーマットにのっとって必要な情報を提供するという場合が少なくない。しかし、当データベースは、一般の新聞や安全・防災関係の雑誌などに公開された情報に基づいてデータを構築しているため、データの質や種類を統一することが難しい。たとえば、事故・災害に関する最も基本的な情報は、発生日時や場所などであるが、そうした基本情報なども入手できない場合もある。海外において大きなエチレン工場の事故が発生した場合でも、その事故の影響でエチレン市場が逼迫しているという報道しかない場合には、事故の正確な日時や内容までは把握できないこともあるのである。また、航空機事故を専門的な立場から検討する上では、その事故を起こした航空機の機種に関する情報が意味を持ってくる場合もある。ところが、実際に一般に報道されたり公開される事故情報では、たとえB747型機の事故があっても、その事故がB747型機のうちのどのタイプの機種なのかわからない場合も多い。

当データベースのキーワードソースは、このように不十分な情報にも対応できるような柔軟なものでなければならない。

### 3. 2 統制キーワードとフリーキーワード

ここまで述べてきたようなデータおよび利用者の多様性に対応し、かつ情報量の不十分なデータについても柔軟な対応を可能とするために、当データベースのキーワード体系は次の2種類に区分して整理する(図3-1)。

#### (1) 統制キーワード

事故・災害事例ごとに、その事例の内容や特徴などを記述するために抽出された単語。使用される単語の種類などがある程度限定されている。これらを同義語、上位・下位語、関連語などのシソーラスで整理することによって、基本的なデータ検索における大きな検索漏れなどを防ぐことができる。

#### (2) フリーキーワード

データベースに登録した文字列の中から、一般に意味を持つ限り制限なく抽出された単語。こうしたキーワードの付与の方法では検索漏れを完全になくすことはできないが、当初必要と考えられていなかった単語に関する検索も可能にするものであり、多様なユーザーの要求に対応していくためには必要不可欠である。

このようなキーワードの付与は、データ登録の際に多くの人的労力が必要となってしまうため、ここではあらかじめ抽出すべきキーワードを辞書として登録し、その辞書にある単語を自動的に抽出するようなシステムを構築する。

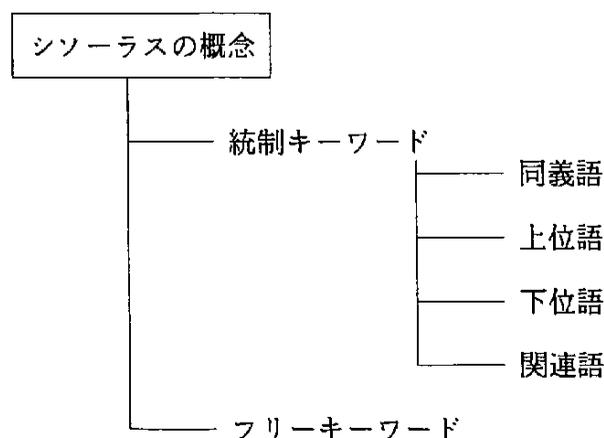


図3.1 キーワードシソーラスの概念

### (3) キーワード用辞書

このデータベースでは、シソーラスの概念にあわせて次の2つの辞書を持つ。

#### ①統制キーワード用辞書 (辞書名: C X)

同義語、上位・下位語、関連語という関連づけを持つ辞書。キーワードフィールドの検索の際に用い、シソーラス検索を可能とする。

#### ②フリーキーワード用辞書 (辞書名: F X)

記事フィールド、抄録フィールド、特記事項フィールド、関連事項フィールドに対して辞書参照型のキーワード自動切り出しを行なうための辞書。

将来的には、文献表題、文献名などのフィールドについてもキーワード自動切り出しを行なう予定であり、その際の辞書にも適用する予定である。

## 4. 統制キーワード

### 4.1 統制キーワードのシソーラス

#### (1) 複合語

これまで、当データベースの検索システムでは、トランケーション機能を用いての検索ができなかった。このため、キーワードとして付与された単語は、複合語も可能な限り分割されている。これは、トランケーション機能を持たない検索システムであるかぎり、検索漏れを防ぐために止むを得ないことであったが、一方で検索された情報にノイズが多くなりすぎるという欠点があった。

データベースの一般公開にあたっては、この点を改良して、トランケーション検索を行えるよう検索システムを変更している。そのため、複合語もそのまま取り入れることが可能になったので、現在分割して登録されているキーワードについては複合語に戻す作業を行った。

例) ポート : モーター+～  
                  ゴム+～  
                  タグ+～  
                  ビニール+～  
                  救命+～  
                  遊覧+～  
  
                  磁気 : ～+観測+衛星  
                          ～+テープ  
                          ～+ディスク  
                          ～+コード  
                          ～+嵐  
                          核+～+共鳴+装置

## (2) 同義語

統制キーワードの中で類似の意味を持つ単語同士は、そのうちのひとつを代表語とし、その他を代表語のファミリーとして同義語関係を作った。

一般のキーワードソーラスでは、ある単語の同義語として関連づけられた単語は「非ディスクリプタ」となり、索引・検索作業には使用できない単語とされる場合がある。しかし、各事例に関する十分な情報が得られないという当データベースの特徴から、ほとんど同義語であると思われながらも完全に一致すると確信できない単語がある場合も多い。このため、当データベースにおいては、同義語は以下の3種類を等しく扱うこととした。

### ①同一の意味で表記方法の違う単語 (⇒一般の同義語)

◦漢字とひらがな 例) こたつ=炬燵

おもちゃ=玩具

◦略称や別称 例) イソシアン酸メチル=M I C

アスベスト=石綿

### ②意味的にほぼ同じだが実際のニュアンスや使われ方がやや異なる単語

・類義語 例) ごみ収集車=清掃車

消防士=消防署員

### ③上下関係にあるが、通常はその下位ランクの情報まで把握できない場合が非常に多く区別がつかない単語

例) B 7 4 7 = B 7 4 7 S R

以上のような関係にある単語同士を全て同義語として関係づけたため、特に②、③の関係にある単語については、キーワードを付与する際に原文献に忠実に可能な範囲で区別することとした。したがって、ある代表語の同義語であっても、それはそれは非ディスクリプタではなく、検索に使用することができるディスクリプタとなる。

また、①のタイプの同義語については、一般ユーザーの利用しやすさを考慮すると、現在キーワードとして登録されていない単語も可能なかぎり登録する必要がある。しかし、前述のように同義語もディスクリプタとして検索可能であることを考えると、このように全く同一の意味を持つ単語に関しては、検索漏れの防止のためにも付与の際の代表語限定はしっかりと行なっておかなければならない。

### (3) 上位・下位語

一般に、単語同士の上下関係は、次の2種類がある。

#### ①包括関係

例) \*\*\*\* 船 \*\*\*\*

[下] . タンカー

. . . LPGタンカー

. . . ケミカルタンカー

\*\*\*\* ライフライン施設 \*\*\*\*

[下] . 電気供給施設

. . . 発電所

. . . . 原子力発電所

. . . . 火力発電所

#### ②部分-全体関係

例) \*\*\*\* 船 \*\*\*\*

[下] . 船室

. . . 1等客室

. . . 操舵室

\*\*\*\* 原子力発電所 \*\*\*\*

[下] . 原子炉

. . . 制御棒

. . . 緊急炉心冷却装置

これらの両方の関係を同時に同一ソースで扱うことは事実上困難であり、また混乱を生じる可能性もあるので避けるべきである。したがって、当データベースのソースでは基本的に包括関係にある単語同士を上位・下位語として関係づけることとした。

部分-全体関係にある単語同士については、全体を表す単語『○○』の関連語（後述）として『○○要素』という単語を関連づけ、部分を表す単語はこの『○○要素』という単語の下位語とした。

例)       \*\*\*\*\* 自動車 \*\*\*\*\*  
          [下] . 一般用車両  
              . . 乗用車  
              . . ワゴン車  
          [関] 自動車要素  
          \*\*\*\*\* 自動車要素 \*\*\*\*\*  
          [下] . アクセル  
              . . ブレーキ  
              . . ブレーキケーブル  
          [関] 自動車

#### (4) 関連語

キーワード間には、同義語関係や上位下位語関係にはならないが、何らかの関係がある単語同士がある。これらのうち、キーワードの付与およびデータ検索の際に、必要な用語を見つけるうえで有用であると考えられる場合には、両者を関連語として関係づける。

関連語の指定は、その性格上ある程度恣意的なものとならざるを得ない。しかし、先に述べた部分—全体関係のほか、たとえば原料と産物、原因と結果、施設・機器とその利用者などは関連語とするものとする。

例)   農薬 ⇨ イソシアン酸メチル (農薬の原料)  
      肺癌 ⇨ アスベスト (肺癌の原因)  
      福祉施設 ⇨ 障害者 (福祉施設の利用者)

## 4. 2 シソーラスの全体構成

統制キーワードは、①事物、②事象、③固有名詞の3種類におおよそ分類し、さらにそれぞれの単語について同義語、上位下位語、関連語の関連づけを行った。

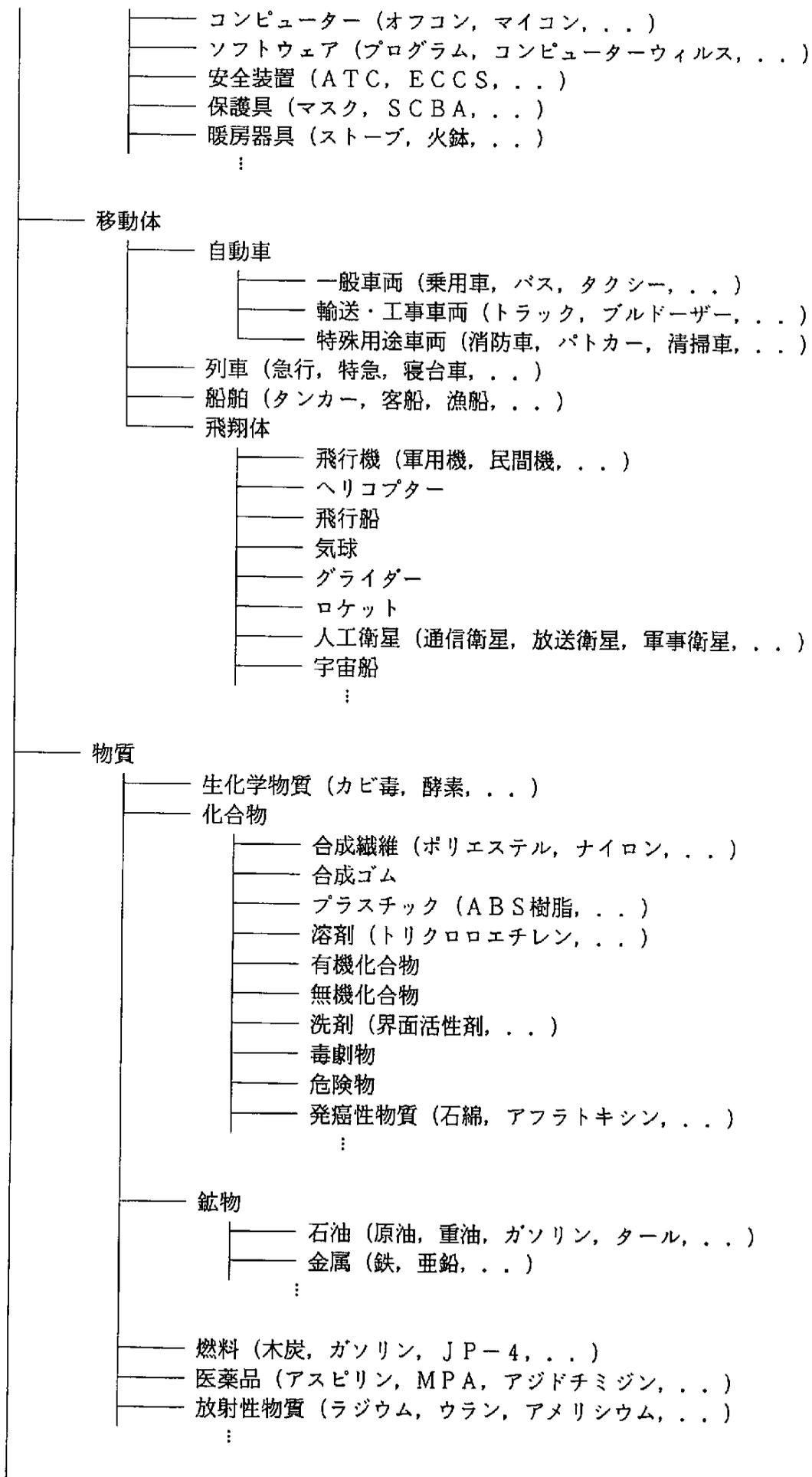
### ① 事物

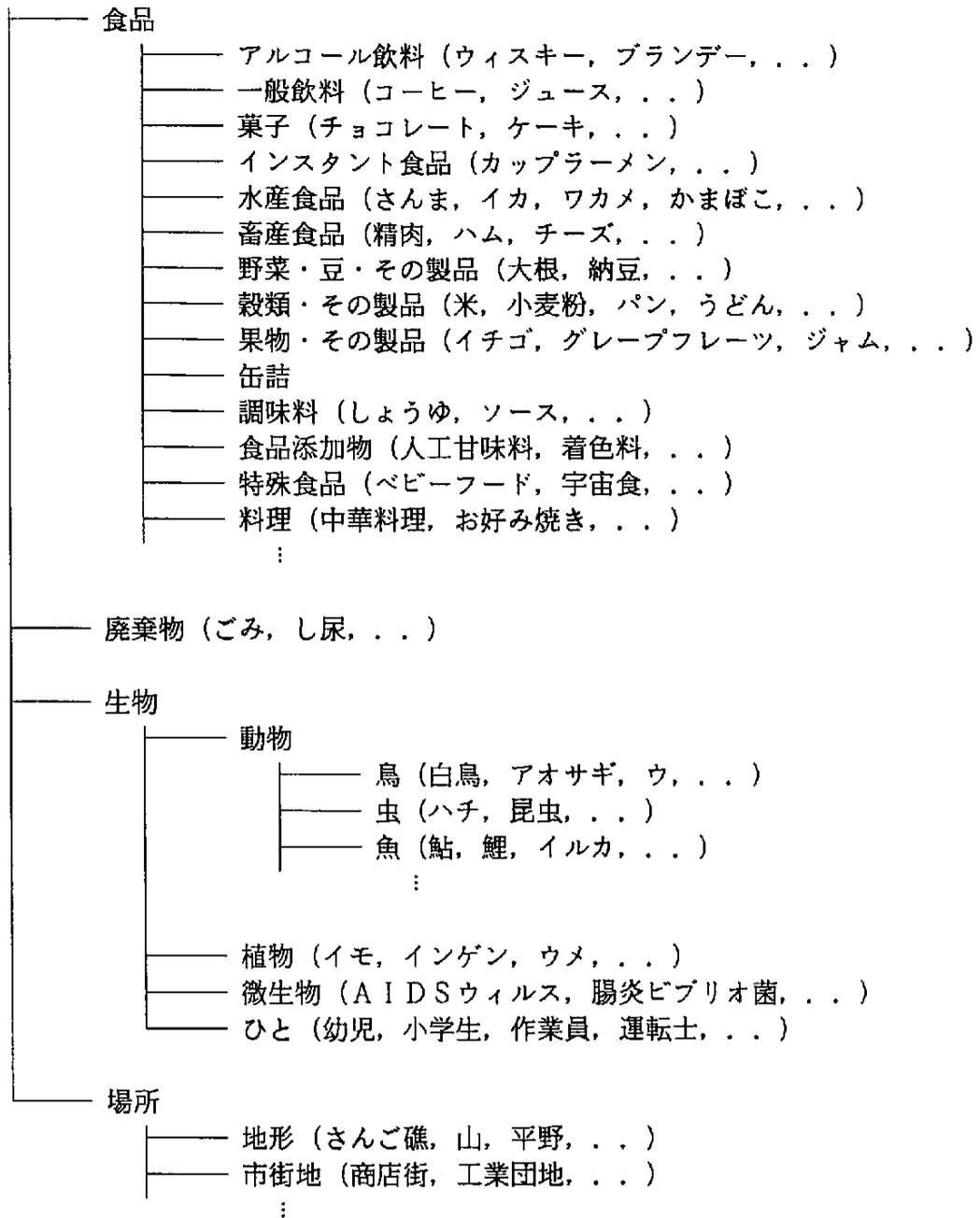
#### 施設

- 農業施設 (用水路, . . .)
- 林業施設
- 漁業施設 (養魚場, . . .)
- 鉱業施設 (炭鉱, 油田, 天然ガス井, . . .)
- 建設施設
- 製造施設 (製造工場, 作業所, . . .)
- ライフライン施設
  - 電気供給施設 (発電所, 送電鉄塔, . . .)
  - ガス供給施設 (ガス管, . . .)
  - 熱供給施設 (スチームパイプ, . . .)
  - 水道施設 (浄水場, 水道管, . . .)
  - 通信施設 (電話局, 通信ケーブル, . . .)
- 運輸施設
  - 鉄道施設 (駅, 線路, . . .)
  - 道路施設 (国道, 高速道路, 交差点, . . .)
  - 港湾施設 (棧橋, 埠頭, . . .)
  - 航空施設 (空港, . . .)
  - 倉庫
- 一般店舗 (デパート, スーパーマーケット, 薬局, . . .)
- 飲食店 (レストラン, バー, 料亭, . . .)
- 宿泊施設 (ホテル, 旅館, 宿舎, . . .)
- レジャー施設 (ゴルフ場, 映画館, 遊園地, . . .)
- 医療施設 (病院, 療養所, . . .)
- 教育・研究施設 (小学校, 中学校, 研究所, . . .)
- 福祉施設 (老人ホーム, . . .)
- 宗教施設 (教会, 寺, . . .)
- 公共施設 (公民館, 公園, . . .)
- 公務施設 (大使館, 市役所, 裁判所, 検察庁, . . .)
- 廃棄物処理施設 (清掃工場, し尿処理場, . . .)
- 住居 (住宅, アパート, 団地, . . .)

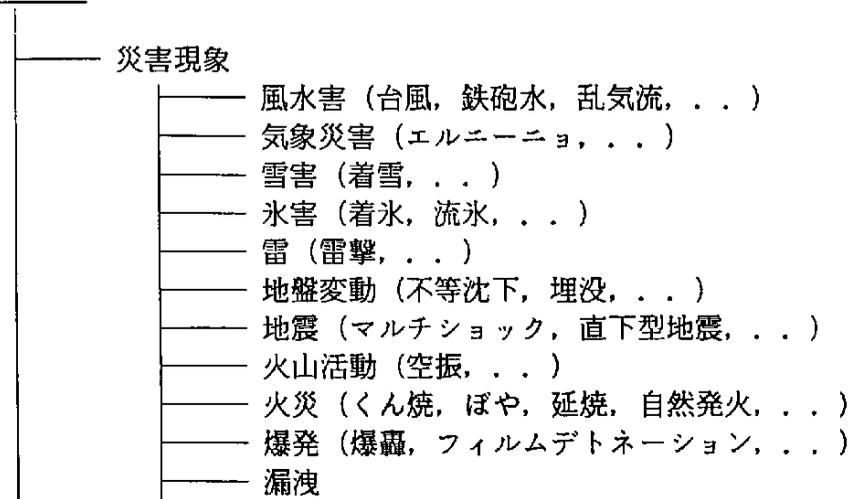
#### 装置・器具・部品

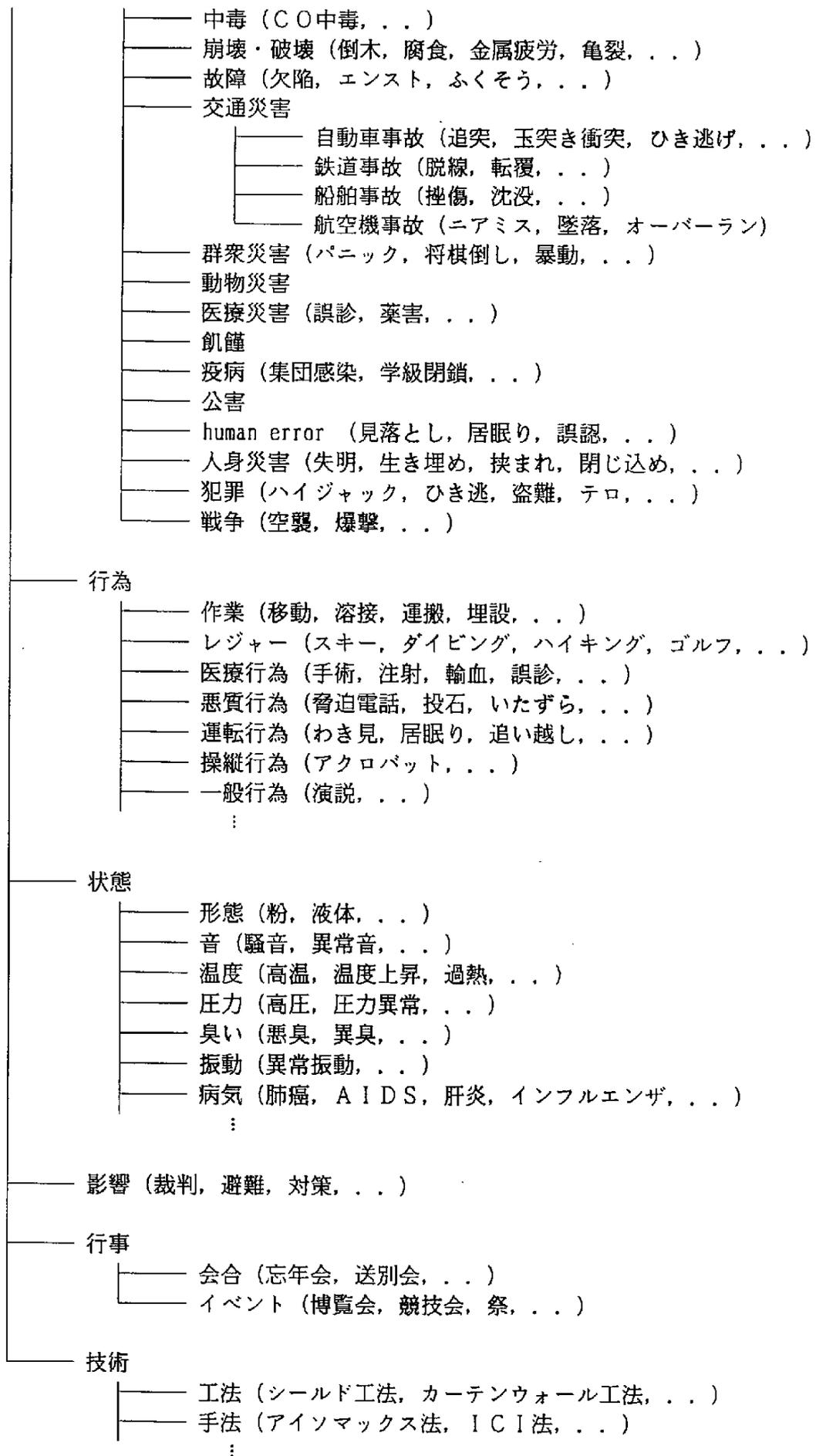
- 製造装置 (蒸留装置, 分解装置, . . .)
- 一般機械 (圧縮機, 溶接機, . . .)
- 工具 (のこぎり, ドライバー, . . .)
- 兵器 (ミサイル, . . .)
- 調理器具 (電気炊飯器, ガスこんろ, . . .)
- 医療器具 (CTスキャナ, 鉗子, . . .)



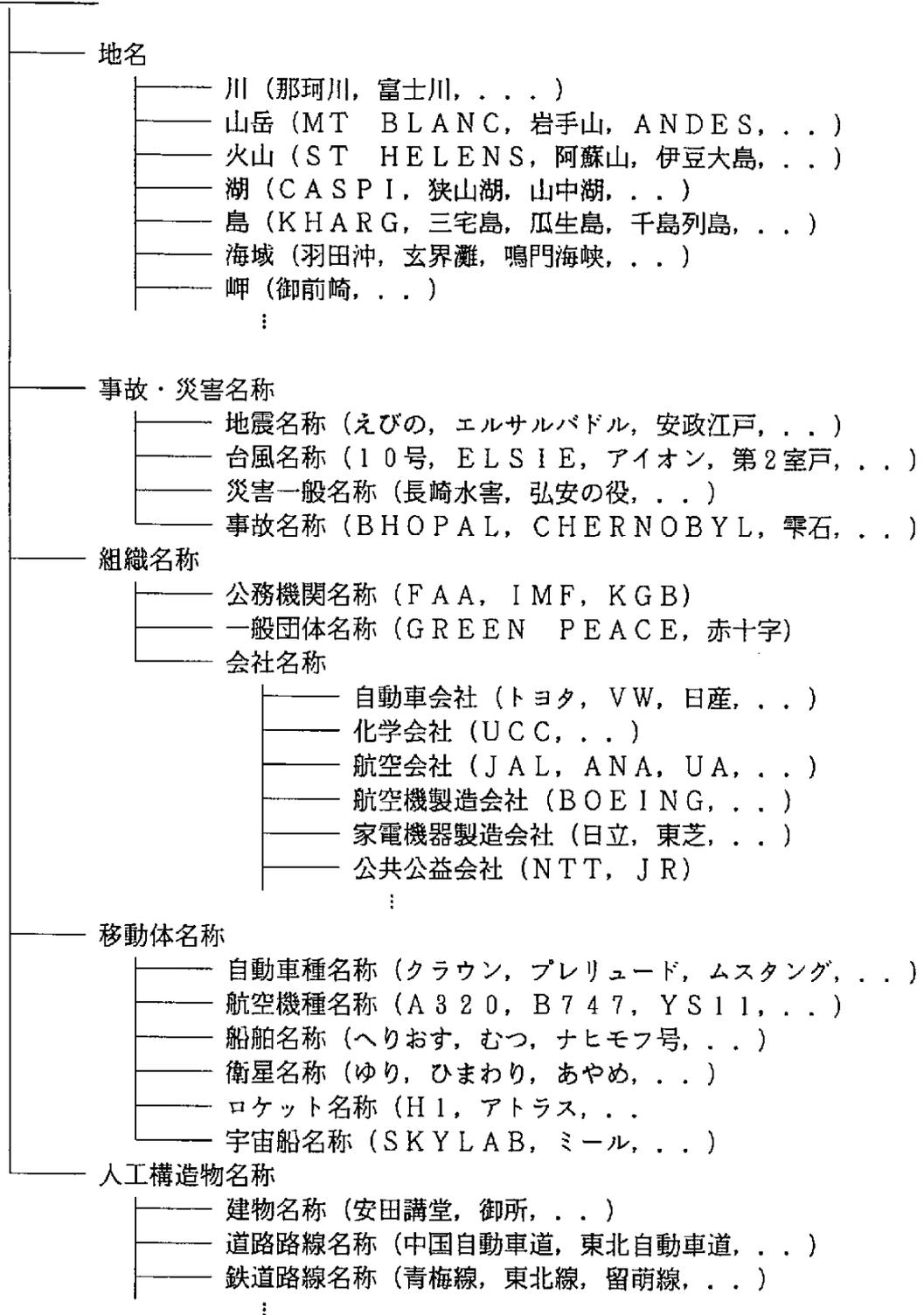


## ② 事象





### ③ 固有名詞



#### 4. 3 シソーラス例

\*\*\*\*\* アーケード \*\*\*\*\*

[関] 商店街

\*\*\*\*\* アースコード \*\*\*\*\*

[代] アース線

\*\*\*\*\* アース線 \*\*\*\*\*

[同] アースコード

接地コード

[上] . 電気部品

\*\*\*\*\* アーム \*\*\*\*\*

[上] . 機械要素一般

[関] クレーン車

クレーン船

ショベルカー

産業ロボット

\*\*\*\*\* アーモンド \*\*\*\*\*

[上] . 果物・その製品

. . . 食品

\*\*\*\*\* アイアタイト \*\*\*\*\*

[同] 通常型爆薬

[上] . 火薬

\*\*\*\*\* アイオン \*\*\*\*\*

[上] . 台風

. . . 事故・災害名称

. . . 熱帯低気圧

\*\*\*\*\* アイスクリーム \*\*\*\*\*

[上] . 菓子

. . . 食品

\*\*\*\*\* アイスバーン \*\*\*\*\*

[上] . 氷害

. . . 災害現象

[関] 凍結

\*\*\*\*\* アイスボックス \*\*\*\*\*

[上] . レジャー用品

\*\*\*\*\* アイソトープ \*\*\*\*\*

[代] RI

\*\*\*\*\* アイソマックス装置 \*\*\*\*\*

[同] 脱硫酸素化分解装置

[上] . 製造装置

[関] 製油所

\*\*\*\*\* アイドリング \*\*\*\*\*

[上] . 運転行為

. . . 行為

[関] エンジン

自動車

\*\*\*\*\* アイルファスト航空 \*\*\*\*\*

[上] . 航空会社

. . . 会社名称

. . . 組織名称

\*\*\*\*\* アイルランド共和国軍 \*\*\*\*\*

[代] IRA

\*\*\*\*\* アイロン \*\*\*\*\*

[上] . 家電製品

\*\*\*\*\* アウターマーカー \*\*\*\*\*

[上] . 航空施設

. . . 運輸施設

. . . 施設

[関] 空港

\*\*\*\*\* アウディ \*\*\*\*\*

[同] AUDI

[上] . 自動車会社

. . . 会社名称

. . . 組織名称

[関] アウディ100

アウディ200

アウディ5000

アウディ80

アウディ90

\*\*\*\*\* アウディ100 \*\*\*\*\*

[上] . 自動車種名称

. . . 移動体名称

[関] アウディ

\*\*\*\*\* M3 \*\*\*\*\*

[代] BMW

\*\*\*\*\* M6 \*\*\*\*\*

[代] BMW

\*\*\*\*\* MA \*\*\*\*\*

[同] MALEV

マレフ航空

[上] . 航空会社

. . . 会社名称

. . . 組織名称

\*\*\*\*\* MAGNETIC TAPE \*\*\*\*\*

[代] MT

\*\*\*\*\* MAKJAN \*\*\*\*\*

[上] . 火山

. . . 地名

\*\*\*\*\* えびの \*\*\*\*\*

[上] . 地震  
    . . 災害現象  
    . . 事故・災害名称

\*\*\*\*\* おがくず \*\*\*\*\*

[代] 木屑

\*\*\*\*\* おしゃぶり \*\*\*\*\*

[上] . おもちゃ

[関] 乳児

\*\*\*\*\* おたふくかぜ \*\*\*\*\*

[同] おたふく風邪  
    流行性耳下腺炎

[上] . 病気  
    . . 状態

\*\*\*\*\* おたふく風邪 \*\*\*\*\*

[代] おたふくかぜ

\*\*\*\*\* おもちゃ \*\*\*\*\*

[同] 玩具

[下] . おしゃぶり  
    . エアソフトガン  
    . 羽子板

[関] STマーク

遊び

\*\*\*\*\* お化 \*\*\*\*\*

[同] お化け

幽霊

\*\*\*\*\* お化け \*\*\*\*\*

[代] お化

\*\*\*\*\* お化け屋敷 \*\*\*\*\*

[代] お化屋敷

\*\*\*\*\* お化屋敷 \*\*\*\*\*

[同] おばけ屋敷

お化け屋敷

幽霊屋敷

[上] . 遊園地

\*\*\*\*\* お好み焼 \*\*\*\*\*

[上] . 料理

. . 食品

\*\*\*\*\* お好み焼店 \*\*\*\*\*

[上] . 飲食店

. . 施設

\*\*\*\*\* かけっこ \*\*\*\*\*

[同] 徒競走

[上] . 遊び

\*\*\*\*\* かじ取り \*\*\*\*\*

[代] 操舵

\*\*\*\*\* かまぼこ \*\*\*\*\*

[上] . 水産食品

. . 食品

\*\*\*\*\* かまぼこ工場 \*\*\*\*\*

[上] . 製造施設

. . 施設

\*\*\*\*\* かみそり \*\*\*\*\*

[上] . 刃物

. . 日用品

\*\*\*\*\* からし \*\*\*\*\*

[同] 辛子

[上] . 香辛料  
    . . 調味料  
    . . . 食品

\*\*\*\*\* からしレンコン \*\*\*\*\*

[上] . 野菜・豆・その製品

. . 食品

\*\*\*\*\* かりふおるにあ丸 \*\*\*\*\*

[上] . 船舶名称

. . 移動体名称

\*\*\*\*\* かんきつ \*\*\*\*\*

[代] かんきつ類

\*\*\*\*\* かんきつ類 \*\*\*\*\*

[同] かんきつ

柑橘類

[上] . 果物・その製品

. . 食品

[下] . オレンジ

. . グレープフルーツ

. . ミカン

. . レモン

\*\*\*\*\* かんしゃく玉 \*\*\*\*\*

[代] 爆竹

\*\*\*\*\* かんてき \*\*\*\*\*

[代] 七輪

\*\*\*\*\* かん子 \*\*\*\*\*

[同] 鉗子

[上] . 医療機器

\*\*\*\*\* がん \*\*\*\*\*

[代] 癌

\*\*\*\*\* ぎ酸 \*\*\*\*\*

[上] . 有機化合物

. . 化合物

. . . 物質

\*\*\*\*\* 案内人 \*\*\*\*\*  
 [同] ガイド  
 [上] . ひと  
       . . 生物  
 \*\*\*\*\* 案内輪 \*\*\*\*\*  
 [同] ガイドタイヤ  
 [上] . 列車要素  
 [関] モノレール  
 \*\*\*\*\* 伊勢湾(タイフウ) \*\*\*\*\*  
 [上] . 台風  
       . . 事故・災害名称  
       . . 熱帯低気圧  
 \*\*\*\*\* 伊勢湾(ワン) \*\*\*\*\*  
 [上] . 湾  
 \*\*\*\*\* 伊東線 \*\*\*\*\*  
 [上] . 鉄道路線名称  
       . . 人工構造物名称  
 \*\*\*\*\* 伊豆急行線 \*\*\*\*\*  
 [同] 伊豆急線  
 [上] . 鉄道路線名称  
       . . 人工構造物名称  
 \*\*\*\*\* 伊豆急線 \*\*\*\*\*  
 [代] 伊豆急行線  
 \*\*\*\*\* 伊豆大島 \*\*\*\*\*  
 [上] . 火山  
       . . 地名  
 \*\*\*\*\* 伊豆大島近海 \*\*\*\*\*  
 [上] . 地震  
       . . 災害現象  
       . . 事故・災害名称  
 \*\*\*\*\* 伊豆半島沖 \*\*\*\*\*  
 [上] . 地震  
       . . 災害現象  
       . . 事故・災害名称  
 \*\*\*\*\* 委員 \*\*\*\*\*  
 [上] . ひと  
       . . 生物  
 \*\*\*\*\* 椅子 \*\*\*\*\*  
 [同] いす  
 [上] . 家具  
 [下] . ソファ  
 \*\*\*\*\* 異臭 \*\*\*\*\*  
 [上] . 臭い  
       . . 状態  
 \*\*\*\*\* 異常音 \*\*\*\*\*  
 [代] 異音

\*\*\*\*\* 移液 \*\*\*\*\*  
 [上] . 作業  
       . . 行為  
 \*\*\*\*\* 移住 \*\*\*\*\*  
 [代] 移転  
 \*\*\*\*\* 移出 \*\*\*\*\*  
 [上] . 作業  
       . . 行為  
 \*\*\*\*\* 移植 \*\*\*\*\*  
 [上] . 医療行為  
       . . 行為  
 \*\*\*\*\* 移設 \*\*\*\*\*  
 [上] . 作業  
       . . 行為  
 \*\*\*\*\* 移送 \*\*\*\*\*  
 [同] 移送作業  
 [上] . 作業  
       . . 行為  
 \*\*\*\*\* 移送作業 \*\*\*\*\*  
 [代] 移送  
 \*\*\*\*\* 移転 \*\*\*\*\*  
 [同] 移住  
 [上] . 影響  
 \*\*\*\*\* 移動タンク貯蔵所 \*\*\*\*\*  
 [代] タンクローリー  
 \*\*\*\*\* 移動住宅車 \*\*\*\*\*  
 [代] MOBILE HOME  
 \*\*\*\*\* MEAL \*\*\*\*\*  
 [代] ME  
 \*\*\*\*\* MEKPO \*\*\*\*\*  
 [同] METHYL ETHYL KETONE PEROXIDE  
       メチルエチルケトンパーオキシド  
 [上] . 有機化合物  
       . . 化合物  
       . . . 物質  
 \*\*\*\*\* MERAPI \*\*\*\*\*  
 [上] . 火山  
       . . 地名

\*\*\*\*\* 医薬品 \*\*\*\*\*

- [上] . 物質
- [下] . MPA
  - . アジドチミジン
  - . アスピリン
  - . アセトアミノフェン
  - . インフルエンザワクチン
  - . エキセドリン
  - . 育毛剤
  - . 風邪薬

[関] MNS

\*\*\*\*\* 医薬品工場 \*\*\*\*\*

[代] 医薬品製造工場

\*\*\*\*\* 医薬品製造工場 \*\*\*\*\*

- [同] 医薬品工場
- [上] . 製造施設
  - . . 施設

\*\*\*\*\* 医療機器 \*\*\*\*\*

[同] 医療器具

- [下] . CT
  - . かん子
  - . アンブル
  - . 赤外線医療装置
  - . 保育器

\*\*\*\*\* 医療施設 \*\*\*\*\*

- [上] . 施設
- [下] . 医院
  - . 治療院
  - . 療養所

\*\*\*\*\* 医療廃棄物 \*\*\*\*\*

[上] . 廃棄物

\*\*\*\*\* 井戸 \*\*\*\*\*

[上] . 水道施設

\*\*\*\*\* 育苗ハウス \*\*\*\*\*

- [上] . ビニールハウス
  - . . 農業施設
  - . . . 施設

\*\*\*\*\* 育毛剤 \*\*\*\*\*

[上] . 医薬品

- . . 物質

\*\*\*\*\* 磯波 \*\*\*\*\*

- [上] . 風水害
  - . . 災害現象

\*\*\*\*\* 一酸化炭素 \*\*\*\*\*

[代] CO

\*\*\*\*\* 一時停止 \*\*\*\*\*

- [上] . 運転行為
  - . . 行為

\*\*\*\*\* 一般飲料 \*\*\*\*\*

- [上] . 食品
- [下] . 飲料水
  - . 清涼飲料
  - . 炭酸飲料

\*\*\*\*\* 一般団体名称 \*\*\*\*\*

- [上] . 組織名称
- [下] . GREENPEACE
  - . IRA
  - . PLO
  - . 過激派
  - . 中央競馬会

\*\*\*\*\* 稲 \*\*\*\*\*

- [上] . 植物
  - . . 生物
- [関] いもち病
  - . 米

\*\*\*\*\* 稲熱病 \*\*\*\*\*

[代] いもち病

\*\*\*\*\* 引き込まれ \*\*\*\*\*

- [上] . 人身災害
  - . . 災害現象

\*\*\*\*\* 引き込み管 \*\*\*\*\*

[代] 引込配管

\*\*\*\*\* 引き込み線 \*\*\*\*\*

[代] 引込線

\*\*\*\*\* 引き込み配管 \*\*\*\*\*

[代] 引込配管

\*\*\*\*\* 引き返し \*\*\*\*\*

[代] Uターン

\*\*\*\*\* 引き揚げ作業 \*\*\*\*\*

- [上] . 作業
  - . . 行為

\*\*\*\*\* 引込線 \*\*\*\*\*

[同] 引き込み線

引込み線

- [上] . 鉄道施設
  - . . 運輸施設
  - . . . 施設

\*\*\*\*\* 飲酒運転 \*\*\*\*\*

- [上] . 運転行為
  - . . 行為

\*\*\*\*\* 飲食店 \*\*\*\*\*

- [上] . 施設
- [下] . お好み焼店
  - . とんかつ店
  - . ウナギ店
  - . 居酒屋
  - . 中華料理店

\*\*\*\*\* MALACCA \*\*\*\*\*

[同] マラッカ海峡

[上] . 海域  
. . 地名

\*\*\*\*\* MALEV \*\*\*\*\*

[代] MA

\*\*\*\*\* MAONA LOA \*\*\*\*\*

[上] . 火山  
. . 地名

\*\*\*\*\* MARBURG \*\*\*\*\*

[同] マールブルグ病

[上] . 病気  
. . 状態

\*\*\*\*\* MARS \*\*\*\*\*

[同] MAGNETIC ELECTRONIC

AUTOMATIC RESERVATION SYSTEM

スーパーマルス

旅客販売総合システム

[上] . システム

[関] JR

コンピューター

\*\*\*\*\* MAS \*\*\*\*\*

[代] MH

\*\*\*\*\* MAYON \*\*\*\*\*

[上] . 火山  
. . 地名

\*\*\*\*\* MD81 \*\*\*\*\*

[上] . 航空機種名称  
. . 移動体名称

\*\*\*\*\* MD87 \*\*\*\*\*

[上] . 航空機種名称  
. . 移動体名称

\*\*\*\*\* MDI \*\*\*\*\*

[同] DIPHENYLMETHANE-4,4'-DIISOCYANATE

ジフェニルメタンジイソシアネート

メチレンジフェニルジイソシアネート

[上] . 有機化合物

. . 化合物

. . . 物質

\*\*\*\*\* ME \*\*\*\*\*

[同] MEA

MEAL

中東航空

[上] . 航空会社

. . 会社名称

. . . 組織名称

\*\*\*\*\* MEA \*\*\*\*\*

[代] ME

\*\*\*\*\* あめ \*\*\*\*\*

[同] キャンディ

ドロップ

飴

[上] . 菓子

. . 食品

\*\*\*\*\* あやめ \*\*\*\*\*

[同] あやめ1号

あやめ2号

[上] . 通信衛星

. . 人工衛星

. . . 飛翔体

. . . . 移動体

\*\*\*\*\* あやめ1号 \*\*\*\*\*

[代] あやめ

\*\*\*\*\* あやめ2号 \*\*\*\*\*

[代] あやめ

\*\*\*\*\* いす \*\*\*\*\*

[代] 椅子

\*\*\*\*\* いすず \*\*\*\*\*

[同] いすず自動車

[上] . 自動車会社

. . 会社名称

. . . 組織名称

[関] アスカ

ジェミニ(クルマ)

トルーパー

ビッグホーン

\*\*\*\*\* いすず自動車 \*\*\*\*\*

[代] いすず

\*\*\*\*\* いたずら \*\*\*\*\*

[同] 悪戯

[上] . 悪質行為

. . 行為

\*\*\*\*\* いたずら電話 \*\*\*\*\*

[同] 無言電話

[上] . 悪質行為

. . 行為

\*\*\*\*\* いもち \*\*\*\*\*

[代] いもち病

\*\*\*\*\* いもち病 \*\*\*\*\*

[同] いもち

稲熱病

[上] . 植物病害

[関] 稲

\*\*\*\*\* いやがらせ \*\*\*\*\*

[上] . 悪質行為

. . 行為

## 5. フリーキーワード

### 5. 1 自動切り出し用辞書の考え方

フリーキーワードは、記事フィールドなどの文字列の中からキーワードを自動的に抽出し、検索を可能にするものである。したがって、当該フィールドの文字列の中にある『単語』のみが対象となり、いかに効果的に必要な『単語』を自動抽出するかが課題となる。

#### (1) キーワード自動抽出機能

当データベース構築にあたって用いた基本ソフトには、キーワード自動切り出しについて次の3種類の機能がある。

##### ①非辞書参照モードによるキーワード自動抽出機能

この機能は、以下のようなロジックを用いて、対象とする文字列の中から単語を自動的に切り出すものである。

- 1)特定の文字を『区切り子』として登録しておき、その文字とひらがなを除くことによって文字列を分解する。
- 2)残った単語の中から、3文字以上のものをキーワードとして自動抽出する。

##### ②辞書参照モードによるキーワード自動抽出機能

この機能は、対象とする文字列の中から、フリーキーワード用辞書に登録されている単語を自動的に切り出すものである。なお、切り出される単語は最長一致をとるようになっている。たとえば文字列中に『原子力発電所』という文字があり、辞書に『原子力』『発電所』『原子力発電所』というキーワードが登録されている場合、『原子力発電所』のみが切り出される。

##### ③共用モードによるキーワード自動抽出機能

この機能は、①②の機能を兼ね合わせたものである。まず、辞書参照モードによりキーワードを切り出し、残った文字列について非辞書参照モードによりキーワードを自動抽出する。

## (2) キーワード自動抽出機能の選択

今回の検討では、まず初めに、当データベース構築にあたって用いた基本ソフトの持つ『非辞書参照モード』により、1千件の事例の記事フィールドに対してキーワード自動抽出を行った。しかしながら、この機能では後述のように十分な『単語』の抽出ができず、『辞書参照モード』によるキーワード自動抽出を行うこととし、最終的には約9千語のフリーキーワード辞書を作成した。

3種類のキーワード自動抽出機能のうち『辞書参照モード』による自動抽出を選択した過程を、次の文章を例にとって示す。

### 【切り出し文章例】

記事：M6.1の地震で住家半壊100棟、木造2階建住宅の台所で天ぷら油から出火。ぼやなどの被害、水道管等にも被害。

まず、①非辞書参照モードによる自動抽出の結果を以下に示す。

#### 〔例1〕非辞書参照モードによるキーワード自動抽出（その1）

区切り子：スペースとカンマ、ピリオド

切り出された単語：M6.1，住家半壊100棟，木造2階建住宅，水道管等

#### 〔例2〕非辞書参照モードによるキーワード自動抽出（その2）

区切り子：スペースとカンマ、ピリオド、数字、アルファベット

切り出された単語：住家半壊，階建住宅，水道管等

例1では、被害棟数など、『単語』というよりは『文節』とも言うべき文字列が膨大に切り出されてしまい、キーワードとして利用することはできない場合が多い。また、例2は、数字・アルファベットも区切り子とした場合の例であるが、このようにすると文字列そのものの意味が不明確な単語が膨大になってしまう。

以上のように、当データベースについては、非辞書参照モードによるキーワード自動抽出機能はあまり実用的ではないことがわかった。

そこで次に、辞書に〔地震，木造，住宅，台所，天ぷら油〕などの単語を登録し、②辞書参照モードによる自動抽出を行ってみた。その結果は以下の通りである。

〔例3〕辞書参照モードによるキーワード自動抽出

辞書登録された単語：木造，地震，住宅，台所，天ぷら油

抽出されたキーワード：木造，地震，住宅，台所，天ぷら油

当然のことながら、このモードでは、辞書に登録されたキーワードのみが自動抽出される。さらに、③共用モードによる自動抽出を行うと、次のようになった。

〔例4〕共用モードによるキーワード自動抽出

辞書登録された単語：木造，地震，住宅，台所，天ぷら油

区切り子：スペースとカンマ、ピリオド、数字、アルファベット

抽出されたキーワード：地震，住家半壊，木造，住宅，台所

天ぷら油，水道管等

例3と例4を比較してみると、例4では多様な単語が抽出されてはいるものの、そのままでは意味が不明確で検索に使用できない文字列（例：水道管等）も少なくない。

以上のような結果をうけて、当データベースでは、3種類のキーワード自動抽出機能のうち②辞書参照モードを採用することとした。その主な理由は以下のとおりである。

- 1)データ登録をバッチ処理で行うため、共用モードでは随時チェック・修正を行うことが難しい。
- 2)共用モードでは、辞書のメンテナンスによって不要な文字列が辞書から削除されても、再度同じ文字列が抽出されてしまい、同じメンテナンス作業を何度も繰り返さなければならない。
- 3)共用モードでは、辞書に登録されていない単語を非辞書参照モードと同様のロジックによって抽出することになるが、それによって必要な単語を適切に抽出できる可能性は低い。むしろ、辞書メンテナンスをする際に自動抽出に頼ってしまい、平仮名を含む単語を見落としがちになったり、辞書を充実させようとする意欲・意識が低下するなどの弊害の方が大きくなる。
- 4)データの不必要な増大を防ぐとともに検索効率を向上させるためには、辞書に登録されるキーワードでは不要な単語をできるだけ減らす必要がある。

5)将来的にはフリーキーワードについてもシソーラス化を検討する方向で進めたいと考えており、そのためにもできるだけキーワードの内容を整理しておきたい。

### (3) 自動抽出用辞書の作成

本調査の開始時点において、当データベースには約5千語のキーワードが登録されていた。自動抽出用辞書の作成に当たっては、まずこれらのキーワードの中から自動抽出には不要と考えられるキーワードを削除し、残りの約4千語を自動抽出用辞書(辞書名:F X)に登録した。

次いで、この辞書F Xを用いて、現在登録されているデータのうち1千件の事例の記事フィールドに対して、辞書参照型のキーワード自動抽出を行った。そして、その自動抽出の対象外となった部分に対して、特に自動抽出に必要と考えられる単語を抽出し、辞書に新規登録した。

一連の作業の過程における単語の取捨選択の基準は、次のとおりである。

①基本的に、単語として意味があり、かつ現在および将来的に複数の事例に共通して使用される可能性を持つという点を重視した。

②複合語に関しては、可能なかぎり分割した。ただし、次のような場合には分割せずにそのまま登録した。

1)分割すると一方が一文字になってしまう場合(一文字の単語が自動抽出用辞書に登録されると、本来不要な部分でヒットしてしまう可能性が多いため)

例)送電線⇒×送電:線

2)分割後の2語が、複合語以外の場合も同時に出現する可能性が高いもの(検索時に2つの単語の積をとっても、目的とする語にヒットしない場合が多いため)

例)安全装置⇒×安全:装置

③既存キーワードには、著名な災害の名称も登録されているが、フリーキーワードとして自動的に抽出する必要はないので、これらは除いた。

④国名や都市名など、場所フィールドによって検索することの可能な地名は除いた。したがって、○○岬、○○沖、○○川、○○断層などの地名は登録した。

⑤そのほかの固有名詞は、基本的に抽出の対象としないことにした。ただし、次のような項目は例外とした。

○大学の名称   ○公共公益会社(国内のみ)   ○高層ビルなど著名施設の名称

## 5. 2 自動抽出例

### 非辞書参照モードによる自動抽出例

- >>[自転車]で坂を下っていた[高校生]が空き缶に[ハンドル]をとられ転倒,[トラック]にひかれる
- >>[陣痛促進剤]の投与を受けていた妊婦が死亡
- >>Kazakhstan [ソ連初]の[スペースシャトル][「アラゴン」][緊急脱出用プラットフォーム]の切り離し失敗,秒読み停止
- >>JR[石北線北見峠付近]で雪のため[タイヤ]乱れる
- >>国道16号線で[ワゴン車]を避けようとした[オートバイ]が転倒,[トラック]にひかれる
- >>[東北道]で[トラック]が[乗用車]と衝突,[落下物]を避けようとした[ワゴン車横転],計5[台]衝突
- >>JR[信越線御代田-横川間]が雪による倒木で停電し不通.特急など14[本]運休
- >>[秋田発羽田行]ANA880便B767[機操縦室下]の[電流交換装置]から煙が出る
- >>冬型の[気圧配置]で各地に降雪.国道などで[通行止]めV
- >>天ぷら油に引火し[木造店舗兼住宅台所]から出火,2[階]部分焼くV
- >>深50km [宇都宮],[水戸],[小名浜]E
- >>CIAの[委託研究]で[精神コントロール]の[人体実験]を[精神病患者]を対象に行う

### 非辞書参照モードによる自動抽出例

- >>[自転車]で[坂]を下っていた[高校生]が[空き缶]に[ハンドル]をとられ[転倒],[トラック]に[ひかれ]る
- >>[陣痛促進剤]の[投与]を受けていた[妊婦]が死亡
- >>Kazakhstan ソ連初の[スペースシャトル][「[プラン]」][緊急脱出]用[プラットフォーム]の[切り離し][失敗],秒読み[停止]
- >>[JR][石北線]北見峠付近で雪のため[タイヤ][乱れ]る
- >>[国道]16号線で[ワゴン車]を避けようとした[オートバイ]が[転倒],[トラック]に[ひかれ]る
- >>[東北道]で[トラック]が[乗用車]と[衝突],[落下物]を避けようとした[ワゴン車][横転],計5[台]衝突
- >>[JR][信越線]御代田-横[川]間が雪による[倒木]で[停電]し[不通].[特急]など14[本][運休]
- >>秋田発[羽田]行[ANA]880便[B767]機[操縦室]下の[電流交換装置]から[煙]が出る
- >>冬型の[気圧配置]で各地に降雪.[国道]などで[通行止め]V
- >>[天ぷら油]に[引火]し[木造店舗兼住宅][台所]から[出火],2階[部分][焼く]V
- >>深50km [宇都宮],[水][戸],[小名浜]E
- >>[CIA]の委託[研究]で[精神コントロール]の[人体実験]を[精神病患者]を対象に行う

非辞書参照モードによる自動抽出例

- >>JR[山手線大塚駅構内]で男性が外回り電車にはねられる
- >>[私立湘南学園小学校]で[インフルエンザ]による集団かぜで[学級閉鎖]
- >>[スピロス社工場火災]
- >>[選挙運動員]の乗った[乗用車]に[火炎瓶]が投げられる
- >>[ジャム州]の[ヒンズー教寺院]へ[巡礼途中]の[長距離バス車内]に仕掛けられた爆弾が爆発
- >>深190km 父島II
- >>深190km 父島I
- >>[区立開進第1]小学校で[インフルエンザ]による集団かぜで[学級閉鎖]
- >>[陸上自衛隊立川駐屯地]で[偵察ヘリコプター]が[訓練中着陸]に失敗.[尾翼部分]を切断
- >>JR[東北線浦和-大宮駅間]の[レール]にひび.[普通電車]2[本運休],上下40本に遅れ
- >>[単純ヘルペスウイルス]の感染による[新生児ヘルペス急増],10年間で265例
- >>[昨年実験]のため購入した[塩化水素ポンペ]から[ガス漏れ],白煙
- >>高さ600mの[テレビ塔]が[建設中突然崩壊]
- >>猛烈な[低気圧]のため[数十年]ぶりの激しい暴風(=22557)

非辞書参照モードによる自動抽出例

- >>[JR][山手線]大塚[駅構内]で[男性]が外回り[電車]に[はねられ]る
- >>[私立]湘南学園[小学校]で[インフルエンザ]による[集団かぜ]で[学級閉鎖]
- >>スピロス社[工場]火災
- >>[選挙][運動員]の乗った[乗用車]に[火炎瓶]が投げられる
- >>ジャム州のヒンズー教[寺院]へ[巡礼]途中の[長距離バス][車内]に[仕掛け]られた[爆弾]が[爆発]
- >>深190km 父島II
- >>深190km 父島I
- >>[区立]開進第1[小学校]で[インフルエンザ]による[集団かぜ]で[学級閉鎖]
- >>[陸上自衛隊]立[川][駐屯地]で[偵察ヘリコプター]が[訓練中][着陸]に[失敗].[尾翼][部分]を[切断]
- >>[JR][東北線]浦和-大宮[駅]間の[レール]に[ひび].[普通電車]2[本][運休],上下40本に[遅れ]
- >>[単純ヘルペスウイルス]の[感染]による[新生児ヘルペス][急増],10年間で265例
- >>昨年[実験]のため[購入]した[塩化水素ポンペ]から[ガス漏れ],[白煙]
- >>高さ600mの[テレビ塔]が[建設中][突然][崩壊]
- >>猛烈な[低気圧]のため数十年ぶりの激しい暴[風](=22557)

非辞書参照モードによる自動抽出例

- >>[新潟西港]で[修理中]の[水産庁調査船]「みずほ丸」全焼
- >>[ヘリコプター]での[乗組員交代]が[不能]な天候のため[補給船]を出したが[接舷時]に[船員被災]
- >>[治山用鋼製枠]の[取付掘削場所]で[火山性爆発]が生じ[作業中]の[男性]が[土砂]に[首]まで埋まる
- >>[信濃川水力発電水路]「ずい道」[コンクリート]作業中[大落盤]が生じ47名が[埋没],3名[救出]後死亡
- >>[工場専用線]で[機関車]が[ブレーキコック]を閉めたまま発車したため[暴走]し[列車脱線転覆]
- >>[玩具ゴム銃]「[シュレットガン]」から[銃弾発射],[マンション]の[窓ガラス]貫通.悪質な[イタズラ]?
- >>[浜岡原子力発電所]2号機で[復水昇圧ポンプ]が[絶縁不良]により故障
- >>[北アルプス]前穂高岳近くの[吊尾根]で[雪崩発生],[縦走中]の[パーティ]が[巻き込まれる]
- >>[玩具ゴム銃]「[シュレットガン]」から[銃弾発射],[マンション]の[窓ガラス]貫通.悪質な[イタズラ]?
- >>[コンピューターネットワーク]「[インターネット]」を通じ[スタンフォード]研究所などに[電子ウイルス]が侵入V
- >>[トラック]の[荷崩れ]で[ライター]から[ガス漏れ]引火,炎上.[電話線]に燃え移り[電話不通]

非辞書参照モードによる自動抽出例

- >>[新潟]西[港]で[修理中]の[水産庁][調査船]「みずほ丸」[全焼]
- >>[ヘリコプター]での[乗組員]交代が[不能]な天候のため[補給船]を出したが[接舷時]に[船員][被災]
- >>[治山用鋼製枠]の取付[掘削]場所で[火山性爆発]が生じ[作業中]の[男性]が[土砂]に[首]まで埋まる
- >>[信濃川][水力発電][水路][ずい道][コンクリート][作業中]大[落盤]が生じ47名が[埋没],3名[救出]後死亡
- >>[工場専用線]で[機関車]が[ブレーキコック]を閉めたまま発車したため[暴走]し[列車][脱線][転覆]
- >>[玩具][ゴム銃]「[シュレットガン]」から[銃弾][発射],[マンション]の[窓ガラス][貫通].悪質な[イタズラ]?
- >>[浜岡][原子力発電所]2号機で[復水昇圧ポンプ]が[絶縁不良]により[故障]
- >>[北アルプス][前穂高岳]近くの[吊尾根]で[雪崩][発生],[縦走中]の[パーティ]が[巻き込まれる]
- >>[玩具][ゴム銃]「[シュレットガン]」から[銃弾][発射],[マンション]の[窓ガラス][貫通].悪質な[イタズラ]?
- >>[コンピューター][ネットワーク]「[インターネット]」を通じスタン[フォード][研究所]などに[電子ウイルス]が[侵入]V
- >>[トラック]の[荷崩れ]で[ライター]から[ガス漏れ][引火],[炎上],[電話線]に燃え移り[電話不通]

自動抽出用辞書の登録単語例

あふれ  
あやめ  
あやめ1号  
あやめ2号  
いす  
いすず  
いたずら  
いぬ  
いもち  
いやがらせ  
いん石  
うどん  
うわさ  
おがくず  
おしゃぶり  
おたふく風邪  
おもちゃ  
おもり  
お化  
お化屋敷  
お好み焼  
お茶の水女子大  
かけこみ  
かけっこ  
かじ取り  
かす  
かばん  
かまど  
かまぼこ  
かまれ  
かみこみ  
かみそり  
からし  
からしレンコン  
かんしゃく玉  
がん  
きのこ  
き酸  
くし  
くじ  
くず  
くずかご  
くり返し  
くん焼  
くん蒸  
くん蒸剤  
けい化  
けい化石灰  
けんか  
けん引  
こたつ  
こんろ  
ごま葉枯  
ごみ  
ごみ屑  
ごみ焼き  
ごみ袋  
ごみ箱  
さくら  
さくら3a

さくら3b  
さくら3号a  
さくら3号b  
さらし粉  
さんご礁  
ざるヶ岳  
しけ  
しみ抜き  
しょう紅熱  
しょう油  
しらす干し  
し尿  
し尿溝  
じん肺  
すい星  
すし  
すす  
すり鉢  
すれ違い  
ずい道  
ずきん  
ずさん  
せき  
せんべい  
せん孔  
そば  
そり  
たき火  
たこ足  
たばこ  
ため池  
たらい回し  
だんじり  
ちょうちん  
つけっ放し  
つまらせ  
つまり  
つめ込み  
つり  
てんかん  
てんぶら  
てん菜  
とう道  
とどヶ崎沖  
とばく  
とばく場  
とんかつ  
なつめやし  
なべ  
ぬいぐるみ  
ねこ  
ねじ  
のこぎり  
のぞき  
のど  
のり面  
はきけ  
はく脱  
はさまれ  
はさみ

はしか  
はしけ  
はずれ  
はねられ  
はり紙  
はれる  
ばい煎  
ひかれ  
ひき逃  
ひび  
ひまわり  
ひまわり1号  
ひまわり2号  
ひまわり3号  
ひまわり4号  
ひも  
ひょう  
ひよこ  
びょう打ち  
ふきん  
ふくそう  
ふく射  
ふげん  
ふそう  
ふ頭  
ほこり  
ぼう硝  
ぼた山  
ぼや  
まき  
まきぞえ  
まくら  
まくら木  
まくれ  
みこし  
みどりの窓口  
めまい  
もく浴  
もち  
もも  
もも1号  
やかん  
やぐら  
やけど  
やしがら  
ゆう菓  
ゆり  
ゆり2a  
ゆり2b  
ゆり2号a  
ゆり2号b  
よごし  
よごれ  
ろうそく  
ろ過  
ろ布  
わい河  
わき見  
わく星  
アーケ

自動抽出用辞書の登録単語例

亜鉛  
 亜酸化窒素  
 亜硝酸  
 亜硝酸ナトリウム  
 亜炭廃坑  
 亜麻仁  
 亜硫酸  
 亜硫酸ソーダ  
 亜硫酸塩  
 阿賀野川  
 阿久根  
 阿蘇山  
 阿多古川  
 阿武隈川  
 悪戯  
 悪臭  
 悪性  
 旭岳  
 旭川  
 芦田川  
 圧延  
 圧延機  
 圧壊  
 圧気  
 圧死  
 圧縮  
 圧縮管  
 圧縮器  
 圧縮機  
 圧縮弁  
 圧接  
 圧雪  
 圧雪車  
 圧送  
 圧送式  
 圧送車  
 圧力  
 圧力バランス  
 圧力釜  
 圧力計  
 圧力上昇  
 圧力鍋  
 圧力弁  
 圧力容器  
 安全  
 安全ベルト  
 安全器  
 安全試験  
 安全栓  
 安全装置  
 安全灯  
 安全板  
 安全弁  
 安定  
 安定器  
 安定剤  
 安田講堂  
 安田生命  
 安倍川  
 暗礁

案内  
 案内人  
 案内輪  
 伊豆急  
 伊豆急行  
 伊豆諸島  
 伊豆大島  
 委員  
 委員会  
 椅子  
 異音  
 異径  
 異材  
 異臭  
 異常  
 異常音  
 異常燃焼  
 異常繁殖  
 異性  
 異物  
 移液  
 移住  
 移出  
 移植  
 移設  
 移送  
 移送中  
 移送弁  
 移転  
 移動  
 移動住宅  
 移民  
 移民船  
 衣服  
 衣料品  
 衣類  
 違反  
 違法  
 遺産  
 遺跡  
 医院  
 医師  
 医者  
 医薬  
 医薬品  
 医療  
 井戸  
 井戸水  
 育成  
 育苗  
 育毛  
 一酸化炭素  
 一酸化窒素  
 一酸化二窒素  
 一時不通  
 一次冷却水  
 一人暮  
 溢水  
 溢流  
 逸脱

稲熱病  
 印刷  
 引き金  
 引き込まれ  
 引き返し  
 引き返す  
 引き揚げ  
 引火  
 引込  
 引抜  
 飲酒  
 飲食  
 飲食店  
 飲料  
 飲料水  
 右折  
 右翼  
 宇原川  
 宇宙  
 宇宙酔い  
 宇宙船  
 宇宙服  
 宇美川  
 鳥賊  
 羽越  
 羽根  
 羽子板  
 羽虫  
 雨戸  
 雨水  
 雨量  
 雨漏り  
 鷓  
 白杵山  
 鯉  
 厩舎  
 浦賀水道  
 瓜生島  
 運河  
 運休  
 運航  
 運行  
 運送  
 運転  
 運転員  
 運転士  
 運転手  
 運動  
 運動員  
 運動会  
 運動場  
 運搬  
 運搬船  
 運搬中  
 運輸  
 運輸省  
 運用  
 雲仙岳  
 営業  
 影響

自動抽出用辞書の登録単語例

アークヒルズ	アクリロニトリル	アルカリイオン
アーケード	アクリロニトリルモノマー	アルカリイオン水
アース	アクロバット	アルカリ骨材
アースコード	アクロレイン	アルキルアルミニウム
アースダム	アグアリコ川	アルキルベンゼン
アースドリル	アコード	アルキル化
アース線	アサリ	アルコール
アーチ	アザラシ	アルゴン
アーム	アシナガバチ	アルシン
アーモンド	アジア風邪	アルディカーブオキシム
アイアマイト	アジドチミジン	アルデヒド
アイス	アジ化ナトリウム	アルト
アイスクリーム	アスカ	アルドックス
アイスバーン	アスパルテーム	アルバイト
アイスボックス	アスピリン	アルバム
アイソトープ	アスファルト	アルファ線
アイソマックス	アスフラル	アルマイト
アイドリング	アスベスト	アルミ
アイドル	アセチルサリケル酸	アルミナ
アイルランド海	アセチレン	アルミニウム
アイロン	アセトアミノフェン	アルミホイール
アウトマーカ	アセトアルデヒド	アレージ
アウディ	アセトニトリル	アレージホール
アウディ100	アセトン	アレスティング
アウディ200	アセトンシアンヒドリン	アレルギー
アウディ5000	アゾルピン	アンカ
アウディ80	アタクチックポリマー	アンカー
アウディ90	アダムオベル	アンカーボルト
アウトバーン	アトマイザー	アングル
アエロゾル	アトモス	アンダーパス
アエロフロート	アトラス	アンチモン
アエロペルー	アトランティス	アンテナ
アオサギ	アドービ	アンデス
アカシア	アドベ	アントノフ
アカハラ	アニサキス	アントノフ12
アカンツアメンバー	アニユラ板	アントノフ124
アキュラ	アニリン	アントノフ2
アキュラインテグラ	アニリン油	アントノフ24
アクアスクーター	アパート	アントノフ24型
アクアラング	アフラトキシン	アントノフ26
アクセサリー	アブローズ	アントノフ32
アクセル	アボジモーター	アントラキノン
アクセルケーブル	アボロ	アンフォ
アクセルペダル	アボロ病	アンブ
アクチュエーター	アマゾン川	アンブル
アクリル	アマチュア	アンモオキシデーション
アクリルアルデヒド	アマルガム	アンモニア
アクリルプラスチック	アミオキシム	アンモ酸化
アクリル酸	アミノオキシム	
アクリル酸2-エチルヘキシル	アミノチアゾール	
アクリル酸2-エチルヘキシルエステル	アミルアルコール	
アクリル酸エステル	アミン	
アクリル酸エチル	アミンヒーター	
アクリル酸エチルエステル	アメリカンフットボール	
アクリル酸メチル	アメリカシウム	
アクリル酸メチルエステル	アメ横	
アクリル酸モノマー	アラル海	
アクリル樹脂	アリーナ	
アクリレート	アリアン	
	アルカリ	

## 6. 成果と今後の課題

### 6. 1 成果

本事業委託では、災害情報データベースを一般に公開するにあたって必要とされたキーワード体系の整備を実施した。キーワード体系は、①統制キーワードと②フリーキーワードの2種類に大きく区別され、そのそれぞれについて以下のような検索システムができあがった。

#### (1) 統制キーワードのシソーラス構築

従来のデータベースの蓄積に際して事故・災害ごとに付与してきたキーワードに対して、次のような観点から整理をした。

##### ①複合語の整理

これまで分割して付与されてきた複合語を接合させ、ひとつひとつの単語として登録しなおした。これによって、検索に際して発生するノイズ（不要データ）を最小限にすることができるようになった。

##### ②同義語による整理

次のような関係にある単語同士は、同義語として関係づけた。

- イ) 同一の意味で表記方法の違う単語
- ロ) 意味的にほぼ同じだが実際のニュアンスや使われ方がやや異なる単語
- ハ) 上下関係にあるが、通常はその下位ランクの情報まで把握できない場合が非常に多く区別がつかない単語

##### ③上位語・下位語による整理

『包括関係』にある単語同士は、上位語・下位語として整理した。同じように上位語・下位語として整理することも可能である『部分－全体関係』の単語同士は、混乱を避けるため、部分を表す単語をひとつの上位語でまとめ、それと全体を表す単語を関連語で関係づけた。

##### ④関連語による整理

同義語、上位・下位語の関係にはないが、意味的に何らかのつながりがある単語同士

は、関連語として整理した。上述の部分—全体関係の語のほか、原料と産物、原因と結果、施設・機器とその利用者などを関連語とした。

## (2) フリーキーワード自動切り出し用辞書の構築

利用者の多種多様なニーズに応じるため、一般に意味をなす単語をほぼ無制限に抽出するフリーキーワード制も採用した。フリーキーワードの付与は基本ソフトの持つ辞書参照型自動抽出機能を用いて行うこととし、その自動抽出用辞書を構築するため1千件のデータに対して記事フィールドを対象とした単語の抽出を行った。その結果、従来付与されていたキーワードも含めて、およそ9千語の単語が辞書に登録された。

## 6. 2 今後の課題

昨年度および今年度の2ヵ年にわたる当事業委託の成果を受けて、当データベースはこれまでのプロトシステムから新しい実用システムへの移行をほぼ完了した。平成2年4月に一般への公開を開始する予定である。しかし、今後ますます利用者のニーズに応え、データベースの活用を図っていくためには、さらに次のような課題がある。

### (1) 読み仮名の入力

当データベースが採用している基本システムの機能のひとつとして、辞書に登録されている単語には語の『読み』を持てるという機能がある。今回のキーワードシソーラスの構築によって2つの辞書（統制キーワード辞書とフリーキーワード自動抽出用辞書）が作成されたが、現段階ではこれらの辞書に登録されている単語は『読み』を持っていない。

このこと自体は、キーワードシソーラスや自動抽出の上で大きな問題となるわけではない。しかし、『読み』が登録されていると、単語のリスト化などの際にJISコード順ではなく『読み方』による五十音順にリストを作成することが可能となり、利用者にとってより使いやすいものとなることは間違いない。

このため、利用者へのより一層の便宜をはかるためにも、今回構築された辞書の単語に対して『読み』を入力する必要がある。

## (2) 残りの3万9千件に対する自動抽出

今回、フリーキーワードの自動抽出用辞書の作成にあたっては、1千件のデータに対して実施し、そこから抽出された単語を辞書登録していく作業を行った。しかし、当データベースに登録されているデータ件数は4万件を越えている。今回の作業によって、比較的多く使用される単語は辞書に登録することができたものの、利用者のニーズに応えるためには今後さらに自動抽出用辞書の充実が求められる。

したがって、今後に残された課題のひとつとしては、残りの3万9千余件のデータに対して自動抽出に必要な単語を抽出し、辞書登録することがあげられる。

## (3) 文献表題、文献名に対するキーワードの自動抽出

今回のフリーキーワード自動抽出は、データの中の記事フィールドを対象として行った。しかし、事故・災害のその後の経緯（たとえば、裁判の結果、原因調査、対策、社会的・経済的影響）などは、徐々に追加されていく文献によって情報が得られていくものが数多くある。

事故・災害のデータをその発生から追跡的に収集・整理しているという当データベースの特徴を最大限に活用するためには、キーワード抽出は記事のみではなく、文献の内容を表す文献表題、文献名に対しても行うことが不可欠である。

## (4) 既往事例の登録

これまでも当データベースでは、毎日発生している事故・災害のほか、折にふれて過去の重要な事故・災害事例の登録を進めてきた。昨年当事業委託によって実施した既往主要事例の登録も、その一環である。

しかし、これまでに発生してきた事故・災害は数多く、防災・安全の専門家の手元にはなお多くの資料が山積している。それらの資料は、各専門家の目で厳選されてきたものであり、今後の事故・災害の防止には欠かすことのできないものである。しかも、古い災害であればあるほど一般には風化していくことを考えると、早急にデータベースに登録して一般利用者の目に触れるようにしていかなければならない。

このように、既往の事故・災害事例に関するデータの登録は、その重要度を鑑みながらもさらに進めていかなければならない重要な課題である。

#### (5) 各種支援データベースの構築

当データベースの構築を進めてきた結果、事故・災害事例に関するさまざまな周辺情報も徐々に蓄積されるようになってきた。これらの周辺情報に関しても、今後はデータベース化を進め、現在のデータベースと関連づけて大いに活用されることが望まれる。

今後、構築することが望ましいデータベースとしては、たとえば次のようなものがあげられる。

##### ①VTRデータベース

事故・災害情報には、当データベースが現在利用している文献情報のほか、テレビニュースなどの映像情報もある。従来、この映像情報に関しては、当データベースの構築に努力してきた一部の関係者が個人的に収集を進めており、現在でもすでにビデオ・テープ 800本余りの映像が蓄積されている。

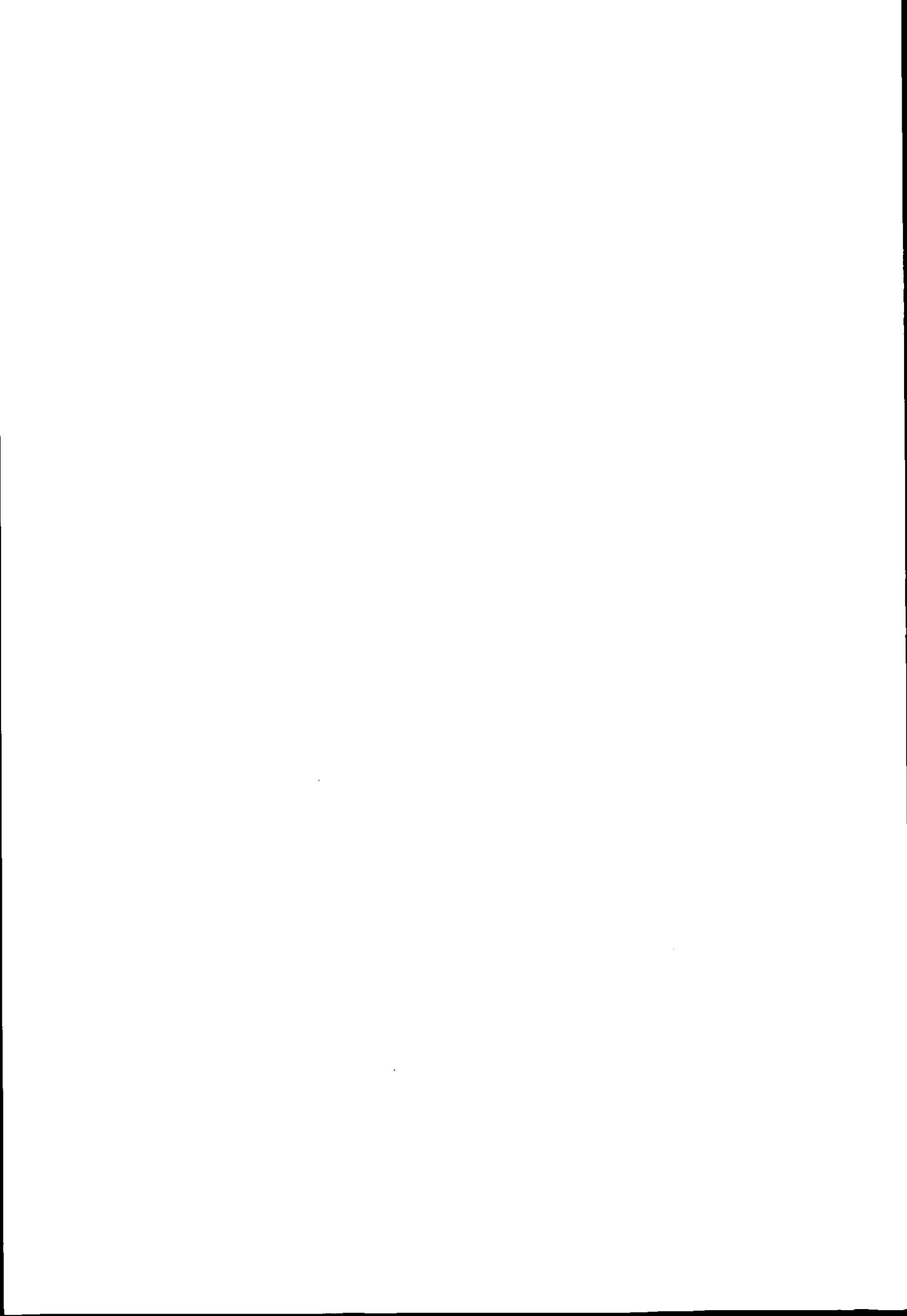
この貴重な資料を活用するために、これらの映像情報に関してもデータベースを構築し、当データベースとのリンクを図るとともに、一般への公開も進めるべきであると思われる。

##### ②事故・災害情報源データベース

現在、当データベースに入力するデータは、毎日の新聞のほか、定期・不定期を問わず発行されている雑誌・書籍類など、多種多様にわたっている。これらは、単に事故・災害事例のみではなく、広く安全・防災に関する情報を得ることができるという点で、重要な情報源であると考えることができる。また、こうした資料の中には、一般にはなかなか入手しにくく、その存在さえ知られていないものもある。

さらに、これまでの当データベースへの問い合わせ状況などを見ると、当データベースの利用者にとっては、ある文献に関する当データベースの処理状況（たとえば、定期刊行物に関して何巻何号以降が登録されているか）などの情報も、検索の際の貴重な判断基準になる例がある。

そのため、これらの情報源に関するデータを系統的に整理し、その内容や所在、入手方法、および当災害情報データベースでの処理状況などについて検索できるようなデータベースを構築することも有用なことである。



—— 禁無断転載 ——

平成2年3月発行

発行 財団法人 データベース振興センター  
東京都港区浜松町二丁目4番1号  
世界貿易センタービル7階  
TEL 03-459-8581

委託先 株式会社 防災都市計画研究所  
東京都千代田区五番町4番地5  
番町第6金井ビル3階  
TEL 03-262-6376

印刷所 株式会社 プリントボーイ  
東京都世田谷区南烏山6-24-13  
TEL 03-309-0230

