

博士論文

WWW 文書検索支援のための
対話的な情報検索手法に関する研究

仲川 こころ

2002年2月15日

奈良先端科学技術大学院大学
情報科学研究科 情報処理学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士(工学)授与の要件として提出した博士論文である。

仲川 こころ

審査委員： 関 浩之 教授
植村 俊亮 教授
石井 信 教授

WWW 文書検索支援のための 対話的な情報検索手法に関する研究*

仲川 こころ

内容梗概

WWW (World Wide Web) サービスの普及と流通する文書の増加に伴い、膨大な情報に誰もが自由にアクセスできる環境が実現された一方で、WWW 上の情報検索システムがきわめて重要な役割を担うようになった。現在数多くの WWW 検索サービスが提供されているが、未だ問題点も多く、専門的な知識や技術に乏しくても必要な情報を簡単に探し出せるような、使いやすい検索手法の必要性が高まっている。

従来の検索支援に関する研究が主に検索の効率や精度 (再現率・適合率) の改善を目的としているのに対し、本研究では特に情報検索作業に対するユーザのメンタルモデルに着目し、それに沿ったユーザにとって無理のない検索手法の考案を目的としている。本論文では、上記の目的に沿って提案した 2 種類の WWW 検索支援手法 (可変なカテゴリ構造を用いた支援手法、調整可能な文書ランキングを用いた支援手法) について、その詳細と評価実験の結果を報告する。

可変なカテゴリ構造を用いた検索支援手法では、検索のたびに、ユーザの検索目的を反映した小規模なディレクトリを対話的に構築することで、検索作業を支援する。既存の検索手法であるキーワード検索手法および統計的分類手法 (クラスタリング) との比較を行った評価実験では、この支援手法が特に難しい検索目的に対して有効であることが確認できた。更に使いやすさに対する被験者の主観的評価は、クラスタリングに大きく勝り、キーワード検索手法に対して遜色ない程度であった。

一方、調整可能な文書ランキングを用いた検索支援手法では、キーワード検索手法の出力である文書のランキングをユーザが直観的な方法で調整できる手段を提供することで、検索作業中に生じる心理的な負担の軽減を図る。既存のキーワード検索システムにこの支援手法を適用し、調整によるランキングの改良効果を調べた評価実験では、調整を行うことによって適合文書を選択的に上位に持ち上げていることが観測できた。これにより、下位に埋もれた未知の適合文書を発見し易くするという点で有効な手法であることが確認できた。

*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DT9961019, 2002 年 2 月 15 日.

本研究で提案する二つの支援手法は互いに独立しており，どちらも既存の検索手法および検索支援手法に対する改良として利用・併用できる技術である．本論文では最後に今後の展望として，二つの提案手法の統合や，従来の各検索支援技術と提案手法との統合の可能性について述べる．

キーワード

WWW 文書検索，情報検索システム，カテゴリ構造，文書ランキング

Interactive Methods for Supporting WWW Document Retrieval*

Kokoro Nakagawa

Abstract

As the number of documents increases explosively on World Wide Web (WWW), the necessity for an effective searching method for WWW retrieval is rapidly growing. More than a hundred Web sites (search services) for aiding document retrieval on WWW exist; however, usability problems still remain, and the load on a user in searching documents is still heavy. Searching documents not only requires much time and a large number of operations, but also puts much frustration on the user.

Many studies direct their attention to increase the efficiency, precision, and recall of retrieval. In contrast to these studies, the author aims to reduce the frustration put on a user, and to provide a more usable search method which agrees with the user's mental-model for information retrieval.

In this study, the author proposes two methods for supporting WWW document retrieval to solve some usability problems in existing search services. One is a method which **constructs a flexible category structure** (called **CATE**), and the other is a method which **allows a user to directly adjust a ranking of documents** (called **UADR**).

In the method CATE, a category structure suitable for the purpose of retrieval is constructed whenever a query is submitted by a user. By using an experimental system, CATE was evaluated with BMIR-J2 test collection to compare with a keyword-based method, and with a clustering method (*K*-means). CATE outperforms the other methods in precision and quality of resultant documents, especially for "difficult" retrieval problems.

In the method UADR, a user can pick any document in a ranked list and insert it into any position of the ranking, whenever she/he thinks that the rank of the document is inappropriately low/high. The system determines a new scoring criterion which fits the user's purpose of retrieval by analyzing the user's operation. By using the experimental system, UADR was evaluated with three methods for modifying scoring criterion based on the adjustment by a user. Results on the experiments show UADR is useful for finding out relevant documents of which ranks were low in the initial ranking before adjustment.

*Doctor's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT9961019, February 15, 2002.

Keywords:

WWW document retrieval, information retrieval, category structure, ranking of documents

目次

第1章	はじめに	1
1.1	WWW利用の現状	3
1.2	WWW検索サービス	5
1.2.1	キーワード入力型検索手法	5
1.2.2	ディレクトリ探索型検索手法	6
1.3	文書検索支援に関する関連研究	7
1.4	本研究の目的	12
第2章	検索支援手法のアプローチ	13
2.1	アプローチ1：可変なカテゴリ構造を用いた 検索支援手法(CATE手法)	13
2.1.1	既存の検索手法における問題点	13
2.1.2	CATE手法の目的と概略	15
2.1.3	関連研究—CATE手法との比較	17
2.2	アプローチ2：調整可能な文書ランキングを用いた 検索支援手法(UADR手法)	19
2.2.1	既存の検索手法における問題点	19
2.2.2	UADR手法の目的と概略	20
2.2.3	関連研究—UADR手法との比較	22
第3章	可変なカテゴリ構造を用いた検索支援手法(CATE手法)	23
3.1	はじめに	23
3.2	CATE手法	23
3.2.1	ベクトル空間モデルに基づく文書の表現	24
3.2.2	システムの動作の概要	25
3.2.3	分類観点の評価	26
3.3	性能評価実験	28
3.3.1	システムの概要	28
3.3.2	実験(1)	31
3.3.3	実験(2)	34
3.4	被験者を用いた評価実験	36

3.4.1	システムの概要と実験環境	36
3.4.2	結果と考察	38
3.5	まとめ	43
第4章	調整可能な文書ランキングを用いた検索支援手法 (UADR 手法)	45
4.1	はじめに	45
4.2	UADR 手法の概要とシステム設計の方針	45
4.2.1	システムの動作概要	45
4.2.2	ランキングの作成と調整	46
4.3	問合せベクトルの推測手法	47
4.3.1	成分補正法 (手法 A)	48
4.3.2	E 尺度に基づく推測手法 (手法 B)	50
4.3.3	順序情報に基づく推測手法 (手法 C)	51
4.4	被験者を用いた評価実験	52
4.4.1	試作システムの概要	53
4.4.2	実験環境	54
4.4.3	測定値	56
4.4.4	結果と考察	56
4.5	性能評価実験	60
4.5.1	システムの概要	60
4.5.2	実験環境	62
4.5.3	結果と考察	63
4.6	まとめ	65
第5章	おわりに	66
5.1	まとめと考察	66
5.2	今後の課題	67
	謝辞	69
	参考文献	70
	参照 URL	74
	研究業績	76
	付録	78
A	WWW 検索サービスの変遷と今後について	78
B	E 尺度に基づく推測手法の導出	78
C	CATE システムに用意した分類観点の一覧	80

目次

1.1	利用者の職業の変遷（第14回CSJ WWW利用者調査結果 ⁴⁾ より）	4
1.2	信頼できるメディア（朝日新聞社第5回新聞読者基本調査結果 ⁵⁾ より）	4
1.3	カテゴリ階層構造の例	7
2.1	ディレクトリサービスの問題点	15
2.2	検索目的に合わせたカテゴリ階層構造の例	16
2.3	ユーザによるランキングの調整	21
3.1	CATEシステム：動作の概要	25
3.2	分類の明確さ	27
3.3	分類のエントロピー	27
3.4	CATEシステムの構成	29
3.5	再現率と適合率（10課題の平均）	33
3.6	検索時間と再現率（補正後）	41
3.7	「減税」に対するカテゴリ構造	44
4.1	UADRシステムの構成	53
B.1	g, b, k の関係	79
B.2	$dk = cg - b$	79

表目次

3.1	CATE: 実験に使用した検索課題	30
3.2	CATE: 性能評価実験に使用したキーワード	30
3.3	E 尺度, $\beta = 2.0$	34
3.4	E 尺度平均値	34
3.5	評価基準と適合率との相関係数	35
3.6	E 尺度, $\beta = 2.0$ (被験者平均)	38
3.7	E 尺度平均値	38
3.8	BMIR-match (全被験者の平均)	39
3.9	検索終了時における適合率 (全被験者の平均)	40
3.10	検索時間 (キロ秒)・再現率間の回帰直線	41
3.11	アンケート結果 (全被験者の平均)	42
4.1	UADR: 実験に使用した検索課題とキーワード	54
4.2	被験者を用いた実験の結果	57
4.3	意図通りの調整ができたか	59
4.4	性能評価実験の結果	63

第1章 はじめに

いまやインターネット上の WWW (World Wide Web) サービスは、あらゆる種類の情報の宝庫として、また一般大衆にも発信可能な双方向メディアとして、従来のテレビやラジオに並んで現代社会に浸透しつつある。利用の拡大状況は数値からも明らかであり、利用者は 1993 年から 2001 年の 8 年間にほぼ 100 倍に増加し、また流通する情報は一日に 150 万ページ、8ヶ月で倍の規模で増加しているとの説もある。

現在、例えばある都市についての観光情報を探したい、あるいはある製品や商品についての様々な意見を資料として集めたい、などのごく日常的な要求に対し、答となる情報がほぼ確実に WWW 空間に存在している。しかし、情報はその創造や流通と同じく、獲得や管理も重要な要素であり、“どこかにある”という状態では役に立たない。どれほど大量の情報が流通していても、利用者にとって重要なことは、「必要な時に必要な情報だけを確実に手に入れることができる」ということである。また、利用者の最終目的は必要な情報をただ見つけることではなく、それを仕事や生活に活用することである。従って情報の獲得は可能なかぎり時間や労力をかけず、また仕事や生活の次の段階に利用しやすい形で行えることが重要である。しかし現在の WWW では、情報量の増加によって情報獲得の高速性や利便性が阻害されているのが実情である。特に近年の WWW は、理科系・技術系といった人々による研究目的中心の利用から、それ以外の人々による娯楽目的中心の利用へと徐々に変化を遂げている。このような状況に際し、専門的な知識や技術を持たなくても必要な情報を簡単に探し出せる、使いやすい情報検索手法の必要性は増すばかりである。

現在提供されている WWW 検索サービスにはキーワード入力型とディレクトリ探索型の 2 種類の検索手法が用いられているが、その使いやすさには未だ問題点が残っている。本研究では、ユーザがより自然かつ容易な方法で必要な情報を得られるような検索支援システムを提案することを目的としている。具体的には、現在の検索手法が抱える問題点を二つの視点から検討し、それぞれに対してその問題の解決を図る新しい WWW 検索支援手法を提案する。二つの支援手法は互いに独立しており、どちらも既存の WWW 検索サービスを改良する一手法として単独または併用で利用できる技術である。

一つめの提案手法である可変なカテゴリ構造を用いた検索支援手法では、検索のたびに、ユーザの検索目的を反映した小規模なディレクトリを対話的に構築することで検索作業を支援する。この支援手法は、キーワード入力型の検索手法では適切なキー

ワードを考えることが難しく、またディレクトリ探索型の検索手法ではユーザの想定する情報の分類方法が検索目的に応じて様々に変化することに対応できない、という問題の解決を図り、ユーザにとってより自然な検索手法を提案することを目的としている。既存の検索手法であるキーワード入力型検索手法およびクラスタリングによる検索支援手法との比較を行うため、筆者の属する研究グループで試作したシステムを用いて2種類の評価実験（性能評価実験と被験者による評価実験）を実施した。

一方二つめの提案手法である調整可能な文書ランキングを用いた検索支援手法では、キーワード型検索手法の出力である文書のランキングをユーザが直観的な方法で調整できる手段を提供することで、検索作業中に生じる心理的な負担の軽減を図る。この支援手法は、システムが提供するランキングの基準、すなわち、検索目的に対する文書の適合/不適合の基準が不適切であると感じても、ユーザはその不満をシステムに直接指摘できる手段を持たない、という問題の解決を図り、より適切なランキングを対話的に構築することを目的としている。本手法では、ユーザの操作に応じて現在のランキングをより適切なものへと調整するための3種類のアルゴリズムを考案した。これらのアルゴリズムによるランキングの改良効果を調べるため、既存のキーワード検索システムにこの支援手法を適用し、2種類の評価実験（被験者による評価実験と性能評価実験）を実施した。

以下、第1章では、WWWおよびWWW検索サービスの現状を紹介し、使いやすい検索手法の必要性と、本研究全体を通じての目的を述べる。また従来の文書検索支援に関する研究についてその概要を述べる。第2章では、2種類の提案手法それぞれについてそのアプローチと概略を述べ、1章で紹介した関連研究との比較を行う。第3章では可変なカテゴリ構造を用いた支援手法について、第4章では調整可能な文書ランキングを用いた支援手法について、それぞれ手法の詳細と評価実験の結果を述べ、有効性について論じる。最後に第5章では、本研究で得られた主な結果をまとめ、今後の課題や展望について総括する。

1.1 WWW利用の現状

1969年、アメリカ国防総省のARPA(Advanced Research Projects Agency: 高等研究計画局)が導入したARPAnetによって始まったインターネットの歴史は、1994年頃からのWWW(World Wide Web)ブームによって、研究者たちのものから一般大衆のものへと大きく変化を遂げた。WWWは1989年に欧州核物理学研究所で開発された分散型データベースシステムであり、1993年にWWW上のデータにアクセスするためのクライアントソフト(WWWブラウザ)を米イリノイ大学の研究グループが無料公開した事をきっかけに、爆発的にその利用が進んでいる。特に近年では、テレビやラジオのような一方性のメディアと異なり、一般人にも発信可能な双方向メディアとして急速にその認識と利用が広がった。単なるデータベースシステムとしてのみならず、株式取引や旅行の予約などWWWを窓口とした商業サービスも急増し、インターネットの中心的なサービスの一つとして現代社会に浸透しつつある。

利用者と流通する情報の増加

利用者の拡大状況は種々のインターネット調査¹⁾からも伺い知ることができる。例えばインターネットに接続されている組織数(ドメイン数)に関する調査²⁾では、2001年7月現在で約1億2589万、1993年の調査開始時に比べるとおよそ100倍の組織がインターネットに接続されているという結果が出ている。利用者の実数を世界的に調査した例はないが、およそ上記組織数に比例すると考えられている。

一方WWWで提供される情報の増加状況については、たとえば1999年2月の時点で5,000サイト以上を有する大型WWWサーバが全世界で430万台³⁾、それらのサーバから提供されたWWW文書は推定約8億ページであったという統計調査報告がNEC北米研究所から発表されている[1]。2002年1月現在では5,000サイト以上を有する大型WWWサーバが3,669万台に増加していることから、現時点で流通しているWWW文書は約68.3億ページと推定されている[1,2]。

利用形態の変化

近年のインターネットおよびWWW利用の拡大は、流通する情報や利用者の増加だけでなく、提供されるサービスの種類やユーザの利用形態にも変化を及ぼしている。例えばWWWの普及によってインターネットの世界が研究者から一般大衆へと開かれたことは前述の通りである。サイバースペースジャパンが半年毎に行うWWW利用者調査⁴⁾によると、調査を開始した1995年には利用者の職業として最も多いのは「理科系学生(36%)」、次に「技術系会社員(30%)」であったが、この二つの割合は1997年をピークにどちらも減少し、かわって「専業主婦」が急増を続け、2001年の調査結果では「専業主婦(23.1%)」「技術系以外の会社員(21.8%)」が1,2位を占める

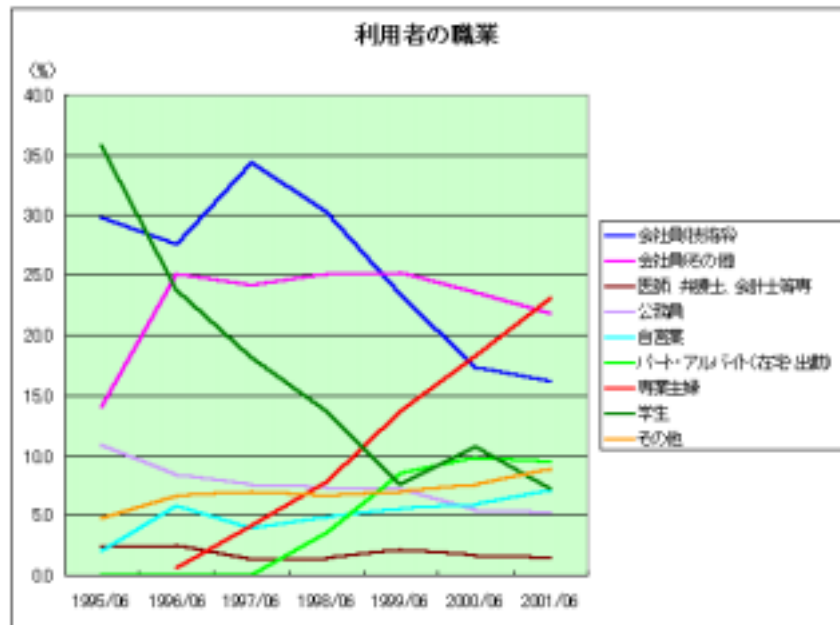


図 1.1 利用者の職業の変遷 (第 14 回 CSJ WWW 利用者調査結果⁴⁾ より)

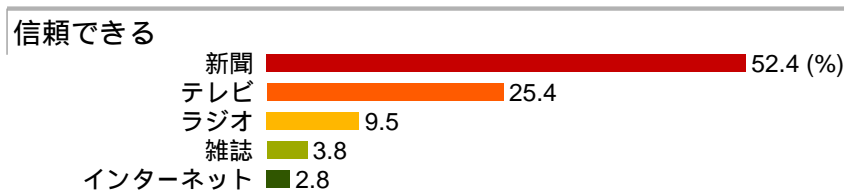


図 1.2 信頼できるメディア (朝日新聞社 第 5 回新聞読者基本調査結果⁵⁾ より)

という結果になっている (図 1.1)。またその利用目的についても「趣味や楽しみのため (72%)」が「仕事や研究の情報収集のため (20%)」を大きく凌駕する結果になっており、非専門家が娯楽目的に利用できるサービスとして広く認識が進んだことを示している。また 2000 年の同調査では“最も時間を費やすメディア”という質問について、前年まで 1 位を占めていた「テレビ」を追い抜いて「ネットサーフィン」が 1 位になり、2001 年もその状況は続いている。

しかしながら、朝日新聞社が行った 2000 年 2 月の新聞読者基本調査⁵⁾では、新聞・テレビ・ラジオ・雑誌・インターネットの 5 メディアに対する印象について「信頼できる」という解答が得られた割合は新聞が最も高く (52%)、次にテレビ、ラジオ、雑誌、の順になり、インターネットが最も低い (3%) という結果が報告されている (図 1.2)。

すなわち、現在の WWW 利用者は、日常的に WWW を利用しながらも、そこから得られる情報は膨大かつ玉石混淆であり、必要な情報を的確に得る手段としては新聞などの従来メディアに及ばないと感じていることが推察される。

より簡便で使いやすい検索手法の必要性

WWW上の検索サービスについては1.2節で詳しく述べるが、実際のところ現在の検索サービスが検索可能なデータとして収集できているのは広大なインターネットの中に流通する情報の一部にすぎない(現在推定68.3億ページの流通に対し、最大規模の検索サービスGoogle⁹⁾の収集量は20.7億ページ)。前項で述べたような流通情報の増大は、WWWをテレビと並ぶ主メディアとして確立させた一方で、ユーザが必要な情報を獲得する際の高速性や簡便性を阻害しているのが実情である。WWW空間には様々な情報があふれているが、その時本当に必要な情報を得るためには多くの時間と労力を費やさねばならない場合も多い。より早く、的確に必要な情報を得たいという要求に対して今、情報検索技術が極めて重要になっている。また、いまやWWW上のサービスは幅広いユーザに利用されるものであるため、専門的な知識や技術に乏しくても簡単に目的を達成できるような、使いやすい仕組みやインタフェースを備えることが最重要と言える。

以下、本章ではまず1.2節で現在提供されているWWW検索サービスについて述べ、次に1.3節で文書検索支援に関して従来から行われている様々な研究について概要を述べる。1.4節では本研究全体を通じての目的を述べる。

1.2 WWW検索サービス

膨大な量の情報が散在し日々増加する現在のWWWでは、必要な情報を的確に探すことは困難である。そこで必要な情報を効率良く見つけるための手段として種々のWWW検索サービスが公開され、広く利用されている。WWW検索サービスとは一般に、WWW文書のアドレス(Uniform Resource Locator: URL)と内容に関する情報(タイトルや本文中に現れる単語のリストなど)を独自に収集し蓄積したデータベースを持ち、それに対する文書検索をユーザに提供するシステムである。

現在運営されているWWW検索サービスは、キーワード入力型とディレクトリ探索型の2種類の検索方法のいずれか、または両方をユーザに提供するものである。

1.2.1 キーワード入力型検索手法

AltaVista⁷⁾、Lycos⁸⁾、Google⁹⁾などで提供されるキーワード入力型の検索手法は、ユーザが探したい情報に関連すると思うキーワードを入力し、システムは入力キーワードに合致する文書をデータベースから取り出しその一覧を出力する、という仕組みである。サービスによって定義は多少異なるが、基本的に入力語を(多数)含む文書が出力される。多くの場合、複数のキーワードをAND,OR,NOTなどの論理演算子で組み合わせたり、フレーズとして出現順序を指定するなどした検索式を入力するこ

ともできる。出力は、データベースから取り出した文書を入力に対する各文書の適合度（スコアと呼ぶ）の降順に並べたリスト形式になっている。以降、キーワード入力型の検索手法を『キーワード検索手法』、キーワード検索手法を提供する情報検索サービス/システムを『キーワード検索サービス/システム』と記す。

キーワード検索サービスでは通常、Web ロボットやスパイダーと呼ばれる WWW 探索プログラムを用いてインターネット内 WWW サーバ上の情報を定期的に収集し、データベースとして蓄積している（このため、ロボット型検索サービスとも呼ばれる）。プログラムによって自動的に全世界のデータを定期的に収集しているため、情報量が多く、かつ、情報が新しいという利点がある。一方、文書の索引付けや要約作成も自動プログラムによって行うため、実際には入力キーワードとの関連が低く検索目的に合わない文書が大量に出力されたり、要約の完成度が低いという欠点がある。例えば、人間ならば「インターネットという単語は出てくるが、この文書は日本の経済について書かれたものである」と判断できる文書でも、プログラムは「インターネットという単語が多いので（インターネットの）関連文書である」と判断してしまうため、本来関係のない文書であるにもかかわらず“インターネット”という入力キーワードに対する検索結果として出力されてしまうことになる。

最近ではこのような欠点をカバーするため、例えば「首相官邸」のように公式サイトがあるキーワードに対しては、通常の（自動的に作成される）データベースとは別に人手で登録した公式サイトデータベースを用意したり、後に 1.3 節で述べるスコア改良技術を取り入れるなど、各サービス毎に様々な工夫がなされている。

1.2.2 ディレクトリ探索型検索手法

Yahoo!¹⁰⁾、iNET Guide¹¹⁾などで提供されるディレクトリ探索型の検索手法は、あらかじめデータベースの全文書を「芸術」「ビジネス」「教育」などのように分野別に整理し、ユーザに対してはその分類結果であるディレクトリを提供する仕組みである。ディレクトリは図 1.3 のようなカテゴリの階層構造を成しており、各カテゴリの名称をハイパーリンクとして表示するインタフェースが標準的である。ユーザは、大分類項目（トップカテゴリ）から小分類項目（サブカテゴリ）へとトップダウンに絞り込んで必要な情報を探す。この探索方法は、図書の分類表や職業別電話帳を使う場合と同じであるため、ユーザにとってはキーワード入力型に比べてより自然で負担の少ない情報検索手法と考えられる。また、検索の目的にちょうど合ったカテゴリを見つけることができれば、正解文書のみが集まった精度の良い結果を得ることができる。以降、ディレクトリ探索型の検索手法を『ディレクトリ探索手法』、ディレクトリ探索手法を提供する情報検索サービス/システムを『ディレクトリサービス/システム』と記す。

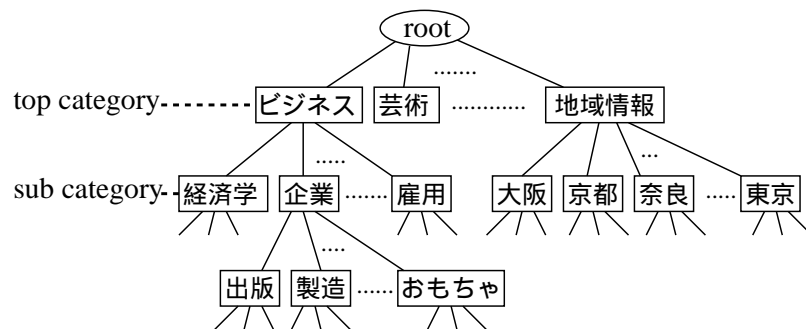


図 1.3 カテゴリ階層構造の例

ディレクトリサービスでは、各 WWW サイトや WWW 文書に対してその評価と登録・分類作業を人手で行うため、ある程度良質なサイト・文書のみが登録されており、索引や要約の完成度も高いという利点がある。一方、ロボットプログラムによって自動的に収集を行うキーワード検索サービスに比べデータ量（文書数）が 2~3 桁少ないという欠点がある。

WWW 検索サービスの分類として、上記 2 種類の他にメタ検索型と呼ばれるサービスを挙げる場合もある。メタ検索型サービスは、自身ではデータベースを持たず、ユーザからの検索要求を複数の検索サービスへ送り、その結果を統合・加工してユーザに提示するサービスである（e.g., SavvySearch [3]¹²⁾, ナレッジエリーサーチ¹³⁾）。ユーザに対しては、利用可能な論理演算子が制限される以外はキーワード検索手法と同じインタフェースを提供するため、本論文ではキーワード検索手法の一種と考える。

現在では、キーワード検索手法を主とするサービスでも定評のある良質なサイトのみを集めたディレクトリを提供し、またディレクトリ探索手法を主とするサービスでもキーワードによるディレクトリ内検索を可能にするなど、一つの WWW 検索サービス内で 2 種類の検索手法を提供するものが主流を占めるようになった。

一般に公開されている WWW 検索サービスに関する最近の話題等について、付録 A で簡単に紹介する。

1.3 文書検索支援に関する関連研究

情報科学分野では、従来、WWW 文書に限らず、新聞記事や論文などの一般的な文書や図書（および図書目録）など種々の文書データベースを対象に、より良い検索システム・検索手法に関する研究がさかんに行われてきた。本節では、1.2 節で述べた既存の検索手法（キーワード検索手法・ディレクトリ探索手法）を背景に、その改良や使いやすさの向上を目的として研究されている文書検索支援に関する最近の研究を (A) 2 次キーワードの提示、(B) 初期文書集合の分類による絞り込み、(C) 検索式の改

良, (D) スコア計算方法の改良, (E) 情報視覚化, (F) 文書タイプによる分類・検索, の6種類に分類して紹介する.

本研究は, あまり検索作業に精通していないユーザの検索作業を容易にすることを最終的な目標とした文書検索支援研究のひとつであり, 以下に挙げる研究と目標を同じくするものである. 本研究で提案する検索支援手法と従来の研究との具体的な比較については後に第2章(2.1.3節, 2.2.3節)で詳しく述べる.

(A) 2次キーワードの提示

キーワード検索手法において, 初期キーワードの入力によって得られた結果を更に絞り込む(あるいは新しくキーワードとして採用する)のに有用な2次キーワードを提供するシステムがいくつか提案され, 実際に運用されている. 検索システムRCAAU [4]は, 初期文書集合に対してデータマイニング技術を応用することで2次のキーワードを抽出し, ユーザに提供する. 同様にAnickらのシステム [5]も, 絞り込みに利用できる2次キーワードや成句を提供するものである. Anickらは, 語彙分散の高い単語(*facet*と呼ばれる)が, 文書集合に含まれる重要概念を表現するために有効であると考えている. そこでユーザに対しては, 初期文書集合から求めた*facet*と*facet*中の名詞句を, 検索式に関係する副次的概念として提示する. またHiB [6]は, 初期文書集合の自然言語解析を行い, 検索式を精練するのに有用な句を抽出してユーザに示すシステムである.

(B) 初期文書集合の分類による絞り込み

同じくキーワード検索手法の支援として, 初期キーワードの入力によって得られた結果(初期文書集合)を分類することでより精度の高い部分集合を得ることを目的としたシステムが数多く研究されている. このタイプの検索支援システムは, 初期文書集合の分類に統計的分類手法を用いる手法と, 先験的知識を用いる手法に大別できる.

統計的な分類手法: 文書集合の統計的性質を利用してもっともらしい分割を行う手法で, 例えばScatter/Gather [7, 8]はクラスタリングによってデータベースの全文書から1個の階層構造を構築するシステムである.

統計的手法による分類は, あらかじめ分類体系やシソーラスのような先験的知識を用意する必要がない, という利点がある. しかし, 統計的手法による分類には, 分類で得られた各部分集合がどのような内容か理解しにくいという問題がある. ユーザに適切な部分集合を選択してもらうためにはその特徴を簡潔に表示する必要があるが, 各部分集合にそのようなラベル(部分集合の名前や要約, ディレクトリ内のカテゴリ名に相当するもの)を自動的に作成することは極めて困難である.

先験的知識を用いる分類： 検索システム HIBROWSE [9] は、先験的知識として *view* と呼ばれる部分的なシソーラスを用いてデータベースの全文書を分類するものである。ユーザは、いくつかの *view* による分類結果を同時に見ながら、*view* 中のキーワードを使って現在の結果を絞り込むことが出来る。また、新しく収集した文書を既に構築されたカテゴリ階層構造のどのカテゴリに分類・登録すべきかを推定する手法 [10-12] などの研究も、先験的知識を用いる分類手法の一つと考えられる。

(C) 検索式の改良

キーワード検索手法において、現在の検索式（問合せベクトル）をより精度の良い結果を得られるように自動/半自動的に洗練する支援手法が古くから研究されている。代表的な手法として、現在の検索式で得た文書集合中の各文書に対し、ユーザが適合・不適合の判定を行い、システムはその判定結果を基に検索式を洗練するという適合フィードバック手法 [13] があり、この手法を応用した検索支援システムが多数考案されている。例えば検索システム ARF [14] (adaptive relevance feedback) では、時間的に変化するユーザの検索目的に適應しながら検索式を改良する。更にクラスタリング手法を ARF に応用し、ユーザの適合・不適合判定作業をクラスタ単位で行えるように拡張したシステム AIQEC [15] など提案されている。

また適合フィードバックとは異なるが、ユーザの入力したキーワードの履歴を基に現在の検索式を自動的に洗練する VOIR [16] や、検索依頼を送るいくつかの WWW 検索サービスそれぞれに特化した検索式の洗練方法を自動獲得するメタ検索システム Inquirus 2 [17] などの研究もある。

(D) スコア計算方法の改良

スコアとは、キーワード検索手法において、ユーザが入力したキーワード（または検索式）と各文書との関連度として計算される値である。キーワード検索システムは、検索要求のたびに、スコアが大きい値よりも大きい文書だけをデータベースから取り出し、それらをスコアの降順に並べたリストをユーザに提示する仕組みになっている。基本的に各文書のスコアは、文書中にキーワードができるだけ多くかつ集中して出現している場合に高得点となるような計算規則によって求められるが、より適切なスコアを与えるための様々な付加的技術が研究されている。現在の WWW 検索サービスの多くはキーワードの珍しさに着目する $tf * idf$ [18] や、文書の構造解析、アンカーテキスト解析、PageRankTM 技術 [19] などを用いてスコアの改良を図っている。

近年脚光をあびている Google 社の PageRank 技術は「他の文書からどれだけリンクされているか」という参照頻度をもとに各文書の重要度を算出し、スコア計算規則に反映させるものである。この際、リンク元の文書についても重要度を考慮し、“重要度

の高い文書からのリンクは高得点になる”という具合に、各文書の重要度をリンク構造によって伝搬する仕組みになっている。従って例えば Yahoo!のような人気サイトからリンクされている文書や、多くの文書からリンクされている文書は重要度が高くなり、スコアも高くなる。PageRank 技術によって計算される各文書の重要度は WWW 上での認知（人気）度を示すようなものであり、入力キーワードに対する関連度とは別の指標である。Google では、入力キーワードに対する関連度計算については、従来のスコア計算規則に加えアンカーテキスト解析（リンク元のアンカーテキストと入力キーワードの適合度を調べる）などを併用して行っている。

(E) 情報視覚化

情報検索を支援するための異なるアプローチとして、文書集合をユーザに分かりやすく提示しようという情報視覚化の技術がある。視覚化技術は、特に巨大な文書集合を一度に表示したり、文書間の関係などをユーザに理解させる際に効力を発揮するものである。WWW に適用した研究例も数多くあり、文書集合の類似度の関係（似ている / 似ていない）を提示するもの [20]、入力キーワードに相当するフィルタを通して全文書集合を提示するもの [21]、WWW サーバ間の関係を提示するもの [22]、リンク構造を提示するもの [23,24] など様々である。また、情報視覚化の種類（テキスト・2次元視覚化・3次元視覚化）によって検索作業の効率やユーザの感じる使いやすさなどのように異なるかを比較した研究 [25] などもある。

一方、寺岡らのシステム [26] では、ユーザは、情報を扱う際の要求や目的を“視点”と呼ばれる多次元空間中の位置で指定することができる。システムは、ユーザの指定した視点に基づいて関連のある情報を適宜取捨選択し、その結果を Relationship Focused Cone (RF-Cone) tree の形で提示する。ユーザは、視点を次々と変化させることで情報間の関連を把握し、より適切な検索や閲覧を行うための指針を得ることができる。ユーザの検索目的を推測するという点で (C) 検索式の改良技術と共通している。

(F) 文書タイプによる分類・検索

WWW に流通する情報はその目的も完成度も様々であり、検索結果として出力される文書集合、特にロボットプログラムによってデータを収集している検索サービスからの出力は、分野・新旧・公式文書か否か、などの特徴が入り混じった雑多な文書集合である場合がほとんどである。特に何らかの問題解決を目指すユーザにとってこのような雑多な検索結果は望ましいものではなく、心理的にも不満の残るものとなる。ディレクトリサービスの場合、問題解決に即した適切なカテゴリを発見することができるれば質の揃った結果を得ることができる。しかし、問題解決に即したキーワードを想定することはできてもそのキーワードを一般的な概念体系に則ったカテゴリ構造中

から探し出すことは難しく，またそのようなキーワードがカテゴリとして用意されていない場合やカテゴリ構造を横断している場合も多い．特定の問題解決に役立つサイト，例えば企業の公式サイトや旅行に関するサイトのみを集めたリンク集なども提供されているが，人手による登録方法では結局のところ WWW 空間の拡大・変化に対応しきれず，高い質を維持するためにはコストの面で困難である．

そこで，主に何らかの問題解決を目指した検索作業の支援として，ロボットプログラムが収集した文書を「リンク集」「求人情報」「商品カタログ」などあらかじめ用意した文書タイプ（問題解決に利用できる情報の種類）へ自動的に分類する技術が研究され [27]，携帯端末向けの WWW 文書検索システム Attayo¹⁴⁾ などに応用されている．またメタ検索サービス Inquirus 2 [17] では，複数の検索エンジンから返された結果を統合し，文書タイプごとに分類して表示するシステムとして実用に向けての研究が進められている．

検索の精度の評価について

通常，検索システムの結果を評価するための基準として再現率と適合率がよく用いられる．再現率とは，データベースに存在する全ての正解に対して，結果として取り出すことの出来た正解の割合である．再現率の低い結果とは，取り漏らした正解が多いことを意味する．従って，正解をもれなく抽出する必要がある場合，例えば関連特許の検索などの場合には，再現率が高いことが重要である．一方適合率とは，結果に含まれる正解の割合である．適合率の低い結果とは，検索目的に対して不適切な要素（ハズレ）が多いことを意味する．従って，ハズレによって新しい検索目的や興味の喚起を望むような特殊な場合を除いて，適合率はどのような検索の場合にも高いことが望ましい．

理想的には，精度の良い検索とは，再現率と適合率が共に高い結果を得られる検索である．しかし現実には再現率と適合率の値はトレードオフの関係になり，どちらの値を重要視するかは検索目的によって左右される．一方，検索の精度を評価するための単一の尺度として再現率と適合率を複合した指標についてもいくつか考案されており，現在は双方の重み付き調和平均で定義される E 尺度 [28] がよく用いられている． E 尺度は以下の式で定義される．

$$E = 1 - \frac{(\beta^2 + 1)PR}{\beta^2P + R} \quad \beta \text{ は非負定数} \quad (1.1)$$

β は適合率と再現率のどちらを重視するかを設定するための非負定数で，両者を同等に見る場合は $\beta = 1$ である． E 尺度の値が小さいほど，ユーザの満足度が高い「望ましい」結果を示す．検索システムの評価には，通常 β を 0.5（適合率重視），1.0（公平），2.0（再現率重視）の 3 通りに設定してそれぞれの場合の検索性能を確認する機会が多い．

1.4 本研究の目的

1.2 節で紹介したとおり，現在，小規模なものも含めると国内だけでも 100 を超える WWW 検索サービスが公開されている．しかし，その使いやすさに関して未だ問題点も多く，どの検索サービスを利用しても満足できる結果を得られないことしばしばである．解答者の過半数が「WWW 利用の大半を検索作業に費やすが，必要な情報を的確に得られるとは感じられない」と答えたという WWW 利用者調査の結果⁶⁾もある．流通する情報の爆発的増加，そして利用者層・利用目的の変化などもあいまって，検索作業におけるユーザの負担は増大しているのが実情である．

本研究の目的は，既存の検索手法が抱える問題の解決を目指し，ユーザがより自然かつ容易な方法で必要な情報を得られるような検索支援システムを提案することである．ここでユーザとは，WWW や WWW 検索サービスを利用するが，その詳しい仕組みやデータベースシステム・論理演算などにはあまり精通していないような，現在の一般的な WWW 利用者を考えている．特に従来の検索支援に関する研究が主に検索の効率や精度（再現率・適合率）改善を目的としているのに対し，本研究では情報検索作業に対するユーザのメンタルモデルに着目し，メンタルモデルに沿った，ユーザにとって無理のない検索支援手法の考案を目的としている．

以下，第 2 章では，本研究で提案する 2 種類の検索支援手法それぞれについて，既存の検索手法の問題点に対するアプローチと手法の概略を述べ，また 1.3 節で紹介した関連研究との比較を行う．

第 3 章では可変なカテゴリ構造を用いた支援手法（CATE 手法）について，第 4 章では調整可能な文書ランキングを用いた支援手法（RADR 手法）について，それぞれ手法の詳細と評価実験の結果を述べ，その有効性について論じる．

第 3 章で述べる可変なカテゴリ構造を用いた支援手法（CATE 手法）では，検索のたびに，ユーザの検索目的を反映した小規模なディレクトリを対話的に構築し提供することで検索作業を支援する．この支援手法は，キーワード検索手法では適切なキーワードを考えることが難しく，またディレクトリ探索手法ではユーザの考える情報の分類方法が検索目的に応じて様々に変化することに対応できない，という問題点の解決を図り，ユーザにとってより自然な検索手法を提案することを目的としている．

一方第 4 章で述べる調整可能な文書ランキングを用いた支援手法（UADR 手法）では，キーワード検索手法の出力である文書のランキングをユーザが直観的な方法で調整できる手段を提供することで，検索作業中に生じる心理的な負担の軽減を図る．この支援手法は，システムが提供するランキングの基準が検索目的に対して不適切である場合に，その不満を直接システムに指摘できる手段がない，という問題の解決を図り，より適切なランキングを対話的に構築してゆくことを目的としている．

最後に第 5 章で，本研究全体の結果および今後の課題や展望について総括する．

第2章 検索支援手法のアプローチ

本研究では，ユーザにとって自然で無理のない検索手法を提供するシステムの構築を目標に，既存の WWW 検索手法が抱える問題を二つの視点から考察し，それぞれに対してその問題の解決を図る新しい WWW 検索支援手法を提案する．2つの支援手法は互いに独立しており，どちらも既存の検索手法に対する改良として，単独または併用で利用できる技術である．

本章では各支援手法ごとにその方針を述べる．具体的にはまず既存の検索手法が抱える問題点を挙げ，その解決を目指して提案する検索支援手法について目的と概略を述べる．また，1.3節で述べた従来の文書検索支援に関する諸研究との比較を行う．各手法の詳細や試作システム，評価実験とその結果については，第3章と第4章でそれぞれ詳しく述べる．

2.1 アプローチ1：可変なカテゴリ構造を用いた検索支援手法（CATE手法）

1.4節で述べたとおり，本研究の目的は，検索作業にあまり精通しておらず，またあまり専門的な調査を目的としない一般的な WWW 利用者を対象に，より自然かつ容易な方法で必要な情報を得られるような検索支援システムを提案することである．上記を目標に，本研究における一つめの提案手法として，検索のたびに，ユーザの検索目的に合わせた小規模なカテゴリ構造を構築するという検索支援手法を提案する [29–31]．

以降，本手法（可変なカテゴリ構造を用いた支援手法）を，後に第3章で述べる試作システムの名を冠して CATE 手法と呼ぶ（CATE は categorize の略）．

2.1.1 既存の検索手法における問題点

1.2節で述べた通り，現在の WWW 検索サービスは，キーワード検索手法・ディレクトリ探索手法の2種類の検索手法のどちらか一方，または両検索手法を併用で提供する形式になっている（検索手法の統合については，WWW 検索サービスの最近の話題として付録 A で簡単に紹介する）．ここではキーワード検索手法・ディレクトリ探索手法それぞれの使いやすさに関して考察し，問題点を挙げる．

キーワード 検索手法

1.2 節で紹介した通り，キーワード 検索手法によるシステムは，ユーザの入力したキーワードに合致する文書をデータベースから取り出し，その各文書をキーワードとの適合度（スコア）の降順に並べたリストを出力する．キーワード 検索手法を主とするサービスのほとんどはロボットプログラムを用いて巨大なデータベースを作成し，また更新も頻繁に行っているため，現在の WWW 空間に対して比較的再現率の高い検索を行うことが出来るという利点がある．また，出力される文書は基本的に「入力したキーワードを含む文書」であり，ユーザにとっては作業の意味が理解しやすいことも利点である．

一方，問題点は，

- 目的の情報だけを的確に抽出できるようなキーワードを考えるのは一般に難しいということである．簡単に思い付くような単語をキーワードとして用いると，大量のハズレ（不適合文書）を含む膨大な数の文書が抽出されてしまい，その中から正解を探すために多くの時間と労力を費やさねばならない場合が多い．より精度の良い検索を行うためには，正解のみを抽出できるようなより適切な単語を考えたり，またそれらの単語を AND,OR,NOT などを組合せた検索式として精錬する必要があり，ユーザにとって慎重さを要する作業になる．更にデータベースに対する詳細な知識や高度な検索式を書く技術なども必要となるため，一般のユーザにとっては困難な作業となる．

ディレクトリ探索型の検索手法

ディレクトリ探索手法では，ユーザは提供されているカテゴリ階層構造の中から目的に合うものを選択するだけでよい．データベースに関する詳しい知識も不要であり，キーワード 検索手法に比べユーザの負担が少ない検索手法である．また，人手によって分類・登録された文書が提供されるため，目的にちょうど合ったカテゴリを見つけることができれば，正解のみが集まった（ハズレの少ない）結果を得ることができる．

しかし，ユーザに提供されるカテゴリ構造は，固定的な分類体系に基づいてデータベースのすべてを表現するものであるため，

- 巨大すぎて全体の把握が難しく，カテゴリ選択の際に迷いや誤解を生じやすい．
- 情報の分類方法（分類体系）は様々あり，一つのカテゴリ構造が必ずしも常にユーザの検索目的に適するとは限らない．

という問題がある．例として，環境問題を扱う公的研究機関に関してある地域（奈良）の情報を検索する場合，ユーザが例えば

root → 奈良 → 環境と自然 → 研究機関

（経路 ♣）

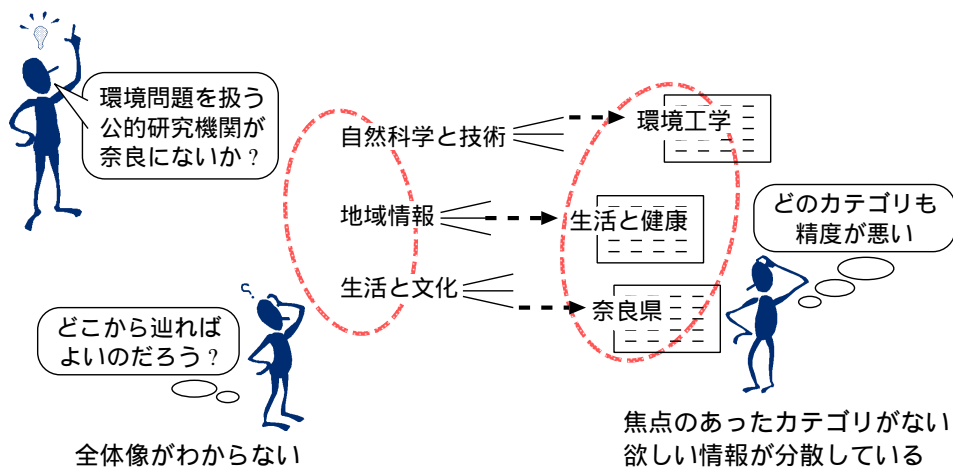


図 2.1 ディレクトリサービスの問題点

のような経路で目的の情報を見つけることを期待しても、提供されるカテゴリ構造では例えば

root → 自然科学と技術 → 研究機関 → 環境工学

root → 地域情報 → 奈良 → 生活と健康

root → 生活と文化 → 地球環境 → 奈良

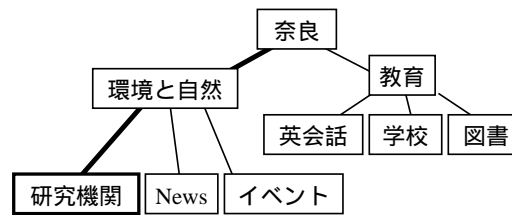
など、複数の経路上に目的の情報分散している場合があり、ユーザはどのカテゴリを選択すればよいのか迷ったり、誤解を生じたりしやすい(図 2.1 中の)。またたとえ全ての経路を辿ったとしても、検索目的に対して焦点があったカテゴリがないので、個々のカテゴリからは精度の悪い結果しか得られない(図 2.1 中の)。この問題は、ユーザの期待する分類方法は検索の目的に応じて様々であり、固定的なカテゴリ構造ではすべての要求に対処できないことに起因している。

ディレクトリサービスの中には、ユーザがカテゴリ構造を自分用にカスタマイズできる機能を提供し問題の解決を図っているものもある。しかし、特に WWW のように様々な種類の情報が混在するデータベースに対しては、1 人のユーザでも様々な検索目的を持つのが通常であり、それぞれの目的に対して適切なカテゴリ構造は本来異なるものと考えられる。つまり、適切なカテゴリ構造は同じユーザに対しても個々の検索毎に構築する必要があり、問題の解消には至っていない。

2.1.2 CATE 手法の目的と概略

CATE 手法の目的は、検索のたびに、ユーザの検索目的に合わせた小規模なカテゴリ構造を構築することにより、検索作業を支援することである。これにより、キーワード検索手法では精度の良い結果を得るための適切なキーワードを考えることが難しく、また固定的な一つのディレクトリを探索する手法ではユーザの考える様々な検索目的

(a) 環境問題を扱う研究機関が奈良にあるか、知りたい場合



(b) 環境問題に関する最近の話題について、知りたい場合

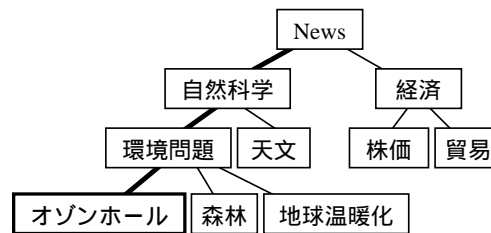


図 2.2 検索目的に合わせたカテゴリ階層構造の例

に対応できない、という前節で述べた問題の克服を図っている。同時に、まずおおよその検索範囲を決め、次に分類方式を参照しながら目的に合った情報を探す、という、例えば図書館や書店での資料検索とも共通する検索戦略を用いることで、一般的なユーザにとってより自然で馴染みやすい検索システムになることを目指している。

小規模なカテゴリ構造の提供とは、ユーザの興味に近い文書だけを対象にした部分的なディレクトリを作成することである。これによりカテゴリ構造が巨大すぎて使いにくいという問題を解決できる。また検索目的に合わせたカテゴリ構造の提供とは、ユーザの検索目的に対して焦点のあったカテゴリ（および経路）を含む階層構造を検索のたびに対話的に構築し提供することである。例えば、環境問題を扱う研究機関が奈良にあるかを調べたい場合は前述の（経路 ♣）を含むカテゴリ構造を提供し（図 2.2 (a)）、環境問題に関する最近の話題を調べたい場合は下記の（経路 ◇）を含むカテゴリ構造を提供する（図 2.2 (b)）。

(a) root → 奈良 → 環境と自然 → 研究機関 (経路 ♣)

(b) root → News → 自然科学 → 環境問題 → オゾンホール (経路 ◇)

このように、検索要求のたびにそれに応じて適切なカテゴリ構造を提供することができれば、前節で述べたディレクトリ探索型の問題を解決できる。さらに、ユーザの望む分類順序に沿ったカテゴリ構造であれば、ユーザが検索結果として選択するカテゴリだけでなくその周辺のカテゴリもユーザのその時の興味に近いものになっていることが期待できる。例えば図 2.2 (a) の構造では「研究機関」の周辺に「News」や「イベント」、(b) の構造では「オゾンホール」の周辺に「森林」や「地球温暖化」というように、目的のカテゴリ周辺にユーザにとって興味に近いものが配置される。

CATE 手法の概要

CATE 手法では、文書の分類に際して分類観点という先験的知識を使用する。分類観点とは、一揃いのカテゴリ名の集合である。例えば「地域」という分類観点は「東京」「大阪」「奈良」などのカテゴリから成り立っており、「趣味」という分類観点は「アウトドア」「写真」「釣り」「読書」などのカテゴリから成り立っている。このような分類観点を複数用意し、それらの組み合わせによって適切なカテゴリ構造を構築することを考えている。システムはまず、ユーザの入力したキーワードを用いて初期文書集合を抽出し、それを全ての分類観点をを用いて分類する。次に各分類観点についてその分類状況を明確さまたはエントロピーという基準で評価し、評価値の高いものをユーザに提示する。ユーザはシステムから提示される評価値を参考に、適切な分類観点を選択する。1回の分類は1階層のカテゴリ構造の構築に相当するため、この時点でユーザは、選択した分類観点の各カテゴリをトップカテゴリとするような1階層のディレクトリを得ることになる。ユーザは、あるカテゴリを上記と同様の手段で更に分類して好みのサブカテゴリ群を構築する(2階層目を構築する)か、既存のディレクトリサービスでの探索と同様に現在のディレクトリの中から適切なカテゴリを選択して閲覧することができる。このような、1) 文書集合の分類 2) 分類観点の選択、という手順をシステムとユーザが再帰的に繰り返すことで、目的の情報のみが集まった適度な大きさのカテゴリ構造を提供できると期待される。さらに、ユーザが検索結果として選んだカテゴリの周辺に配置されるのは同じ分類観点に属するカテゴリであるため、ユーザが自分の望む分類方法に沿ってカテゴリ構造を適切に構築してゆけば、自動的に周辺カテゴリもユーザの興味に近いものになっていると期待できる。

また CATE 手法では、キーワード入力は分類のための前準備として簡単な単語を列挙するだけでよいとため、従来のキーワード検索手法のように慎重に検索式を考える必要がない。ユーザにとっては、簡単な(適合度が低くてもよい)キーワード入力と評価値を参考にした分類観点の選択、および小規模なディレクトリ内の探索だけで良いため、キーワード検索・ディレクトリ探索の双方と比べ、作業の種類は増加するが使いやすさの点では遜色ない手法と考えられる。

2.1.3 関連研究 — CATE 手法との比較

CATE 手法は、精度の良くない1次的な検索結果を分類することによって検索支援を行うものであるため、1.3 節で述べた関連研究の中では(B)初期文書集合の分類による絞り込み に属する一手法になる。更に初期文書集合の分類による絞り込みを行う研究には、統計的分類手法を用いるものと先験的知識を用いる手法があり、CATE 手法は分類観点という先験的知識を用いるため、後者にあたる。

CATE 手法と同じく、先験的知識を用いる分類の研究として 1.3 節で挙げた検索シ

システム HIBROWSE は、先験的知識として *view* と呼ばれる部分的なシソーラスを用いている。しかし HIBROWSE は医学文献の検索に用いられるため、*view* の選択は医学の専門家であるユーザに任されており、検索目的に対して適切な *view* を提示する方法については考えられていない。CATE 手法では分類の状況を明確さやエントロピーという基準で評価し、ユーザが分類観点を選択する際の負担軽減を図るという点で HIBROWSE と異なる。また、新しく収集した文書を既に構築されたカテゴリ構造のどのカテゴリに分類・登録すべきかを推定する手法は、ユーザの検索目的に合った小規模なカテゴリ構造を構築しようとする本手法とは本質的に異なるアプローチである。ただし、この技術を、CATE 手法で個々の分類観点を用いた分類を行う段階に応用することは可能である。

統計的分類手法による分類の研究として 1.3 節で挙げた Scatter/Gather は、クラスタリングによって階層構造を構築するシステムである。Scatter/Gather はデータベースの全文書から 1 個の階層構造を構築することを目的としているが、最初に検索式によって文書集合を限定し、それに合わせた階層構造を構築するように改変すると、CATE 手法と似たものになる。このような統計的手法は、本研究で使用する分類観点のような先験的知識が不要という利点がある。しかし、統計的手法による分類には、分類で得られた各部分集合がどのような内容が理解しにくいという問題がある。ユーザに適切な部分集合を選択してもらうためにはその特徴を簡潔に表示する必要があるが、そのようなラベル(部分集合の名前や要約)を自動的に作成することは極めて困難である。

関連研究 (A) 2 次キーワードの提示による検索支援は、初期文書集合の絞り込みという点において、CATE 手法のような分類による絞り込み手法(関連研究 (B)) と目的を同じくするものである。しかし、(A) による絞り込みは、初期文書集合の中から 2 次キーワードに関連する部分集合のみを取り出すことを目的としているため、ユーザが 2 次キーワードの候補の中から一つを選択し絞り込みを行った場合、その 2 次キーワードを含まない文書(絞り込みによって削減される文書)についての情報は失われてしまうことになる。NOT 演算子を用いることでそのキーワードに対する補集合(キーワードを含まない文書の集合)を得ることができるシステムもあるが、ユーザが 2 次キーワードの選択に際して「このキーワードを含む場合と含まない場合で結果はどう変わるのか?」「キーワードを含まない集合の中にも適合文書はあるのではないか?」などの疑問を抱いた場合に直観的に比較できる仕組みにはなっていない。2 次キーワードによる絞り込みは、非常に適合率の悪い大量の結果を扱いやすい分量になるまで削減するという目的に適しており、分類による絞り込みは、それほど大量ではない結果を様々な角度から更に検討するという目的に適していると考えられる。

また、関連研究 (D) スコア計算方法の改良、(E) 情報視覚化、(F) 文書タイプによる分類などは、いずれも CATE 手法とアプローチは異なるが、CATE 手法との統合が可能な技術である。

2.2 アプローチ 2：調整可能な文書ランキングを用いた 検索支援手法（UADR 手法）

1.2 節の最後に述べたとおり，キーワード検索手法はすでに文書検索における標準的な検索手法として定着し，ディレクトリ探索を中心とする検索サービスにおいても標準的に提供されるようになった．そこで本研究における二つめの提案手法では，キーワード検索システムの改善として，システムが提示する文書ランキングを，ユーザが自分の検索目的に沿うように調整できる検索システムを提供することでユーザの心理的負担を軽減することを目指す [32]．

以降，本手法（調整可能な文書ランキングを用いた支援手法）を，後に第 4 章で述べる試作システムの名を冠して UADR 手法と呼ぶ（UADR は User-Adjusted Document Ranking の略）．

2.2.1 既存の検索手法における問題点

いまやキーワード検索手法はユーザにとって簡便な検索方法として広く受け入れられているが，その使いやすさに関しては以下のような問題がある．

(1) 適切なキーワードを考える難しさ

2.1.1 節でも述べた通り，必要な情報のみを的確に取り出すようなキーワードを考えることは一般に難しい．精度の良い（適合率・再現率の高い）文書集合を結果として得るためには，適切なキーワードや検索式を考案するか，または精度の悪い一時的な結果を追加キーワードなどによってさらに絞り込んでゆく必要があり，ユーザにとって慎重さを要する作業となる．専門的な知識や技術が必要になる場合もあり，検索作業に不慣れなユーザにとって難しい作業である．キーワード検索サービスでは平均 2 個弱のキーワードしか使われないという調査報告 [33] もあり，一般のユーザにとってキーワードを増やすことは困難であることがわかる．

(2) 一方的な結果表示によるストレス

多くのキーワード検索サービスは，検索要求のたびに，データベース中の各文書に対して問合せ（ユーザが入力したキーワードや検索式）との適合度を計算し，適合度がしきい値よりも大きい文書を適合度の降順に並べてユーザに提示する仕組みになっている．本論文では以降，文書と問合せとの適合度をスコア，文書をスコアの降順に並べたリストのことをランキングと呼ぶ．

システムが抽出する文書全てが必ずしもユーザにとっての適合文書ではないので，ユーザは，提示されたランキングを上位から下位へと順にチェックして，検索目的に

適合した文書を探すことになる。この際、ランキングは非常に長い場合がほとんどであるため、ユーザが実際に見るのは上位からある程度までの文書だけになる。従って、たとえユーザにとって適合度の高い文書を多く抽出することができても、それらがランキングの上位に位置していなければユーザにとっては精度の悪い結果になる。つまり、単に再現率・適合率が高いだけでなくユーザにとっての適合文書が上位に集まったランキングを提供することが重要である。しかし入力キーワードから正確な検索目的を把握するのは困難であり、スコアの計算方法をいかに工夫しても、常に適切なランキングを提供することは難しい。

例えば「奈良市、旅行」というキーワードは、奈良市内の旅行社を探す場合や、奈良市へのパッキングツアーを探す場合、奈良市の旅行者向け観光案内を探す場合など、様々な検索目的に対して用いられると考えられる。このようにキーワードとは検索目的を非常に端的に表現したものであるため、たとえスコアの計算方法がいかに優れていても、同じ入力キーワードに対して常に同じランキングを出力するがぎり、異なる検索目的を持つ全てのユーザに対して適切なランキングを提示することは不可能である。

その結果、ユーザにとって適合度が高い文書でもシステムが計算したスコアが低い場合ランキングの下位に配置されてしまうような場合がしばしば生じるのが現状である。このような場合に、ユーザはシステムの決めるランキング基準を信用できず、『他の(もっと良い)適合文書が下位にあるかもしれない』という疑惑を抱くことになる。(さらに、通常キーワード検索サービスにおけるスコアの計算規則はユーザにとって未知であり、ランキング基準に対する不信感を生じやすいという背景もある。)

しかし、現在の検索サービスではこのような場合にユーザがランキングを変更・修正する手段は提供されておらず、ユーザはシステムの提供する不適切なランキングを受け入れ続けなければならない。これにより、適合文書を得るまでに余計な時間や手間を要するのみでなく、ランキングの不適切さをシステムに指示することができないという心理的なストレスも引き起こしていると考えられる。

2.2.2 UADR 手法の目的と概略

UADR 手法の目的は、キーワード検索システムの改良として、ユーザが、直観的かつ簡単な操作で、提示されたランキングを自分の適合度に沿うように調整できる手段を提供することである。これにより前節で述べた問題(2)の解決を図り、不適切なランキングを我慢し続けねばならないというユーザの心理的なストレスを軽減することを目指す。

				検索目的との適合度
1.	score:99	http://www.aaa.com/	✓ yes
2.	score:98	http://www.bbb.com/	no
3.	score:96	http://www.ccc.com/	no
4.	score:90	http://www.ddd.com/	✓ yes
5.	score:89	http://www.eee.com/	no
	⋮			
11.	score:68	http://www.kkk.com/	✓ yes

図 2.3 ユーザによるランキングの調整

UADR 手法の概要

UADR 手法では、ユーザはまず従来通りのランキング出力を得、その中で不当にランクが低いと感じる文書を上位に押し上げたり、逆に自分にとっては不適合文書（ハズレ）なのにランクが高い文書を下位に押し下げたりすることで、システムに対して「現在のランキングはこのように不適切である」という指摘を行うことができる（図 2.3）。システムは、ユーザの行った調整操作（任意文書の順位上げ / 下げ）に応じてユーザの検索目的、すなわちユーザにとっての各文書の適合度を推測し、それに沿った新しいランキングを作成し提示する。具体的には、ユーザが指示した文書についてのみ順位を変更するのではなく、ユーザの指示した文書を適合 / 不適合文書のサンプルと考えて現在のスコア計算規則を修正し、ランキング中の全文書についてスコアを再計算する。これにより、ユーザにとっての適合文書が上位に集まったより適切なランキングを作成し提供することを目指している。本手法では、現在のスコアの計算規則をより適切なものへと変更するための 3 種類のアルゴリズムを考案した。

UADR 手法で提供する任意文書の順位上げ / 下げという指定方法は、“適合文書が下位、不適合文書が上位になっているのはおかしい”という、ユーザが現実的に直面している状況をそのまま表現する事に近い。従ってユーザにとっては、初期キーワードを補足しシステムに対してより正確な検索目的を伝える手段として、キーワードを追加・変更するよりも直観的で無理のない方法と考えられる。

2.2.3 関連研究 — UADR 手法との比較

ここではユーザの感じる心理的ストレスを軽減し使いやすさの向上をはかるという観点から，1.3 節で紹介した従来の文書検索支援手法と UADR 手法を比較する．

関連研究 (A) 2 次キーワードの提示，(B) 初期文書集合の分類による絞り込み，(F) 文書タイプによる分類，で挙げた研究は，基本的に精度の悪い初期検索結果を絞り込んでより精度の良い文書集合を得ることを目的としており，文書集合の提示段階においてユーザの負担を軽減しようとする UADR 手法とは異なるアプローチである．しかし，不適合な文書がほとんどなくなるまで文書集合を絞り込むことは難しい（現実的ではない）ため，絞り込みによる検索支援手法においても適切なランキングを提示するための技術は必要である．この点から UADR 手法は上記検索支援手法との統合が期待される技術と言える．関連研究 (D) スコア計算方法の改良，(E) 情報視覚化 についても同様に，それぞれ UADR 手法との組合せが可能な技術である．

関連研究 (C) 検索式の改良 で挙げた適合フィードバックは，現在の検索式で得た文書集合中の各文書に対し，ユーザが適合・不適合の判定を行い，システムはそれを基に検索式を洗練する手法である．UADR 手法では，ユーザの調整作業に応じてスコアを再計算するが，文書集合の変更は行わない．一方適合フィードバックでは，検索式が変更されるたびに文書集合とスコアの両方が変更されるが，文書集合はそのままスコアだけを変更するように実装することも可能である．この場合，調整作業のインタフェースは異なるが，ユーザの検索目的に適応した新しいスコア付けを目指すという点で UADR 手法と適合フィードバックは似たアプローチになる．

AIQEC ではユーザが判定作業をクラスタ単位で行えるようにすることで適合・不適合判定作業の負担軽減を目指しているが，各クラスタの内容を簡潔に表現するラベルを付けることは難しいため，判定作業の難しさが増すことも考えられる．UADR 手法では，ユーザは文書の順位上げ/下げという単純な操作を行うだけでよい．更に，現在の WWW 検索サービスにおいて最終的な出力は常にランキング形式であるため，検索サービスのユーザは文書空間に対して，ベクトル空間モデルのような空間上の分布ではなく，スコア順に並んだリストに近いメンタルモデルを持っていると考えられる．ユーザにとって精度の悪い結果とは，“正しい順序に並んでいないだけでどこか（下位）には適合文書がまぎれている” ような物であり，UADR 手法の提供する『このような文書は上位に移動せよ』という指定方法は上記のメンタルモデルに沿った無理のない表現形式と考えられる．

第3章 可変なカテゴリ構造を用いた検索支援手法(CATE手法)

3.1 はじめに

本章で提案する検索支援手法は，検索のたびに，ユーザの検索目的に適した小規模なカテゴリ構造を構築し提供するものである．本手法ではあらかじめ分類のための先験的知識（分類観点）を用意し，ユーザが簡単なキーワード入力と分類観点の選択を行うことで，1階層ずつ対話的にカテゴリ階層構造を構築する．これにより，キーワード検索手法では精度の良い結果を得るための適切なキーワードを考えることが難しく，また固定的な1つのディレクトリを探索する手法ではユーザの考える様々な検索目的に対応できない，という問題の克服を図っている．同時に，まずおおよその検索範囲を決め，次に分類方式を参照しながら目的に合った情報を探すという，例えば図書館や書店での資料検索とも共通する検索戦略を用いることで，一般的なユーザにとって自然で馴染みやすい検索支援システムとなることを目指している．

本手法を実装した試作システムと BMIR-J2 テストコレクションを用いて，従来のキーワード検索手法・クラスタリング手法 (*K*-means) との比較実験を行い，(1) 検索結果の質 (2) 検索の効率 (3) 使いやすさ の3点から本手法の有効性を評価した．

3.2 CATE 手法

本手法は以下の2段階の処理から成り立っている．

[I] ユーザの検索目的を大まかに反映する文書の集合を決める．

[II] [I] で求めた文書集合に依存してカテゴリ構造を作る．

[II] では，カテゴリ構造を構築するために，設計者がシステムに予め与える分類観点という知識を用いる．分類観点とは，一貫性のある一揃いのカテゴリの集合である．例えば「地域」という分類観点は「京都」「大阪」「奈良」などのカテゴリから成り立っており，「News」という分類観点は「経済」「自然科学」「スポーツ」などのカテゴリから成り立っている．このような分類観点を複数用意し，それらの組み合わせによって適切なカテゴリ構造を構築することを考える．

システム設計者が与える分類観点の集合を

$$S = \{S_j \mid 1 \leq j \leq m\}$$

とする。ただし、分類観点 $S_j = (l_j, W_j)$ は、分類観点名 l_j とカテゴリ名の集合 $W_j = \{w_{j1}, \dots, w_{jk_j}\}$ の2字組である。ここで k_j は S_j に属するカテゴリの数である。各 W_j は、3.2.1 節で述べる W (システムが扱う全単語) の部分集合と仮定する。カテゴリは、そのカテゴリ名である単語 w_{ji} で参照する。

各分類観点は、ディレクトリサービスでいうと、カテゴリ階層構造に現れる兄弟ノードの集合 (1 ノードに対する子の集合) に相当する (図 2.2)。本手法では、システムが適切な分類観点をいくつか提示し、ユーザが分類観点 (カテゴリ) を選択することで、システム側とユーザ側の双方が協力して適切なカテゴリ構造を対話的に構築してゆくことを目指している。

3.2.1 ベクトル空間モデルに基づく文書の表現

文書とカテゴリの関連を定義するために、ベクトル空間モデル [34] を用いる。このモデルでは、各文書の特徴は、各単語との関連度を並べたベクトルで表現されるとする。つまり、システムが扱う単語の集合を

$$W = \{w_1, w_2, \dots, w_n\}$$

とすると、文書 d の特徴ベクトルは

$$fv(d) = (c_{w_1,d}, c_{w_2,d}, \dots, c_{w_n,d})$$

の形で表される。ただし $c_{w_i,d}$ は、単語 w_i と文書 d との関連度であり、文書 d 中の単語 w_i の出現頻度と、 w_i が出現する文書の少なさを表す値 (inverse document frequency [3]) との積を、 d の長さで正規化したものと定義する。簡単のため、本章では $d = fv(d)$ とし、以降 d の特徴ベクトルを単に d と表記する。

特徴ベクトル u, v の類似度 $\text{sim}(u, v)$ を、

$$\text{sim}(u, v) = \frac{u \cdot v}{|u| |v|}$$

と定義する。ここで $|u|$ はベクトル u の長さ、 $u \cdot v$ は u と v の内積である。

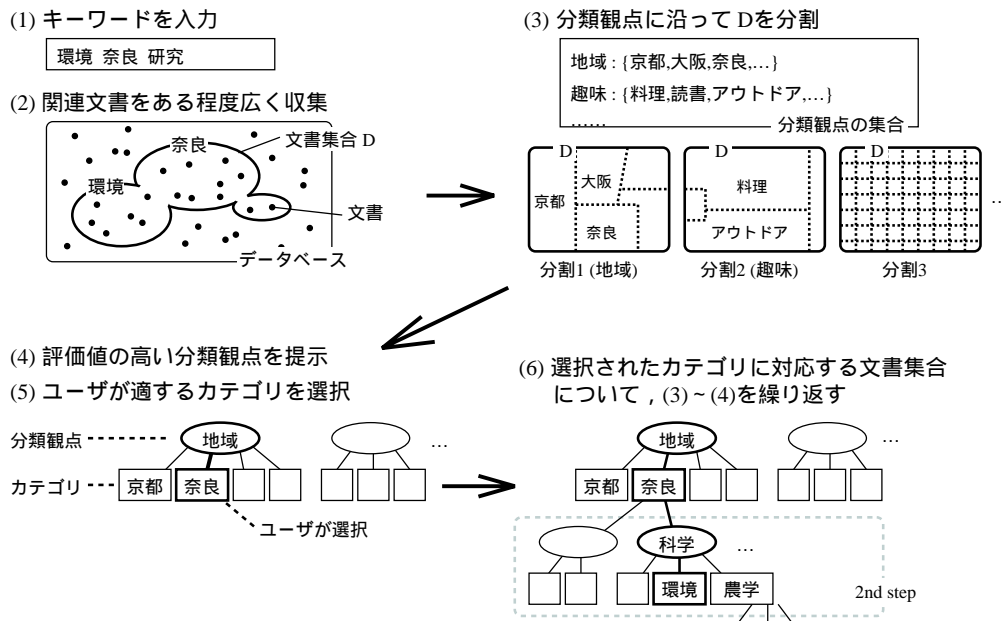


図 3.1 CATE システム：動作の概要

3.2.2 システムの動作の概要

システムの動作手順は以下のとおり（図 3.1）。

- (1) ユーザーがいくつかのキーワードを入力する．ここでは検索対象とする文書を大まかに限定することだけが目的なので，キーワードの選択はあまり慎重に考えなくても良い．
- (2) 文書データベースに対して入力キーワードの OR 検索を行い，得られた文書の集合を D とする．
- (3) システムに用意した各分類観点 S_j に沿って， D を分割する．すなわち，各文書 $d \in D$ について， S_j のカテゴリ集合 $W_j = \{w_{j1}, \dots, w_{jk_j}\}$ の中から d が属すべきカテゴリ $w_j^{(d)}$ を一つ選ぶ． $w_j^{(d)}$ は， W_j の要素のうち d との類似度が最大のものとする．すなわち

$$\text{sim}(w_j^{(d)}, d) = \max_{1 \leq h \leq k_j} \text{sim}(w_{jh}, d) .^1$$

S_j による文書集合 D の分割 $(D_{j1}, \dots, D_{jk_j})$ を以下で定義する．

$$D_{ji} = \{d \in D \mid w_j^{(d)} = w_{ji}\}$$

このように求めた D_{ji} を，カテゴリ w_{ji} に関連する文書の集合と呼ぶ．

¹簡単のため， $\text{sim}(w_{jh}, d)$ を最大にする h が複数ある場合は，任意の 1 個を選んで $d \in D_{jh}$ とする．

w_{ji} はカテゴリ w_{ji} の特徴ベクトルである．これを w_{ji} のカテゴリベクトルと呼ぶ． w_{ji} の初期値を，カテゴリ名である単語 w_{ji} に対応する成分が 1 で他の成分がすべて 0 の特徴ベクトルと定義する ($w_{ji} = (\delta_1, \delta_2, \dots, \delta_n)$ ，ただし $w_{ji} = w_k \in W$ の時 $\delta_k = 1$ ，それ以外の時 $\delta_k = 0$)．この初期カテゴリベクトルを使った分割は，各文書の属するカテゴリをカテゴリ名との関連度のみに基づいて決めることに相当する (d と初期カテゴリベクトル w_{ji} の類似度は，文書 d と カテゴリ名である単語 w_{ji} の関連度 $c_{w_{ji},d}$ を正規化した値 $\text{sim}(w_{ji}, d) = c_{w_{ji},d}/|d|$ になるため)．しかしこの初期カテゴリベクトルによる分類では，1 つの分類観点の中でどのカテゴリにも属さない文書 (すべての i について $\text{sim}(w_{ji}, d) = 0$ となる文書) が多く生じてしまう．そこで，初期カテゴリベクトルに関連する文書の特徴ベクトルを平均し，これをカテゴリベクトルとして定義し直す．すなわち

$$w_{ji} = \sum_{d \in D_{ji}} d / |D_{ji}|$$

ただし $|D_{ji}|$ は文書集合 D_{ji} の要素数

とし，分類を平滑化する．

すべての分割を行った後 (すなわち，用意されたすべての分類観点 S_j ($1 \leq j \leq m$) に対して $D_j = \{D_{j1}, \dots, D_{jk_j}\}$ を求めた後)，各分類観点について D に対する適切さを評価する (3.2.3 節)．

- (4) 評価値の高い (複数の) 分類観点と，それぞれの要素であるカテゴリ集合をユーザに提示する．
- (5) ユーザが，提示されたカテゴリの中から検索目的に適すると思うものを選ぶ．ユーザによって選択されたカテゴリに関連する文書の集合を D' とする．ユーザは， D' を検索結果として検索を終了するか， D' をさらに分割するために $D = D'$ として (3) へ進む．

3.2.3 分類観点の評価

分類観点の評価値を決めるため，分類の明確さとエントロピーという 2 種類の評価基準を考えた．2 つの評価基準はどちらも，文書集合 D に対して分類観点 S_j がどの程度妥当であるかを， D の分割状況 (D 中の全文書が S_j に属する各カテゴリにどのように振り分けられたか) から推測しようとするものである．文書集合 D に対する分類観点 S_j の評価値 $e_D(S_j)$ を，明確さを用いる場合は式 (3.1)，エントロピーを用いる場合は式 (3.2) でそれぞれ定義する．これら 2 つの評価基準の有効性を，後に述べる評価実験で比較した (3.3 節 性能評価実験, 3.4 節 被験者を用いた実験) ．

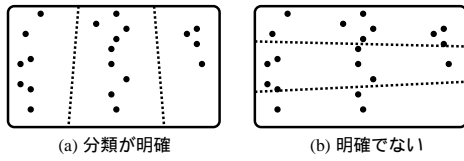


図 3.2 分類の明確さ

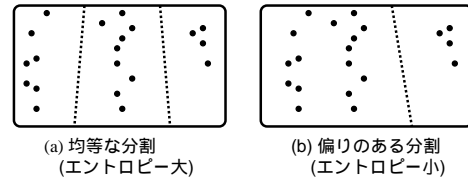


図 3.3 分類のエントロピー

分類の明確さに基づく評価

文書集合 D に対して明確な分類を行う分類観点とは、各文書 $d \in D$ が属すべきカテゴリ $w_j^{(d)}$ が明確である、すなわち、 d とカテゴリ $w_j^{(d)}$ との類似度が d と $w_j^{(d)}$ 以外のカテゴリとの類似度よりも十分高いということである（図 3.2 (a)）。分類が明確な分類観点がユーザの検索意図を反映するとは限らないが、文書集合を明瞭に分割できるもってもらいたい分類という点で、ユーザにとって有益と考えられる。文書集合 D に対する分類観点 S_j の明確さを、 D 中の各文書とその文書が属するカテゴリとの類似度の平均値で定義する：

$$e_D(S_j) = \sum_{d \in D} \text{sim}(w_j^{(d)}, d) / |D|. \quad (3.1)$$

分類のエントロピーに基づく評価

文書集合 D に対する分類観点 S_j の分類のエントロピーを以下の式で定義する。

$$H_D(S_j) = - \sum_{i=1}^{k_j} P_i \log P_i, \quad (3.2)$$

ただし、

k_j : S_j に属するカテゴリの数,

$$P_i = \frac{|D_{ji}|}{|D|},$$

D_{ji} : カテゴリ w_{ji} に関連する文書の集合.

エントロピーの大きい分類観点とは、文書集合をほぼ均等に、より多くのカテゴリに分割するものである。つまり、ユーザにとっては1回の分類で文書数をほぼ $1/(\text{カテゴリ数})$ に減らすことができ、文書数を削減する効率が良い分類と言える（図 3.3 (a)）。

3.3 性能評価実験

本手法と従来の検索手法を比較する 2 種類の評価実験を行った (3.3 節:性能評価実験, 3.4 節:被験者を用いた実験) . 性能評価実験では, ある理想的なふるまいをするユーザを仮定し, その行動パターンを機械的に実行して, システムのベンチマークを獲得し比較する. ここで理想的なユーザとは, 最も評価値の高い分類観点を選択し, その分類観点の中で適合率の高い順にカテゴリを選択するユーザである.

3.3.1 システムの概要

研究グループで試作したシステムは, ユーザによる検索が行われる前に予め (1 度だけ) 実行しておくデータベース構築部と, ユーザの問い合わせに応じる問い合わせ処理部からなる (図 3.4) . 図 3.4 では各部分プログラムとその入出力ファイルの関係を簡単に表している. 本システムは全文キーワード検索システムである Freya [35] を拡張する形で実現した. また, システムが扱う単語の集合 (3.2.1 節の W) を決めるため, 日本語形態素解析システム茶筌 [36] に同梱の茶筌ライブラリを用いて全文書の形態素解析を行っている. 問い合わせ処理部は CGI プログラムとして実装した.

実験に用いたシステムを次のように定義する.

- システム A: 分類観点の評価に分類の明確さを用いたシステム (CATE 手法 1) .
- システム B: 分類観点の評価に分類のエントロピーを用いたシステム (CATE 手法 2) .
- システム C: クラスタリングによる分類を行うシステム (先験的知識を用いないシステム) .
- システム D: 全文キーワード検索システム Freya .

分類観点という先験的知識を用いる本手法との比較のため, システム C には代表的な統計的分類手法 (クラスタリング手法) である K -means 法 [37] を実装した. クラスタリングを用いた文書分類の研究としてよく引用される Scatter/Gather [7] システムにおいてクラスタ数が 10 であったことと, ユーザがクラスタを選択する際の利便性の点から, 今回の実験ではクラスタ数を 10 とし, 100 回反復した時点でクラスタの重心が収束していなければその時点の結果を出力するものとした.

システム D (Freya) の検索方式について簡単に述べる. Freya は, 原則としてすべての 2 字組 (bigram) を索引語として登録する. また検索によって抽出された文書は, 2.1 節で述べた単語と文書の関連度 $c_{w_i,d}$ によってスコア付けされ, スコアの降順に出力される. 複数のキーワード w_{i_1}, \dots, w_{i_n} が入力された場合は, $\sum_{i \in \{i_1, \dots, i_n\}} c_{w_i,d}$ が d のスコアとなる. システム A, B, C, D が使う $c_{w_i,d}$ の値はすべて同じである.

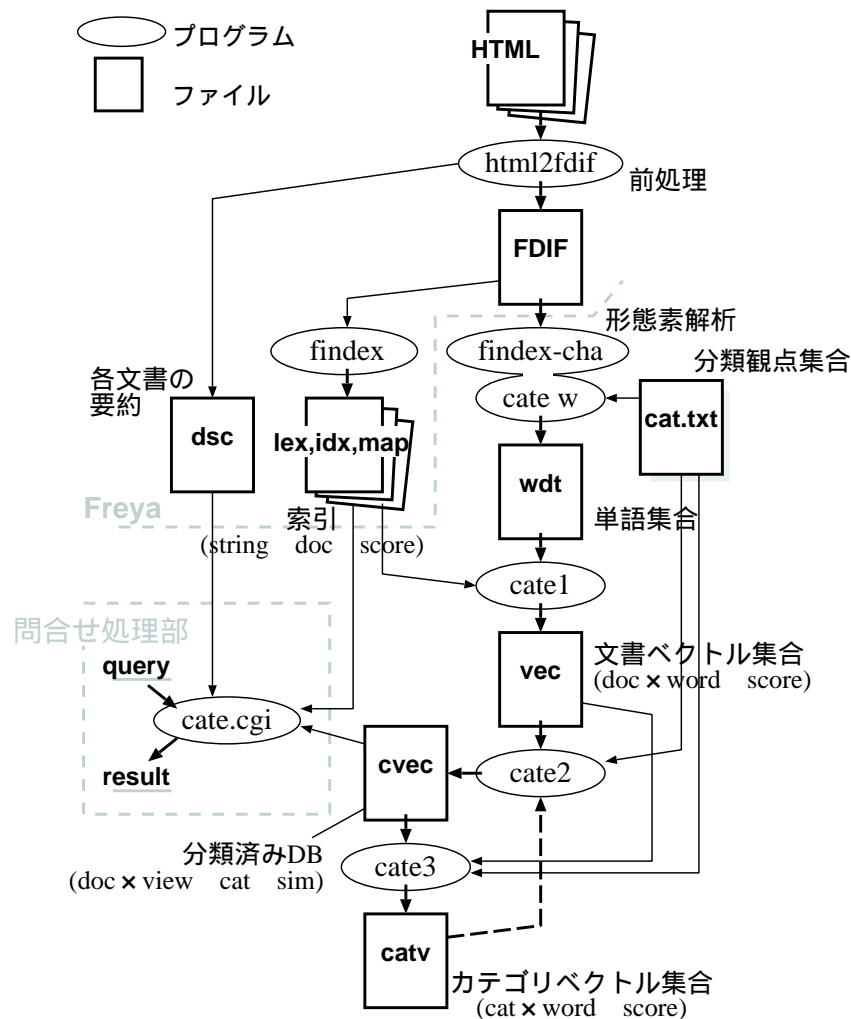


図 3.4 CATE システムの構成

文書集合と検索課題：文書集合にはBMIR-J2テストコレクション [38]を用いた(5,080記事)。BMIR-J2は、(社)情報処理学会・データベースシステム研究会が、新情報処理開発機構との共同作業により、毎日新聞 CD-ROM'94 データ版を基に構築した情報検索システム評価用テストコレクションである。

実験では、BMIR-J2が提供する50件の検索課題のうち、本手法の本質に関わりのないBタイプ(数値・レンジ機能を必要とする)とCタイプ(構文解析を中心とする)の課題を除外し、また平均や精度を議論する際の統計的な信頼性の点から正解件数が15件以下の課題も除外し、残りの中から正解件数があまり偏らないよう配慮して表3.1に挙げた10課題を選んだ²。また各課題に対して、BMIR-J2が提供するA,B両レベルの正解を正解文書とした。

²今回は各課題の正解件数をその課題の難易度の一種(正解件数の多い課題は、正解文書を見つけ易い)と考えたため、正解件数があまり偏らないように配慮した。

表 3.1 CATE : 実験に使用した検索課題

番号	課題名	正解文書の数 ^{*1}
i	農薬	26
ii	飲料品	60
iii	減税	302
iv	核兵器	98
v ^{*2}	教育産業	17
vi	株価動向	37
vii	安売りをを行う流通業者	36
viii	映画	21
ix	東南アジアから日本への輸出	59
x	女性の雇用問題	80

*1: BMIR-J2 が提供する A レベル正解と B レベル正解の和

*2: 3.4 節の実験では“ビデオデッキ（正解文書の数 33）”に変更。

表 3.2 CATE : 性能評価実験に使用したキーワード

課題	キーワード ^{*1}
i	農薬 農業 薬品 農作物 殺虫剤 除草剤
ii	飲料 食品 ジュース コーヒー サントリー キリン
iii	減税 税金 税制改革 政策
iv	核兵器 核爆弾 核 ミサイル 戦争 軍隊 軍備 原爆 放射能
v	教育 学校 生徒
vi	株価 取引
vii	安売り
viii	映画
ix	東南アジア 日本
x	女性 雇用 労働 職業 就職

*1: 課題名と補足説明（BMIR-J2 で提供される，課題に関する簡単な説明文）の中から主要な名詞を抜き出して作成。

ただしシステム D では，各行最初の 2 単語の AND 検索と OR 検索を試行して E 尺度が小さくなる方を採用した。

分類観点集合：既存の大手ディレクトリサービスが提供するカテゴリ階層構造を1階層ごとに分解し，各カテゴリ名とそのサブカテゴリ集合を，分類観点名 l_j とカテゴリ集合 W_j として定義するという手順で第1階層（ディレクトリサービスの最初のページにあたる）から第3階層までを順に降下しながら作成した．その後別のディレクトリサービスに対しても同様の手順を行い，類似したものがない分類観点やカテゴリを追加した（計68分類観点，延べカテゴリ数1,723）．以下に分類観点集合の一部を挙げる（詳細は付録C参照）．

$$\begin{aligned}
 l_1 &= \text{都道府県} : W_1 = \{ \text{北海道, 青森, 岩手, ...} \} \\
 l_2 &= \text{教育} : W_2 = \{ \text{学校, 塾, 大学, ...} \} \\
 l_3 &= \text{メディア} : W_3 = \{ \text{テレビ, ラジオ, 番組, ...} \}
 \end{aligned}$$

クラスタリングでは文書集合を常に10クラスタに分割するのに対し，分類観点は平均して25個のカテゴリに分類するため，一見クラスタリングの方が荒い分類になり精度が悪くなるように思える．しかし，分類した結果所属する文書が0になったカテゴリは無視されるので，実際には分類観点による分類もそう細くならない．3節の実験においては，1分類観点あたり平均11.2カテゴリに分類されており，クラスタ数の10に対して特筆すべき差はないと考えられる．

3.3.2 実験(1)

本手法を用いた場合の検索の精度を調べる．この実験ではA,B,C,Dの4システムについて検索課題ごとに再現率と適合率の関係を求め，同再現率に対する適合率を比較する．システムAとBでは，表3.1の各検索課題 q について，以下の(1)から(4)の手順を各目標再現率 $r = 0.1, 0.2, \dots, 1.0$ に対して行う．

- (1) $Result = \emptyset$ とする．検索課題 q から思いつく適当なキーワード（表3.2）を入力し，初期文書集合 D_0 を求める． $D = D_0$ とする．
- (2) D の分類を行い，各分類観点の評価値（システムAでは明確さ，Bではエントロピーの値）を求める．全分類観点 S_1, S_2, \dots, S_m ($m = 68$)の中で最も評価値の高い分類観点を選択し， S_j とする．分類観点 S_j に属するカテゴリの集合を $W_j = \{w_{ji} | 1 \leq i \leq k_j\}$ ，カテゴリ w_{ji} に属する文書集合を D_{ji} と表記する．
- (3) W_j の中で，まだ選択されていない最も適合率の高いカテゴリ w_{jh} を選択する． $|D_{jh}| \leq 30$ であれば， D_{jh} を $Result$ に追加する． $|D_{jh}| > 30$ の場合は， $D = D_{jh}$ として(2)に戻る．
- (4) (2),(3)を $Result$ の再現率が r 以上になるまで繰り返した後， $Result$ の再現率，適合率を求める．

手順(3)中の再分類を実行するための文書数のしきい値(30)は,“一覧表示した時に2画面で収まる量”を基準に設定している.既存の文書検索サービスのほとんどが文書一覧を2~2.5画面づつ区切って表示していることから,2画面程度がユーザにとって一度に閲覧しやすい文書数の限界と考えた.

測定値 検索課題 q に対して BMIR-J2 が規定している正解文書の集合を $Rel(q)$ とする. $Result$ の再現率, 適合率は以下のように定義される:

$$\text{適合率 } P = \frac{|Result \cap Rel(q)|}{|Result|},$$

$$\text{再現率 } R = \frac{|Result \cap Rel(q)|}{|D_0 \cap Rel(q)|^3}.$$

システム C の場合は, (2) で分類観点の代わりに K -means 法を用いて分類を行い, (3) で適合率の高いクラスタを選ぶ. システム D の場合は, まず各検索課題に対してキーワード(表 3.2)を用いて検索を行い, 得られた文書リストに対して TREC [39] で用いられる計算方法に従って再現率と適合率の値を計算した. すなわち,

- (1) 表 3.2 に挙げた検索式を入力する.
- (2) 検索結果である各文書について, システムが出力した順位に従って1文書ずつ出力文書集合に加える. 各時点での再現率と適合率を求める.

(2) で得られた結果は, 同一再現率に対して一般に複数の適合率を与えるので, 以下のように再現率 r に対する適合率を決める. (2) で得られた再現率と適合率の組の集合を $Result$ とする. 再現率 r に対する適合率 $p(r)$ を次のように定義する.

$$p(r) = \max(\{p \mid (r', p) \in Result \text{ for some } r' \geq r\} \cup \{0\}).$$

また上記方法によって求めた各再現率 R ・適合率 P のペアに対し, 2つを複合した指標である E 尺度 [28] を求める. これは以下の式で定義される (P.11 1.3 節).

$$E = 1 - \frac{(\beta^2 + 1)PR}{\beta^2 P + R}. \quad (3.3)$$

E 尺度の値が小さいほど, ユーザの満足度が高い「望ましい」結果を示す. ここで, β は評価において再現率と適合率のどちらを重視するかを設定するためのパラメータで, 通常 β を 0.5 (適合率重視), 1.0 (公平), 2.0 (再現率重視) の3通りに設定してそれぞれの場合の検索性能を評価する.

³再現率の分母はシステム A,B,C,D で共通の値である.

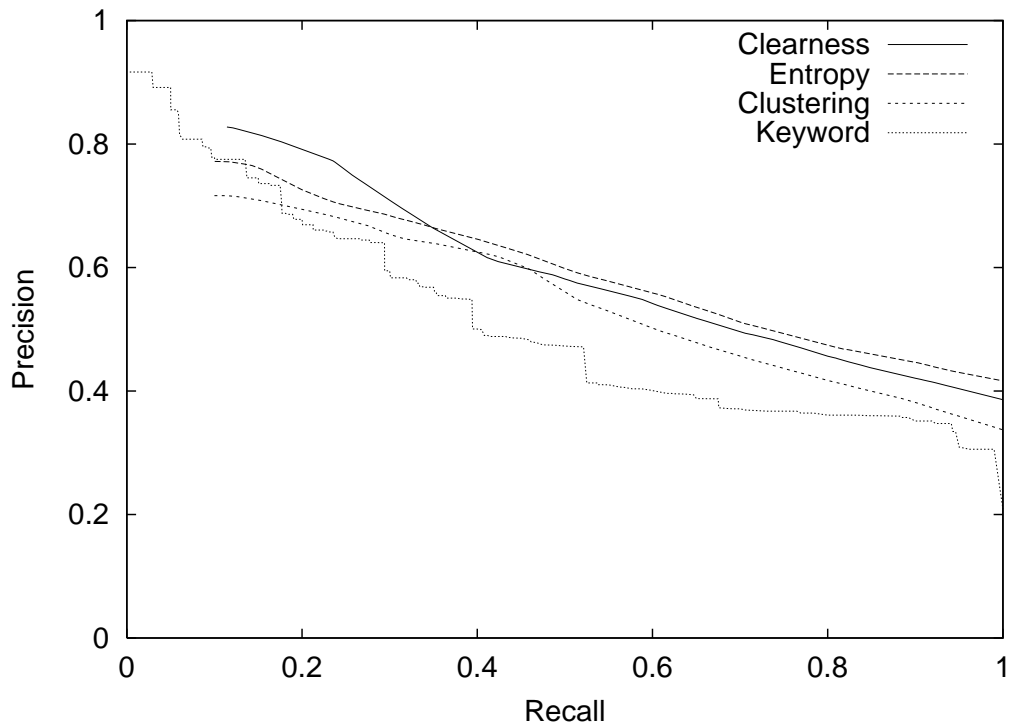


図 3.5 再現率と適合率 (10 課題の平均)

結果

再現率と適合率の関係 (10 課題の平均) を図 3.5 に示す . ただし , 図中の Clearness, Entropy, Clustering, Keyword はそれぞれシステム A, B, C, D に対応する . また E 尺度 ($\beta = 0.5, 1.0, 2.0$) について分析を行い , システムごとの E 尺度平均値に有意差があるかどうかを最小有意差法によって検定した (表 3.3, 表 3.4 . どちらも再現率 0.1, 0.2, ..., 1.0 の平均) .

分散分析の結果 , $\beta = 0.5, 1.0, 2.0$ すべての場合において , E 尺度の値はシステムの違いによって影響を受けていた ($\beta = 0.5$ の場合 5% 有意 , $\beta = 1.0, 2.0$ の場合 1% 有意で , システム間に差がある) . またシステムごとの平均値を見ると , $\beta = 0.5$ の場合は $B < A < D < C$ の順になっており , B と C, D , A と C の間に有意差があった (5% 有意) . このことから , 適合率を重視して比較すると , CATE 手法 (A, B) がクラスタリング手法 (C) に比べて優れていると言える . $\beta = 1.0, 2.0$ の場合はいずれも平均値が $A < B < C < D$ の順になっており , A と D , B と D の間に有意差があった (1% 有意) . このことから , 再現率を重視する場合と公平な場合のどちらにおいても , CATE 手法 (A, B) がキーワード手法 (D) に比べて優れていると言える .

一方システム A, B 間では , $\beta = 0.5, 1.0, 2.0$ すべての場合において , E 尺度の平均値に有意差はなかった . 比較的正確な正解文書数の少ない課題においては明確さ , 正解文書数

表 3.3 E 尺度, $\beta = 2.0$

課題	A	B	C	D
i	0.492	0.494	0.447	0.398
ii	0.449	0.395	0.433	0.661
iii	0.443	0.453	0.454	0.487
iv	0.421	0.418	0.458	0.621
v	0.510	0.491	0.675	0.655
vi	0.562	0.575	0.441	0.518
vii	0.371	0.399	0.396	0.198
viii	0.245	0.414	0.460	0.360
ix	0.519	0.527	0.708	0.878
x	0.463	0.534	0.511	0.656
平均	0.447	0.470	0.498	0.543

表 3.4 E 尺度平均値

	A	B	C	D
$\beta = 0.5$	0.489	0.486	0.531	0.526
$\beta = 1.0$	0.490	0.500	0.540	0.558
$\beta = 2.0$	0.447	0.470	0.498	0.543

の多い課題においてはエントロピーの結果が良い傾向にあるが、正解文書数と E 尺度の間の相関はほとんどなかった（正解文書数・ E 尺度の相関係数 A: 0.03, B: -0.10）。そこでシステム A と B をより詳しく比較するため、次の実験 (2) を行った。

3.3.3 実験 (2)

本手法では、ユーザに対して評価値の高い順に分類観点を提示するので、分類観点の評価値が検索精度を反映していることが望ましい。ここではシステム A と B について、分類観点の評価値と結果 ($Result$) の適合率との間に正の相関関係があるかどうかを調べる。各課題について、以下の (1) から (3) の手順を、各目標再現率 $r = 0.2, 0.5, 0.8$ に対して行う。

- (1) $Result_j = \emptyset$ ($1 \leq j \leq m$) とする。検索課題 q から思いつく適当なキーワード（実験 (1) と同じ）を入力し、初期文書集合 D を求める。
- (2) D に対する分類を行い、各分類観点の評価値（明確さ、エントロピー）を求める。各分類観点 S_j ($1 \leq j \leq m$) について、 W_j の中から適合率の高い順にカテゴリを選択し、それに属する文書を $Result_j$ に加える。この作業を、 $Result_j$ の再現率が r 以上になるまで繰り返す。
- (3) 各分類観点 S_j ($1 \leq j \leq m$) ごとに評価値と $Result_j$ の適合率を求め、全分類観点における評価値・適合率間の相関係数を求める。

表 3.5 評価基準と適合率との相関係数

評価基準	目標再現率	正の相関がある課題数	\bar{R} *1	平均カテゴリ数 *2
A：明確さ	0.2	9	0.496	2.2
	0.5	8	0.595	4.1
	0.8	8	0.646	7.3
B：エントロピー	0.2	10	0.596	2.3
	0.5	10	0.675	4.9
	0.8	10	0.684	8.7

*1: \bar{R} は 評価値・適合率間の相関係数（正の相関がある課題に関する平均値）。

*2: 評価値の高い上位 10 個の分類観点について、目標再現率の達成に要するカテゴリ数の平均（10 課題の平均）。

結果

全課題のデータをまとめて見ると、システム A（明確さ）とシステム B（エントロピー）の両システムとも、目標再現率 0.2, 0.5, 0.8 すべての場合において、分類観点の評価値と適合率との間に正の相関が認められた。10 個の検索課題を個別に見た場合の分類観点の評価値と適合率の相関を表 3.5 に示す。システム A では 10 個のうち 8 ないし 9 個、システム B ではすべての検索課題について正の相関が認められた。

また評価値の高かった上位 10 個の各分類観点について、目標再現率 r を得るために必要なカテゴリ数を調べた。表 3.5 中の「カテゴリ数」は、この値の 10 課題についての平均である。その結果、システム A の方がシステム B よりも平均カテゴリ数がやや少ないことがわかった。平均カテゴリ数に関して、評価基準（分類の明確さ、エントロピー）と検索課題（10 個）を因子に分散分析を行ったところ、 $r = 0.5$ および $r = 0.8$ のそれぞれについて、評価基準による差があることが言えた（5% 有意）。ユーザにとっては、選択しなければならないカテゴリの数は少ないほうが望ましいので、この点では、明確さの方が使いやすさを提供していると考えられる。

キーワードの影響について：3.3.2 節および 3.3.3 節の図表はすべて表 3.2 のキーワードを用いた場合のベンチマークであるが、各システムの性能は入力するキーワードによって左右されるため、更に異なる 2 通りのキーワード（3.4 節の実験に参加した被験者 2 人にそれぞれ指定してもらったもの）を用いて実験 (1) と実験 (2) を実施した。3 種類のキーワードに対する結果を比較すると、具体的な数値（ E 尺度など）には違いがあるが、システム間の性能の差については同じ傾向が見られた。3.3.2 節および 3.3.3 節では、3 種類のキーワードを用いて得られた共通の傾向として結果を述べた。

3.4 被験者を用いた評価実験

実際にユーザがシステムを使用する際の使いやすさを評価するため、被験者を用いた評価実験を行った。実験では性能評価実験と同じ A,B,C,D の 4 システムを用い、各被験者に、検索課題に対して正解だと思う文書をいずれか 1 つのシステムを使って検索してもらった。それぞれの課題をどのシステムを用いて検索するかについては、各 (課題, システム) の組合せが異なる 5 人の被験者によって試行されるように設計した。

3.4.1 システムの概要と実験環境

検索課題 検索課題は、性能評価実験と同じ 10 課題 (表 3.1) を用いた。ただし被験者には「正解と思う文書を 20 個探してください」という形で実験手順を指示するので、正解文書数が 20 未満の課題 (v:「教育産業」) は混乱を招く可能性があると考え、「ビデオデッキ」に変更した。以後、各課題は表 3.1 中の課題番号で参照する。

被験者 情報科学専攻の大学院生 20 人。実験に対する謝礼として千円を支払った。事前アンケートによると、被験者全員が少なくとも 1 年以上 WWW 上のキーワード検索サービスを日常的に使用していた。このことから仮説として、ほとんどの被験者はキーワード検索のユーザインタフェースを情報検索作業のメンタルモデル [40] として強く確立していると考えられる。

実験環境 被験者は DELL Dimension XPS D300 上の Microsoft Internet Explorer version 5.0 を用いて実験作業を行った。オペレーティングシステムは Microsoft Windows98 Second Edition である。実験中の全ての操作を画面録画ソフトウェア Lotus ScreenCam97 を用いて録画し、後の分析に用いた。

手順 実験前：実験の目的と概要を説明する。また各被験者ごとに、実験すべき (課題, システム) の組合せを指示する。たとえば被験者 1 番は (課題 iii, システム C), (vii, C), (ii, B), (x, B), (vi, B), (iv, D), (viii, D), (i, A), (ix, A), (v, A) の順に実験を行う。

実験：被験者に、各 (課題, システム) に対する検索作業をそれぞれ 10 分間行わせる。ただしマークをつけた文書の数が 20 を超えた時点で、その課題に対する検索作業は終了とする。被験者は適当なキーワードを入力し、課題に対して正解と考える文書を探し出してその文書にマークをつける、という作業を繰り返す。

事後アンケート：各システムの使いやすさ・満足度や、良い点・悪い点などについて問う。

その他，実験に関する条件は以下のとおり：

- 被験者には事前に各システムの練習（練習用の課題を検索する）を十分に行ってもらい，使用方法について問題がないことを実験者が確認する．
- 被験者の特性による影響を避けるため（課題, システム）の組合せ系列は全員異なるように設定．
- 被験者の学習による影響を避けるため，4つのシステムを使用する順番はランダムに設定．ただし同じシステムは連続して使用する．
- 被験者は，正解と思う文書にマークをつける作業を以下のどちらの場面でも行うことが出来る：(a) 選択したカテゴリ（クラスタ）に関連する文書集合，または検索式によって抽出した文書集合のタイトル一覧を表示した時，(b) タイトル一覧の中からある文書を選択し，その文書の内容を表示した時．
- 実験の様子を (1) PC 上の画面録画ソフトウェア，(2) 背後からのビデオカメラ，の2種類の方法で録画する．

測定値

課題 q に対する検索作業において，タイトル一覧に現れた文書の集合を $Title(q)$ とする．この $Title(q)$ を，被験者が検索作業によって取り出した全文書（結果文書集合）と考える． $Title(q)$ は課題 q の検索開始時に空に初期化され，以下の2種類の場合に増加する：(1) 被験者が新しいタイトル一覧を開いた時⁴，(2) 被験者がタイトル一覧を今迄よりも下にスクロールした時．

課題 q に対する検索作業において，被験者が課題に対して適切と考えマークをつけた文書の集合を $Mark(q)$ とする． $Mark(q)$ は $Title(q)$ の部分集合である．課題 q に対して，BMIR-J2 が規定する正解文書の集合を $Rel(q)$ とする．適合率・再現率・BMIR-match を以下のように定義する：

$$\text{適合率 } P = \frac{|Title(q) \cap Mark(q)|}{|Title(q)|},$$

$$\text{再現率 } R = \frac{|Mark(q)|}{|Rel(q)|},$$

$$\text{BMIR-match} = \frac{|Mark(q) \cap Rel(q)|}{|Mark(q)|}.$$

⁴タイトル一覧が以前に開かれたものと全く同じである場合は， $Title(q)$ に追加しない．

3.4.2 結果と考察

(a) E 尺度 (検索システムの性能)

各課題ごとの E 尺度と, 10 課題の E 尺度平均値をそれぞれ表 3.6, 表 3.7 に示す (全被験者の平均) . E 尺度は各再現率・適合率のペアから (3.3) 式に従って計算され, 値が小さい程性能が良いことを意味する .

分散分析の結果, $\beta = 0.5, 1.0, 2.0$ すべての場合において課題間に有意差が認められた (1% 有意) が, 4 つのシステム間に有意差は認められなかった . そこで, (b) BMIR-match , (c) 適合率, (d) 再現率それぞれに関する詳しい考察を以下に述べる .

表 3.6 E 尺度, $\beta = 2.0$ (被験者平均)

課題	A	B	C	D
i	0.535	0.657	0.530	0.622
ii	0.813	0.678	0.755	0.691
iii	0.920	0.919	0.920	0.919
iv	0.776	0.796	0.796	0.815
v	0.649	0.788	0.829	0.797
vi	0.631	0.487	0.597	0.463
vii	0.595	0.743	0.576	0.656
viii	0.837	0.778	0.657	0.816
ix	0.853	0.871	0.881	0.839
x	0.768	0.831	0.791	0.730
平均	0.738	0.755	0.733	0.735

表 3.7 E 尺度平均値

	A	B	C	D
$\beta = 0.5$	0.760	0.773	0.754	0.712
$\beta = 1.0$	0.764	0.776	0.757	0.740
$\beta = 2.0$	0.738	0.755	0.733	0.735

(b) BMIR-match (結果の質)

各課題ごとの BMIR-match と, 10 課題の平均 BMIR-match を表 3.8 に示す (全被験者の平均値) . BMIR-match が高いということは, 被験者がマークをつけた文書集合の中に BMIR-J2 規定の正解文書がより多く含まれることを意味する . よって, BMIR-match の値は, 被験者がマークをつけた文書集合の質の良さを反映すると考えられる .

分散分析の結果, 課題間には有意差が認められた (1% 有意) が, 4 つのシステム間に有意差は認められなかった . しかし, 10 課題の平均 BMIR-match に注目すると, CATE 手法 (A,B) はその他の手法 (C,D) よりも高い値になっていることがわかる .

課題ごとの BMIR-match に注目すると, B は 10 課題中 8 課題で C,D よりも高い値

表 3.8 BMIR-match (全被験者の平均)

課題	A	B	C	D
i	0.947	0.984	0.958	0.982
ii	0.968	0.938	0.908	0.883
iii	0.990	0.990	0.980	1.000
iv	0.580	0.458	0.560	0.726
v	0.829	0.907	0.872	0.817
vi	0.471	0.606	0.578	0.505
vii	0.537	0.516	0.476	0.479
viii	0.929	0.900	0.607	0.719
ix	0.651	0.727	0.667	0.600
x	0.835	0.900	0.963	0.890
平均	0.774	0.793	0.757	0.760

になっている。一方課題 iii と課題 iv においては、D (キーワード検索) が CATE 手法 A,B の両方に勝っている。この 2 課題には BMIR-J2 が規定する正解文書が多く存在し (表 3.1) , また BMIR-match が他の 8 課題よりも有意に高い値 (1% 有意で他の 8 課題と差がある) になっていることから、被験者にとっては正解を見つけ易い、比較的「簡単な」課題であると考えられる。

(c) 適合率 (結果の質)

適合率が高いということは、被験者が検索作業中に見るハズレ (被験者の考える不正解) の文書が少ないということである。従って、一般に検索時間が長くなれば適合率は下がる傾向にあるが、理想的には検索時間にかかわらず常に高い値であることが望ましい。そこで各課題ごとに検索終了時の適合率を求めたものを表 3.9 に示す (全被験者の平均値) 。

分散分析の結果、適合率の値はシステムによる影響 (1% 有意) と課題による影響 (5% 有意) の両方を受けていることがわかった。全課題の平均値を見ると、D がその他に比べて有意に高く (1% 有意) , C,A,B が僅差で続いているが C,A,B の間に有意差はない。つまり、CATE 手法 (A,B) はキーワード検索手法 (D) に比べ劣っている。

しかし各課題ごとの値をみると、10 課題中 5 課題で、A と B のどちらかが D よりも高い値になっている。この 5 つの課題 (ii, iv, v, vi, vii) には、(1)BMIR-J2 が規定する正解文書が比較的少ない、(2) 課題名が 1 単語ではないため、そのまま初期キーワードとして使いにくい、また例えキーワードに使っても見つかる正解文書が少ない、という特徴があり、被験者にとっては「難しい」課題であったと考えられる。

表 3.9 検索終了時における適合率（全被験者の平均）

課題	A	B	C	D
i	0.277	0.182	0.348	0.352
ii	0.214	0.413	0.252	0.393
iii	0.586	0.492	0.627	0.844
iv	0.435	0.262	0.446	0.433
v	0.296	0.161	0.121	0.083
vi	0.235	0.558	0.258	0.536
vii	0.277	0.155	0.320	0.236
viii	0.105	0.087	0.156	0.124
ix	0.187	0.103	0.177	0.304
x	0.212	0.191	0.249	0.484
平均	0.282	0.260	0.295	0.379

(d) 再現率（検索の効率）

全被験者の全課題に対する検索時間（キ口秒）と再現率の 2 値分布と，検索時間・再現率 回帰直線および決定係数 R^2 を図 3.6 に示す．

オリジナルのデータは傾き・切片ともに課題間の差が大きく（1%有意で課題間に差がある），そのまま平均するとシステム間の差が分かりにくくなるため，図 3.6 では課題の影響を排除するように以下に述べる補正を行っている．全被験者全課題分のデータから求めた検索時間 (*Time*)・再現率 (*Recall*) 間の回帰直線を

$$Recall = a \times Time + b \quad (3.4)$$

とし，また課題 q に対する全被験者分のデータから求めた検索時間・再現率間の回帰直線を

$$Recall = a_q \times Time + b_q \quad (3.5)$$

とする．以下の式にしたがって，補正後の再現率 $Recall'$ を求める：

$$Recall' = \frac{a}{a_q}(Recall - b_q) + b. \quad (3.6)$$

通常，検索時間が長くなれば被験者がマークをする文書の数が多くなり，すなわち再現率は高くなる．したがって，検索時間・再現率間の回帰直線は右肩上がりのグラフになる（図 3.6）．更に，回帰直線の傾き（回帰係数）が大きい方が，短い時間でより多くの正解が見つかる，すなわち検索の効率が良い結果と言える．

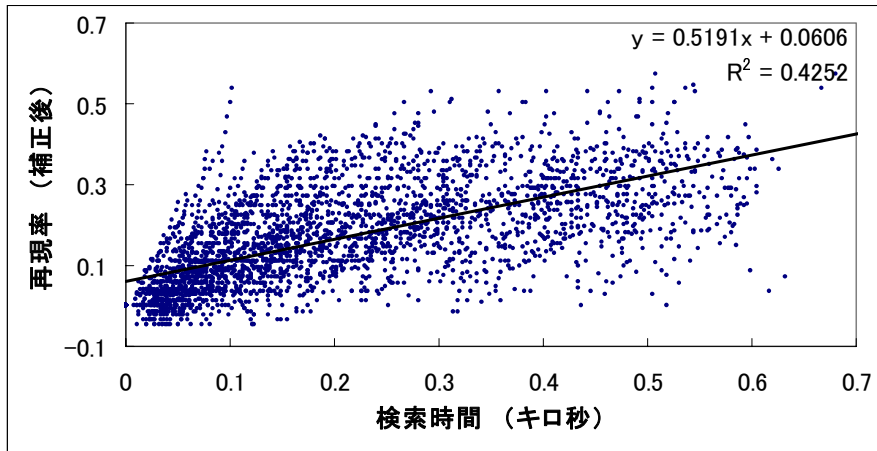


図 3.6 検索時間と再現率 (補正後)

表 3.10 検索時間 (キロ秒)・再現率間の回帰直線

	A	B	C	D	平均
傾き	0.6617	0.5029	0.5263	0.4904	0.5191
切片	0.0373	0.0528	0.0372	0.0914	0.0606
R^2 *1	0.5426	0.4436	0.5188	0.3020	0.4252

*1: R^2 は検索時間・再現率 回帰直線の決定係数

表 3.10 に、各システムに対する補正後の検索時間・再現率 回帰直線の傾き、切片、決定係数 R^2 をそれぞれ示す (全被験者・全課題の平均値)。すなわち

$$Recall' = a_X \times Time + b_X. \quad (3.7)$$

傾きの大きさに注目して 4 つのシステムを比較すると、A が最も傾きが大きく、続いて C, B, D の順になっている。このことから、CATE 手法 (A, B) はキーワード検索手法 (D) よりも効率が良いといえる。

上記 (b)–(d) から以下の考察が得られる

本手法を用いることにより、短い時間で効率良く正解を見つけることができる。特に正解が少なく、また課題名がキーワードとして有効でないような「難しい」課題に対して、本手法は有効な検索手法である。一方正解が多く、課題名をそのままキーワードに使用すれば多くの正解が得られるような「簡単な」課題に対しては、従来のキーワード検索手法が有効である。

表 3.11 アンケート結果 (全被験者の平均)

	A	B	C	D
使いやすさ (0 - 10)	6.50	6.30	4.85	6.65
満足度 (0 - 10)	5.75	5.65	4.85	6.35
総合評価 (1 - 4)	2.65	2.65	1.75	3.10

(e) 事後アンケートより (使いやすさ)

《 4システムと比較 》 各システムの使いやすさと満足度を、0から10までの11段階で評価してもらった (評価値が大きいほど被験者が「使いやすい」と感じたことを意味する)。また、4つのシステムの総合評価を4段階で評価してもらった (一番使い易かったシステムが4、一番使いにくかったシステムが1、同点なし)。全被験者の解答を平均したものを表3.11に示す。

表3.11の3種類の平均評価値は、いずれもD(キーワード検索)が最も高く、続いてCATE手法のAとB、大きく離れてC(クラスタリング)という結果になった。ここから、先験的知識を用いて分類を行う本手法は、統計的手法を用いて分類を行うクラスタリング手法よりも使いやすいものであると推測できる。

一方、本手法はキーワード検索に多少劣る結果になった。しかし被験者に関する節でも述べた通り (P.36)、被験者のほぼ全員が、少なくとも1年以上はキーワード検索サービスを日常的に使用しているのに対し、本手法によるシステムについてはせいぜい数時間の使用経験しかない。このように、被験者がキーワード検索手法に慣れ親しみ、情報検索のメンタルモデルとして強く確立している状況を考えると、本手法の有効性はキーワード検索と遜色ないと推測できる。

《 システムAとBの比較 》 A(明確さ基準)とB(エントロピー基準)の両システムに対して、表示された上位10個の分類観点が課題に対して適切であったかどうかを、0から10までの11段階評価で答えてもらった。全被験者の評価値を平均すると、AがBよりもわずかに高い結果になった (A:平均6.00, B:平均5.35)。

事後アンケートでは、被験者の約半分が、システムによって検索した課題が違うのでAとBの違いを感じ取れなかったと答えている。ただし、2つのシステムの違いを指摘する意見も少数ながらあった：

- 文書集合を2回以上分類した時 (3.2.2節 ステップ(6))、Bは前回とは全く違う分類観点を提示してくれるので、Bの方が便利。
- Aの方が課題に対して適切な分類観点を上位に提示していたので、Aの方が使い易い。

(f) 明確さ (A) とエントロピー (B) についての考察

(a)–(d) の量的な結果においても、(e) のような主観的評価においても、A と B の差はわずかである。(a)–(d) では、いずれもシステム A, B 間に有意差は認められなかった。ただし、(c) 適合率について、システム A では正解文書数との間に強い正の相関があり、正解文書数の多い課題ほど適合率が良い傾向があったが、システム B にはそのような傾向は認められなかった(正解文書数・適合率の相関係数 A: 0.85, B: 0.49)。

今回の実験からは、明確さとエントロピーのどちらが優れているとはいえないが、適合率については明確さの方が課題の難しさに影響されやすい傾向にあるといえる。

3.5 まとめ

本章では、可変なカテゴリ構造を構築することでユーザの文書検索作業を支援する手法を提案し、実験によってその有効性を確認した。

性能評価実験の結果からは、適合率を重視すると本手法がクラスタリング手法よりも優れており、再現率を重視すると本手法がキーワード手法よりも優れていることが言えた。また分類の結果を評価する基準としては、エントロピーの方が明確さよりもやや精度が良い一方で、明確さと同程度の検索精度を得るためにより多くのカテゴリを閲覧しなければならないことが分かった。

被験者による評価実験では、本手法が特に「難しい」検索課題に対して有効であることが確認できた。更に本手法の使いやすさに対する被験者の主観的評価は、クラスタリング手法に大きく勝り、キーワード検索手法にやや劣る程度であった。被験者全員が日常的にキーワード検索サービスを利用し、情報検索のメンタルモデルとして既に確立している状況を考えると、本手法の使いやすさはキーワード検索手法と遜色ないものであった。一方分類の評価基準である明確さとエントロピーについては、どちらが優れているとも言い難い結果となった。

現実のユーザが検索システムを用いる場合は、分類観点やカテゴリの名前から正解の多いカテゴリを推測する必要があるため、それらが検索課題に対して妥当かどうかは重要である。3.3 節および 3.4 節で述べた実験結果とは別に、評価値の高い分類観点や適合率の高いカテゴリの名前を調べたところ、明確さを用いたほうが検索課題に対して妥当と思われるものが多かった(図 3.7)。一方、明確さはその定義から、2 回以上分類を行った際に前回選択した分類観点が再び高い評価値を得る場合が多い(すなわち、同じ分類観点が繰り返し上位に提示されやすい)。逆にエントロピーの場合は、2 回以上分類を行うと前回とは全く異なる分類観点が高い評価値を得る場合がほとんどである。どちらをより使いやすく感じるかは、検索目的や状況、あるいはユーザの好みなどによって左右されると考えられ、評価実験の事後アンケートでも意見が別れた。

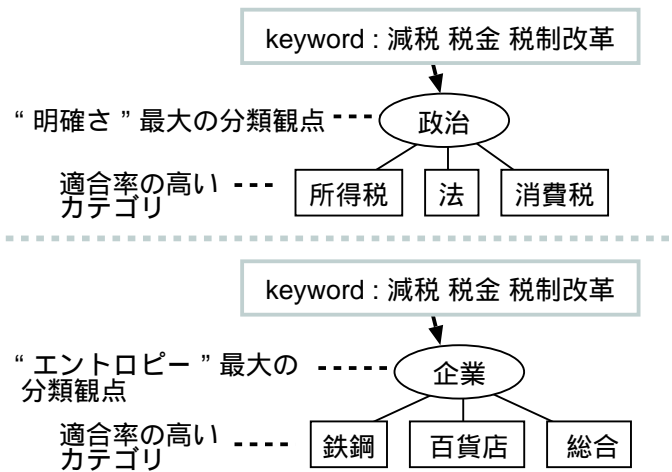


図 3.7 「減税」に対するカテゴリ構造

今後、同じ課題を明確さとエントロピーの両方を用いて被験者に検索してもらうなどの方法で、2つの評価基準をより詳しく比較検討する予定である。更に本手法の改善として、2つの評価基準をユーザの好みやその時々を検索状況に応じて使い分けるための指針を求めたい。また検索課題（課題名と説明文）という形の検索目的に対してだけでなく、以前に見たちょっとした情報、などの多少漠然とした検索を行う場面における本手法の有効性を評価したいと考えている。

実験では、利用可能な WWW 文書のテストコレクションがなかったので、新聞データのテストコレクションを実験に使用した。新聞記事と WWW 文書は、検索の目的が雑多で、検索の専門家でなくても必要な情報を早く取り出せることが重要という共通点がある。一方、新聞記事は完成された商業用出版物であり、ある程度文書の質が揃っているのに対し、WWW 文書は文書の質がバラバラという相違点がある。

今回の実験では、文書の質に関する比較を行っていないので、本手法を WWW 文書データベースに適用した際の結果については議論していない。しかし、関連研究として挙げた文書タイプによる分類技術 [27] を前処理として組み合わせ、文書の質が揃ったデータベースに対して本手法を適用した場合には、少なくとも実験と同等の性能が得られると期待できる。

また本手法では先験的知識として与える分類観点によって性能が左右されるため、分類観点の作成方法やカテゴリベクトルの求め方についても今後検討を重ね、指針を得たい。本手法をシステムとして実現する際、検索対象である文書集合の特徴に合わせて分類観点を調整することで、様々なタイプの文書集合に対してより精度のよい検索を提供できると考えている。

第4章 調整可能な文書ランキングを用いた 検索支援手法(UADR手法)

4.1 はじめに

本章で提案する検索支援手法は、キーワード検索手法の改良として、システムが提示したランキングをユーザが自分の検索目的に沿うように調整する手段を提供するものである。これにより、たとえ現在のランキングが検索目的に対して不適切であるという不満を持ったとしても、ユーザはそれをシステムに示す手段を持たない、ということから生じる心理的なストレス軽減を図る。同時に、『このような文書は上位に移動せよ』というランキング調整手段を提供することで、ユーザにとって直観的で無理のない方法でシステムに対してより正確な検索目的を伝えることのできる検索システムとなることを目指している。

本手法ではユーザの調整操作をスコアの計算規則に反映するための3通りのアルゴリズムを提案し(4.3節)、調整によるランキングの改良効果を実験によって評価した。

4.2 UADR手法の概要とシステム設計の方針

4.2.1 システムの動作概要

UADR手法によるシステムの動作手順は以下のようになる：

- (1) ユーザがキーワードを入力する。
- (2) システムは、データベース中の各文書ごとにキーワードに対する適合度を計算し、その降順に並べた文書リスト(初期ランキング)を出力する。
- (3) ユーザは、不当に順位が低いと感じる文書を上位に上げるか、または自分にとって不適合と感じる文書を下位に下げる操作を行う(図2.3)。
- (4) システムは、ユーザの調整操作(文書順位の上げ/下げ)からユーザの調整意図を推測し、スコア計算規則を変更する(4.3節)。
- (5) システムは、手順(4)で推測した新しい計算規則を用いて、手順(2)で抽出した各文書のスコアを再計算し、新しいランキングをユーザに提示する。

一回の調整作業ごとにランキング中の全文書のスコアが再計算されるため，ユーザが直接調整しなかった文書も順位が変化することになる．ユーザにとっての適合度を手順(4)でうまく推測できれば，新しいランキングでは適合文書が（調整前よりも）上位に集まると期待できる．以下，本手法の各要素について説明する．

4.2.2 ランキングの作成と調整

本章の提案手法では第3章と同様に，文書の表現方法やスコアを定義するためにベクトル空間モデル [34] を用いている．3章 3.2.1 節での定義を用い，有限集合 S に対し， S の要素数を $|S|$ と書く．同様に，ベクトル v_1 と v_2 の内積を $v_1 \cdot v_2$ ，ベクトル v の長さを $|v| = \sqrt{v \cdot v}$ とそれぞれ表記する．

本手法では，文書 d を3章 CATE 手法での特徴ベクトル ($fv(d)$) ではなく，以下の縮約特徴ベクトル ($rfv(d)$) で表す．ユーザが手順(1)で入力したキーワードの集合を $KW = \{kw_1, kw_2, \dots, kw_p\}$ とすると，文書 d の KW に対する縮約特徴ベクトル $rfv(d)$ は以下で定義される $p+m$ 次元ベクトルである．

$$rfv(d) = \underbrace{(c_{kw_1,d}, \dots, c_{kw_p,d})}_p, \underbrace{(c_{f_1,d}, \dots, c_{f_m,d})}_m \quad (4.1)$$

ここで， $c_{kw_u,d}$ ($1 \leq u \leq p$) はキーワード kw_u と文書 d の関連度 ($tf * idf$ [18] など)， $c_{f_v,d}$ ($1 \leq v \leq m$) は単語以外の文書の特徴を示す値（例えばリンク密度や更新の新しさなど）である．簡単のため，本章では $d = rfv(d)$ とし，以降 d の縮約特徴ベクトルを単に d と表記する．

文書の表現として CATE 手法で用いた特徴ベクトル ($fv(d)$) は，データベースに出現する全単語に対する関連度を要素に持つものであるため，非常に高次元のベクトル (3.2.1 の W に対して $|W|$ 次元) になる．CATE 手法では，実装上の工夫により，各文書に対して必要なベクトル計算 (3.2.2 節手順(3)での類似度計算) を予め (実際に検索式が入力される前に) 実行しておく事が可能であったため，そのような高次元の特徴ベクトルを用いることができた．しかし UADR 手法では，各文書に対して必要なベクトル計算 (4.3 節で述べる問合せベクトルの計算) は調整要求のたびに必ず実行しなければならないため，特徴ベクトルのように高次元のベクトルを用いることは応答時間の点から不可能である．そこで，データベースに出現する全単語ではなく，ユーザが入力したキーワードのみに対する関連度 ($c_{kw_u,d}$) と，文書の特徴を簡単に表現できる要素 ($c_{f_v,d}$) を付け加えたベクトルを考え，縮約特徴ベクトルとして用いた．

2つのベクトル v_1 と v_2 間の類似度 $\text{sim}(v_1, v_2)$ を以下のように2種類定義する．

$$\text{sim}(v_1, v_2) = v_1 \cdot v_2 \quad (4.2)$$

$$\text{sim}(v_1, v_2) = \frac{v_1 \cdot v_2}{|v_1| |v_2|} = \cos \theta \quad (4.3)$$

式 (4.3) の定義は，CATE 手法で用いた類似度と同じであり，2 つのベクトルの成す角の余弦を用いる．一方，式 (4.3) の定義はベクトルの内積を用いるものである．この 2 種類の類似度はどちらもベクトル空間モデルにおいてよく利用される定義である．本手法では 4.3 節で述べる推測手法によってこの 2 種類の類似度を使い分ける．

システムが手順 (4) で推測するスコア計算規則とは，縮約特徴ベクトルの各成分に対する重みを成分に持つベクトル $\mathbf{k} = (k_1, k_2, \dots, k_{p+m})$ (問合せベクトルと呼ぶ) である． \mathbf{k} の x 番目の成分 k_x は， d の x 番目の成分に対する重みを示している．システムが手順 (2) で抽出した文書集合を D ，ユーザの i 回目の調整操作に応じてシステムが作る問合せベクトルを $\mathbf{k}^{(i)}$ とする． i 回目の調整後の文書 d のスコア $\text{score}(d, i)$ ($i \geq 0$) を，

$$\begin{aligned} \text{score}(d, 0) &= \text{sim}(d, \mathbf{k}^{(0)}) \\ \text{score}(d, i) &= \text{sim}(d, \mathbf{k}^{(i)}) + \alpha (\text{sim}(d, \mathbf{k}^{(i)}) - \text{sim}(d, \mathbf{k}^{(i-1)})) \end{aligned} \quad (4.4)$$

とする (α は，今回の調整による変化をどの程度強調するかを決める非負定数． α が 0 の場合， $\text{score}(d, i)$ は $\text{sim}(d, \mathbf{k}^{(i)})$ と等しい)．ユーザの i 回目の調整後の D のランキングとは，以下の不等式を満たす文書リスト $d_1, d_2, \dots, d_{|D|}$ である．

$$\text{score}(d_1, i) \geq \text{score}(d_2, i) \geq \dots \geq \text{score}(d_{|D|}, i).$$

ただし，初期ランキングの作成 (手順 (2)) には $\mathbf{k}^{(0)} = (\underbrace{1, 1, \dots, 1}_p, \underbrace{0, 0, \dots, 0}_m)$ で表される初期間合せベクトルを用いる¹． $\mathbf{k}^{(i)}$ ($i \geq 1$) の作成方法を 4.3 節で述べる．

4.3 問合せベクトルの推測手法

4.2.1 節手順 (3) において，ユーザが文書 d_h を文書 d_l ($1 \leq l < h \leq |D|$) のすぐ上に移動する調整操作を行った場合，ユーザの意図を次のように仮定する．

- d_h は d_1, d_2, \dots, d_{l-1} よりも検索目的に適していない．
- d_h は $d_l, d_{l+1}, \dots, d_{h-1}$ よりも検索目的に適している．

上記の条件を満たす新しい問合せベクトルを作成するために，本研究では，以下に述べる 3 種類の問合せベクトル推測手法を提案する (4.3.1 節 ~ 4.3.3 節)．ただしこれらの推測手法は，その定義より，必ずしも文書 d_h をちょうど d_{l-1} と d_l の間に配置するとは限らない．4.2.1 節で述べたとおり，ユーザの調整意図に対して適切な問合せベクトルを推測することができれば，新しいランキングでは適合文書が上位に集まることが期待できる．

¹手順 (1) でユーザがキーワード以外の要素についても指定できるように実装すると，後半の m 成分についても 0 以外の数値を持ちうる初期間合せベクトルになる．

4.3.1 成分補正法（手法 A）

文書 d_i の縮約特徴ベクトルの第 j 成分 ($1 \leq j \leq p+m$) を $d_{i,j}$ と書く．すなわち $d_{i,1} = c_{kw1,d_i}$, $d_{i,2} = c_{kw2,d_i}$, ..., $d_{i,p+m} = c_{fm,d_i}$ (式 (4.1) 参照)．また $d_1, d_{l+1}, \dots, d_{h-1}$ の平均ベクトルを d_{lows} とする．すなわち $d_{lows} = \sum_{i=l}^{h-1} d_i / (h-l)$ ． d_{lows} の第 j 成分も $d_{lows,j}$ と表す．本手法では，新しい問合せベクトルの第 j 成分 $k_j^{(i)}$ を，現在の問合せベクトルの第 j 成分 $k_j^{(i-1)}$ と，2つの縮約特徴ベクトル d_{lows} , d_h の第 j 成分の差 $d_{h,j} - d_{lows,j}$ に基づいて決定する．具体的には，まず，スカラー y を未知数とする次のような 1 次不等式を考える．

$$\text{sim}(\mathbf{md}_h, \mathbf{k}^{(i-1)} + y(\mathbf{md}_h - \mathbf{md}_{lows})) \geq \text{sim}(\mathbf{md}_{lows}, \mathbf{k}^{(i-1)} + y(\mathbf{md}_h - \mathbf{md}_{lows})) \quad (4.5)$$

ここでは類似度として式 (4.3) を用いる．また， \mathbf{md}_h , \mathbf{md}_{lows} はそれぞれ d_h , d_{lows} に対して後で述べる補正を行ったベクトルである．式 (4.5) は，補正後のベクトル $\mathbf{md}_h, \mathbf{md}_{lows}$ において， \mathbf{md}_h と $\mathbf{k}^{(i-1)} + y(\mathbf{md}_h - \mathbf{md}_{lows})$ の類似度が \mathbf{md}_{lows} と $\mathbf{k}^{(i-1)} + y(\mathbf{md}_h - \mathbf{md}_{lows})$ の類似度以上になることを意味している．そして，式 (4.5) を満たす最小の y に対して，

$$\mathbf{k}^{(i)} = \mathbf{k}^{(i-1)} + y(\mathbf{md}_h - \mathbf{md}_{lows})$$

を新しい問合せベクトルとする．

成分補正法を用いた問合せベクトル変更の例を示す．1 位から順に d_1, d_2, d_3, d_4 である下記のランキングに対し，ユーザが i 回目の調整操作として文書 d_4 を文書 d_1 のすぐ上に移動したとする．

$$\begin{aligned} \mathbf{k}^{(i-1)} &= (1, 1, 0), \\ \mathbf{d}_1 &= (0.9, 0.2, 0.4), \\ \mathbf{d}_2 &= (0.8, 0, 0.4), \\ \mathbf{d}_3 &= (0, 0.6, 0.1), \\ \mathbf{d}_4 &= (0.3, 0.2, 0.5). \end{aligned}$$

この例の場合， $d_h = d_4$ であり， d_{lows} は d_1, d_2, d_3 の平均である．補正後のベクトル $\mathbf{md}_h, \mathbf{md}_{lows}$ はそれぞれ以下ようになる．

$$\begin{aligned} \mathbf{md}_h &= (0.3, 0.2, 0.5) \\ \mathbf{md}_{lows} &= (0.567, 0.267, 0.3) \end{aligned}$$

y の値を 0.01 刻みで増加させて式 (4.5) を試行することで，式 (4.5) を満たす最小の $y = 2.33$ が求まる．よって

$$\mathbf{k}^{(i)} = (0.379, 0.845, 0.466)$$

が導き出され，これを新しい問合せベクトルとしてスコアを再計算すると，式 (4.4) で $\alpha = 1$ の場合，新しいランキングは d_1, d_4, d_3, d_2 となる．

補正について $d_i = (d_{i,1}, d_{i,2}, \dots, d_{i,p+m})$ に対し, ベクトル md_i を以下のように定義する.

$$md_i = (\eta_1 d_{i,1}, \eta_2 d_{i,2}, \dots, \eta_{p+m} d_{i,p+m})$$

すなわち md_i は, d_i の第 j 成分に下記の補正係数 η_j を掛けたものである. d_{lows} に対しても同様に md_{lows} を定義する. η_j ($1 \leq j \leq p+m$) は次のように定義する².

$$\eta_j = \begin{cases} 0 & \text{条件 (a) が真} \\ 0.5 & \text{条件 (b) が真} \\ 1 & \text{otherwise} \end{cases}$$

$$(a) = \begin{cases} d_{h,j} > d_{lows,j} \wedge d_{l-1,j} < d_{lows,j} \\ \text{または} \\ d_{h,j} < d_{lows,j} \wedge d_{l-1,j} > d_{lows,j} \end{cases}$$

$$(b) = \begin{cases} d_{h,j} > d_{lows,j} \wedge d_{l-1,j} < \frac{d_{lows,j} + d_{h,j}}{2} \\ \text{または} \\ d_{h,j} < d_{lows,j} \wedge d_{l-1,j} > \frac{d_{lows,j} + d_{h,j}}{2} \end{cases}$$

この補正は, 問合せベクトルを作成する際に, ユーザが変更を希望していないと思われる成分をシステムが誤って大幅に変更してしまうことを避けるために行う. すなわち,

- 条件 (a) が真となるのは, 第 j 成分に関して, 上位に移動した d_h が d_{lows} より値が大きいかにも関わらず, d_{l-1} が d_{lows} よりも小さい場合, もしくは d_h が d_{lows} より値が小さいにも関わらず, d_{l-1} が d_{lows} よりも大きい場合である. この場合, ユーザの調整意図は第 j 成分とは無関係と考え, 補正係数 η_j を 0 とする.
- 条件 (b) が真となるのは, 第 j 成分に関して, d_{l-1} が d_h よりも d_{lows} に近い場合である. この場合, ユーザの調整意図と第 j 成分の関係は, d_{l-1} が d_{lows} よりも d_h に近い場合 ($\eta_j = 1$ の場合) に比べて薄いと考え, 補正係数 η_j を 0.5 とする.

²以下, 特別な場合として $l=1$ の時は $d_{l-1} = d_h$ とみなす.

4.3.2 E 尺度に基づく推測手法 (手法 B)

4.3.1 節の手法 A と同様に, ユーザが i 回目の調整で文書 d_h を文書 d_l ($1 \leq l < h \leq |D|$) のすぐ上に移動した場合について考える. 本手法では次の 2 つのベクトル g (good) と b (bad) を用いる.

$$\begin{aligned} \mathbf{g} &= (\mathbf{d}_{l-1} + \mathbf{d}_h) / 2 \\ \mathbf{b} &= \mathbf{d}_{lows} = \sum_{i=l}^{h-1} \mathbf{d}_i / (h-l) \quad (4.3.1 \text{ 節}) \end{aligned}$$

ユーザの調整意図を反映した新しい問合せベクトル $\mathbf{k}^{(i)}$ を求めるため, 本手法では 1.3 節 (P.11) で述べた E 尺度 [28] の概念を用いる. E 尺度は検索結果の適合率 P と再現率 R の重み付き調和平均から求められ, 値が小さいほど良い検索結果であることを意味する.

$$E = 1 - \frac{(\beta^2 + 1) PR}{\beta^2 P + R} \quad \beta \text{ は非負定数} \quad (4.6)$$

直観的に, 新しい問合せベクトル \mathbf{k} は, 出来るだけ \mathbf{g} に類似し, かつ, 出来るだけ \mathbf{b} に類似していない事が求められる. 本手法はこの関係を再現率と適合率のトレード・オフに見立てることで, \mathbf{k} の最適化問題を E 尺度の最小化問題と定式化する. 本手法では類似度として式 (4.3) を用いる. また, ベクトル間の非類似度 $\text{dsim}(\mathbf{v}_1, \mathbf{v}_2)$ を以下のように定義する.

$$\text{dsim}(\mathbf{v}_1, \mathbf{v}_2) = \sqrt{1 - \left(\frac{(\mathbf{v}_1, \mathbf{v}_2)}{|\mathbf{v}_1| |\mathbf{v}_2|} \right)^2} = \sin \theta \quad (4.7)$$

次に, 式 (4.6) の P, R を $\text{sim}(\mathbf{g}, \mathbf{k}), \text{dsim}(\mathbf{b}, \mathbf{k})$ で置き換えた式

$$E = 1 - \frac{(\beta^2 + 1) \text{sim}(\mathbf{g}, \mathbf{k}) \text{dsim}(\mathbf{b}, \mathbf{k})}{\beta^2 \text{sim}(\mathbf{g}, \mathbf{k}) + \text{dsim}(\mathbf{b}, \mathbf{k})}$$

を最小にする \mathbf{k} を求め, $\mathbf{k}^{(i)}$ とする. 簡単のため $\beta = 1$ とすると, 与えられたベクトル \mathbf{g} と \mathbf{b} に対して, 以下の r が最大になるような \mathbf{k} を求めれば良い.

$$r = \frac{\text{sim}(\mathbf{g}, \mathbf{k}) \text{dsim}(\mathbf{b}, \mathbf{k})}{\text{sim}(\mathbf{g}, \mathbf{k}) + \text{dsim}(\mathbf{b}, \mathbf{k})} \quad (4.8)$$

式 (4.8) の r を最大にする長さ 1 のベクトル \mathbf{k} は, 次式 (4.9) で与えられる (導出過程は付録 B 参照).

$$\mathbf{k} = \frac{c\mathbf{g} - \mathbf{b}}{|c\mathbf{g} - \mathbf{b}|} \quad \text{ただし} \quad c = \frac{1 + \sin \theta}{\cos \theta} \quad (4.9)$$

以下に E 尺度に基づく推測手法を用いた問合せベクトル変更の例を示す．4.3.1 節の手法 A と同様に，ユーザが文書 d_4 を文書 d_1 のすぐ上に移動したとする．

$$\begin{aligned} \mathbf{k}^{(i-1)} &= (1, 1, 0), \\ \mathbf{d}_1 &= (0.9, 0.2, 0.4), \\ \mathbf{d}_2 &= (0.8, 0, 0.4), \\ \mathbf{d}_3 &= (0, 0.6, 0.1), \\ \mathbf{d}_4 &= (0.3, 0.2, 0.5). \end{aligned}$$

d_4 を正規化して \mathbf{g} を求め， d_1, d_2, d_3 の平均を正規化して \mathbf{b} を求めると以下のようになる．

$$\begin{aligned} \mathbf{g} &= (0.487, 0.324, 0.811) \\ \mathbf{b} &= (0.816, 0.384, 0.432) \end{aligned}$$

ここで $\cos \theta = (\mathbf{g}, \mathbf{b}) = 0.872$ であることから \mathbf{g} と \mathbf{b} のなす角 θ が求まり，式 (4.9) から

$$\begin{aligned} c &= 1.71, \\ \mathbf{k} &= (0.016, 0.176, 0.984) \end{aligned}$$

が導き出せる．この \mathbf{k} を新しい問合せベクトルとしてスコアを再計算すると，式 (4.4) で $\alpha = 1$ の場合，新しいランキングは d_4, d_2, d_1, d_3 となる．

4.3.3 順序情報に基づく推測手法 (手法 C)

4.3.1 節の手法 A，4.3.2 節の手法 B と同様に，ユーザが i 回目の調整で文書 d_h を文書 d_l ($1 \leq l < h \leq |D|$) のすぐ上に移動した場合について考える．4.3 節冒頭で述べた条件を満たすランキングを生成するために，下記の不等式を満たす問合せベクトル $\mathbf{k} = (k_1, \dots, k_{p+m})$ を求める．ただし，本手法では類似度として式 (4.3) を用いる．

$$\text{sim}(\mathbf{d}_h, \mathbf{k}) \leq \text{sim}(\mathbf{d}_j, \mathbf{k}) \quad \text{for } 1 \leq j \leq l-1 \quad (4.10)$$

$$\text{sim}(\mathbf{d}_h, \mathbf{k}) \geq \text{sim}(\mathbf{d}_j, \mathbf{k}) \quad \text{for } l \leq j \leq h-1 \quad (4.11)$$

式 (4.10)，式 (4.11) を満たす解は一般に複数存在するため， k_1, \dots, k_{p+m} に対する別の線形制約式として

$$k_1 + \dots + k_{p+m} = k_1^{(i-1)} + \dots + k_{p+m}^{(i-1)} \quad (4.12)$$

を考える．ここで $\mathbf{k}^{(i-1)} = (k_1^{(i-1)}, \dots, k_{p+m}^{(i-1)})$ は前回の問合せベクトルである．以上の制約式 (4.10)，(4.11)，(4.12) を満たし，かつ， $\mathbf{k}^{(i-1)}$ と \mathbf{k} の類似度が最大になるようなベクトル \mathbf{k} を線形計画法 [41] によって求め，新しい問合せベクトル $\mathbf{k}^{(i)}$ とする．ただし，そのような解 \mathbf{k} が存在しない場合は $\mathbf{k}^{(i)} = \mathbf{k}^{(i-1)}$ とする．

線形計画法を用いた問合せベクトル変更の例を示す。4.3.1 節の手法 A，4.3.2 節の手法 B と同様に，ユーザが文書 d_4 を文書 d_1 のすぐ上に移動したとする。

$$\begin{aligned} \mathbf{k}^{(i-1)} &= (1, 1, 0), \\ \mathbf{d}_1 &= (0.9, 0.2, 0.4), \\ \mathbf{d}_2 &= (0.8, 0, 0.4), \\ \mathbf{d}_3 &= (0, 0.6, 0.1), \\ \mathbf{d}_4 &= (0.3, 0.2, 0.5). \end{aligned}$$

線形制約式は以下の通り。

$$\begin{aligned} -0.6k_1 + 0.1k_3 &\geq 0, \\ -0.5k_1 + 0.2k_2 + 0.1k_3 &\geq 0, \\ 0.3k_1 - 0.4k_2 + 0.4k_3 &\geq 0, \\ k_1 + k_2 + k_3 &= 2 \end{aligned}$$

上記制約式を満たすベクトルの中で， $k_1 + k_2$ が最大になるものを求めればよい。線形計画法のアルゴリズムを用いて以下の解が導き出せる。

$$\mathbf{k} = \left(\frac{8}{55}, \frac{54}{55}, \frac{48}{55} \right)$$

この \mathbf{k} を新しい問合せベクトルとしてスコアを再計算すると，式 (4.4) で $\alpha = 1$ の場合，新しいランキングは d_4, d_3, d_1, d_2 となる。

4.4 被験者を用いた評価実験

UADR 手法によってランキングを調整することで検索結果の質が向上するかどうかを調べるため，本手法を実装した試作システムを用いて以下の評価実験を行った。まず 6 名の被験者を用いた実験を行って本手法の有効性を評価した（本節）。次に，本節の実験結果をもとに 3 種類の問合せベクトル推測手法を改良した上で，各手法をより詳しく比較検討しベンチマークを獲得するための実験を行った（4.5 節）。

本節で述べる実験では，被験者に検索課題とその初期ランキングを与え，任意回数の調整操作を行ってもらう。実験結果の評価は以下に注目して行った。

- 調整によってより多くの適合文書を上位に配置することができたか。
- 調整によって適合文書の順位を選択的に上げることができたか。すなわち，適合文書の順位は上昇し，不適合な文書の順位は下降しているか。
- ユーザの調整意図をうまく推測し，新しい問合せベクトルに反映させることができたか。また，ユーザの調整意図を正しく推測できた場合とそうでない場合で結果にどのような影響があったか。

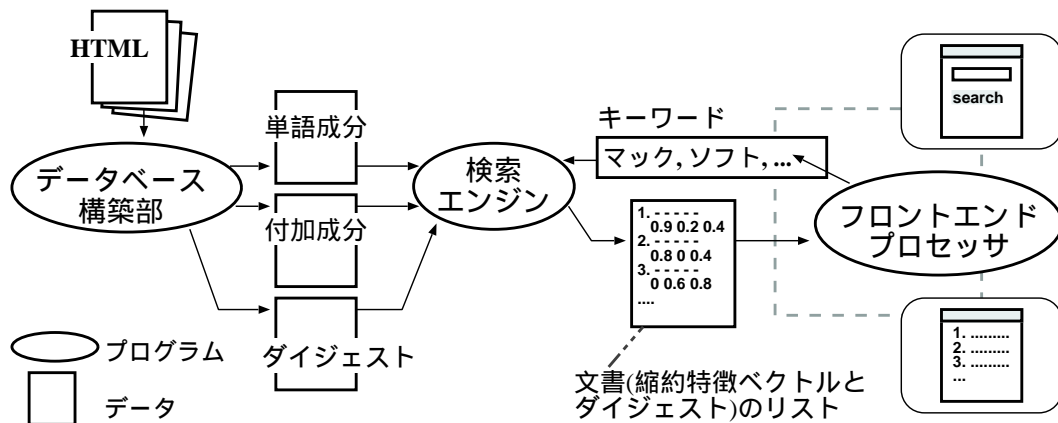


図 4.1 UADR システムの構成

4.4.1 試作システムの概要

試作システムは、データベース構築部、検索エンジン、フロントエンドプロセッサからなる(図 4.1)。データベース構築部と検索エンジンは、既存のキーワード検索システムである *Freya* [35] を拡張して実装した。これらはフロントエンドプロセッサからキーワードを入力として受け取り、以下を出力する。

- 指定されたキーワードに関連する文書の集合
- 上記文書集合中の各文書に対する縮約特徴ベクトルとダイジェスト文

フロントエンドプロセッサはユーザの入力したキーワードを検索エンジンに渡し、検索エンジンから返された情報と現在の問合せベクトルを用いてランキングを作成・表示する。検索エンジンから返された情報はキーワードが変更されない限り保持しておき、ユーザがランキングの調整操作を行った場合には再度検索エンジン側と通信することなくランキングの変更を行う³。

フロントエンドプロセッサには 4.3 節で述べた各問合せベクトル推測手法をそれぞれシステム A,B,C として実装し、実験に用いた。

- システム A：成分補正法 (4.3.1 節)
- システム B：E 尺度に基づく手法 (4.3.2 節)
- システム C：順序情報に基づく手法 (4.3.3 節)

また単語以外の文書の特徴を示す成分(縮約特徴ベクトルの $c_{f_v,d}$)として、“文書中のリンク数”、“更新日の新しさ”の 2 成分を用いた。式 (4.4) 中の α は両実験とも 0.25 に設定した。

³ 今回の実験では、システム A,B,C とも計算時間は 100 文書の再ランキングに対して 1 秒程度。

4.4.2 実験環境

検索課題 検索課題として表 4.1 に挙げた 5 課題を設定した。今回の実験では、4.2.1 節手順 (1) で入力するキーワードはあらかじめ実験者が指定したものをを用いる (表 4.1 各課題の 2 行目)。被験者に対しては各課題の説明文 (100 字程度) とキーワードを提示し、課題の内容について十分理解したことを確認のうえ、実験を行った。これら 5 つの課題は、問合せベクトルの各成分の中で特に入力キーワードに対する重みがランキングの適切さに影響すると思われる課題 (課題 ii, iv, v) と、“文書中のリンク数” や “更新日の新しさ” など単語以外の成分に対する重みが影響すると思われる課題 (課題 i, iii) から成るよにという意図で決定した。

表 4.1 UADR : 実験に使用した検索課題とキーワード

番号	課題名 キーワード	適合文書の数 ^{*1}
i	マック用ソフトウェアに関するリンク集 マック, ソフト, ダウンロード, リンク	36
ii	2001 年に国内で開催される国際会議 国際, 会議, 2001	28
iii	大阪府内で最近国際会議に使用された会場 国際会議, 大阪, 会場	24
iv	奈良先端大周辺のラーメン屋 ラーメン, 大阪, 奈良	45
v	禁煙を成功させるための工夫 禁煙, 工夫	47

*1: あらかじめ実験者が設定した正解文書の数

文書集合 文書データベースは、課題ごとに表 4.1 のキーワードを既存の WWW 検索サービスに入力し、上位から 1000 文書分の URL を収集して作成した。実験では、各課題について、上記データベースと表 4.1 のキーワードを試作システムの検索エンジン⁴に入力した場合の上位 100 文書までを操作対象文書集合とした。また、各課題の操作対象文書 (100 文書) 全てに対し、課題に適合する () か不適合 (x) かを実験者が設定した (被験者には非通知)。

⁴キーワード検索システム *Freya* と同等。tf*idf を用いたキーワードと文書の関連度に基づく。

被験者 情報科学研究科の大学院生 6 人で、いずれも既存の WWW 検索サービスを日常的に使用している。実験に対する謝礼として千円を支払った。

手順 各被験者ごとに以下の手順で実験を行った。

- (1) 実験者が、実験すべき（課題，システム）の組合せを指示する。1 人の被験者は 5 つの課題をちょうど 1 回ずつ，A,B,C のいずれかのシステムを用いて行うよう指示される。たとえば被験者 1 は (v, A), (i, A), (iv, C), (iii, B), (ii, B) の順に実験を行う。各課題ごとにまず実験者がキーワードを入力し，初期ランキングを表示して，被験者に課題の説明を行う。
- (2) 被験者は各課題ごとに以下の作業を 10 分間繰り返し行い，初期ランキングをより適切なものに改良してゆく。ただし被験者が現在のランキングに満足した場合は，10 分未満でも作業を終了する。
 - (a) 現在のランキング中で調整を行いたい（順位を上げ / 下げたい）と感じる文書を探し，調整によってどのような変化を望むのか（調整意図と呼ぶ）を用紙に記述する。調整意図は，問合せベクトルの各成分に対し，その成分の重要度を「上げたい / 下げたい / 今回の調整には関係ない」の 3 択形式で記述する。
 - (b) 実際の調整操作を行う。
 - (c) 調整の結果表示されたランキングがあまりにも調整意図から離れていると感じた場合は，『もどる』ボタンによって操作を取り消すことが出来る。
- (3) 5 課題全てに対する作業が終了した後，被験者はシステムの使いやすさなどに関する事後アンケートに答える。

その他，実験に関する条件は以下のとおり：

- 被験者には実験前日までに実験の目的やシステムの操作方法をまとめたドキュメントを配布し，事前に読んでおいてもらった。また，実験当日にはまず練習用の課題を用いて実験と同じ手順で練習を行い，システムの使用方法について問題がないことを実験者が確認した後に実験を行った。
- 被験者の特性による影響を避けるため，（課題，システム）の組合せ系列は全員異なるように設定した。
- 被験者の学習による影響を避けるため，課題およびシステムの使用順序は全員異なるように設定した。ただし，同じシステムは連続して使用する。

4.4.3 測定値

各課題 q ごとに，操作対象文書の集合を D ($|D| = 100$)，適合文書の集合を D_{rel} ($D_{\text{rel}} \subset D$) とし，全体適合率 P を以下のように定義する：

$$\text{全体適合率 } P = \frac{|D_{\text{rel}}|}{|D|} = \frac{|D_{\text{rel}}|}{100} .$$

また各課題に対する作業ごとに，被験者 s が i 回目の調整操作を行った後に表示されるランキングにおいて上位 20 位⁵までの文書の集合を $D_{(\leq 20)}^i$ とする．同様に，被験者 s が i 回目の調整操作を行った場合に，その調整によって新しく上位 20 位以内にランクインした文書の集合を D_{new}^i ，その調整によって順位が上がった文書の集合を D_{up}^i ，順位が下がった文書の集合を D_{down}^i とする．すなわち，第 i 回目調整後のランキングにおける文書 d の順位を $\text{Rank}(d, i)$ とすると，

$$\begin{aligned} D_{(\leq 20)}^i &= \{d : \text{Rank}(d, i) \leq 20\} , \\ D_{\text{new}}^i &= \{d : \text{Rank}(d, i-1) > 20 \text{ かつ } \text{Rank}(d, i) \leq 20\} , \\ D_{\text{up}}^i &= \{d : \text{Rank}(d, i) < \text{Rank}(d, i-1)\} , \\ D_{\text{down}}^i &= \{d : \text{Rank}(d, i) > \text{Rank}(d, i-1)\} . \end{aligned}$$

各課題 q ・各被験者 s の作業に対し，各 i 回目の調整ごとに以下の 4 種類の適合率を定義する：

$$\begin{aligned} \text{20 位以内適合率} \quad P_{(\leq 20)}^i &= \frac{|D_{(\leq 20)}^i \cap D_{\text{rel}}|}{|D_{(\leq 20)}^i|} , \\ \text{新規 20 位以内適合率} \quad P_{\text{new}}^i &= \frac{|D_{\text{new}}^i \cap D_{\text{rel}}|}{|D_{\text{new}}^i|} , \\ \text{up 適合率} \quad P_{\text{up}}^i &= \frac{|D_{\text{up}}^i \cap D_{\text{rel}}|}{|D_{\text{up}}^i|} , \\ \text{down 適合率} \quad P_{\text{down}}^i &= \frac{|D_{\text{down}}^i \cap D_{\text{rel}}|}{|D_{\text{down}}^i|} . \end{aligned}$$

ただし分母が 0 である適合率（例えば $|D_{\text{down}}^i| = 0$ のときの P_{down}^i ）は，以降での分析対象としない．

4.4.4 結果と考察

(a) 20 位以内適合率

20 位以内適合率が調整前（初期ランキング）よりも高くなっていけば，調整によってランキングの上位部分をより適切なものに改良できたと考えられる．そこで，各 i 回目調整後の 20 位以内適合率と初期ランキングにおける 20 位以内適合率との比率

⁵結果表示の 1 ページ目．既存の WWW 検索サービスのほとんどが 10~20 文書を 1 ページにまとめて表示していることから，試作システムも 20 文書ずつ表示するよう設計した．

表 4.2 被験者を用いた実験の結果

	(a) $\frac{P_{(\leq 20)}^i}{P_{(\leq 20)}^0}$	(b) $\frac{P_{\text{new}}^i}{P}$	(c) $\frac{P_{\text{up}}^i}{P_{\text{down}}^i}$
	平均 > 1 *	平均 > 1 *	平均 > 1 *
A	0.89 N(1%)	1.00	1.13
B	0.80 N(1%)	1.08	1.08
C	0.99	1.34 Y(1%)	1.42 Y(1%)
課題 i	1.05	1.49 Y(1%)	1.00
課題 ii	0.85 N(1%)	0.73 N(5%)	1.14
課題 iii	0.76 N(1%)	0.81	0.93
課題 iv	0.97	1.09	1.54 Y(1%)
課題 v	0.87 N(1%)	1.43 Y(1%)	1.36 Y(1%)
被験者 1	0.90 N(1%)	1.35 Y(1%)	1.03
被験者 2	0.97	1.04	1.13
被験者 3	0.80 N(1%)	0.67	1.17
被験者 4	0.77 N(1%)	0.82	1.35
被験者 5	1.03	1.28 Y(1%)	1.29
被験者 6	0.84 N(1%)	1.25	1.41 Y(1%)
全体	0.90 N(1%)	1.14 Y(1%)	1.22 Y(1%)

* 平均値が 1.0 よりも有意に高い場合を Y, 低い場合を N, 有意差が認められない場合を空欄で示す. カッコ内は有意水準.

$P_{(\leq 20)}^i / P_{(\leq 20)}^0$ を全課題・被験者について求め, 平均値が 1.0 より高いかどうかの検定を行った. 各システム・課題・被験者ごとの平均値と検定結果を表 4.2 の (a) に示す.

全体の結果(表の最下行)を見ると, $P_{(\leq 20)}^i / P_{(\leq 20)}^0$ の平均値は 1.0 よりも有意に低い値となっており, 初期ランキングよりも 20 位以内適合率は悪くなっているという結果になった. 分散分析の結果システム A,B,C の間には有意差が認められ(1%有意), 平均値はシステム C が最も高いが, 1.0 よりも高いとは言えなかった. 同様に課題間・被験者間にも有意差はあるが, どの課題・被験者についても平均値が 1.0 より高いとは言えなかった.

(b) 新規 20 位以内適合率

たとえ上位 20 位の適合率が同じでも, 調整によって新しく未知の適合文書(調整前には 20 位以内に入っていなかった適合文書)を多く見つけることができれば, ユーザにとって有益であると考えられる. そこで, 調整によって適合文書を選択的に上位

20 位以内に上昇させることができたかどうかを調べるため、各 i 回目調整後の新規 20 位以内適合率と全体適合率との比率 P_{new}^i / P を全課題・被験者について求め、平均値が 1.0 より高いかどうかの検定を行った (表 4.2 (b)) .

全体の結果を見ると、 P_{new}^i / P の平均値は有意水準 1% で 1.0 よりも高い値になっており、調整によって下位に埋もれていた適合文書を選択的に上位に移動することができたと言える。分散分析の結果システム間には有意差が認められ (1% 有意)、平均値が 1.0 よりも有意に高いと言えたのはシステム C のみであった。また課題間・被験者間には有意差は認められなかったが、課題 i,v と被験者 1,5 が他に比べ良い結果になった。

(c) up 適合率 vs. down 適合率

上位 20 文書に限らず、全体として適合文書の順位を上げ、不適合文書の順位を下げる事が出来ていれば、適切な問合せベクトル変更が行えたと考えられる。そこで、調整によって順位の上がった文書の適合率が、順位の下がった文書の適合率よりも高いかどうかを調べる。各 i 回目調整後の up 適合率と down 適合率との比率 $P_{\text{up}}^i / P_{\text{down}}^i$ を全課題・被験者について求め、平均値が 1.0 より高いかどうかの検定を行った (表 4.2 (c)) .

全体の結果を見ると、 $P_{\text{up}}^i / P_{\text{down}}^i$ の平均値は有意水準 1% で 1.0 よりも高い値になっており、調整によって起こったランキング (問合せベクトル) の変化は適切であったと言える。分散分析の結果システム間には有意差が認められ (1% 有意)、平均値が 1.0 よりも有意に高いと言えたのはシステム C のみであった。また課題間・被験者間には有意差は認められなかったが、課題 iv,v と被験者 6 が他に比べ良い結果になった。

(d) 意図通りの調整ができたか

システムが実際に行った問合せベクトル変更が、被験者が実験手順 (2-a) で記述した調整意図 (問合せベクトルの各成分に対する重要度上げ / 下げ / 無関係) に沿っていたかどうかを調べるため、以下で定義する値 F, A を考える。被験者 1 人が 5 つの課題に対して行う全調整操作において、

$F(\text{Follow}) : P_F(i, j)$ が真となる組 (i, j) の総数

$A(\text{Against}) : P_A(i, j)$ が真となる組 (i, j) の総数

と定義する。ここで $P_F(i, j)$ は、被験者が第 i 回目の操作において、実験手順 (2-a) で問合せベクトルの第 j 成分の値を大きく (小さく) したいと回答し、かつ $k_j^{(i)} > k_j^{(i-1)}$ ($k_j^{(i)} < k_j^{(i-1)}$) であった場合に真となる述語である。すなわち、 F とは、1 人の被験者が行う調整操作全体を通じて、被験者の意図通りに問合せベクトルが変更された回数を意味する。一方 $P_A(i, j)$ は、被験者が第 i 回目の操作において、実験手順 (2-a) でクエリーベクトルの第 j 成分の値を大きく (小さく) したいと回答し、かつ $k_j^{(i)} \leq k_j^{(i-1)}$ ($k_j^{(i)} \geq k_j^{(i-1)}$) であった場合に真となる述語である。すなわち、 A とは、被験者の意図と逆方向に問

表 4.3 意図通りの調整ができたか

	F	A*
A	30	17
B	24	11
C	16	46(17)
課題 i	14	9
課題 ii	9	16(10)
課題 iii	7	22(7)
課題 iv	23	12(10)
課題 v	17	15(9)
被験者 1	19	10(9)
被験者 2	8	13(8)
被験者 3	6	3
被験者 4	8	27(8)
被験者 5	14	16(14)
被験者 6	15	5(3)
全体	70	74(45)

* カッコ内は、システム C で「解なし」になった場合を除いた値。

合せベクトルが変更された回数である。表 4.3 に各システム・課題・被験者ごとの F 、 A の値を示す。

表 4.3 より、(b) または (c) の結果において平均値が 1.0 よりも有意に高かった課題 (i, iv, v) や被験者 (1, 5, 6) は F の値が大きい課題・被験者であったことがわかり、問合せベクトルをユーザの調整意図に沿って適切に変更することができれば、適合文書を選択的に上位に移動することができると考えられる。

一方、システム C は (b) および (c) の結果において平均値が 1.0 よりも有意に高い結果が得られたが、 F の値はシステム A, B に比べて小さく、逆に A の値が大きい。手法 C (順序情報に基づく推測手法) では、その定義より、連立不等式の解が存在しない場合が生じる。4.3.3 節で述べたように、今回の実験システムでは、 i 回目の調整で「解なし」となった場合は $k^{(i)} = k^{(i-1)}$ とした。このとき任意の j について $k_j^{(i)} = k_j^{(i-1)}$ であるため、 A の値が大きくなったと考えられる。「解なし」は今回の実験においておよそ 3 割弱の割合で生じており、前回と同じ問合せベクトルを用いる方法では問題があると思われたため、4.5 節に述べる性能評価実験では「解なし」の場合でもある程度調整意図に沿った変更ができるよう改良を行った (4.5.1 節)。

(e) アンケート結果より

システムの使いやすさを1~10の10段階で評価してもらった(評価値が大きいほど被験者が「使いやすい」と感じたことを意味する)。その結果、全被験者の解答の平均値は6.6であった。また、システムA,B,Cの違いについて尋ねたところ、6人中3人が「特に違いを感じなかった」と解答した。これは、システムの違いがあまり顕著ではないと同時に、システムによって検索した課題が異なるために比較が難しかったためと考えられる。しかし残り3人からは「Aはランキングが少ししか変化しなかった」「Bは変化しすぎて使いにくかった」「Cは全く変化しない事があって不便」という解答が得られた。この指摘をもとに、4.5節に述べる性能評価実験ではそれぞれの推測手法の改良を目指したパラメータ調整を行った(4.5.1節)。

上記(a)–(d)から以下の考察が得られる

- 本手法を用いることによって下位に埋もれた適合文書を選択的に上位に持ち上げることが可能であり、未知の適合文書を発見しやすくなるという点でユーザにとって有効であると考えられる。
- 全体として適合文書の順位を上げ、不適合文書の順位を下げているため、問合せベクトルを適切に変更していると考えられる。
- ユーザの調整意図をうまく推測し問合せベクトルに反映できた場合は良い結果が得られる。しかし調整意図をうまく推測できない場合が半分程度生じており、各手法の改良が必要である。

4.5 性能評価実験

本節で述べる性能評価実験では、被験者の代わりにある調整戦略に従ってランキングの調整を行い、各種適合率を求める。同時に、各問合せベクトル推測手法A,B,Cの性能を、4.4節の実験結果をもとにそれぞれ改良・調整を行った上でより詳細に比較することが目的である。実験結果の評価は4.4節と同じく、(a) 20位以内適合率・(b) 新規20位以内適合率・(c) *up* および *down* 適合率によって行い、調整によってランキングを適切に変更できたか、4.4節の実験結果と違いはあるか、より良いシステム(推測手法)とはどのようなものか、などに注目して考察を行う。

4.5.1 システムの概要

4.4節の実験結果から得られた指針をもとに、問合せベクトルの推測手法A,B,Cをそれぞれ以下のように改良・調整し、実験に用いた。

システム A：成分補正法 事後アンケートで「(調整しても)変化が小さい」という指摘があった。この原因として、ベクトル md_i を作成する際の補正が効きすぎている(補正值 η_j が小さすぎる)のではないかと考え、 $\eta_j \in \{0, 0.5, 1.0\}$ を以下の3通りに変更しそれぞれシステム A1, A2, A3 とした：

- システム A1： $\eta_j \in \{0.1, 0.5, 1.0\}$
- システム A2： $\eta_j \in \{0.25, 0.75, 1.0\}$
- システム A3： $\eta_j \in \{0.5, 0.75, 1.0\}$

システム B：E 尺度に基づく推測手法 事後アンケートで「変化が大きすぎる」という指摘があった。この原因として、他の2手法では新しい問合せベクトル $k^{(i)}$ を求める計算中に前回の問合せベクトル $k^{(i-1)}$ が用いられているのに対し、E 尺度に基づく手法では $k^{(i-1)}$ を全く使用しないという事が考えられる。そこで、4.3.2 節の式 (4.9) によって k を求めた後に前回の問合せベクトル $k^{(i-1)}$ との重み付き平均をとり、これを $k^{(i)}$ とする改良を行った。すなわち

$$k^{(i)} = \xi k + (1 - \xi) k^{(i-1)} \quad (4.13)$$

とし、 ξ を 0.25, 0.5, 0.75, 1.0 の4通りに設定してそれぞれシステム B1, B2, B3, B4 とした。システム B4 は $k^{(i-1)}$ に対する重みが 0 になるため、従来のシステム B と同等である。

システム C：順序情報に基づく推測手法 「解なし」が高頻度に発生し、事後アンケートで「全く変化しないのは不便」という指摘があったため、「解なし」の場合に以下の方法で新しい問合せベクトルを作成するよう改良を行った。

- (1) 4.3.3 節の式 (4.11) の代わりに次式 (4.14) を用いて再度計算を行う。ただし d_{lows} は $d_l, d_{l+1}, \dots, d_{h-1}$ の平均ベクトル (4.3.1 節)。

$$\text{sim}(d_h, k) \geq \text{sim}(d_{lows}, k) \quad (4.14)$$

- (2) 上記手順 (1) でも「解なし」になった場合、E 尺度に基づく手法 (4.3.2 節) で定義したベクトル $g(\text{good})$ (d_h と d_{l-1} の平均ベクトル) を求め、新しい問合せベクトル $k^{(i)}$ とする。これは、システム C で前回と異なる問合せベクトルを作成できない場合、システム B を簡略化して用いることを意味する。

上位固定と上位変化 事後アンケートにて「調整の後、前のランキングで上位にあった文書を見失ってしまい、探すのが大変だった」という指摘があった。このことから、ユーザが文書 d_h を文書 d_l ($1 \leq l < h \leq |D|$) のすぐ上に移動する調整操作を行った場合、文書 d_l から d_{l-1} は調整の前後で同順位 (調整によって順位が

変化しない)とした方が利便性が高いと考えられる。そこで、先に述べた 8 システム (A1-A3, B1-B4, C) 全てに対し、文書 d_1 から d_{l-1} まではスコアの再計算を行わず前回と同じ順位で出力するもの (上位固定版) を用意し、上位も変化する従来のシステム (上位変化版) と比較を行った⁶。

性能評価実験は以上の 16 システム (上位固定版 A1 ~ C, 上位変化版 A1 ~ C) に対して行った。

4.5.2 実験環境

検索課題と文書集合 検索課題には、4.4 節の被験者実験と共通する 4 課題 (表 4.1 の課題 i,ii,iv,v) を用いた⁷。各課題に対する文書集合と適合文書集合は 4.4 節の実験と同じである。ただし、精度をより厳密に議論するため、課題ごとの適合文書 () を更に“優”と“良”に分類した。ここで優文書とは、検索課題に対して非常によく適合している文書であり、被験者の違いによって適合・不適合の判断が左右されないと考えられるものである。後の分析では D_{rel} の代わりに優文書の集合 $D_{優}$ を用いて各種適合率を求めた。

手順 この実験では被験者を用いず、各課題ごとに以下の作業を機械的に行った。

- (1) 現在のランキングを上位から順にチェックし、不適合文書 (×) が 2 つ以上連続している箇所を見つける。
- (2) その下に現れる最初の適合文書 d_h () を、連続した×の上に持ち上げる。

ユーザが調整を行いたいと感じるのは、システムの考える適合基準と自分自身の適合基準とが明らかに異なると感じる時であると考え、そのような状況を「不適合文書の連続」と定義した。

測定値 測定値は 4.4 節の実験と同様に、20 位以内・新規 20 位以内・*up*・*down* の 4 種類の適合率である。上位固定版のシステムにおいては、調整によってスコアを再計算した文書のみを対象として各適合率を求めた。すなわち、 i 回目の調整で文書 d_h を文書 d_l のすぐ上に移動した場合、1 位から $l-1$ 位までの文書についてはスコア再計算を行わないため調整による変化が生じない。よって新しいランキングでは第 l 位を 1 位と考え、20 以内文書集合 $D_{(\leq 20)}^i$ と新規 20 位以内文書集合 D_{new}^i の定義を以下のように変更した。

$$D_{(\leq 20)}^i = \{d : Rank(d, i) \leq l + 19\}$$
$$D_{new}^i = \{d : Rank(d, i-1) > l + 19 \text{ かつ } Rank(d, i) \leq l + 19\}$$

⁶上位固定版と上位変化版は、4.2.1 節手順 (4) までは完全に同じ動作を行う。

⁷課題 iii には優文書がなかったため、性能評価実験では除外した。

表 4.4 性能評価実験の結果

	(a) $\frac{P_{(\leq 20)}^i}{P_{(\leq 20)}^0}$		(b) $\frac{P_{\text{new}}^i}{P}$		(c) $\frac{P_{\text{up}}^i}{P_{\text{down}}^i}$	
	平均	> 1 *	平均	> 1 *	平均	> 1 *
固定	0.63	N(1%)	1.23	Y(5%)	1.54	Y(1%)
変化	1.18	Y(1%)	1.12		1.35	Y(1%)
A1	0.99		1.40	Y(1%)	1.63	Y(5%)
A2	0.87	N(5%)	1.30		1.14	
A3	0.93		1.32	Y(5%)	1.10	
B1	0.88	N(1%)	1.00		1.35	Y(5%)
B2	0.87	N(1%)	1.16		1.68	Y(1%)
B3	0.92		1.12		1.37	
B4	0.91		0.76		1.62	Y(1%)
C	0.89		1.31		1.59	Y(1%)
全体	0.91	N(1%)	1.17	Y(1%)	1.44	Y(1%)

* 平均値が 1.0 よりも有意に高い場合を Y, 低い場合を N, 有意差が認められない場合を空欄で示す。かっこ内は有意水準。

4.5.3 結果と考察

(a) 20 位以内適合率

各 i 回目調整後の 20 位以内適合率と初期ランキングにおける 20 位以内適合率との比率 $P_{(\leq 20)}^i / P_{(\leq 20)}^0$ を全課題について求め、平均値が 1.0 より高いかどうかの検定を行った (表 4.4 (a))。

全体の結果 (表の最下行) を見ると、 $P_{(\leq 20)}^i / P_{(\leq 20)}^0$ の平均値は 1.0 よりも有意に低い値となっており、初期ランキングよりも 20 位以内適合率は悪くなっている。上位固定版と上位変化版の間には有意差が認められ (1% 有意), 上位変化の方が上位固定に比べ平均値が高く、また 1.0 よりも有意に高い値となった。A1~C の 8 システム間に有意差は認められず、また平均値が 1.0 よりも高いシステムはなかった。

(b) 新規 20 位以内適合率

各 i 回目調整後の新規 20 位以内適合率と全体適合率との比率 P_{new}^i / P を全課題について求め、平均値が 1.0 より高いかどうかの検定を行った (表 4.4 (b))。

全体の結果を見ると、 P_{new}^i / P の平均値は有意水準 1% で 1.0 よりも高い。つまり、調整によって下位に埋もれていた適合文書を選択的に上位に移動することができたと言

える．上位固定と上位変化の間には有意差は認められなかったが，上位固定のみ平均値が 1.0 よりも有意に高かった．システム間にも有意差は認められなかったが，システム A1, A3 は平均値が 1.0 よりも有意に高く，その他の 6 システムについては 1.0 よりも高いとは言えなかった．

(c) *up* 適合率 vs. *down* 適合率

各 i 回目調整後の *up* 適合率と *down* 適合率との比率 P_{up}^i / P_{down}^i を全課題について求め，平均値が 1.0 より高いかどうかの検定を行った（表 4.4 (c)）．

全体の結果を見ると， P_{up}^i / P_{down}^i の平均値は有意水準 1% で 1.0 よりも高い値になっており，調整によって起こったランキング（問合せベクトル）の変化は適切であったと言える．上位固定と上位変化の間に有意差は認められず，どちらの結果も平均値は 1.0 よりも有意に高かった．システム間にも有意差は認められなかったが，システム A1, B1, B2, B4, C について平均値が 1.0 よりも有意に高いと言えた．

上記 (a)–(c) から以下の考察が得られる

- 全体の傾向として，調整によって上位 20 文書の適合率を上げることは出来ないが，下位に埋もれた適合文書を選択的に上位に持ち上げることが可能であり，未知の適合文書を上位に発見できるという点で有効な手法と言える．また，調整によって起こった問合せベクトルの変化は適切であることが確認できた．
- 上位固定と上位変化を比較すると，上位変化の方が上位固定を少々上回る結果となった．ただしその差はあまり顕著ではなく，またユーザの感じる使いやすさや負担の少なさという観点からの評価も必要であり，双方を使い分けるインタフェースなども考慮すべきと思われる．
- システム間の比較については，今回の実験では (a), (b), (c) いずれの結果においても有意差は認められなかった．ただし問合せベクトル変更の適切さを表す (c) の結果においてシステム B2, A1, B3, C の順に平均値が高いことから，システム A, B, C で比較すると $B > A > C$ という傾向が窺える．パラメータの違いによる比較では，システム A の 3 種類中では A1，システム B の 4 種類中では B2 が最も良い結果になった．
- 10 回の調整を比較すると，全てのデータに共通して，1, 2 回目の調整後に最大値となる傾向が見られた．このことから，調整によるランキング改良は高々 2 回程度の操作で最も効率よく行えることがわかり，ユーザにとって無理のない方法と考えられる．

4.6 まとめ

本章では，WWW 検索作業におけるユーザの心理的負担の軽減を目的に，文書順位の上げ / 下げ操作によってユーザが対話的にランキングを調整できる手法を提案し，実験によってその有効性を検証した．

被験者を用いた実験の結果から，本手法によって上位文書の適合率を上げることは出来ないが，下位に埋もれた適合文書を選択的に上位に持ち上げることが可能であり，未知の適合文書を発見しやすいという点で有効な手法であると確認できた．また，全体として適合文書の順位を上げ，不適合文書の順位を下げており，提案した推測手法によって問合せベクトルを適切に変更できたと考えられるが，特にユーザの調整意図をうまく推測できた場合にはその効果が大きく，推測手法の精度向上が重要であることがわかった．

性能評価実験では被験者を用いた実験から得られた知見をもとに3推測手法それぞれを改良し，より厳密に比較評価を行った．その結果，被験者を用いた実験と同じ傾向として，下位に埋もれた適合文書を選択的に上位に持ち上げていること，全体として問合せベクトルを適切に変更できていることを確認した．また，調整によるランキング改良手法は高々2回程度の操作で最も効率よく行えることがわかり，被験者実験でのアンケート結果からも，本手法がユーザにとって無理のない方法であることが確認できた．手法間の比較では，あまり顕著な差はないがシステム B2, A1, B4, C が優れていることが分かった⁸．一方，調整操作が行われた部分よりも上位の文書の順位を固定する方法は，従来の（固定しない）方法に比べて良いとは言えない結果であったが，今後被験者による評価によって再検討する必要があると感じられた．

今後，ユーザの調整意図をより正確に推測するための改良を続けると同時に，被験者を用いた更なる評価実験を行う予定である．特に，従来の適合フィードバック手法や統計的分類手法との比較を，性能・使いやすさの両面から行いたいと考えている．

ユーザの調整意図を正確に推測するための改良としては，縮約特徴ベクトルの成分の追加を検討している．現在はユーザの入力したキーワード，リンク数，更新の新しさをを用いているが，今後はキーワードの類義語やキーワードとの共起性の高い語なども加えることで推測精度の向上を目指す予定である．また各成分の重み間には現在特に制限を設けていないが，リンク数に対する重みはキーワードに対する重みの最大値を超えない範囲にする，などの補正についても検討したい．またインタフェース面の改良として，4.4節の事後アンケートにおいて「問合せベクトルの各成分値を直接上げ / 下げ指定する機能も欲しい」という解答が6人中4人から得られた．問合せベクトルの成分を直接操作すると，ランキングがその成分に鋭敏になりすぎるという問題が予想されるが，今後実用的な手法として実装できるよう検討していく予定である．

⁸ただし，改良後のシステム C は「解なし」の場合にシステム B の手法を簡略化して実行しているため，純粋に A, B との比較はできない．

第5章 おわりに

本研究では，急速に変貌するインターネット上の WWW サービスにおいて重要な課題である使いやすい情報検索システムの構築を目指し，既存の検索手法が抱える問題の解消を図る 2 種類の検索支援手法を提案した．本章では，本論文の総括として，提案手法および試作システムによる評価実験を通じての考察，今後の課題，そして本研究と従来の関連研究の今後の展望について述べる．

5.1 まとめと考察

現在の WWW 検索サービスにおける標準的な検索手法であるキーワード検索およびディレクトリ探索の両手法は，その使いやすさに関して未だ問題があり，様々な方法による検索支援が検討されている．それらの関連研究の中での本研究の特色は，従来の研究が主に検索の効率や精度向上を目的としているのに対し，特に専門的な知識をあまり持たない一般的なユーザの情報検索におけるメンタルモデルに着目し，それに沿った無理のない方法による検索支援を目指していることである．本研究の提案手法は，場合によっては従来の検索手法よりも作業数や作業時間を増すことも考えられるが，ユーザにとって容易でわかりやすい作業のみで検索作業を行うことができる，心理的な負担の少ない検索方法となることを期待している．

本研究ではまず，キーワード検索およびディレクトリ探索の両検索手法についてその利点と欠点を考察し，双方の抱える問題を解消することを目指した検索支援手法（可変なカテゴリ構造を用いた検索支援手法：CATE 手法）を提案した．CATE 手法については 2.1 節でその目的と概要，関連研究との比較を述べ，第 3 章で手法の詳細と評価実験の結果について論じた．BMIR-J2 テストコレクションを用いて行った評価実験では，従来の検索手法であるキーワード検索手法および統計的分類手法（クラスタリング）と比較して，CATE 手法は特に難しい検索課題に対して精度の良い結果を提供できることが確認できた．また使いやすさに関する被験者の主観的評価は，クラスタリング手法に大きく勝り，キーワード検索手法にやや劣る程度であった．しかし，分類の結果を評価するために考案した 2 種類の基準（明確さとエントロピー）について，今回実施した実験の結果からは明確な結論を得られず，更なる評価，および双方を使い分けるインタフェースについても検討する必要性が示唆された．

次に，サービスの種類を問わず，現代の WWW における標準的な検索機能として広く普及しているキーワード検索手法に焦点をあて，その改善を目指した検索支援手法（調整可能な文書ランキングを用いた検索支援手法：UADR 手法）を提案した．本手法については 2.2 節でその目的と概要，関連研究との比較を述べ，第 4 章で手法の詳細と評価実験の結果について論じた．既存のキーワード検索システムに UADR 手法を適用してランキングの改良効果を調べた評価実験では，本手法を用いてランキングの調整を行うことで適合文書を選択的に上位に持ち上げていることが確認できた．これにより，下位に埋もれた未知の適合文書を発見し易くするという点で有効な手法であると考えられる．また，ユーザの調整操作を新しいランキングに反映させるアルゴリズムを 3 種類考案し，問合せベクトル変更の適切さという点から比較評価を行った．その結果，被験者を用いた評価実験では順序情報に基づく推測手法（手法 C），性能評価実験では E 尺度に基づく推測手法（手法 B）が他の 2 手法に比べ優れているという結論を得た．しかし手法間の差はあまり顕著ではなく，また被験者を用いた評価実験の結果を基に手法の改良を試みた後もユーザの調整意図をうまく推測できない場合が依然として生じていることから，いずれの手法もよりユーザの視点に近付くための更なる改良が求められる．加えて，従来の適合フィードバック手法や統計的分類手法と UADR 手法との比較についても，性能・使いやすさの両面から実験評価するべきと感じられた．

5.2 今後の課題

本研究で提案した二つの検索支援手法について，今後に残された課題を述べる．

実施した評価実験の結果からは，それぞれの支援手法についてその有効性を確認できた一方で，どちらの手法についても今後更なる改良の余地があることがわかった．数値的なデータの傾向からだけでなく，被験者による実験後アンケートの解答などからも様々な知見が得られ，これらをもとに，それぞれの手法を以下のような面について再検討し改善を行うことが今後の重要な課題である．

(1) 計算アルゴリズムやシステム設計の改良

- CATE 手法：分類の結果を評価するための 2 基準（明確さ，エントロピー）それぞれの改良
- CATE 手法：分類観点集合の改良および，分類観点集合の作成方法の開発
- UADR 手法：新しい問合せベクトルを推測するための 3 手法（成分補整法， E 尺度に基づく推測手法，順序情報に基づく推測手法）それぞれの改良

(2) 上記計算アルゴリズムのより詳細な比較

(3) 上記計算アルゴリズムの使い分けに関する指針の獲得

(4) インタフェース面の改良

- CATE 手法：構築したカテゴリ階層構造のわかりやすい表示方法
- UADE 手法：問合せベクトルの成分値を直接上げ / 下げ指定できる機能
- UADR 手法：調整によって生じた変化がどのようなものかをユーザにわかりやすく提示するための方法

また手法の改良とは別に（あるいは改良の後に）、両提案手法の使いやすさに関して、よりユーザの視点に立った具体的な評価を行うことが必要だと考えている。提案手法と従来の検索支援手法との比較（CATE 手法とクラスタリング手法、UADR 手法と適合フィードバック手法など）においても、実際に運用することを念頭におくと、ユーザにとっての習熟しやすさ・間違いの少なさ・使用感のおもしろさ、などの尺度からの評価が重要である。さらに、検索作業を通じての心理的負担を軽減することができたかどうかについても、本研究の評価実験で実施したようなアンケート形式だけではなく、定量的な評価も試みたい。

今後の展望として、まず第一に、CATE 手法と UADR 手法は互いに独立した検索支援手法であるため、両手法の統合が可能である。最も簡単な統合例としては、CATE システムにおいてユーザが選択したカテゴリに属する文書を表示する際、従来の固定的なランキング（検索システム *Freya* が出力するランキング）ではなく UADR 手法を適用して調整可能なランキングを提供するシステムが考えられる。更に、CATE システムにおいてシステムが提示する分類観点のランキングを、UADR 手法を用いてユーザが調整し、明確さやエントロピーなどの分類観点に対する評価値（分類観点に対するスコア）を検索目的に沿うよう改良できるようにする、というような応用も考えられる。

さらに、1.3 節で述べた種々の検索支援手法と CATE 手法・UADR 手法との統合も可能である。例えば、CATE 手法によって構築するカテゴリ階層構造を情報視覚化技術を用いてわかりやすく表示するなど、それぞれの手法の改良としての統合はもちろん、ユーザにとってより使いやすく性能の良い検索システムの構築を目標に様々な可能性を考えることが出来る。例えば、同一の検索システムにおいて、初期ランキングを適合フィードバックと UADR 手法のどちらの方法でも調整できるようなインターフェースを備えること、UADR 手法や適合フィードバック手法における問合せベクトル再計算の際に文書タイプによる分類技術や 2 次キーワード抽出技術を応用することなど、ユーザに対して様々な機能を統合して提供する方向での各手法の応用が期待される。

謝辞

奈良先端科学技術大学院大学情報科学研究科 情報基礎学講座にて本研究を行う機会を与えていただいた関浩之教授に厚く御礼申し上げます。主指導教官として研究の全過程において直接細やかな御指導御鞭撻を賜ったばかりでなく、情報科学の基盤なく入学した私に広い視野から様々なことを御教授いただきました。どんな小さな思い付きや疑問にも常に真摯な態度で接していただき、納得ゆくまで議論につきあってくださいました。心から感謝しております。

学内や学外での発表を通じて、終始有益な御助言や的確な御指摘をいただいた同研究科 マルチメディア統合システム講座の植村俊亮教授に深謝いたします。

同研究科 論理生命学講座の石井信教授には、研究の初期段階から有意義な御意見や御助言をいただき、また4章で述べた提案手法の改良・拡張に関して、御多忙な中お時間を割いて様々な可能性を示唆していただきました。深く御礼申し上げます。

同研究科 自然言語処理学講座の松本裕治教授には修士論文の副指導教官を引き受けていただき、ゼミナールなどの機会に鋭い御指摘、御指導を賜りました。深く御礼申し上げます。

情報科学の基礎勉強に端を発し、本研究の発足から論文の執筆に至るまでの全過程において終始熱心な御指導と御協力をいただいた同研究科 情報基礎学講座の高田喜朗助手に心から感謝いたします。研究内容のみならず、研究生活全般にわたって数限りない御助言と御指導をいただき、特に自分の中のアイデアを他人に伝えることの大切さと難しさを学ばせていただきました。

本研究は、常に暖かい御支援をいただき、熱心な議論や様々な御意見・御忠告、そして評価実験においても快い御協力をいただいた情報基礎学講座の皆様方の御力添えにより遂行することができました。特に4章の提案手法については岩崎正秀氏、5章の提案手法については木下敦史氏とともに研究を進めさせていただきました。直接的、間接的両面において様々な支援していただき、研究に関してもそれ以外に関しても実りの多い5年間を過ごさせていただきましたこと、この場を借りて深く御礼申し上げます。本当にありがとうございました。

最後になりましたが、試作システムの実装には、日本語形態素解析『茶釜』のライブラリ、検索システム『Freya』のクラス群を利用させていただきました。これらの優れたフリーソフトウェアを公開してくださっている作者の方々へ深く敬意を表します。

参考文献

- [1] Lawrence, S. and Giles, C. L.: Accessibility of Information on the Web, *Nature*, Vol. 400, pp. 107–109 (1999).
- [2] Lawrence, S. and Giles, C. L.: Searching the World Wide Web, *Science*, Vol. 280, No. 5360, pp. 98–100 (1998).
<http://www.sciencemag.org/cgi/content/full/280/5360/98?>
- [3] Dreilinger, D. and Howe, A. E.: Experiences with Selecting Search Engines Using Metasearch, *ACM Trans. Information Systems*, Vol. 15, No. 3, pp. 195–222 (1997).
- [4] 河野浩之, 長谷川利治: WWW データ資源に対する重み付き相関ルール導出アルゴリズムの適用, 重点領域研究「高度データベース」松江ワークショップ 講演論文集, Vol.1, pp. 90–99 (1996).
- [5] Anick, P. G. and Tipirneni, S.: The Paraphrase Search Assistant: Terminological Feedback for Iterative Information Seeking, *Proc. SIGIR 99*, pp. 153–159 (1999).
- [6] Bruza, P., McArthur, R. and Dennis, S.: Interactive Internet Search: Keyword, Directory and Query Reformulation Mechanisms Compared, *Proc. SIGIR 2000*, pp. 280–287 (2000).
- [7] Pirolli, P., Shank, P., Hearst, M. and Diehl, C.: Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection, *Proc. CHI 96*, pp. 213–220 (1996).
- [8] Hearst, M. and Pedersen, J. O.: Re-examining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, *Proc. SIGIR 96*, pp. 76–84 (1996).
- [9] Pollitt, S. A.: The Key Role of Classification and Indexing in View-based Searching, *Proc. 63rd IFLA General Conf.* (1997).
<http://www.ifla.org/IV/ifla63/63cp.htm>.
- [10] Schutze, H., Hull, D. A. and Pedersen, J. O.: A Comparison of Classifiers and Document Representations for the Routing Problem, *Proc. SIGIR '95*, pp. 229–237 (1995).
- [11] 村本達也, 鷲崎誠司: 階層型知識体系を用いた WWW 情報の自動カテゴリ推定方法, 情報処理学会第 56 回全国大会講演論文集, pp. 3-205–3-206 (1998).
- [12] 伊藤篤, Anwyl, P., 大野亜矢子: WWW の分類・検索システム Crowww, 情報処理学会第 56 回全国大会講演論文集, pp. 3-221–3-222 (1998).
<http://crowww.ricoh.co.jp/search/>.

- [13] Rocchio Jr., J.: Relevance Feedback in Information Retrieval, *The SMART Retrieval System : Experiments in Automatic Document Processing* (Salton, G.(ed.)), Prentice Hall Inc., pp. 313–323 (1971).
- [14] 江口浩二, 伊藤秀隆, 隈元昭: ユーザへの適応性を考慮した適合フィードバックによる WWW 情報検索, 電気学会論文誌電子・情報・システム部門誌, Vol. 117-C, No. 11, pp. 1643–1649 (1997).
- [15] 江口浩二, 伊藤秀隆, 隈元昭, 金田彌吉: 漸次的に拡張されたクエリを用いた適応的文書クラスタリング法, 電子情報通信学会論文誌, Vol. J82-D-I, No. 1, pp. 140–149 (1999).
- [16] Golovchinsky, G.: Queries? Links? Is There a Difference?, *Proc. CHI 97*, pp. 407–414 (1997).
- [17] Glover, E. J., Flake, G. W., Lawrence, S., Birmingham, W. P., Kruger, A., Giles, C. L. and Pennock, D. M.: Improving Category Specific Web Search by Learning Query Modifications, *Symposium on Applications and the Internet (SAINT2001)*, San Diego, CA, pp. 23–31 (2001).
- [18] Witten, I. H., Moffat, A. and Bell, T. C.: *Managing Gigabytes: Compressing and Indexing Documents and Images*, Von Nostrand Reinhold, New York (1994).
- [19] Page, L., Brin, S., Motwani, R. and Winograd, T.: The PageRank Citation Ranking: Bringing Order to the Web (1998).
<http://citeseer.nj.nec.com/345917.html>.
- [20] Lokuge, I., Gilbert, S. A. and Richards, W.: Structuring Information With Mental Models: A Tour of Boston, *Proc. CHI 96*, pp. 413–419 (1996).
- [21] Fishkin, K. and Stone, M. C.: Enhanced Dynamic Queries via Movable Filters, *Proc. CHI '95*, pp. 415–420 (1995).
- [22] 早川和宏, 福永博信, 鈴木達郎: ユーザの利用履歴に基づく WWW サーバの地図型ディレクトリ, 情報処理学会研究報告, 97-HI-70, pp. 17–24 (1997).
- [23] Lamping, J., Rao, R. and Pirolli, P.: A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies, *Proc. CHI '95 Conf. Human Factors in Computing Systems*, pp. 401–408 (1995).
<http://www.inxight.com/>.
- [24] 塩澤秀和, 相馬隆宏, 野田純也, 松下温: 切り取り操作による柔軟な情報選択ができる WWW 視覚化, 情報処理学会研究報告, 97-HI-72, pp. 61–66 (1997).

- [25] Sebrechts, M. M., Vasilakis, J., Miller, M. S., Cugini, J. V. and Laskowski, S. J.: Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces, *Proc. SIGIR 99*, pp. 3–10 (1999).
- [26] Teraoka, T. and Maruyama, M.: Adaptive Information Visualization Based on the User's Multiple Viewpoints - Interactive 3D Visualization of the WWW, *Proc. IEEE Symposium on Information Visualization (Info Vis 97)*, pp. 25–28 (1997).
- [27] 松田勝志, 福島俊一: 文書タイプ分類による問題解決向き WWW 検索システムの開発と評価, 情報処理学会研究報告, 99-FI53-2, pp. 9–16 (1999).
- [28] van Rijsbergen, C.: *Information Retrieval*, Butterworths, 2nd edition (1979).
- [29] Takata, Y., Nakagawa, K. and Seki, H.: Flexible Category Structure for Supporting WWW Retrieval, *Proc. ER2000 Conference Workshop on the World Wide Web and Conceptual Modeling (WCM2000)*, LNCS1921, pp. 165–177 (2000).
- [30] 仲川こころ, 高田喜朗, 関浩之: 可変なカテゴリ構造を用いた文書検索支援手法, 情報処理学会論文誌, Vol. 42, No. 10, pp. 2441–2453 (2001).
- [31] 仲川こころ, 高田喜朗, 関浩之: 検索目的を反映したカテゴリ構造に基づく WWW 検索支援, 情報処理学会研究報告, 99-HI-82, pp. 59–64 (1999).
- [32] 木下敦史, 仲川こころ, 高田喜朗, 関浩之: 対話的に文書ランキングを調整できる WWW 検索支援手法, 情報処理学会研究報告, 2001-FI-64, pp. 63–70 (2001).
- [33] 岩山真, 徳永健伸: 確率的クラスタリングを用いた文書連想検索, 自然言語処理, Vol. 5, No. 1, pp. 101–118 (1998).
- [34] Salton, G., Singhal, A., Buckley, C. and Mitra, M.: Automatic Text Decomposition Using Text Segments and Text Themes, *Proc. Hypertext 96*, pp. 53–65 (1996).
- [35] 原田昌紀: *Freya version 0.92* (1998).
<http://odin.ingrid.org/freya/>.
- [36] 松本裕治, 北内啓, 山下達雄, 平野善隆, 今一修, 今村友明: 日本語形態素解析システム『茶筌』version 1.5 使用説明書, 技術報告 NAIST-IS-TR97007, 奈良先端科学技術大学院大学 (1997).
- [37] Tou, J. T. and Gonzalez, R. C.: *Pattern Recognition Principles*, Addison-Wesley, pp. 89–97 (1974).
- [38] 木谷強ほか: 日本語情報検索システム評価用テストコレクション BMIR-J2, 情報処理学会研究報告, 98-DBS-114, pp. 15–22 (1998).

- [39] Voorhees, E. M. and Harman, D. K.: Evaluation Techniques and Measures, *The Seventh Text REtrieval Conference (TREC 7)*, National Institute of Standards and Technology (NIST), pp. A-1 (1998).
- [40] Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T.: *Human-Computer Interaction*, Addison-Wesley (1994).
- [41] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P.: *Numerical Recipes in C*, Cambridge University Press, 2nd edition, pp. 430-444 (1992).
- [42] 坂野幸臣, 原田昌紀, 高野元, 鈴木基久: 検索エンジン解体新書, *INTERNET magazine*, No.79, pp. 222-235 (2001).
- [43] Bowman, C. M., Danzig, P. B., Hardy, D. R., Manber, U. and Schwartz, M. F.: The Harvest Information Discovery and Access System, *Computer Networks and ISDN Systems*, Vol. 28, pp. 119-125 (1995).
<http://harvest.transarc.com/>.
- [44] Mauldin, M. L. and Leavitt, J. R.: Web Agent Related Research at the Center for Machine Translation, *Proc. SIGNIDR '94* (1994).
<http://fuzine.mt.cs.cmu.edu/mlm/signidr94.html>.
- [45] Yamana, H., Tamura, K., Kawano, H., Kamei, S., Harada, M., Nishimura, H., Asai, I., Kusumoto, H., Shinoda, Y. and Muraoka, Y.: Experiments of Collecting WWW Information using Distributed WWW Robots, *Proc. SIGIR '98*, pp. 379-380 (1998).
- [46] 山名早人, 森英雄, 田村健人, 河野浩之, 村岡洋一: 分散型 WWW ロボットの予備評価と高速化の検討, 日本ソフトウェア科学会 The Third Workshop on Internet Technology (WIT2000) (2000).
<http://www.yama.info.waseda.ac.jp/publications/gifimage/papers3/Yamana-wit2000.pdf>.
- [47] 田原康之, 大須賀昭彦, 本位田真一: ビヘイビアパターンに基づくモバイルエージェントシステム開発手法, *情報処理学会論文誌*, Vol. 40, No. 12, pp. 4319-4332 (1999).

参照 URL

2002 年 2 月 14 日 確認

- 1) インターネット調査リンク :
<http://www.fish-u.ac.jp/~kajitori/ichosa.html>
- 2) Internet Domain Survey :
<http://www.isc.org/ds/>
- 3) Netcraft Web Server Survey :
<http://www.netcraft.co.uk/survey/>
- 4) サイバースペースジャパン 第 14 回 CSJ WWW 利用者調査 (2001 年 6 月) :
<http://www.csj.co.jp/www14/index.html>
- 5) 朝日新聞社総合研究センター 第 5 回新聞読者基本調査 (2000 年 2 月) :
<http://adv.asahi.com/2001/index.html>
- 6) GVU's Tenth WWW User Sruvey by GVU Center :
http://www.gvu.gatech.edu/user_surveys/survey-1998-10/
- 7) Alta Vista by Digital Equipment Corporation :
<http://altavista.digital.com/>
- 8) LYCOS by Lycos Inc :
<http://www.lycos.com/>
- 9) Google by Google :
<http://www.google.com/>
- 10) Yahoo! by Yahoo! Inc. :
<http://www.yahoo.com/>
- 11) iNET Guide by CyberSpaceJapan Inc. :
<http://www.inetg.com/>
- 12) SavvySearch by Savvy Search Limited. :
<http://www.savvysearch.com/>
- 13) ナレッジクエリーサーチ by Dataware Technologies Inc. :
<http://kd.iws.ne.jp/kms/>

- 14) BIGLOBE サーチ Attayo! by NEC Corporation :
<http://attayo.jp/>
- 15) BIGLOBE サーチ by NEC Corporation :
<http://search.biglobe.ne.jp/>
- 16) Goo by NTT-X :
<http://www.goo.ne.jp/>
- 17) FreshEye by FreshEye Corporation :
<http://www.fresheye.com/>
- 18) infoseek by Infoseek Japan K.K. :
<http://www.infoseek.co.jp>
- 19) Mobeet by 東芝研究開発センター , 国立情報学研究所 :
<http://mobeet.exrd.nacsis.ac.jp/>

研究業績

1. 学術論文誌

- 1-1 仲川こころ, 高田喜朗, 関浩之: 可変なカテゴリ構造を用いた文書検索支援手法, 情報処理学会論文誌, Vol.42, No.10, pp.2441–2453 (2001).
- 1-2 仲川こころ, 高田喜朗, 関浩之: 対話的に調整可能な文書ランキング — WWW 検索支援の一手法, 情報処理学会論文誌, 投稿中.

2. 国際会議論文集

- 2-1 Takata, Y., Nakagawa, K. and Seki, H.: Flexible Category Structure for Supporting WWW Retrieval, *Proceedings of ER2000 Conference Workshop on the World Wide Web and Conceptual Modeling (WCM2000)*, Lecture Notes in Computer Science 1921, pp.165–177 (2000).

3. 研究会およびシンポジウム発表

- 3-1 仲川こころ, 高田喜朗, 関浩之: 可変なカテゴリ構造を用いた WWW 検索支援方法の提案, 電子情報通信学会 第9回データ工学ワークショップ (DEWS'98), 論文番号 22 (1998).
- 3-2 仲川こころ, 高田喜朗, 関浩之: 検索目的を反映したカテゴリ構造に基づく WWW 検索支援, 情報処理学会研究報告, 99-HI-82, pp.59–64 (1999).
- 3-3 岩崎正秀, 仲川こころ, 高田喜朗, 関浩之: 可変なカテゴリ構造を用いた文書検索支援手法の実験的評価, 情報処理学会研究報告, 2000-DBS-120, pp.1–8 (2000).
- 3-4 木下敦史, 仲川こころ, 高田喜朗, 関浩之: 対話的に文書ランキングを調整できる WWW 検索支援手法, 情報処理学会研究報告, 2001-FI-64, pp.63–70 (2001).

4. 全国大会での口頭発表

- 4-1 仲川こころ, 高田喜朗, 関浩之: 可変なカテゴリ構造を用いた WWW 検索支援方法, 情報処理学会 第57回全国大会講演論文集, 分冊3, pp.3-153–3-154 (1998).

- 4-2 仲川こころ, 木下敦史, 高田喜朗, 関浩之: ユーザによる文書ランキングの調節が可能な対話的 WWW 検索支援手法の提案, 情報処理学会 第 63 回全国大会講演論文集, 分冊 3, pp.3-131-3-132 (2001).

5. テクニカルレポート

- 5-1 Nakagawa, K., Takata, Y. and Seki, H.: Flexible Category Structure for Supporting Document Retrieval and Its Evaluation, Technical Report NAIST-IS-TR 2000013, Nara Institute of Science and Technology (2000).

付録

A WWW 検索サービスの変遷と今後について

WWW やその利用状況の変化に伴い、検索サービスもインターネット普及の初期段階から現在に至るまでいくつかの変遷を遂げてきた。「人手による収集・分類」を第 1 世代、「ロボットによる自動収集」を第 2 世代とし、現在は「検索機能の分化・統合」と「文書のランク付け技術」を特徴とする第 3 世代とも言われており [42]、この二つの特徴は本研究で提案する検索支援手法の着眼点とも一致している。

「検索機能の分化・統合」に着目した例として、国内では、一般の WWW 文書・ディレクトリ内のカテゴリ・地図や音楽等の専門データベースの 3 種類の検索を融合し、様々な入力補助機能を備えた BIGLOBE サーチ¹⁵⁾ や、ロボットによって収集した文書を半自動半手でディレクトリに分類・登録し、キーワード入力型とディレクトリ探索型のどちらの方法を用いても同じデータベースを検索できる goo¹⁶⁾ (2001 年 7 月より実施) などが注目されている。「文書のランク付け技術」については、1.3 節 (D) で紹介した Google 社の PageRankTM 技術 [19] やアンカーテキスト解析技術、より新しい情報に高いランクをつける FreshEye¹⁷⁾、日本語の表記ゆれ(“問合せ”と“問い合わせ”、“スパゲッティ”と“スパゲティ”など)を類義語辞書によって吸収する infoseek¹⁸⁾ などが話題となっている。

また新しい話題として、分散型 Web ロボットにより、検索サービス毎に個々のロボットプログラムを作動させることによるネットワークや WWW サーバへの負荷を減らす試み (Harvest [43], WebAnts [44], 広域分散協調 Web ロボット [45,46]) や、ユーザの代理人として WWW 上の情報を収集し、次にユーザがインターネットに接続した際に専用サーバに保管しておいた検索結果をユーザの利用端末 (パソコン, 携帯電話, 情報端末など) に合わせた形で報告する検索エージェント Mobeet [47]¹⁹⁾ などがあり、次世代の検索技術となり得るかどうか注目されている。

B E 尺度に基づく推測手法の導出

4.3.2 節の式 (4.8) から式 (4.9) を導出する手順を説明する。

一般性を失うことなく、ベクトル g, b, k は正規化されているものと仮定する (すな

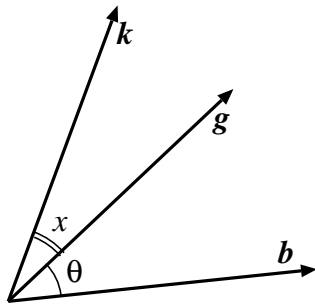


図 B.1 g, b, k の関係

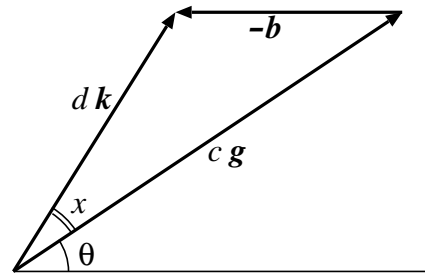


図 B.2 $dk = cg - b$

わち $|g| = |b| = |k| = 1$) . g と b のなす角を θ , g と k のなす角を x とすると , 式 (4.8) を満たす k が g と b が作る平面上に存在することは明らかである . 従って b と k のなす角は $x + \theta$ となる (図 B.1) .

式 (4.3), (4.7) より , 式 (4.8) の r は以下のような x の関数として表される .

$$f(x) = \frac{\cos x \sin(x + \theta)}{\cos x + \sin(x + \theta)} \quad (5.1)$$

(5.1) 式の $f(x)$ を最小にする x を求めるため $df/dx = 0$ とすると , 以下の関係が成り立つ .

$$(\cos(x + \theta) - \sin x)(1 + \sin x \cos(x + \theta)) = 0 \quad (5.2)$$

(5.2) 式を満たす解は以下の通りである .

$$\cos x = \sqrt{\frac{\sin \theta + 1}{2}} \quad \text{or} \quad x = \frac{1}{2}(\frac{\pi}{2} - \theta) \quad (5.3)$$

今 , ある正の実数 c と d を用いてベクトル k を式 (5.4) のように表すと , 式 (5.5), (5.6) が成り立つ (図 B.2) .

$$dk = cg - b \quad (5.4)$$

$$c \sin \theta = d \sin(x + \theta) \quad (5.5)$$

$$c \cos \theta = 1 + d \cos(x + \theta) \quad (5.6)$$

(5.3), (5.5), (5.6) 式より ,

$$c = \frac{1 + \sin \theta}{\cos \theta} . \quad (5.7)$$

E 尺度を用いて k の最適解を求める本手法は , 必ずしも文書 d_h をちょうど d_{l-1} と d_l の間に配置できるものではないが , 高々数回のベクトル正規化と内積だけの計算で済むため高速に実行できるという利点がある .

C CATEシステムに用意した分類観点の一覧

評価実験のために試作した CATE システムには、以下の 68 分類観点（延べカテゴリ数 1,723）を先験的知識として用意した。3 章 3.3.1 節システムの概要でも述べた通り、これらの分類観点集合は既存の大手ディレクトリサービス（2 サービス）が提供するカテゴリ階層構造を 1 階層ごとに上位 3 階層まで分解して作成したものである。作成には 6 時間程度を要した。以下に分類観点名とそこに属するカテゴリ名の一覧を挙げる。

分類観点名：属するカテゴリ名

都道府県：北海道，青森，岩手，宮城，秋田，山形，福島，茨城，栃木，群馬，埼玉，千葉，東京，神奈川，新潟，富山，石川，福井，山梨，長野，岐阜，静岡，愛知，三重，滋賀，京都，大阪，兵庫，奈良，和歌山，鳥取，島根，岡山，広島，山口，徳島，香川，愛媛，高知，福岡，佐賀，長崎，熊本，大分，宮崎，鹿児島，沖縄，日本，全国，地方，地域

日本の都市：札幌，仙台，千葉，川崎，横浜，名古屋，京都，大阪，神戸，広島，北九州，福岡

世界の地方：アジア，アフリカ，アメリカ，北アメリカ，南アメリカ，オセアニア，カリブ，アジア，ヨーロッパ，中近東スカンジナビア，地中海，北極，南極，欧州，北欧，南欧

世界の国々：アジア，インド，インドネシア，ベトナム，ヴェトナム，カンボジア，シンガポール，スリランカ，スリ・ランカ，タイ，韓国，中国，ネパール，パキスタン，バングラデシュ，フィリピン，ブルネイ，マレーシア，ミャンマー，モンゴル，ラオス，北米，アメリカ，カナダ，中南米，アルゼンチン，ベネズエラ，ヴェネズエラ，ウルグアイ，ウルグアイ，エクアドル，エル・サルバドル，キューバ，グアテマラ，コスタ・リカ，コスタリカ，コロンビア，ジャマイカ，チリ，ドミニカ，ニカラグア，ハイチ，パナマ，パラグアイ，パラグアイ，ブラジル，ペルー，ボリビア，ホンジュラス，メキシコ，アイルランド，イタリア，ウクライナ，ウズベキスタン，エストニア，オーストリア，オランダ，カザフスタン，ギリシャ，クロアチア，スイス，スウェーデン，スペイン，スロバキア，スロヴェニア，チェコ，デンマーク，ドイツ，ノルウェー，ハンガリー，フィンランド，フランス，ブルガリア，ベラルーシ，ベルギー，ポーランド，ポルトガル，ユーゴスラビア，ルーマニア，ルクセンブルグ，イギリス，英国，ロシア，オーストラリア，ニュージーランド，パプアニューギニア，パプア，ニューギニア，フィジー，マーシャル，ミクロネシア，中近東，アフガニスタン，アラブ，アラブ首長国連邦，イエメン，イスラエル，ジョルダン，シリア，トルコ，レバノン，イラク，イラン，オマーン，カタール，クウェート，クウェイト，サウジアラビア，アフリカ，アルジェリア，ウガンダ，エジプト，エチオピア，ガーナ，ガボン，カメルーン，ギニア，ケニア，ナイジェリア，ザンビア，ジブチ，ジンバブエ，スーダン，セネガル，コートジボワール，象牙海岸，タンザニア，チュニジア，コンゴ，ブルキナ・ファソ，ブルキナ，ファソ，ボツワナ，マダガスカル，マラウイ，南アフリカ，モーリタニア，モザンビーク，モロッコ，リビア，リベリア，ルワンダ

料理：お節，宴会，会席，懐石，割烹，精進，前菜，薬膳，朝，昼，夕，夜，給食，弁当，おやつ，ジャンクフード，肉，魚，野菜，卵，豆，飲，離乳食，幼児食，病人食，老人食，ダイエット，和，洋，中華，フランス，アジア

ジャンルNTT-Top：マルチメディア，インターネット，コンピュータ，ショッピング，旅行，レジャー，宿泊，アウトドア，教育，文化，メディア，テレビ，ラジオ，本，雑誌，グルメ，飲食店，映画，企業，求人，クイズ，ゲーム，スポーツ，野球，サッカー，競馬，趣味，生活，写真，車，バイク，ペット，占い，みんなの広場，日記，おしゃべり，サークル，医療，病院，プレゼント，携帯電話，PHS，音楽，芸能，アーティスト，俳優，女優，男優，新聞，ニュース，都道府県，地域，地方自治体，官公庁

マルチメディア：CD-ROM，サウンド，通信カラオケ，コンピュータ，CATV，CG，インターネット，画像，メーリングリスト，通信，LAN

インターネット：プロバイダ，ホームページ，素材，サーバ，インターネットカフェ，ディレクトリ，検索エンジン，リンク集，ドメイン，プラグイン，ブラウザ，WWW，HTML，Shockwave，JAVA，RealAudio，VRML

通信・LAN：携帯，PHS，ネットワーク，イントラネット，セキュリティ，ISDN，ATM，LAN，パソコン通信

コンピュータ：ハード，周辺機器，MODEM，DSU，プリンタ，ボード，カード，ディスプレイ，記憶装置，ソフト，OS，ツール，スクリーンセーバ，データベース，表計算，ワープロ

ショッピング：通信販売，チケット，輸入，食品，コンピュータ，ソフトウェア，ハードウェア，菓子，健康食品，名産，嗜好品，米，茶，酒，海産物，畜産物，野菜，果物，加工食品，コーヒー，紅茶，電化製品，本，雑誌，雑貨，住宅，カタログ，サービス，携帯電話，PHS，ビデオ，CD，おもちゃ，ゲーム，アクセサリ，ファッション，美容，健康，インテリア，植物，ペット，スポーツ用品，NIKE，靴，時計，個人売買，フリーマーケット

食品：名産，嗜好品，野菜，果物，米，酒，加工食品，茶，海産物，健康食品，菓子，畜産物，コーヒー，紅茶，飲料

チケット：コンサート，スポーツ，乗り物，映画，舞台，ミュージカル，歌舞伎，講習会

旅行&レジャー：旅行，レジャー，宿泊施設，アウトドア，ホテル，交通，天気

教育・文化：教育，研究，小学校，中学校，高等学校，高校，高専，専門学校，大学，短大，文化施設，博物館，美術館，図書館

メディア：テレビ，ラジオ，CATV，番組，衛星放送，本，雑誌，出版社，書店

テレビ・ラジオ：テレビ，ラジオ，衛星放送，PerfecTV，テレビ局，NHK，日本テレビ，フジテレビ，TBS，テレビ朝日，テレビ東京，地方テレビ，CATV，TV番組，アニメ，放送局，アナウンサー，DJ，CM，ラジオ局，AM，FM，ラジオ番組

本・雑誌：NEWS，一般誌，コンピューター，女性誌，就職誌，車，バイク，ビジネス，生活，不動産，アウトドア，書籍，少女漫画，少年漫画，漫画，詩，漫画家，インターネット小説，オンラインマガジン，小説，エッセイ，文学，新刊

Holiday & AFTER5：スポーツジム，習い事，音楽鑑賞，映画，飲食，遊園地，動物園，水族館，文化施設，ギャラリー，イベント，チケット，スポーツ，旅行，アウトドア

企業紹介： 企業，水産，農林，鉱業，建設，食品，繊維，パルプ，紙，化学，ゴム製品，窯業，鉄鋼，非鉄金属，金属製品，機械，電気機器，輸送用機器，精密機器，製造，商業，金融，保険，証券，不動産，倉庫，運輸関連，通信，電力，ガス，サービス，放送，報道，飲食業，調査，研究，広告，印刷，出版，ファッション，医療，医薬，人材紹介，代行サービス，学習塾，稽古，旅行業，専門サービス，商店，商社，百貨店，スーパー，本屋，リース，レンタル，写真，デザイン，理容，美容，趣味，娯楽，インターネット接続，ホームページ作成，インターネット，パソコン販売，ソフトハウス，コンピュータ，外国企業，団体，総合，財団法人，青年会議所，ボランティア，ボーイスカウト，環境保護団体

クイズ・ゲーム： インターネットゲーム，Shockwave，JAVA，テレビゲーム，セガサターン，任天堂，プレイステーション，コンピュータゲーム，クイズ

スポーツ： 野球，サッカー，バスケットボール，バスケット，バレーボール，テニス，ゴルフ，ラグビー，アメリカンフットボール，アメフト，球技，ヨット，スキューバダイビング，ウィンドサーフィン，サーフィン，スキー，スケート，アイスホッケー，スノーボード，スカイスports，F1，バイク，相撲，格闘技，武道，陸上，体操，オリンピック，アウトドア，カヌー，カヤック，クライミング，ソーラーカー，ダイビング，ダンス，トライアスロン，バドミントン，レーシングカー，競馬，競輪，水上スキー，釣り

趣味： アウトドア，写真，車，バイク，釣り，映画，Art，ギャラリー，漫画，ファッション，コンピュータ，アマチュア無線，無線，クイズ，ゲーム，おもちゃ，プラモデル，ぬいぐるみ，人形，オルゴール，ラジコン，競馬，競輪，競艇，パチンコ，麻雀，宝くじ，電車，航空機，楽器，カラオケ，ダンス，占い，オカルト，手品，俳句，書道，茶道，発明，ライセンス，天文，将棋，園芸，タバコ，切手，ハーブ，手芸，彗星

生活 (all)： 医療，税金，結婚，不動産，子育て，料理，飲食物，法律，福祉，学習，ペット，お花見，花粉症，ヘアケア，マナー，環境，健康，宗教，住宅，人生，政治，ペットショップ，動物病院，獣医，犬，猫，熱帯魚，魚，ハムスター，小動物，カブト虫，クワガタ

アウトドア： キャンプ，ハイキング，サイクリング，サークル，団体，グッズ，ログハウス，別荘，キャンプ場

写真： フォト，ギャラリー，風景，動物，人物，旅行記，国内，海外，カメラ，技術，企業，サークル，写真家，モデル，タレント，写真展，セクシータレント

車・バイク： 自動車販売，カー用品，ドライブ，ツーリング，外車，修理，車検，ロードマップ，中古車，販売，買取り，情報誌，サークル，教習所，新車紹介，レンタカー，レーシングカー

Art： 映像，音響，演劇，生け花，伝統芸能，陶芸，工芸，美術

ファッション： ジュエリー，ヘアデザイン，ランジェリー，リフォーム，眼鏡，香水，刺青，小物，着物，美容，服飾

ペット： 犬，猫，熱帯魚，魚，ハムスター，小動物，カブト虫，クワガタ，ペットショップ，動物病院，獣医

みんなの広場： 日記，おしゃべり，サークル，出会い，同窓会，ファンクラブ

音楽・芸能界：音楽，ポップス，ロック，ジャズ，フュージョン，クラシック，アマチュアバンド，合唱団，楽団，楽器，コンピュータミュージック，音楽情報，芸能界情報，アーティスト，男，女，ジャニーズ，小室，俳優，女優，男優，グループアーティスト，お笑い，セクシータレント

新聞・ニュース：，，全国紙，ブロック紙，朝日新聞，毎日新聞，読売新聞，日本経済新聞，産経新聞，北海道新聞，中日新聞，西日本新聞，地方紙，ミニコミ誌，タウン誌，天気予報，スポーツ新聞，プロ野球ニュース，交通情報，政治，経済，災害，専門紙，大学新聞，個人新聞，海外新聞

ニュース：新聞，天気予報，スポーツ，交通，政治，経済，災害，地震，洪水，火事

官公庁：内閣，総理府，法務省，外務省，大蔵省，文部省，厚生省，農林水産省，通商産業省，運輸省，郵政省，労働省，建設省，自治省，政府機関

専門分野：理系，コンピュータ，電気，電子，工学，機械，理学，医学，科学，化学，生物，建築，土木，農学，地質学，地学，物理，文系，経済，政治，法律，外国語，心理，教育，宗教，歴史，社会，文学，国語，芸術

検索サーバー：国内検索，海外検索，一覧，登録，全文検索，新着情報，インデックス，キーワード，メガインデックス，リスト，サーチエンジン，ロボット，メタサーチ

マルチメディア (all)：CD-ROM，サウンド，通信カラオケ，コンピュータ，CATV，CG，インターネット，画像，メーリングリスト，インターネットカフェ，LAN，インターネット，プロバイダ - ，プロバイダ，ホームページ，ホームページ作成，素材，サーバ，ディレクトリ，通信，検索エンジン，リンク集，ドメイン，プラグイン，ブラウザ，WWW，HTML，Shockwave，JAVA，RealAudio，VRML，携帯電話，携帯，PHS，ネットワーク，イントラネット，セキュリティ，ISDN，ATM，LAN，パソコン通信，コンピュータ，ハード，PC，DOS/V，Macintosh，UNIX，モバイル，周辺機器，MODEM，DSU，プリンタ，ボード，カード，ディスプレイ，記憶装置，ソフト，OS，FTP，Telnet，エディタ，音声ツール，音声，画像ツール，画像，電子メール，通信，スクリーンセーバ，データベース，表計算，ワープロ

分野 Yahoo!-Top：，芸術，人文，写真，建築，美術館，画廊，歴史，文学，メディア，ニュース，テレビ，ラジオ，番組，新聞，雑誌，ビジネス，経済，企業，雇用，マーケット，投資，求人，趣味，スポーツ，アウトドア，ゲーム，ペット，車，旅行，コンピュータ，インターネット，WWW，資料，情報源，図書館，辞書，郵便，電話番号，教育，学校，大学，受験，言語，講座，地域，都，道，府，県，日本，海外，世界，エンターテインメント，映画，音楽，芸能，グルメ，プレゼント，自然科学，技術，政治，行政，社会科学，経済学，社会学，言語学，健康，医学，生活，文化，子供，環境，自然，博物館，展示会

芸術：芸大，デザイン，マスコミ，理容，美容，ファッション，服飾，家政，劇，人文，建築，彫刻，手芸，写真，書，陶芸，美術史，芸術史，演劇，文学

人文：言語学，宗教，哲学，倫理，人智学，数学，錬金術，中国哲学，文学，古典，神学，歴史

文学：文学情報，出版，外国文学，脚本，伝統文学，児童文学，漢詩，短歌，俳句，書評，古典，作家，詩，批評，文学理論，執筆，文献，目録

文学ジャンル：SF, ファンタジー, ホラー, 詩, エッセイ, エッセイスト, 小説, 漢詩, 神話, 民話, 民俗学, 短歌, 俳句, 歴史, 時代, 外国文学, 戯曲, ノンフィクション, フィクション, ミステリ, ロマン, 児童文学, 童話, アダルト, ユーモア

伝統芸術：浮世絵, 錦絵, 演芸, 大道芸, 伝統文学, 雅楽, 陶芸, 歌舞伎, 能, 狂言, 茶道, 書, 文楽, 浄瑠璃, 民芸品, 和服, きもの

演劇：演技, 演出, 監督, 戯曲, 喜劇, 劇場, 舞台, 劇団, 即興, 古典劇, 人形劇, 能, 狂言, 俳優, 舞台俳優, 舞台美術, ミュージカル

ビジネスと経済：企業, 教育, 経営情報, 金利, 株価, 銀行, 経済学, 経済史, 労働, 経済指標, 国際経済, 貿易, 雇用, コンベンション, 小規模事業, 消費者, 消費者経済, 税, 知的所有権, 電子商取引, デジタルマネー, 不動産, 貿易, マーケット, 投資

雇用：企業, 求人, 失業, 就職, 障害者, 女性, 体験記, 労働法, 労働組合

マーケットと投資：シミュレータ, 株式, 決算, 財務情報, 経済指標, 債券, 先物, オプション, 市況, 証券取引, 為替, 投資信託, マーケットサマリー

業種：運輸, 通信, 海運, 空運, 陸運, 金融, 保健, 銀行, 証券, 建設, サービス, 製造, エネルギー, 不動産

教育：オンライン教育, 学習指導, 学習塾, 教育工学, 教育理論, 言語, 高校, 小学校, 小中高校, 中学校, 職業訓練, 専門学校, 大学, 短大, 予備校, 幼稚園, 保育園, 留学, 資格, 奨学制度, 社会人教育, 生涯学習, 日本人学校, 日本語, 特殊教育, 数学, 科学, 社会, 環境, 政治

政治：軍事, 兵器, 研究機関, 国際機関, 国連, 国会, 参議院, 衆議院, 議員, 司法府, 裁判所, 最高裁判所, 税, 消費税, 所得税, 税理士, 政治, 行政, 政策, 政党, 選挙, 法, 法律, 法案, 法令, 刑法, 憲法, 裁判, 民法, 労働法, 六法, 知的所有権, プライバシー, 法学

医学：病院, 診療所, 治療, 生体工学, 医療倫理, 緊急治療, 疫学, 解剖, 血液, 神経学, 心臓, 腎臓病, スポーツ医学, 精神学, 聴覚学, 生理学, 毒物学, 皮膚科学, 病理学, 放射線学, 麻酔学, 免疫学, 呼吸器, 耳鼻咽喉, 消化器, 小児, 整形, 精神, 神経, 臓器移植, 輸血, 理学療法, 作業療法, 鍼灸, 遠隔治療, 眼計測法, メディカルイメージング, 整体, 指圧, マッサージ, アイバンク, 白内障, 緑内障

病院：診療所, 病院, 眼科, 外科, 肛門科, 呼吸器科, 産科, 婦人科, 歯科, 耳鼻科, 咽喉科, 消化器科, 胃腸科, 小児科, 神経科, 整形外科, 精神科, 内科, 総合, 泌尿器科, 皮膚科, 美容外科, 腎臓病, 放射線, 鍼灸師, 治療院, 獣医, カイロプラクティック, 整体, 医療経営

情報源：カレンダー, 暦, 国旗, 時間, 辞書, シソーラス, 測定, 計量, 地図, 電話番号, 頭字語, 略語, マーケット, 投資, マナー, エチケット, 名言, 格言, 郵便, 図書館, 特許, 検索, 文献, 目録

自然科学と技術：, 調査, 測定, 計量, 単位, 標準, 宇宙, 超心理学, 超常現象, エネルギー, ライフサイエンス, 植物, 生物, 生物物理, 動物, 化学, 工学, 数学, 地球科学, 地理学, 天文学, 複雑系, 物理学, 生物学, 生態学, エコロジー, 水産学, 海洋, 農学, 気象, 天気, 情報処理, 人工生命, コンピュータサイエンス, 心理学, 認知科学

エネルギー: , , ガソリン, 自動車, 原子力, 石油, 石油工学, 天然ガス, 太陽, 地熱, 電力工学, 電力, 電気, 燃料電池, 燃料, 風力

情報処理科学: , , コンピュータ, OS, スーパーコンピュータ, 並列処理, モバイル, ソフトウェア, アルゴリズム, オブジェクト指向, ソート, 人工知能, 自然言語, データベース, 文字認識, 手書き, ニューラル, 言語学, 知識科学, ロボット工学, セキュリティ, 暗号, ネットワーク, 通信, インタフェース, バーチャルリアリティ, グラフィックス

宇宙: 宇宙物理学, 天文学, 天文台, プラネタリウム, 望遠鏡, 衛星, スペースシャトル, ロケット, 太陽, 月, 月食, 彗星, 流星, 日食, 惑星, 火星, 水星, 木星, 土星, 天王星, 冥王星, 銀河系, 銀河, 写真, 画像

化学: 生化学, 有機, 無機, 電気, 分光技術, ポリマー, 元素, 原子, 分子, 元素周期表, 錬金術, コンピュータ解析

工学: 化学, 海洋, 環境, 機械, 経営, 人間, 構造, 航空, 材料, 生産, 石油, 造船, 電気, 電力, 土木, 農業, 医用生体, 電子情報, 原子力, 自動車, 信頼性, 音響学, 経営科学, 光学, 核, 生物材料科学, 流体力学, 通信, 都市計画, 度量衡学, 力学, 品質管理, 溶接, 製造, 製品, 建築, 測量, 採鉱, 視覚, レーザー, ロボット, 紙, 半導体, エネルギー, ガソリン, 自動車, エンジン

心理学: 心理テスト, オンラインテスト, 音楽, 行動, 色彩, 社会心理, 性格, 筆跡, 認知科学, メンタルヘルス, 睡眠, 発達

数学: 数, 数学史, 幾何, 形式論理, 数値解析, 整数, 相乗作用, 数理, 統計, 複雑系, 保険数理, カオス, セキュリティ, 暗号化, フラクタル

生物学: 生化学, 遺伝, ゲノム, 解剖, 海洋, 寄生虫, 原生, 古生物, 細胞, 獣医, 植物, 進化, 神経, 生態, エコロジー, 気象, 生物物理, 動物, 鳥類, 昆虫, 霊長類, バイオテクノロジー, 発生, 微生物, 細菌, ウイルス, 分子生物, 免疫, 血管

地球科学: , 地球, 環境, 海洋, 気象, オゾン層, 温暖化, 古生物, 化石, 恐竜, 水文学, 生物気候, 地質, 火山, 地震, 鉱物, 火災, 震災, 地球物理, 陸水

農学: 農業, 畜産, 水産, 園芸, 草花, 栽培, 植物, 病理, 生物, 生物材料, 動物, 造園, 緑地, 農家, 作物, 有機, 森林, 土壌

博物館: 古生物, 自然史, 自動車, 鉄道, 飛行機, 乗り物, スポーツ, 植物園, 水族館, 動物園, プラネタリウム, 科学博物館, 昆虫, 宇宙, エネルギー, 歴史, 恐竜

物理学: 物理, 物質, 宇宙物理, X線, 核, プラズマ, レーザー, 磁性物理, エネルギー, 素粒子, 超電導, 生物物理, 相対性理論

社会科学: 科学, 技術, 社会, 経営, 経済, 社会, 言語, コミュニケーション, 女性, 心理学, 人類学, 考古学, 政治, 民族, 哲学, 都市研究, 法学, 歴史, 和平, 紛争