

## 日本語情報検索システム評価用テストコレクションの構築

木本晴夫<sup>†1</sup> 小川泰嗣<sup>†2</sup> 石川徹也<sup>†3</sup>  
 増永良文<sup>†4</sup> 福島俊一<sup>†5</sup> 田中智博<sup>†6</sup>  
 中渡瀬秀一<sup>†7</sup> 芥子育雄<sup>†8</sup> 豊浦潤<sup>†9</sup>  
 宮内忠信<sup>†10</sup> 上田良寛<sup>†11</sup> 松井くにお<sup>†12</sup>  
 木谷強<sup>†13</sup> 三池誠司<sup>†14</sup> 酒井哲也<sup>†15</sup>  
 徳永健伸<sup>†16</sup> 鶴岡弘<sup>†17</sup> 安形輝<sup>†18</sup>

本論文では、情報処理学会内のワーキンググループ活動によって構築し、1996年3月からモニタ公開している日本語情報検索システム評価用テストコレクション BMIR-J1 の設計・構築・利用・評価について報告する。BMIR-J1 は日本語情報検索システムのための初めてのテストコレクションであると同時に、情報検索システムに求められる、自然言語で表現された種々の検索要求を解釈・処理するためのファンクションの観点から検索要求を分類して提供している点が特長である。ファンクション分類したことにより、システム間の比較評価として、全検索要求を用いた総合的/平均的な比較だけでなく、個別のファンクションごとの分析的な比較も行えるようになった。同時に、高精度な情報検索システムの実現へ向けて、自然言語処理技術/知識処理技術を活用した取組みに期待される課題を、具体的な検索要求文の形で示したものである。公開後、多くのモニタユーザによる活発な研究発表がなされ、アンケート回答も多く寄せられている。これらから BMIR-J1 の有効性と次期バージョンへの課題などが明らかになった。

### Construction of a Test Collection for the Evaluation of Japanese Information Retrieval Systems

HARUO KIMOTO,<sup>†1</sup> YASUSHI OGAWA,<sup>†2</sup> TETSUYA ISHIKAWA,<sup>†3</sup>  
 YOSHIFUMI MASUNAGA,<sup>†4</sup> TOSHIKAZU FUKUSHIMA,<sup>†5</sup>  
 TOMOHIRO TANAKA,<sup>†6</sup> HIDEKAZU NAKAWATASE,<sup>†7</sup> IKUO KESHI,<sup>†8</sup>  
 JUN TOYOURA,<sup>†9</sup> TADANOBU MIYAUCHI,<sup>†10</sup> YOSHIHIRO UEDA,<sup>†11</sup>  
 KUNIO MATSUI,<sup>†12</sup> TSUYOSHI KITANI,<sup>†13</sup> SEIJI MIIKE,<sup>†14</sup>  
 TETSUYA SAKAI,<sup>†15</sup> TAKENOBU TOKUNAGA,<sup>†16</sup> HIROSHI TSURUOKA<sup>†17</sup>  
 and TERU AGATA<sup>†18</sup>

This paper describes BMIR-J1 (Benchmark for Information Retrieval Systems for Japanese texts Ver.1), which is designed as a test collection for Japanese information retrieval systems. The design principles, construction, user reports and the evaluation of BMIR-J1 are described. BMIR-J1, which was released for monitoring use in March 1996, was developed by a working group under the Special Interest Group on Database Systems in Information Processing Society of Japan. BMIR-J1 is the first test collection for Japanese information retrieval systems, although there are a couple of test collections for English information retrieval systems. One of the unique features of BMIR-J1 is that the queries are classified into five groups according to interpretation levels needed for processing these queries which are represented as natural language sentences. These query groups enable more analytical comparison of systems, while total comparison of systems using whole queries is possible. English test collections do not have these groups in themselves. The query groups presented by their queries themes to be studied and functions to be developed using natural language processing techniques and knowledge processing techniques for a next generation information retrieval system which will be more effective and user friendly. Since the release of BMIR-J1, many researchers and developers are using BMIR-J1, and there are a couple of research reports on the evaluation of each information retrieval system using BMIR-J1 as a test collection. Many BMIR-J1 user opinions show that BMIR-J1 is effective as a test collection for Japanese information retrieval systems, and also these opinions show future works for developing a next version of BMIR-J1.

## 1. はじめに

最初に、テストコレクション構築の背景と意義を述べる。情報検索技術は従来から、論文などの文献検索、図書検索、新聞記事検索、特許検索などの分野で必要とされ、種々の技術開発がなされ、サービス提供がされてきた。これらの技術動向、技術解説はサーベイ論文にて報告されている<sup>1)~5)</sup>。近年、インターネットが爆発的に全世界に普及するにともなって、一般のユーザがインターネット上の情報を検索するのが日常的となり、このために多くのサーチエンジンが開発されサービス提供された。一般のユーザが情報検索技術を

利用する場面が従来に比べて格段に増加した。サーチエンジンを含めて多くの情報検索システムが世に出まわっているが、日本においては、それらのシステムの検索精度評価が共通の評価用データに基づいてなされたことはない。このためにユーザはどの検索システムを選べばよいかの判断するのが困難であった。

共通の評価用データを使って、情報検索システムを相互比較することを目的として、評価用テストコレクションが、欧米では古くから作成されて、一般に提供されている<sup>6)~8)</sup>。近年は、フルテキスト形態の大規模なテストコレクションの作成が進められている<sup>8)</sup>。特に、米国規格協会(NIST)が1992年から開催している情報検索システムの評価会 TREC (Text REtrieval Conference)<sup>9)</sup>では、分野を限定しないギガバイト規模のテストコレクションが構築・使用されている。これら欧米のテストコレクション(基本的に英語\*)は、TRECに代表されるような複数システム間の優劣・特徴の公正な比較や、論文などで提案する手法の客観的な有効性検証、システムの設計・改良のための分析評価などにすでに活用されている。

しかし、日本語については、これまでテストコレクションが存在しなかった。これまで日本では、情報検索システムの開発元が独自に準備したテキスト集合と基準に基づいて、方式の精度やシステム性能が評価されており、必ずしも客観的な評価や公正な比較とはいいにくかったのが事実である。情報検索システムの客観的な評価、公正な比較のためにテストコレクションは必要なものであり、情報検索研究開発の分野全体のための共通基盤である。

以上の背景をふまえて、我々は、日本語情報検索システム評価用テストコレクションを開発し、一般に公開する必要性を痛感し、このために1993年2月に情報処理学会データベースシステム研究会の下部組織として「情報検索システム評価用データベース構築ワーキンググループ」(以下ではWGと略記)を設立し、検討を重ねてきた<sup>10)~12)</sup>。その結果、1996年3月に、中間成果物としてWG外からも広く意見を吸い上げるためのテスト版という位置付けで、小規模ではあるが、日本語としては初めてのテストコレクション BMIR-J1 (Benchmark for Information Retrieval Systems for Japanese texts Ver.1)をモニタ公開するに至った<sup>13)~15)</sup>。情報検索システムの性能として一般には処理速度と検索精度の2面が考えられるが、BMIR-J1

- 
- †1 NTT サイバースペース研究所  
NTT Cyber Space Laboratories
  - †2 リコーソフトウェア研究所  
Software Research Center, Ricoh
  - †3 図書館情報大学図書館情報学部  
University of Library and Information Science
  - †4 お茶の水女子大学理学部  
Faculty of Science, Ochanomizu University
  - †5 日本電気株式会社ヒューマンメディア研究所  
Human Media Res. Labs., NEC Corp.
  - †6 NTTコミュニケーションズ株式会社メディア技術開発センター  
Media Technology Development Center, NTT Communications Co.
  - †7 NTT東日本法人営業本部  
Multimedia Business Department NIPPON TELEGRAPH AND TELEPHONE EAST Co.
  - †8 シャープ株式会社情報商品開発研究所  
Information Systems Product Development Labs., Sharp Corp.
  - †9 三菱電機株式会社産業システム研究所  
Industrial Electrics & Systems Laboratory, Mitsubishi Electric Corp.
  - †10 富士ゼロックス IT 事業開発部  
IT Business Development Dept., Fuji Xerox Co., Ltd.
  - †11 富士ゼロックス総合研究所  
Corporate Research Labs., Fuji Xerox Co., Ltd.
  - †12 富士通研究所コンピュータシステム研究所  
Computer Systems Laboratories, Fujitsu Laboratories Ltd.
  - †13 株式会社 NTT データ技術開発本部  
Research and Development Headquarters, NTT DATA Corporation
  - †14 株式会社東芝デジタルメディア機器社情報サービス事業推進室  
Digital Media Equipment & Services Company, Toshiba Corporation
  - †15 株式会社東芝研究開発センター  
Toshiba R&D Center
  - †16 東京工業大学大学院情報理工学研究科  
Department of Computer Science, Tokyo Institute of Technology
  - †17 東京大学地震研究所  
ERI, University of Tokyo
  - †18 亜細亜大学教養部  
Faculty of Liberal Arts, Asia University

---

\* TRECでは1994年(TREC3)からスペイン語のテストコレクションも作成されている。

は主に検索精度面の評価のために用いられることを目的としている。

BMIR-J1は、基本的には欧米のテストコレクションの考え方を踏襲しているが、従来のものにはない特徴として、検索要求の表現において種々の表現方法を取り入れ、これらを検索要求のファンクションとして分類している。このようなファンクションの観点を導入したことによって、システム間の比較評価として、全検索要求を用いた総合的/平均的な比較だけでなく、個別のファンクションごとの分析的な比較（あるシステムは数値レンジ機能を備えていないとか、内容解析機能が強化されているなど）も行えるようになる。また、これらは、高精度な情報検索システムの実現へ向けて、自然言語処理技術/知識処理技術を活用した取組みが期待される課題を、具体的な検索要求文の形で示したものと見ることもできる。

今後は、欧米並の大規模なテストコレクションが必要とされると考えられるが、本報告がその開発のための一助となれば幸いである。

本論文では、まず、2章でテストコレクションの構成条件を示し、それらの構成条件が具備すべき要件を、それらに対する要求条件として3章で明らかにする。次に4章で、これらの要求条件に対するテストコレクションの設計を述べる。この設計に基づいて作成したテストコレクションの内容と具体的な作成手順を5章で説明する。テストコレクションは、1996年3月からモニタに配布して試用されているが、その利用状況を6章で述べ、モニタからのBMIR-J1に対する評価を7章にまとめた。8章で今後の課題を述べる。

## 2. テストコレクションの構成条件

テストコレクションは、検索対象テキスト集合・検索要求・適合性判定結果から構成される。検索対象テキスト集合は、システムが検索の対象とするテキストの集合である。検索要求は、検索するための入力であり、通常はキーワード（単語）や自然言語文が使われる。適合性判定結果は、検索要求に対して、検索対象テキスト集合の中で、適合しているテキストを示すものである。

## 3. テストコレクションへの要求条件

### 3.1 実用的要求条件

まず、第1に要求されるのは、テストコレクションとしての実用性である。テストコレクションは世の中の実用システムの評価のために用いられる。したがって、基本的には実用システムを模擬できる環境でな

ければならない。環境として考慮しなければならないのは、検索対象テキスト集合の内容、規模と、検索要求の表現方法と、適合性判定におけるその基準、レベルである。ここで大きな問題となるのは検索対象テキスト集合の規模である。これ以外のものは実用システムを模擬できる環境は実現可能である。実用システムの検索対象テキスト集合の規模は、たとえば、WWWのURLを検索するサーチエンジンでは、1億件を越す。新聞記事検索システムでは、300万件に及ぶものがあり、年々、50万件の規模でそれが増加している。特許検索システムも500万件の規模である。これらの件数を模擬できる件数は何件であるかが問題になるが、この問題に対して定量的に検討して報告された例は、日本国内はもとより、海外でもない。ACM SIGIRなどの国際会議やワークショップでも、情報検索システムの評価というテーマのもとにしばしば、この問題が議論されるが、結論めいたものはない。現実には可能な限り多くの件数を検索対象テキスト集合として作るというのが日本国内でも、海外でも共通の認識である。

可能な限り多くの件数を入手するためには、費用がかかる。我々は情報処理学会データベースシステム研究会のご理解によって、費用支援を受けて日経新聞社の新聞記事の使用権を得る契約をした。この契約で41843件の新聞記事の使用権契約をしたが、より多くの件数の新聞記事の使用権を得るには多くの費用が必要である。我々の使用できる費用の限界から、使用できる件数は41843件が最大であった。

TREC<sup>9)</sup>で使用される検索対象テキスト集合の規模は約70万件である。テキスト集合の内容は大部分が新聞記事やプレスリリースで、特許も含まれている。TRECではこの70万件のテキスト集合は、評価会に参加したグループにCD-ROMで配布される。TRECはNISTが主催するものであり、参加費用は不要で、新聞記事などの使用権も無償である。今後、日本でも米国の例のように公的機関によって評価会が開催され、新聞記事などのテキストが無償で使用できるようになることが必要である。

### 3.2 信頼度要求条件

テストコレクションを使用して実用システムを評価する場合には、その評価の信頼性が重要な問題となる。テストコレクションの検索対象テキスト集合の大きさは一般に実用システムの検索対象テキスト集合の大きさよりも小さく、テキスト集合の大きさによって評価結果の信頼度が左右される。この信頼度は高ければ高いほど良く、使用できるテキスト集合の大きさが、大きいほど信頼度は高くなる。使用できる検索対象テキ

スト集合の大きさはそれに充当することのできる費用の制約から自由に大きくはできない。費用の許す限り大きなテキスト集合を入手して、そのテキスト集合を使用する場合の評価の信頼度を明確にする必要がある。

### 3.3 正確性要求条件

テストコレクションの適合性判定結果は100%の正確性が要求される。また、だれが見ても同じ判定結果であり、客観的であることも要求される。しかし、適合性判定は人が検索要求を読み、理解し、そして数多くの件数を持つ検索対象テキスト集合を1つ1つ読んで、適合性判定を行うので、検索要求の理解や、テキストでの多様な表現の解釈において主観が入る余地があり、人によって判定結果が異なる場合もある。また、膨大な数で、かつ、長いテキストを読む過程において、読み間違いなどが発生する可能性もある。これらの主観の相違や読み間違いをなくし、100%の正確性を実現する方策が必要である。

### 3.4 検索要求表現からの要求条件

検索対象がテキストである場合は、検索要求はキーワードもしくは自然言語文で表現されるのが一般的である。キーワードを使用する場合は、AND、ORなどを使って論理表現を作って検索要求とすることも一般的である。論理表現による検索要求ばかりでなく、ユーザからは、自然言語文を使って検索要求を作りたいという要求がある。自然言語文で検索要求を作成すること自体は手間のかかることであるが、よりきめ細かで多様な表現が可能となる。情報検索システムはこの要求に応えることが求められる。自然言語文を使って多様に表現された検索要求を解析するためには高度な自然言語解析技術が必要である。自然言語文によって検索要求を表現することによって、これを解析するための自然言語処理技術のよりいっそうの進展を促すことは本テストコレクションの設計思想の1つでもある。

## 4. テストコレクションの設計

### 4.1 種々の要求条件

本章では、BMIR-J1の設計について述べる。テストコレクション設計への要求条件としては、前章で述べた、実用的要求条件、信頼度要求条件、正確性要求条件、検索要求表現からの要求条件がある。以下の各節では、これらの各要求条件に対するBMIR-J1の設計を述べる。

### 4.2 実用的要求条件・信頼度要求条件と設計

実用的要求条件と信頼度要求条件は、具体的にはともに検索対象テキスト集合の規模に対する要求で、どちらもその件数が多ければ多いほど要求条件が満たさ

れることになる。このことから、これらの両者を本節であわせて論じて設計に反映させる。

テキスト集合の件数を多くするうえでの問題点は2つある。1つはテキスト集合の使用権とそれにともなう使用料費用の問題である。もう1つは多くのテキストを読んで適合性判定作業をするのであるが、多すぎると作業量的に実行不可能になるという問題である。

まず、最初にどのようなテキスト集合を使うか決めなければならない。テキスト集合としては、WWWのホームページ、新聞記事データベース、特許データベース、論文・研究発表資料データベースなどがある。これらの、各々のテキスト集合は固有の性質を持っているので、以下で論ずる。WWWのホームページはいろいろな人が自由に作成するので、文書の記述方法が多様で、テキストの品質が一定でない。たとえば、「ほんとにかっこいいで〜す」のようなくだけた口語表現で書かれている場合もある。また表記のゆれが多い。たとえば、「ジャンパー」と「ジャンパ」、「マリンプルー」と「マリーンプルー」のように表記がゆれるケースが多く見られる。これらのことから、自然言語処理による文章解析が困難であったり、検索キーが一定しないなどの問題がある。特許、論文・研究発表資料は、その大部分が科学技術文献としての性質を持っている。つまり、専門用語が多く使用されていたり、数式、図表そのものが文献中で中心的な意味を持っている場合が多い。また、特許などのように、独特の文章表現をするものもある。数式、図表の扱いは、図表中のキャプションや文字を除いて、自然言語処理技術の範疇を越えるものであり、かつ、数式、図表を検索のためにどう扱うかはいまだ明らかではない。また、特許、論文などは一般の人の目に触れる機会が少ない。新聞記事は上記のような種類の文書と比べて、文体、用語、用語表現が安定しており、専門用語の割合が小さく、数式、図表も例外的に扱えるほどに出現頻度が小さい。また、多くの人が読むものであり、テストコレクションの一般性を考えると、テストコレクションとして適している。テキスト集合を決めるときに考慮した条件は、より多くの、かつ、一般の人を含めてより広範な人が検索したいものであること、そしてテキストとしての一定の品質を備えているものであることである。新聞記事は前記の2つの条件を満足しているので、これをテキスト集合として選んだ。

新聞記事の使用権は新聞社と契約して購入しなければならない。このときの契約料は、同じ新聞記事であっても、契約者や利用目的によってケースバイケースである。一般企業が研究目的で使用する場合での契約で

も契約料は非常に高額であり、新聞記事1年分(約25000件)に対して数千万円が相場である。我々は情報処理学会データベースシステム研究会のワーキンググループとして活動している、非営利のボランティアグループであるが、新聞社との契約によって、50万円で41843記事の使用権を購入できた。これ以上の費用支出は不可能であった。このように費用の制約からテキスト集合の件数を多くすることには、自ずと上限がある。

適合性判定の作業量からの制約もあるが、今回は費用の制約からテキスト集合の大きさは41843件となったために、適合性判定の作業量からの制約にはひっかからなかった。ただし、後の節でも述べるが41843件の新聞記事からサンプリングを行って抽出した600件の新聞記事に対して60件の検索要求の適合性判定作業を、1つ1つの新聞記事を読んで行う作業を2人1組で合計6組で行ったが、この作業量は、作業の限界に近かった。

#### 4.2.1 検索対象テキスト集合の規模とテストコレクションとしての精度評価能力

BMIR-J1では、検索対象テキスト集合を41843件の新聞記事として、この中から600件を無作為抽出して、適合性判定を行った。この無作為抽出した新聞記事をサンプル文書集合と呼ぶ。評価される検索システムはこのサンプル文書集合を対象として検索を行い、適合性判定結果に対する再現率・適合率を算出して、これらを検索精度とする。この600件のサンプル文書をベースとした検索精度評価が何百万件、何千万件のテキストを検索対象とする検索システムの検索精度評価にどの程度の信頼度があるかが問題である。この問題はシステムの検索精度評価の尺度である再現率、適合率の推定問題と、検索要求による適合文書の絞り込み率の問題とから成る。

再現率、適合率の推定問題はともにサンプル文書集合のサンプル数から母集合での再現率・適合率を推定する母比率推定問題になる。しかし、適合率は個々の実用システムによって異なり、そもそもテストコレクションを作成する時点では検索は行われなかったのでまったく不明である。したがって、適合率からの母比率推定はできない。再現率の母比率推定問題は、母集合である検索対象テキスト集合での適合文書数を、サンプル文書の中での適合文書数から推定する問題である。

検索要求に起因する問題は、サンプル文書数の中での適合文書数の比率の問題で、この比率を絞り込み率と呼ぶことにする。絞り込み率を0.05と仮定すると、これは100件のサンプル文書に検索を行った場合に適

合する文書が5件であることを示している。100件のサンプル文書の検索において50件が適合するような場合は通常の場合とは異なり、許容できる絞り込み率の最大の値として、最大許容絞り込み率を導入して、これを設定し、適合文書数からサンプル文書数を逆算することになる。

この信頼度推定問題の検討については、当ワーキンググループからすでに報告<sup>10),11)</sup>されているが、ここではその検討結果を示す。つまり、サンプル文書中での適合文書数を67として、最大許容絞り込み率を0.1とすると、サンプル文書数は670件となり、この場合、再現率の推定において、無限母集団に対して10%の許容誤差のもとに90%の信頼区間を有する。BMIR-J1でのサンプル文書数は600件である。

ここに示されたことから分かるとおり、600件程度の規模のサンプル文書集合では、許容誤差も大きく、信頼区間もあまり大きくないことから、より大規模のサンプル文書集合を使ったテストコレクションが作成されることが望まれる。

#### 4.3 正確性要求条件と設計

適合性判定作業においては正確性と客観性が必須である。このために2つの方策を講じた。

1つは、正解判定基準を明確にし、個々のテキストの適合性判定において、判定根拠を記述するようにした。もう1つは、2人の判定者が別々に、同じテキストに対する判定作業を行い、作業を終えてから結果を報告しあい、相互にチェックし、結果が一致したものはそのままとし、一致しないもの、あるいは結果が一致しても疑問点があるものについては、12人のワーキンググループメンバー全員で議論して判定結果を決めた。

このように正確性要求条件についてはテストコレクションの開発工程においてその対策を織り込んだ。開発工程については5.2節で詳述する。

#### 4.4 検索要求表現からの要求条件と設計

検索要求の記述において、自然言語文を使ってよいこととし、多様な検索要求に対応することにした。BMIR-J1での検索要求の自然言語文表現の種類としては、基本表現(キーワード表現とシステムでのソースラス展開)、数値レンジ表現・構文表現・言語知識表現・世界知識表現であり、これらの自然言語表現を解析・処理する機能をファンクションと呼ぶ。表現方法の詳細は5.3節で説明する。検索システム側には、これらの表現がなされた検索要求を解析したり、システム内部で持っている辞書・ルールベースなどを使用して、検索要求の表現変換や展開をして検索を行うことが求められる。

## 5. 日本語情報検索システム評価用テストコレクション BMIR-J1

### 5.1 構成

BMIR-J1 は、(1) 検索対象テキスト集合、(2) 検索要求、(3) 検索要求に対する適合性判定結果（正解テキスト集合）を備える<sup>13)~15)</sup>。

#### 5.1.1 検索対象テキスト集合

日本経済新聞の朝刊経済面 4 カ月分（1993 年 9 月 1 日～12 月 31 日）\*の 41843 件（ただし、人事異動などの記事は除いている）を母集合とし、そこから無作為に 600 記事をサンプリングしたものを BMIR-J1 の検索対象テキスト集合とした。

テキスト件数については、検索システムの比較評価に符号検定を用いることなどを含めて統計的な観点から議論を行い<sup>10),11)</sup>、WG で作成する最終的な日本語テストコレクションでは少なくとも 6000 件規模を目標とすることにした<sup>12)</sup>。中間成果物として WG 外からも広く意見を吸い上げるためのテスト版である BMIR-J1 では、テキスト件数を最終版の 1/10 である 600 件とした。また、テキストの形態はフルテキストとし、その内容は、テスト版の段階では単一分野、実際のビジネスシーンでの利用ニーズが高いと考えられる理由から経済面を選んだ。

各記事は、本文テキストに加えて、記事 ID\*\*、見出し文、書誌事項、日経新聞社で付与したキーワードリストなどが、以下の例のように SGML 形式\*\*\*で記述されている。

```
<NEWS id=nikkei-EDN-9309010003>
<見出し部>
<見出し文>NTT、希望退職 1 万人募集。</見出し文>
<クレジット>
<発行社>日本経済新聞社</発行社>
</クレジット>
</見出し部>
<記事部>
<記事>
<文章>
<段落>
```

日本電信電話（NTT）は三十一日、経営合理化の

\* 日経データと情報処理学会の間で本記事データの利用に関する契約を結んだ。本新聞記事データをご提供いただいた日経データに深謝する。

\*\* 5.1.3 項の正解テキスト集合の記述では、NEWS タグに記述された id 属性値から“nikkei-EDN-93”を除いた 8 桁の数字列を記事 ID に用いている。

\*\*\* 毎日新聞社と NTT データが新聞記事向けに定義した DTD を用いた。本 DTD の BMIR-J1 での使用を快諾して下さった毎日新聞社メディア事業部の養田正彦部長に深謝する。

一環として十月一日から一般社員を対象に約一万人の希望退職者を募集すると発表した。対象者は満四十歳以上五十七歳以下の勤続十年以上の社員で、規定の退職金に加え基本給の九カ月または十二カ月分の「特別一時金」を支給する。すでに同社の組合である全電通とは実施について合意している。同社は管理職についても希望退職者の募集を検討している。（解説 10 面に）

</段落>

<段落>

希望退職者の募集は今年十月一日、来年四月一日と十月三日からそれぞれ半月間の三期に分けて実施する。退職日は今年十二月末、来年六月末、同十二月末となる。特別一時金は今年末の退職者が基本給の十二カ月分、来年は九カ月分となる。退職後三カ月以内に故郷などへ転居する人には旅費も支給する。

</段落>

<段落>

</文章>

</記事>

</記事部>

<出稿部>

<クレジット>

<発行社>日本経済新聞社</発行社>

</クレジット>

</出稿部>

<メディア部>日経新聞 朝刊 1 ページ</メディア部>

<重要度>default prio = 5</重要度>

<キーワード>A-1:日本電信電話(ニホンテ ンシンテ ンワ)</キーワード>

<キーワード>E-1:希望退職(キホ ータイシヨク)</キーワード>

..... (中略) .....

<キーワード>T-1:特別(トクヘ ー ツ)</キーワード>

<ジャンル kind="経済">経済</ジャンル>

<開放日時>93 年 9 月 1 日</開放日時>

</NEWS>

#### 5.1.2 検索要求

検索要求は、自然言語 1 文で記述するものとし、「～に関する記事がほしい」という形式で統一した。「～」の部分にあたる名詞句のみを列挙する形式で、60 件用意した。キーワード論理式の形式では検索要求が適切に表現できないことや、実際の利用場面やシステムへの入力のしやすさの観点から自然で簡潔な表現が好ましいことから、上記の表現形式とした。ただし、5.1.3 項で述べる適合性判定について、その基準（検索意図）を明確にするためには、どうしても補足説明が必要になり、検索要求文（名詞句）とあわせて、補足説明も付記することになった。

下記の例は、BMIR-J1の検索要求のNo.13である。最初の1行が「千人以上の人員削減を行なう企業」に関する記事が欲しいという検索要求文で、後半の3行が補足説明文である。

0013-0:Q:F=○○○○x:「千人以上の人員削減を行なう企業」  
 0013-0:Q:N-1:千人以上の人員の削減を予定している企業に関する記事を探す。  
 0013-0:Q:N-2:「行なう」の解釈として、今後実施予定のものを正解とする。  
 0013-0:Q:N-3:すでに行なわれたもの(実績)は、不正解とする。

各検索要求文には、その検索要求を正確に解釈するのに情報検索システムに求められるファンクションの種類も付与されている。上記の例の先頭行に含まれる“F=”以下の5つのo/xが順に、F1:基本機能(キーワード照合、シソーラス展開)、F2:数値レンジ機能、F3:構文解析機能、F4:内容解析機能(言語知識利用)、F5:知識処理機能(世界知識利用)という5つのファンクションに対応している。これらのファンクションについては、5.3節で詳述する。ここで、oはそのファンクションが必要であり、xは必要でないことを表している。?となっている場合はその必要性はファンクションとしてはポイントではないことを表している。

### 5.1.3 適合性判定結果

各検索要求に対する適合性判定結果として、正解記事IDの集合と各々の判定理由(コメント)を用意した。正解はAランクとBランクの2種類に分けた。

**A ランク** 検索要求を主題とする記事。

**B ランク** 検索要求が主題にはなっていないが、検索要求の内容を少しでも記述している記事。

情報検索システムの性格や利用場面によって、Aランクのみを正解とするか、Aランク+Bランクを正解とするかが異なってくると考えたためである。また、不正解としたものでも、正解か否かの判定が微妙だったものや、分析の参考になりそうなものについては、Cランクを付けて不正解判定の理由をコメントに残した。A、Bランクにしたものについてもその判定の根拠となるコメントを付けた。

下記の例は、5.1.2項で例にあげた検索要求No.13に対応する適合性判定結果の記述例である。検索要求番号(この例では0013-0)と記号'R'に続く8桁の数字列が記事IDであり、その次のA/B/Cがランクを表している。

0013-0:R:09270080:A:沖電気による2000人削減計画  
 0013-0:R:09300222:B:IBMによる2万5000人の削減計画、主題は日立の話し  
 0013-0:R:09060035:C:新日鉄による2万人削減実績  
 0013-0:R:09090140:A:工場従業員の3000人削減計画(工場従業員≠技術系社員)  
 0013-0:R:09170117:A:コマツによる1000人の人員削減計画  
 0013-0:R:09230007:A:ビクターによる2000人の人員削減計画  
 0013-0:R:09270080:A:沖電気による2000人削減計画  
 0013-0:R:09280143:C:松下冷機による400人削減計画  
 0013-0:R:09290032:A:複数企業による1000人以上の削減計画  
 0013-0:R:09290140:C:秩父セメント約400人の削減計画  
 0013-0:R:09290221:C:人数が不明  
 0013-0:R:09300002:A:マツダによる3000人削減計画  
 0013-0:R:09300222:B:IBMによる2万5000人の削減計画、主題は日立の話し  
 0013-0:R:10020121:C:川崎汽船による約120人削減  
 ……(以下略) ……

## 5.2 作成手順

BMIR-J1は、以下のような手順で作成した。この手順は、いっさい外部へは委託せずに、WGメンバ(第1期の12人)自身で実行した。テストコレクションの作成においては、適合性判定基準の妥当性や、複数人で作業を行うことによる効率性と不均一性のバランスなどが問題になる。以下の手順は、この問題の議論と並行して進めていった。

- (1) 検索対象テキスト集合の決定 5.1.1項で述べたように、まず、テスト版としてのテキスト件数を600件と定めた。その一方、著作権の問題があるので、日経データと記事データ(日本経済新聞の経済面4カ月分)の利用契約を交わした。4カ月分の記事データから600件を無作為にサンプリングして、それを検索対象テキスト集合に決定した。
- (2) 検索対象テキストの割当て 600件のテキストを100件ずつの6セットに分割し、各セットに作業者を2人ずつ割り当てた。各作業者の分担するテキスト件数を絞ることで作業の効率化を図るとともに、1つの分割単位を複数人(2人)とすることで適合性判定で生じる個人差の問題が軽減できることを期待した。
- (3) 検索要求候補の作成 割り当てられた100件のテキストを参考に、各自15件ずつの検索要求候補を作成した。この15件は、5.1.2項で述べた5つのファンクションの各々について3件ずつを目安と

することで、検索要求のバリエーションを確保した（作業員間での重複がないように調整はした）。作業員は12人なので、結果として、 $15 \times 12 = 180$ 件の検索要求候補が集まった。なお、この段階では、「～に関する記事がほしい」（～は名詞句）という検索要求文の部分のみで、補足説明は付けていない。

- (4) 適合性の予備判定 作業員の各々が自分で作成した15件ずつの検索要求候補について、自分の分担している100件のテキストのみを対象に適合性判定を行った。その際に、A/B/Cなどの判定ランクや判定理由（コメント）も記述した。また、判定が微妙なケースを参考に、必要であれば、検索要求の補足説明文を書き加えた。
- (5) 検索要求の選択 (3)で作成した180件の検索要求候補から、最終的な60件の検索要求を選択した。その選択では次のような点を考慮した。まず、正解テキスト件数は、少なすぎると統計的な議論が困難であり、一方、多すぎると検索要求としては現実的でない。そこで、(4)の予備判定における正解テキスト件数（A/Bランクの総数）が1～5件程度に収まっている検索要求を選ぶようにした。次に、ファンクションのタイプや検索要求の作成者などの観点から、検索要求のバリエーションが偏らないように配慮した。
- (6) 適合性判定 各作業員は、(5)で決定された60件の検索要求のうち自分自身で作成したもの以外についても、自分の担当する100件のテキストを対象に適合性判定を行った。その際、すでに(4)で付与されている判定理由や補足説明を参考にした。
- (7) 議論・調整 (2)で述べたように、検索対象テキスト集合の1つの分割単位100件あたり2人の作業員が割り当てられている。この2人の間で適合性判定結果に食い違いが生じたものについては調整を行った。その際、必要であれば、2人の間にとどまらず、WG全体で議論した。
- (8) 検索要求と適合性判定結果の確定 (5)で触れたような正解テキスト件数の範囲に関して統計的に吟味すると、600件の検索対象テキスト集合に対して、正解テキスト件数が5～30件に収まるのが望ましい<sup>11),12)</sup>。そこで、この条件を満たした検索要求とそれ以外を区別し、条件を満たさなかった検索要求については、検索要求を微修正して(4)以降の作業を繰り返した。
- (9) 検索要求の分類 (3)や(5)の段階でファンクションを意識した作成・選択を行っているが、途

中過程で検索要求の修正も加えられているので、最終段階で、改めて各検索要求に対応するファンクションを判断した。さらに、そのファンクションの観点から、5.3節で述べるような検索要求の分類を行った。

- (10) モニタ公開の準備 検索対象テキスト集合を、SGML形式に変換し、検索要求や適合性判定結果のデータ形式、補足説明・コメントの内容や表記法なども整えた。モニタユーザ向けのドキュメント類も準備した。

### 5.3 ファンクションとその分類の提示

自然言語で様々な表現される検索要求に対して的確な検索を行うためには、基本的なキーワードマッチやシソーラス展開だけでなく、日本語テキストの深い解析や多様な知識の利用も望まれる。BMIR-J1では、各検索要求文に下記のような5種類のファンクションを対応付けることで、自然言語処理技術/知識処理技術に期待される今後の研究課題を、具体的な検索要求の形で示している。

**基本機能 (F1)** キーワードの存在確認、あるいは、キーワードのシソーラスによる展開語の存在確認。および、それらの語の存在に関する論理式（ANDやORなど）の充足判定など。

**数値レンジ機能 (F2)** 数の数え上げや、数値などの範囲を正しく解釈する。数値の大小比較や単位の理解・変換なども含む。

**構文解析機能 (F3)** 複数のキーワードの間の係り受け関係を判断する（構文解析する）。

**内容解析機能 (F4)** 通常の内容解析に必要とされるよりも深い言語知識を利用する。文脈を理解することや、言葉の深い意味を理解することを含む。

**知識処理機能 (F5)** 世界知識を利用する。常識的な判断や、蓄積された事実からの推論などを含む。

このようなファンクションの観点を導入して、各検索要求を各ファンクションに分類することによって、システム間の比較評価として、全検索要求を用いた総合的/平均的な比較だけでなく、個別のファンクションごとの分析的な比較も行えるようになる。また、個々のシステムの評価・改良にとっても、様々な難易度の検索要求をひとまとめで扱うよりは、ファンクション分類に基づき、各システムの特長・実状に合った検索要求セットが選べた方が、テストコレクションとして使いやすい。

以下には、BMIR-J1の全検索要求（60件）を、ファンクションの観点に基づいて6グループに分類して示す。

### 5.3.1 グループ A：基本機能のみ

F1 (キーワード照合・シソーラス展開などの基本機能)のみで、正解テキスト(Aランク+Bランク)が検索できるタイプ(F=oxxxx)である。

たとえば、検索要求 No.2 の「国内航空大手 3 社」は、「全日空」「日本航空」「日本エアシステム」に展開してキーワード照合することで正解テキストが得られる。以下に例を示す。

0001-0:Q:F=oxxxx:「菓子メーカー」  
 0002-0:Q:F=oxxxx:「国内航空大手 3 社」  
 0003-0:Q:F=oxxxx:「任天堂」  
 0004-0:Q:F=oxxxx:「農業」  
 0005-0:Q:F=oxxxx:「飲料品」  
 0006-0:Q:F=oxxxx:「液晶」  
 0007-0:Q:F=oxxxx:「ビデオデッキ」  
 0008-0:Q:F=oxxxx:「携帯電話またはパーソナル  
 ハンディホン」  
 0009-0:Q:F=oxxxx:「会社更生法」  
 0010-0:Q:F=oxxxx:「減税」

### 5.3.2 グループ B：数値・レンジ機能必要

ファンクションとして F2 (数値レンジ機能)を含むタイプ(F=?o???)である。F2 以外のファンクションも必要であるが、F2 に関する特別の対応の有無がポイントになる。

検索要求 No.11 では、表記のバリエーションを含めた時刻表現の抽出が必要になる。検索要求 No.12～No.15 では、「～以上」で指定された数値範囲の解釈が必要になる。これらのほかに「リットル」と「CC」のような単位換算が必要になるケースも考えられる。以下に例を示す。

0011-0:Q:F=ooxxo:「開始時間が午前 10 時の日経  
 ビジネススクール」  
 0012-0:Q:F=ooxxx:「3 期以上連続の減益企業」  
 0013-0:Q:F=ooxxx:「千人以上の人員削減を行なう  
 企業」  
 0014-0:Q:F=ooxxx:「中国にある資本金五億円以上の  
 合併企業」  
 0015-0:Q:F=ooxxx:「1 ドル=105 円以上の円高」

### 5.3.3 グループ C：構文解析機能中心

F2 (数値レンジ機能)・F4 (内容解析機能)・F5 (知識処理機能)などは必要ではないが、F3 (構文解析機能)が必要なタイプ(F=?xox)である。

たとえば、検索要求 No.16 の「半導体製品の生産」は、単に「半導体製品」の展開語と「生産」の同義語が出現すればよいわけではなく、精密に検索しようとするならば、「生産」の対象物が「半導体製品」であることの判断が求められる。以下に例を示す。

0016-0:Q:F=oxox:「半導体製品の生産」  
 0017-0:Q:F=oxox:「電話通話料金の値下げ」  
 0018-0:Q:F=oxox:「所得税の減税」  
 0019-0:Q:F=oxox:「コンピュータメーカーの  
 人員削減」  
 0020-0:Q:F=oxox:「非製造業による現地法人設立」  
 0021-0:Q:F=oxox:「電力業界の自民党に対する献  
 金」

### 5.3.4 グループ D：言語知識利用中心

F2 (数値レンジ機能)や F5 (知識処理機能)は必要ではないが、F4 (内容解析機能)が必要なタイプ(F=?x?ox)である。

たとえば、検索要求 No.23 の「円高による物価の低下」は、「円高」が原因になって「物価」に影響を及ぼしたという因果関係を、テキスト内容から把握する必要がある。同様に、検索要求 No.24, No.30 などでも因果関係の把握が必要である。また、検索要求 No.26 や No.27 における「動向」、No.28 における「実績」、No.29～No.31 における「事例」などの語も、テキスト中にそのままの言葉で現われるとは限らず、各々の語の意味に文章内容が合致するかの判断が必要になってくる。以下に例を示す。

0022-0:Q:F=oxoox:「製販一体化」  
 0023-0:Q:F=oxoox:「円高による物価の低下」  
 0024-0:Q:F=oxoox:「冷夏の被害」  
 0025-0:Q:F=oxoox:「メーカーの減益に対する対策」  
 0026-0:Q:F=oxoox:「株価動向」  
 0027-0:Q:F=oxoox:「ファクシミリの市場動向」  
 0028-0:Q:F=oxoox:「日本製品の対米輸出の実績」  
 0029-0:Q:F=oxoox:「企業における情報共有の  
 導入事例」  
 0030-0:Q:F=oxoox:「業績悪化を原因とする企業の  
 合併の事例」  
 0031-0:Q:F=oxoox:「日本の製造業における生産性  
 向上またはコスト・ダウン  
 の事例」  
 0032-0:Q:F=oxoox:「銀行の経営計画」  
 0033-0:Q:F=oxoox:「リエンジニアリングカリスト  
 ラの定義」

### 5.3.5 グループ E：世界知識利用中心

F2 (数値レンジ機能)や F4 (内容解析機能)は必要ではないが、F5 (知識処理機能)が必要なタイプ(F=?x?xo)である。

たとえば、検索要求 No.34 の「多角化事業の低迷」において「多角化事業」とは何かとか、検索要求 No.35 の「異業種会社間の共同経営」において会社 A と会社 B は「異業種」なのかといった判断は、言葉の辞書的な意味(言語知識)ではなく、むしろ世界知識的なものを必要とする。以下に例を示す。

- 0034-0:Q:F=oxoxo:「多角化事業の低迷」  
 0035-0:Q:F=oxoxo:「異業種会社間の共同経営」  
 0036-0:Q:F=oxoxo:「特殊装備自動車の新会社設立」  
 0037-0:Q:F=oxoxo:「電気メーカーの中国への投資」  
 0038-0:Q:F=oxoxo:「外国企業の日本への進出」  
 0039-0:Q:F=oxoxo:「権限の役員への委譲」  
 0040-0:Q:F=oxoxo:「管理部門の統廃合と営業部門の強化を行なう会社」  
 0041-0:Q:F=oxoxo:「アジア諸国による物資または製品の日本への輸出」  
 0042-0:Q:F=oxoxo:「北陸地方の会社」  
 0043-0:Q:F=oxoxo:「第三セクター事業運営」

### 5.3.6 グループ F: 言語知識と知識処理の併用

F2 (数値レンジ機能) は必要ではないが, F4 (内容解析機能) と F5 (知識処理機能) の両方が必要なタイプ (F=?x?oo) である。以下に例を示す。

- 0044-0:Q:F=oxooo:「現地調達」  
 0045-0:Q:F=oxooo:「流通改革」  
 0046-0:Q:F=oxooo:「経営陣刷新」  
 0047-0:Q:F=oxooo:「女性の雇用問題」  
 0048-0:Q:F=oxooo:「企業の社会貢献」  
 0049-0:Q:F=oxooo:「企業の低価格競争」  
 0050-0:Q:F=oxooo:「第3次産業のサービス向上」  
 0051-0:Q:F=oxooo:「逆輸入を行なう日本企業」  
 0052-0:Q:F=oxooo:「安売りを行なう流通業者」  
 0053-0:Q:F=oxooo:「トップの不況対策に関する発言」  
 0054-0:Q:F=oxooo:「業績不振の責任を取った経営者」  
 0055-0:Q:F=oxooo:「企業による配下企業の再編成」  
 0056-0:Q:F=oxooo:「円高対策のためのメーカーの海外進出」  
 0057-0:Q:F=oxooo:「行政機関が関係する不況対策」  
 0058-0:Q:F=oxooo:「不況におけるディスカウンターの台頭」  
 0059-0:Q:F=oxooo:「買い取り制による低価格ブランドの成長」  
 0060-0:Q:F=oxooo:「経営多角化の事例」

### 5.4 情報検索における日本語固有の問題

本節では, 情報検索における日本語固有の問題を以下にまとめた。これらの問題に対処するために, 高精度な形態素解析技術, 大規模な同義語辞書, 多義性の解消技術などが必要である。

- (1) 日本語は文の分かち書きが必要である。
- (2) 表記の揺れがある。

1つの概念を表すときに, その表記が揺れる場合がある。例:「生け花」と「生花」と「活花」,「プリンター」と「プリンタ」

- (3) 語の定義が曖昧な場合がある。

英語に比べて, 日本語では単語の単位が不明確で

ある。

例:「研究所長」の単語分割のしかたは, 以下のよう複数通りあって, 一意ではない。

- (A)「研究(サ変名詞)/所長(名詞)」
- (B)「研究(サ変名詞)/所(接尾語)  
/長(接尾語)」

この例では, テキストを(A)でインデクシングしたとき, キーワード「研究所」で検索するとヒットしない。キーワード「所長」ではヒットする。一方, テキストを(B)でインデクシングすると, キーワード「所長」で検索するとヒットせず, キーワード「研究所」ではヒットする。

このように, 単語分割の方針, 単語の定義のしかたで検索結果が変わってくる。

- (4) 同義語(異表記同義)が多数存在する。

ひらがな・カタカナ・漢字・英数字といった複数の文字セットを併用して, このために, 同義語(異表記同義)が多数存在する。

例:「英国」,「イギリス」,「United Kingdom」,「UK」  
 たとえば,「英国」を検索キーとして検索した場合は,「イギリス」,「United Kingdom」,「UK」はヒットせず, 検索漏れが多くなる。

- (5) 漢字表記によって多義性がある場合がある。

例:「米」は, 国家としての「アメリカ」を意味する場合と, 穀物の「お米」を意味する場合とがある。

多義性(同表記異義)がある場合は, 検索結果に, 要求していないものが含まれることになる。

## 6. 日本語情報検索研究における BMIR-J1 の利用状況

### 6.1 モニタ配布状況

BMIR-J1 は, 情報処理学会誌にてモニタ利用を募り, 1996年3月からモニタに配布を開始した<sup>\*</sup>。現時点でのモニタ配布先の数は50カ所である(WGメンバの所属機関も含む)。

モニタ公開後, アンケート(利用報告)の提出を依頼し, これまでに32カ所から回答が集まっている(WGメンバ自身による回答も11件含まれている)。

### 6.2 BMIR-J1 を用いた研究事例

モニタ公開後, BMIR-J1 を用いた研究発表もされ始めた。表1に, それらの研究発表<sup>16)~35)</sup> におい

<sup>\*</sup> 検索対象テキスト集合に用いた新聞記事データの利用契約により, BMIR-J1 の配布先数は制限されている。テスト版という位置付けの BMIR-J1 は, 限定モニタへの配布という形をとっている。

表1 BMIR-J1を用いた研究事例  
Table 1 A list of research reports using BMIR-J1.

研究者 (論文)	検索要求数	正解ランク	BMIR-J1の利用目的
小川 <sup>16)</sup>	正解数5~30の 47個	A + B	複合語分割やランキングのパラメータ比較
高木ら <sup>17)</sup>	2語以上の48個	A + B	重要度計算の提案法(単語出現共起利用)と従来法とを比較
伊藤ら <sup>18),19)</sup>	60個	A + B	各正解集合をフォルダとしてフォルダのスコア計算法を比較
菅井ら <sup>20),21)</sup>	表層的な10個	A + B	適合性フィードバックの3手法を比較
山田ら <sup>22),27),28)</sup>	60個	A + B	重要度計算の提案法(複合語マッチング)と従来法とを比較
木谷ら <sup>23)</sup>	60個	A + B	検索式の作成方法やフルテキスト/抽出キーワードの違いによる精度の違いを比較, ファンクション分類の視点でも比較
野口ら <sup>24)</sup>	構文要素内 共起関係を抽出できるもの34個 独自の基準19個	A + B	単語統計情報と言語情報を併用した検索モデルの評価
隅田ら <sup>25)</sup>	基本機能のみ 10個	Aを中心	統計的手法を用いて文単位抄録を作成し, これを対象として検索するシステムの評価
塩見ら <sup>26)</sup>	A正解を有する 58個	A	フロー型文書をシソーラスを用いて分類するシステムの評価
佐藤ら <sup>29)</sup>	基本機能のみ 10個	記述なし	キーワードベース分散型検索システムと高再現率検索システムの組合せの評価
酒井ら <sup>30)</sup>	60個	A + B	検索要求文からプロファイルを生成する各種手法の比較
酒井ら <sup>31)</sup>	A正解を有する 58個	Aまたは A + B	多様な検索条件ベクトルにより文書ランキングを行うシステムの評価
中島ら <sup>32)</sup>	評価用30個 パラメータ 決定用30個	A + B	Relevance Feedbackの提案法(単語の文書頻度を利用するID3法)と従来法とを比較
高木ら <sup>33)</sup>	2語以上でシソーラスに関するもの 39個	A + B	文書スコアリングの提案法(シソーラス掲載語を重要視)と従来法とを比較
小川ら <sup>34)</sup>	60個	A + B	インデクシングの提案法(統計的手法を用いた重複単語インデクシング)のパラメータ比較
小川ら <sup>35)</sup>	60個	A + B	文書ランキングでの各種のスコア合成法の比較

て, BMIR-J1がどのように活用されたかを簡単にまとめた。

これらの研究において, BMIR-J1は検索手法の精度(再現率, 適合率)を計算・比較するために用いられるのが基本である。しかし, そのような使われ方に限らず, 正解テキスト集合をクラスタと見なした評価や, 適合性フィードバックのようなユーザとのインタラクションを含めた評価などにも利用されている。

また, 木谷らの研究<sup>23)</sup>では, BMIR-J1のファンクション分類の視点に基づく分析がなされているのが興味深い。木谷らは, BMIR-J1の検索要求からキーワード条件式を作成する方法として, 計算機による自動生成(形態素解析)と人手(研究者/専門家)による作成とを比較し, 専門家による条件式作成の有効性(特に再現率の高さ)を示した。5.3節のファンクション分類ごとの精度分析では, グループD・E・Fがグルー

プA・B・Cよりも低く, 知識を必要とする検索要求は難度が高い傾向が現れた。

## 7. BMIR-J1の評価

### 7.1 欧米のテストコレクションとの比較

表2に, 欧米の古典的なテストコレクション8種と欧米の最近のテストコレクション8種に関するHarmanの比較表<sup>8)</sup>を引用し, それにBMIR-J1の諸元を並べて示した。

この比較により, BMIR-J1は検索対象テキスト集合がかなり小規模なことが分かる。

BMIR-J1の検索要求については, 中核の名詞句部分はシンプルな部類かもしれないが, 補足説明まで含めて考えると, 詳細に記述されているといえる。表2には現れていないが, ファンクション分類した検索要求を提供しているのはBMIR-J1のみであることが最

表2 テストコレクションの比較  
Table 2 Comparison of test collections.

コレクション名	テキスト件数	平均テキ スト長 (*1)	検索 要求数	平均検索 要求長 (*1)	平均正解件数
Cranfield	1398	53.1 語	225	9.2 語	7.2
ADI	82	27.1 語	35	14.6 語	9.5
MEDLARS	1033	51.6 語	30	10.1 語	23.2
TIME	423	570 語	24	16.0 語	8.7
CACM	3204	24.5 語	64	10.8 語	15.3
CISI	1460	46.5 語	112	28.3 語	49.8
NPL	11429	20.0 語	100	7.2 語	22.4
INSPEC	12684	32.5 語	84	15.6 語	33.0
OSHUMED	348566	~250 語	101	~10 語	17/19.4 (*2)
Cystic Fibrosis	1239	49.7 語	100	6.8 語	6.4~31.9 (*3)
FSupp	11953	1823 語	44	17 語	35
Fed	410883	1235 語	44	17 語	56
TREC-1	741856	444.4 語	50	83 語	277
TREC-2	741856	444.4 語	50	105 語	210
TREC-3	741856	444.4 語	50	60 語	196
TREC-4	567529	842.0 語	50	10 語	130
BMIR-J1	600	733.8 字 (*6)	60	10.9 字 /94.5 字 (*5)	5.5/10.1 (*4)

最初の 8 つは欧米の古典的なテストコレクション (TIME のみフルテキストを含むが、他は抄録)。

次の 8 つは欧米の最近のテストコレクション (OSHUMED 以外はフルテキスト)。

これら欧米の 16 コレクションの情報は文献 8) からの引用である。

(\*1) 欧米のテストコレクションでは語数でカウントしているが、有効語数は stoplist や stemming によって変化するので、ここに示した語数は比較のための目安でしかない。

(\*2) 正解が 2 ランクある。

(\*3) 正解が 6 ランクある。

(\*4) 正解が 2 ランクある。

(\*5) 検索要求の名詞句部分のみの長さとして、補足説明まで含めた長さ。

(\*6) 見出し文タグと文章タグの領域を合わせた長さ。

大の特長である。

表 2 において正解判定のランク分けは、BMIR-J1 が A/B の 2 ランクであるほか、OSHUMED も 2 ランク、Cystic Fibrosis は 6 ランクに分けて提供している。その他のものはランク分けをしていない。

## 7.2 モニタによる利用評価

6.1 節で述べたように、BMIR-J1 の 50 の配布先 (モニタ) のうち 32 カ所からアンケート (利用報告) が集まっている。アンケートでは、次のような事項を質問として用意した。

- Q1: 評価に利用した情報検索システムの概要について
- Q2: 正解のランク分けについて
- Q3: 検索要求数の多少について
- Q4: 検索要求について
- Q5: 正解集合について
- Q6: 検索要求と正解集合の両方に関連することについて
- Q7: BMIR-J1 を利用して役に立ったことについて
- Q8: ベンチマークを拡充する方法について
- Q9: その他

ここでは、Q2~Q9 の回答をもとに、モニタの評価・

感想の傾向を述べる。

今回のアンケートの回答は、選択式でなく、自由記述形式としたため、数値化による集計には不便であったが、集計可能な項目 (Q2・Q3・Q8) は表 3 にまとめてみた。

Q4・Q5・Q6 では、個々の検索要求に対して、適合性判定基準や正解追加/ランク修正に関する意見が寄せられている。それらの各々について WG 内で議論を行ったが、5.2 節で説明した作業手順に不備があったが生ずるような大きなバグの類は見られない。適合性判定の解釈に関する個人差 (ゆれ) の範囲におさまるものが多いように見受けられる。

Q7・Q9 では、BMIR-J1 の有用性に関する感想・意見や今後の要望意見などが多数寄せられた。Q7・Q9 の回答で特に多かったものや目に付いたものを以下に示す。

- 「様々な手法の比較ができるようになった」「日本で初めて検索システムの精度を評価するためのデータを使えるようになった」など比較評価や論文発表に使える点での有効性を述べたもの (WG

表3 アンケート集計の一部  
Table 3 Summary of BMIR-J1 users' opinions.

Q2: 正解のランク分け			Q3: 検索要求数の多少		
A/B/Cの3段階でよい	21	(15)	現状では少ない	16	(8)
A/Bの2段階でよい	5	(3)	現状で十分	10	(7)
1段階でよい	2	(1)	現状では多い	1	(1)
可能な限り多段階がよい	1	(0)	その他	1	(1)
その他	2	(1)	合計	28	(17)
合計	31	(20)			

  

Q8-1: 拡充してほしいテキストの種類			Q8-2: 拡充してほしいテキストの分野		
特許明細書	14	(11)	技術・コンピュータ	16	(9)
論文・抄録	15	(5)	政治・国際	4	(3)
雑誌	3	(2)	法律	2	(2)
百科辞典	3	(2)	社会・文化	2	(2)
マニュアル	3	(3)	趣味・家庭	2	(1)
WWW ページ	3	(1)	広告・新製品紹介	2	(1)
調査記録・修正履歴	2	(2)	論説・解説記事	1	(1)
判例文	2	(2)	雑多な分野	5	(2)
外国語・多言語	2	(2)	限定した分野	1	(0)
合計	47	(30)	合計	35	(21)

カウント値はWGメンバを含めたアンケート回答者の合計。

括弧内のカウント値はWG外のモニタ回答者数。

Q8の方は複数項目の重複回答あり。

外: 10件, WG内: 9件)。

- 「自分の作っているシステムに欠けている機能が明らかになった」「情報検索システムに要求される機能が5種類に分けて明示されているのは素晴らしい」など今後の研究課題を認識したというもの (WG外: 4件, WG内: 4件)。
- 「内容的にかなり精査されたベンチマークなので利用していて不安が少ない」「検索要求・正解集合についてはよく吟味されており」などテストコレクションとしての信頼性を支持するもの (WG外: 2件, WG内: 1件)。
- 「複数のシステム間の性能 (検索精度) の優劣を比較検証するにはテストコレクションとしての規模 (記事数, 検索要求数) が小さいと思う」「本格版では可能な限りデータ規模の拡充をはかってほしい」などテストコレクションの規模拡大を望むもの (WG外: 4件, WG内: 6件)。
- 「幅広い分野のベンチマーク」「インタラクションをとまなう検索システムや画像検索システムのためのベンチマーク」などテストコレクションの分野や用途の拡大を望むもの (WG外: 2件, WG内: 2件)。
- 「BMIRを使って (日本なりの) TRECやMUCのような討論の場ができるといい」など情報交換・議論の機会を望むもの (WG外: 2件)。

## 8. BMIR-J1の意義と今後の課題

6章で述べたBMIR-J1の利用状況, 7.1節における欧米のテストコレクションとの比較, 7.2節に示したモニタからのアンケート結果などをもとに, 本章ではBMIR-J1の意義と今後の課題をまとめる。

情報検索システムの比較評価と提案手法の有効性検証に使える日本で初めてのテストコレクションとしてのBMIR-J1の意義は, BMIR-J1の利用状況やモニタからの反響から確認された。

情報検索システムに求められる機能の観点を導入し, 検索要求を分類して提供したことは, 従来の欧米のテストコレクションにも見られないBMIR-J1オリジナルの特長である。これは, 今後の情報検索研究の具体的な課題提示としての役割も持つことは, モニタユーザにも認識してもらっている。

規模 (特に検索対象テキスト集合の件数) が小さいことは, 欧米のテストコレクションと比較して明らかである。モニタからの規模拡大の要望も強い。WGとしても, もともとBMIR-J1は小規模なテスト版という位置付けで考えており, 5.1.1項でも述べたように最終版 (本格版) では最低でも検索対象テキスト集合は6000件規模にはしたいと考えている。第2期WG活動においても, 規模拡大を最大の課題ととらえて検討を進めている。前記の6000件の位置付けについて述べる。検索対象テキスト集合は, 何の制約もなけれ

ば、実文書集合そのものを使うのが理想である。実文書集合の規模は、たとえば、新聞記事では、おおむね、1つの新聞社で年間10万件程度である。また、特許・実用新案データベースでは、年間約40万件のペースで増加している。テストコレクションにおいて、検索対象テキスト集合を扱ううえでの制約は大きくは2種類である。1つは、すでに述べたが、それを使用するための著作権からの制約である。もう1つは、検索対象テキスト集合に対する適合性判定作業の作業量からの制約である。今回の適合性判定作業では、判定を客観的なものにするために、2人の判定者で判定を相互チェックしている。また、精度評価の信頼性向上のために、検索要求に対する正解件数の下限を設けていて、判定作業の結果で正解件数がこの下限未満の検索要求は、テストコレクションに含めていない。これらの含まれていない検索要求に対する正解判定作業や相互チェック作業を含めて、各判定者は15個の検索要求の各々に対して、100新聞記事の適合性判定を行っている。判定者にもよるが、基本的には作業が研究のためのボランティアとして行われるケースがほとんどであり、この作業量は限界に近かった。BMIR-J1では、検索対象テキスト集合として、600文書を正解判定作業の対象としているが、その次のバージョンであるBMIR-J2では、5080文書を対象としている（前記で6000件規模と述べたが、文書を選ぶ段階で新聞記事の分野分類を利用して選んだことから、5080件となった）。この5080文書を対象とする網羅的な正解判定作業は、人が行う作業量としては限界であった。つまり、作業量の制約からも、今回扱った文書量は限界に近かった。しかしながら、何人かの判定者の協力による文書の全数チェックと判定者間の相互チェックにより、適合性判定結果は網羅的で、信頼性の高いものとなっている。英語のテストコレクションと比較すると、今回のBMIR-J1の検索対象テキスト集合の規模は小さい。英語のテストコレクションの適合性判定作業はTREC<sup>9)</sup>のようにプーリング法と呼ばれる方法によって行われている。プーリング法は、複数の検索システムを使って、それらの個々の検索システムの検索結果の上位の文書を集めて、その集合に対して人が正解判定をする方法である。つまり、複数の検索システムを使って正解候補文書を集めている。このような方法だと、検索対象文書数は機械的に処理可能な数、たとえば、TRECでは、70万件もの数を扱うことが可能である。しかし、システムの検索結果をもとにしている以上、正解の検索モレがある可能性がある。日本でも、最近、IREX<sup>36)</sup>やNTCIR<sup>37)</sup>といった、プーリング法

を使ったテストコレクション作りが開始された。IREXでの検索対象テキスト数は、211853件、NTCIRのそれは、339483件である。BMIR-J1は、検索対象テキスト集合の規模は小さいが、適合性判定において網羅的で精度が高いものであり、一方、IREX、NTCIRは検索対象テキスト集合の規模は大きい、適合性判定においては網羅的でない。今後は、両者がその特徴によって補いあって、検索システムの精度評価に利用されてゆくの理想的な形と考える。BMIR-J1における機能分類についても、検索対象テキスト集合の規模が大きくなった場合は、その適合性判定作業が困難になるが、やはりプーリング法による適合性判定をしたテストコレクションと、BMIR-J1のような網羅的な適合性判定をしたテストコレクションを相互補完的に利用して検索システムの精度評価をするのが、よりきめ細かな精度評価につながると考える。

## 9. おわりに

日本語情報検索システム評価用の最初のテストコレクションであるBMIR-J1の設計・作成について述べた。さらに、BMIR-J1のモニタ利用の状況と評価や欧米のテストコレクションとの比較に言及し、BMIR-J1の意義と課題を論じた。

BMIR-J1は、情報処理学会データベースシステム研究会の下部組織として設立された「情報検索システム評価用データベース構築ワーキンググループ」の第1期活動（1993年2月～1996年3月）の成果物である。小規模ではあるが、新聞記事（経済分野）600件を検索対象テキストとし、60件の検索要求文と、それに対する適合性判定結果（正解テキスト集合）を用意した。WG外からも広く意見を吸い上げるためのテスト版という位置付けで、1996年3月からモニタ公開をした。

BMIR-J1は、日本語情報検索システムのための初めてのテストコレクションであることに加えて、情報検索システムに求められるファクションの観点から検索要求を分類して提供している点が、従来の欧米のテストコレクションにもない新しい特長である。ファクションの観点を導入したことにより、システム間の比較評価として、全検索要求を用いた総合的/平均的な比較だけでなく、個別のファクションごとの分析的な比較も行えるようになった。また、これらは、高精度な情報検索システムの実現へ向けて、自然言語処理技術/知識処理技術を活用した取り組みが期待される課題を、具体的な検索要求文の形で示したものと見ることもできる。

BMIR-J1 を使ったモニタユーザによる研究発表やアンケート回答などから、初めての日本語テストコレクションとしての意義・役割が理解され、情報検索分野の研究発展のために有効に活用されていることが確認できた。

BMIR-J1 に対するモニタからの最大の要望は、テストコレクションの規模拡大である。現在進めている第2期 WG 活動では、テストコレクションの大規模化を最大の課題ととらえ、その他にも BMIR-J1 のモニタユーザからのフィードバックを参考にしながら、取り組みを継続している。

謝辞 情報検索システム評価用データベース構築ワーキンググループの活動をご支援いただいた情報処理学会データベースシステム研究会の田中克己前主査に深謝します。本論文を作成するにあたって議論に加わり、有益なご意見をいただいた学術情報センターの神門典子助教授に感謝の意を表します。最後に、モニタユーザとして BMIR-J1 を使用し、アンケート調査において、改良意見を含めてご回答をいただいた多くの研究開発者に深謝いたします。

### 参考文献

- 小川泰嗣：テキストデータベース技術の最近の動向，情報処理学会 ADBS '93, pp.153-162 (1993).
- 藤澤浩道，絹川博之：情報検索における自然言語処理，情報処理，Vol.34, No.10, pp.1259-1265 (1993).
- 住田一男，三池誠司：知的情報検索の動向，人工知能学会誌，Vol.11, No.1, pp.10-16 (1996).
- 仁木和久，田中克己：ニューラルネットワーク技術の情報検索への適用，人工知能学会誌，Vol.10, No.1, pp.45-51 (1995).
- 森田昌宏，速水治夫：情報フィルタリングシステム—情報洪水への処方箋，情報処理，Vol.37, No.8, pp.751-758 (1996).
- Fox, E.A.: Characterization of Two New Experimental Collections in Computer and Information Science Containing Textual and Bibliographic Concepts, Technical Report 83-561, Cornell University (1983).
- 木本晴夫：自動索引システムと情報検索システムの評価用共通データベースの事例，情報処理学会研究報告，DBS-90-10, pp.83-92 (1992).
- Harman, D. (Moderator): Panel: Building and Using Text Collections, *Proc. ACM SIGIR '96*, pp.335-337 (1996).
- Harman, D.: Overview of the Fourth Text REtrieval Conference (TREC-4), pp.1-23, National Institute of Standards and Technology (1995).
- 木本ほか：情報検索システム評価用データベースの構築の提案，情報処理学会研究報告，FI-32-1, pp.1-8 (1993).
- 石川ほか：情報検索システムの評価のためのベンチマークデータベースの構築，情報処理学会 ADBS '93, pp.217-226 (1993).
- 小川ほか：日本語情報検索システムのためのベンチマークの構築，情報処理学会研究報告，DBS-100-16, pp.145-152 (1994).
- 芥子ほか：情報検索システム評価用ベンチマーク Ver.1.0 (BMIR-J1) について，情報処理学会研究報告，DBS-106-19, pp.139-145 (1996).
- Matsui, K., et al.: Test Collection for Japanese Information Retrieval Systems from the viewpoint of evaluating system functions, *Proc. IROL '96*, pp.42-47 (1996).
- 福島ほか：日本語情報検索システム評価用テストコレクション BMIR-J1，自然言語処理シンポジウム「大規模資源と自然言語処理」(1996).
- Ogawa, Y.: Effective and Efficient Document Ranking without Using a Large Lexicon, *Proc. VLDB '96*, pp.192-202 (1996).
- 高木 徹，木谷 強：単語出現共起関係を用いた文書重要度付与の検討，情報処理学会研究報告，FI-41-8, pp.61-68 (1996).
- 伊藤史朗，大谷紀子，柴田省吾，上田隆也，池田裕治：フロー情報収集・活用のための知的検索システム Fit (2) 処理方式，第53回情報処理学会全国大会論文集，2T-9, pp.3-185-186 (1996).
- 大谷紀子，伊藤史朗，柴田省吾，上田隆也，池田裕治：フロー情報収集・活用のための知的検索システム Fit (3) 類似度判定，第53回情報処理学会全国大会論文集，2T-10, pp.3-187-188 (1996).
- 菅井 猛，和田光教，森田幸伯：WWW 上の電子新聞に対する情報フィルタリング，第53回情報処理学会全国大会論文集，4T-8, pp.3-223-224 (1996).
- 菅井 猛，和田光教：WWW 上の電子新聞に対する情報フィルタリングとその評価，情報処理学会研究報告，FI-43-13, pp.89-96 (1996).
- 山田剛一，森 辰則，中川裕志：情報検索のための複合語マッチング，情報処理学会研究報告，FI-43-5 (NL-115-13), pp.91-97 (1996).
- 木谷 強，高木 徹，木原 誠，関根道隆：フルテキストと抽出キーワードを利用した情報検索，情報処理学会研究報告，FI-43-10 (NL-115-18), pp.71-76 (1996).
- 野口直彦，稲葉光昭，野本昌子，菅野祐司：単語統計情報と言語情報とを併用した新しい文書検索のモデル，情報処理学会研究報告，FI-44-5, pp.33-40 (1996).
- 隅田英一郎，飯田 仁：統計的な抄録法を使った情報検索，言語処理学会第3回年次大会発表論文集，pp.353-356 (1997).

- 26) 塩見隆一, 徳田克巳, 青山昇一, 柿ヶ原康二: ソーラスを用いた文書データの自動分類法, 情報処理学会研究報告, NL-117-14, pp.99-104 (1997).
- 27) 山田剛一, 斎藤公一, 森 辰則, 中川裕志: 複合語マッチングによる情報検索, 言語処理学会第3回年次大会発表論文集, pp.369-372 (1997).
- 28) 山田剛一, 斎藤公一, 森 辰則, 中川裕志: 複合語マッチングによる情報検索, 第54回情報処理学会全国大会論文集, 4K-3, pp.3-27-28 (1997).
- 29) 佐藤進也, 神林 隆, 清水 奨, ポールフランシス: 広域分散検索と高再現率検索の結合について, 信学技報 DE96-77 (1997-01), pp.19-24 (1997).
- 30) 酒井哲也, 梶浦正浩, 住田一男: 情報フィルタリングシステム NEAT のための検索要求文からのプロファイル生成, 情報処理学会研究報告, FI-47-12 (NL-121-20), pp.83-88 (1997).
- 31) 酒井哲也, 梶浦正浩, 三池誠司, 佐藤 誠, 住田一男: ベンチマーク BMIR-J1 を用いた情報フィルタリングシステム NEAT の評価, 第54回情報処理学会全国大会論文集, 1S-11, pp.3-301-302 (1997).
- 32) 中島浩之, 木谷 強: 単語の文書頻度を利用した決定木学習アルゴリズムによる relevance feedback の高精度化, 情報処理学会研究報告, FI-45-2, pp.7-12 (1997).
- 33) 高木 徹, 木谷 強, 関根道隆, 出口信吾: ソーラス掲載語の重要性を考慮した文書スコアリング, 情報処理学会研究報告, FI-47-13, pp.89-94 (1997).
- 34) Ogawa, Y. and Matsuda, T.: Overlapping statistical word indexing: A new indexing method for Japanese documents, *Proc. 20th ACM SIGIR Conf.*, pp.226-234 (1997).
- 35) 小川泰嗣, 松田 透: ランキング文書検索におけるスコア合成法の評価, 情報処理学会研究報告, FI-47-14, pp.95-100 (1997).
- 36) 関根 聡, 井佐原均: IREX: 情報検索, 情報抽出コンテスト, 情報処理学会研究報告, FI-51-15, pp.109-116 (1998).
- 37) 神門典子: 情報検索システムの評価を巡って—テストコレクションとコンペティションを中心に, 1999年情報学シンポジウム, pp.129-136 (1999).
- 38) 喜連川優, 伏見信也, 田中 茂, 丸山光行: データベース処理におけるベンチマーク, 情報処理, Vol.31, No.3, pp.328-342 (1990).

(平成 10 年 4 月 9 日受付)

(平成 11 年 7 月 1 日採録)

#### 木本 晴夫 (正会員)

1949 年生. 1973 年大阪大学基礎工学部制御工学科卒業. 1975 年同大学院基礎工学研究科物理系専攻修士課程修了. 同年日本電信電話公社 (現 NTT) 入社. 現在, NTT サイバースペース研究所主幹研究員. 博士 (工学). 1993~1996 年情報処理学会データベースシステム研究会情報検索システム評価用データベース構築ワーキンググループリーダー, 1997~1998 年同幹事. 電子情報通信学会, 人工知能学会, ACM 各会員.

#### 小川 泰嗣 (正会員)

1985 年東京大学工学部計数工学科卒業. 1987 年同大学院工学系研究科情報工学修士課程修了. 同年 (株) リコー入社. 情報検索・データベース等の研究に従事. 1997~1998 年情報処理学会データベースシステム研究会情報検索システム評価用データベース構築ワーキンググループリーダー.

#### 石川 徹也 (正会員)

1971 年慶應義塾大学文学研究科修士課程修了. 現在, 図書館情報大学教授. 工学博士. 1993~1998 年情報検索システム評価用データベース構築ワーキンググループ幹事. 人工知能学会, ACM 各会員.

#### 増永 良文 (正会員)

1941 年生. 1970 年東北大学大学院工学研究科電気及通信工学専攻博士課程修了. 工学博士. 現在, お茶の水女子大学理学部情報科学科教授. データベースシステムの研究・開発に従事. 情報処理学会データベースシステム研究会主査, ACM SIGMOD 日本支部長を歴任. 現在情報処理学会監事. 電子情報通信学会, ACM, IEEE CS 各会員.

#### 福島 俊一 (正会員)

1958 年生. 1982 年東京大学理学部物理学科卒業. 同年, 日本電気 (株) 入社. 現在, 同社ヒューマンメディア研究所主任研究員. 博士 (工学). 情報検索・自然言語処理分野の研究に従事. 言語処理学会, 人工知能学会, 情報知識学会, 情報科学技術協会, ACM 各会員.

**田中 智博 (正会員)**

1987年大阪大学工学部産業機械工学科卒業。1989年同大学院工学研究科産業機械専攻修士課程修了。同年日本電信電話(株)入社。1999年NTTコミュニケーションズ(株)転籍。現在、NTTコミュニケーションズ(株)メディア技術開発センタ担当課長。1994～1996年情報処理学会データベースシステム研究会情報検索システム評価用データベース構築ワーキンググループ幹事。人工知能学会会員。

**中渡瀬秀一 (正会員)**

1967年生。1992年神戸大学大学院修士課程修了。同年、日本電信電話(株)入社。現在、NTT東日本法人営業本部。セキュリティ通信システム、XMLアプリケーションの開発に従事。

**芥子 育雄 (正会員)**

1957年生。1981年大阪大学工学部電子工学科卒業。1983年同大学院工学研究科電子工学専攻修士課程修了。同年シャープ(株)入社。現在、同社情報商品開発研究所主任研究員。情報検索、ユーザインタフェース等の研究開発に従事。電子情報通信学会、人工知能学会、ACM各会員。

**豊浦 潤 (正会員)**

1986年東京大学理学部物理学科卒業。1988年同大学院修士(物理)課程修了。1991年三菱電機(株)入社。言語処理関係の研究開発に従事。1993年より、技術研究組合新情報処理開発機構の自己組織型情報ベースの研究に従事。

**宮内 忠信 (正会員)**

1964年生。1987年東京農工大学工学部数理情報工学科卒業。同年富士ゼロックス入社。自然言語処理、情報検索等の研究開発に従事。

**上田 良寛 (正会員)**

1980年京都大学工学部電気工学第二学科卒業。1982年同大学院工学研究科修士課程修了。同年富士ゼロックス入社。1988年から1991年までATR自動翻訳電話研究所出向。機械翻訳、推敲支援、情報検索、要約等の研究に従事。言語処理学会、人工知能学会、ACL各会員。

**松井くにお (正会員)**

1957年生。1980年静岡大学工学部情報工学科卒業。同年(株)富士通研究所入社。1986年から1988年まで(株)日本電子化辞書研究所出向。自然言語処理、文書情報処理の研究開発に従事。

**木谷 強 (正会員)**

1960年生。1983年慶應義塾大学工学部電気工学科卒業。同年日本電信電話公社(現NTT)入社。1988年NTTデータ通信(株)転籍。現在、(株)NTTデータ技術開発本部部長。博士(工学)。

**三池 誠司 (正会員)**

1958年生。1982年九州大学工学部情報工学科卒業。1984年同大学院総合理工学研究科情報システム学専攻修士課程修了。同年(株)東芝入社。

**酒井 哲也 (正会員)**

1968年生。1993年早稲田大学大学院理工学研究科工業経営学専門分野修士課程修了。同年(株)東芝入社。自然言語処理、情報検索・フィルタリングの研究開発に従事。

**徳永 健伸 (正会員)**

1961年生。1983年東京工業大学工学部情報工学科卒業。1985年同大学院理工学研究科修士課程修了。現在、同大学大学院情報理工学研究科助教授。博士(工学)。自然言語処理、計算言語学、情報検索等の研究に従事。

**鶴岡 弘 (正会員)**

1966年生。1990年東北大学理学部天文および地球物理学科第二卒業。1995年同大学院理学研究科博士課程修了。現在東京大学地震研究所助手。博士(理学)。地震情報システム等の研究に従事。

**安形 輝 (正会員)**

1971年生。1993年慶應義塾大学文学部図書館・情報学科卒業。1995年同大学院文学研究科修士課程修了。現在、亜細亜大学教養部専任講師。