可変なカテゴリ構造を用いた文書検索支援手法

仲川 こころ† 高田 喜朗† 関 浩之†

WWWなどで提供される情報が爆発的に増加するのにともない，文書検索サービスはより多くの，そして幅広いユーザに必要とされるようになった．そのため，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．そこで本論文では，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提供することで検索作業を支援する手法を提案する．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，カテゴリ構造を構築する．これにより，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，またディレクトリ型サービスの固定的なカテゴリ構造ではユーザの様々な要求に対応できない，という既存の検索手法が持つ問題を解決している．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点から提案手法の有効性を評価した．

A Document Retrieval Method Based on Flexible Category Structure

Kokoro Nakagawa,† Yoshiaki Takata† and Hiroyuki Seki†

A method for supporting document retrieval by constructing a flexible category structure is proposed. In this method, a category structure suitable for retrieval by the user is constructed whenever a query is submitted. By using the experimental system which has 68 categorization viewpoints, the proposed method was evaluated with BMIR-J2 test collection to compare the proposed method, a keyword-based method, and a clustering method (K-means). The proposed method outperforms clustering method in precision and it also outperforms keyword-based method in recall. An experiment using human subjects was also performed. Based on the experimental results, the proposed method is evaluated in its quality, efficiency, and usability.

1. はじめに

WWW（World Wide Web）サービスなどの普及にともない，電子化された文書を検索するためのサービスは幅広いユーザから必要とされるようになった．そこで，専門的な知識や技術を持たなくても必要な情報を探し出せる，使いやすい検索方法を実現することが重要となっている．現在キーワード検索型とディレクトリ型の2種類の文書検索サービスがあるが，その使いやすさに関して次のような問題がある．

*Alta Vista, goo*に代表されるキーワード検索型のサービスは，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，その一覧を出力するサービスである．出力は基本的に「入力したキーワードを含む文書」であり，ユーザにとって理解しやすいという利点がある一方，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．

*Yahoo!*に代表されるディレクトリ型のサービスでは，データベースの全情報をあらかじめ人手で分類し，ユーザに対してはその分類結果を提供する．ユーザは，提供されるカテゴリの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ため，キーワード検索型に比べ負担が少ない．しかし，提供されるカテゴリ構造はデータベースのすべてを表現するものであり，巨大すぎて全体の把握が難しい．また，カテゴリ構造は必ずしもユーザの検索目的に対して適切ではない，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，ユーザがたとえば，

root → 奈良 → 環境と自然 → 研究機関, (1)

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たとえば，

root → 自然科学 → 環境工学,

† 奈良先端科学技術大学院大学情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

root → 地域情報 → 奈良 → 生活と健康,
root → 生活と文化 → 環境と自然→ 研究機関.

など，複数の経路上に目的の情報が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．またたとえすべての経路をたどったとしても，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．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．

本研究では，**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提供する**という検索支援手法を提案する[6),11),12),18)]．**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構築することである．たとえば，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 (1) の経路を含むカテゴリ構造を提供し（**図 1** (a)），環境問題に関する最新の情報を調べたい場合は (2) の経路を含むカテゴリ構造を提供する（図 1 (b)）．

root → *News* → 自然科学 → 環境問題 → オゾンホール (2)

提案手法では，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．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，それをすべての分類観点を用いて分類する．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し，分類を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，カテゴリ構造がユーザの望む分類方法に沿っていれば，ユーザが検索結果として選んだカテゴリの周辺のカテゴリもユーザの興味を反映するものになっていると期待できる．また提案手法では，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ため，キーワード検索のように慎重に検索式を考える必要がない．

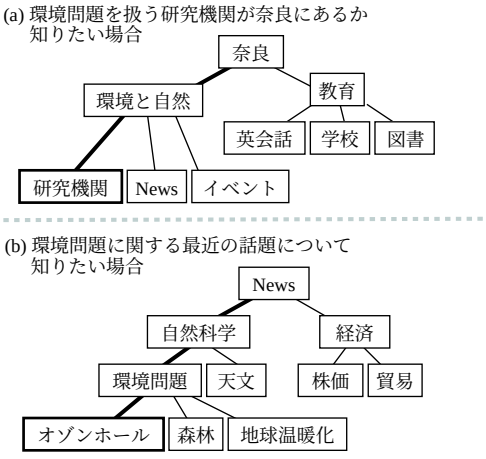


図 1 カテゴリ階層構造の例
Fig. 1 Examples of category structures.

関 連 研 究

ユーザの検索目的を推測し，それに適応することによって文書検索を支援しようという研究は従来より多くなされている．検索システム RCAAU[7)]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，検索結果を絞り込むのに有用と思われる 2 次的キーワードを提供する．同様に Anick らのシステム[1)]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．また検索システム VOIR[4)]では，ユーザの入力したキーワードの履歴を基に，現在の検索式をより精度の良い結果を得られるように自動的に洗練する．これらの研究は，精度の悪い初期検索の結果を絞り込む手法として有効であるが，ユーザの検索目的にあった小規模なカテゴリ構造を構築しようとする本研究とは異なるアプローチである．

一方，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も研究されている．たとえば Scatter/Gather[13)]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，本研究の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，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きわめて困難である．

本研究と同じく，検索システム HIBROWSE[14]で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．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 の中のキーワード使って現在の結果を絞り込むことができる．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．

またアプローチは異なるが，情報視覚化技術に関する研究[3),16)]は，本研究や上述の各手法との統合が可能である．

2. 提 案 手 法

提案手法は以下の 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あらかじめ与える**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．たとえば「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，それらの組合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

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

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 $W_j = \{w_{j1}, \ldots, w_{jk_j}\}$ 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，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 1）．

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**[17]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

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

とすると，文書 d の特徴ベクトルは

$$c(d) = (c_{w_1,d}, c_{w_2,d}, \ldots, c_{w_n,d})$$

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 中での単語 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[2)]）との積を，d の長さで正規化したものと定義する．

特徴ベクトル u，v の**類似度** $sim(u, v)$ を，

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

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．

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 2**）．

(1) ユーザ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よい．

(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 $W_j = \{w_{j1}, \ldots, w_{jk_j}\}$ の中から d が属すべきカテゴリ $w_j^{(d)}$ を 1 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．S_j による文書集合 D の分割 $(D_{j1}, \ldots, D_{jk_j})$ を，

$$D_{ji} = \{d \in D \mid sim(c(w_{ji}), c(d)) = \max_{1 \leq h \leq k_j} sim(c(w_{jh}), c(d))\}^{\star}$$

と定義する．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．

$c(w_{ji})$ 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．$c(w_{ji})$ 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（$c(w_{ji}) = (\delta_1, \delta_2, \ldots, \delta_n)$，ただし $w_{ji} = w_k \in W$

☆ 簡単のため，$sim(c(w_{jh}), c(d))$ 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．

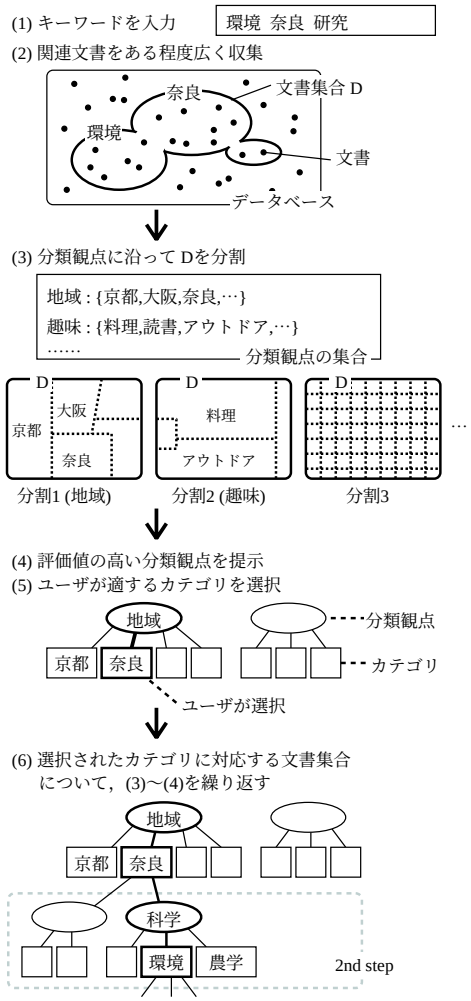


図 2 動作の概要
Fig. 2 Behavior of the system.

のとき $\delta_k = 1$，それ以外のとき $\delta_k = 0$)．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（文書 d と初期カテゴリベクトル $c(w_{ji})$ 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 $c_{w_i,d}$ を正規化した値 $sim(c(w_{ji}), c(d)) = c_{w_{ji},d}/|c(d)|$ 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（$sim(c(w_{ji}), c(d)) = 0$ 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

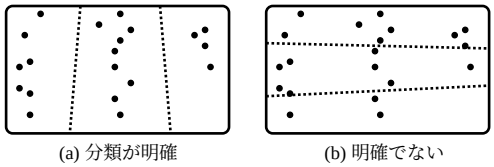


図 3 分類の明確さ
Fig. 3 Clearness of categorization.

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

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．

分割を行った後，すべての分類観点について D **に対する適切さ**を評価する（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適するものを選ぶ．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．

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．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 (4)，エントロピーを用いる場合は式 (5) でそれぞれ定義する．これら 2 つの評価基準の有効性を，評価実験により比較した（3, 4 章）．

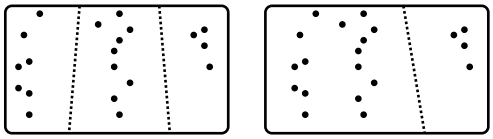
2.3.1 分類の明確さに基づく評価

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** (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**もっともらしい分類**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：

$$e_D(S_j) = \sum_{d \in D} sim(c(w_j^{(d)}), c(d))/|D| \,. \qquad (4)$$

2.3.2 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．



(a) 均等な分割 (エントロピー大)　(b) 偏りのある分割 (エントロピー小)

図 4 分類のエントロピー
Fig. 4 Entropy of categorization.

$$H_D(S_j) = -\sum_{i=1}^{k_j} P_i \log P_i, \qquad (5)$$

ただし,

k_j : S_j に属するカテゴリの数,

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

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/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**といえる（**図 4**）．

3. 性能評価実験

提案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 章：性能評価実験，4 章：被験者を用いた実験）．性能評価実験では，ある理想的な振舞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．

3.1 システムの概要

試作したシステムは，ユーザによる検索が行われる前にあらかじめ実行しておくデータベース構築部と，ユーザの問合せに応じる問合せ処理部からなる．本システムは，全文キーワード検索システムである Freya [5] を拡張する形で実現した（詳細は文献 12) を参照）．また，システムが扱う単語の集合（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』[10] 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．

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（提案手法 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（提案手法 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提案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 [19] 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 [13] 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 $w_{i_1}, \ldots, w_{i_n}$ が入力された場合は，$\sum_{i \in \{i_1,\ldots,i_n\}} c_{w_i,d}$ 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．

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 [8] 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，提案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 1** にあげた 10 課題を選んだ☆．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．

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

☆ 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，正解文書を見つけやすい）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 1 実験に使用した検索課題
Table 1 Topics used in the experiments.

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
*2：4 章の実験では “ビデオデッキ（正解文書の数 33）” に変更.

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あげる．

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

クラスタリングでは文書集合をつねに 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かくならない．3 章の実験においては，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．

3.2 実 験 (1)

提案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 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 $r = 0.1, 0.2, \ldots, 1.0$ 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 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 $S_1, S_2, \ldots, S_m$ $(m = 68)$ 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，“一覧表示したときに 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 画面ず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．

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：

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

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

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 2）を用いて検索を行い，得られた文書リストに対して TREC[21] 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（再現率の分母はシステム A，B，C，D で共通の値）．

また各再現率 R ・適合率 P のペアに対し，2 つを複合した指標である E-尺度を求める．これは以下の式で定義される[20]．

表 2 実験に使用したキーワード
Table 2 Keywords used in the experiments.

課題	キーワード *
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

*：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）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．

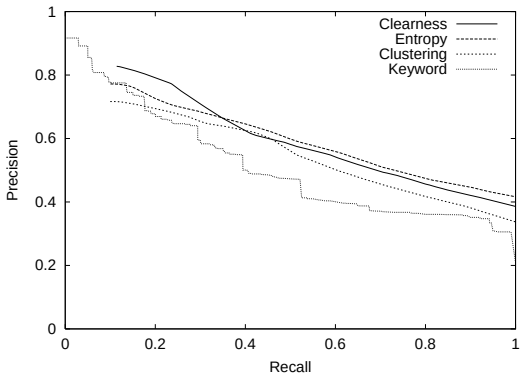


図 5 再現率と適合率（10 課題の平均）
Fig. 5 Average precision-recall curves.

$$E = 1 - \frac{(\beta^2+1)PR}{\beta^2 P + R} \ . \qquad (6)$$

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．

結　　果

再現率と適合率の関係（10 課題の平均）を**図 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（β = 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**，**表 4**）．

分散分析の結果，β = 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，β = 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，提案手法（A，B）がクラスタリング手法（C）に比べて優れているといえる．β = 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，提案手法（A，B）がキーワード手法（D）に比べて優れているといえる．

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．比較的正解文書数の少ない課題においては明確さ，正解文書数の多い課題においてはエントロピーが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い（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．

表 3 E-尺度 $\beta = 2.0$（再現率 0.1,0.2,〜，1.0 の平均）
Table 3 E-measure ($\beta = 2.0$).

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 4 E-尺度平均値
Table 4 Average of E-measure.

	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

3.3 実 験 (2)

提案手法では，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) $Result_j = \emptyset$ $(1 \le j \le m)$ 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 $(1 \le j \le m)$ 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 $(1 \le j \le m)$ 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 5 評価基準と適合率との相関係数
Table 5 Correlation coefficients between the score and the precision.

評価基準	r*1	正の相関がある課題数	$\overline{R}$*2	平均カテゴリ数*3
明確さ	0.2	9	0.496	2.2
	0.5	8	0.595	4.1
	0.8	8	0.646	7.3
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：r は目標再現率
*2：$\overline{R}$ 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）
*3：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

10 個の検索課題についての分類観点の評価値と適合率の相関を**表 5** に示す．システム A（明確さ）では，10 個のうち 8 ないし 9 個の検索課題で正の相関が認められた．システム B（エントロピー）では，全検索課題について正の相関があった．

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 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（明確さ）の方がシステム B（エントロピー）よりも平均カテゴリ数がやや少ないことが分かった．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いえた（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： 3.2，3.3 節の結果図表はすべて表 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，さらに異なる 2 通りのキーワード（4 章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．3.2，3.3 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．

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．

検索課題

検索課題は，性能評価実験と同じ 10 課題（表 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 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[15]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 Microsoft Internet Explorer version 5.0 を用いて実験作業を行った．実験中のすべ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．実験後数日のうちに提出してもらった．

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できる：(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**を表示したとき，(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**を表示したとき．

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 $\emptyset$ で初期化され，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**を開いたとき☆，(2) 被験者が**タイトル一覧**を今までよりも下にスクロールしたとき．

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：

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

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

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

4.2 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 6**，**表 7** に示す（全被験者の平均）．E-尺度は各再現率・適合率のペアから式 (6) に従って計算され，値が小さいほど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．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．

☆ **タイトル一覧**が以前に開かれたものとまったく同じである場合は，$Title(q)$ に追加しない．

表 6 E-尺度 $\beta = 2.0$（被験者平均）
Table 6 E-measure ($\beta = 2.0$).

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 7 E-尺度平均値
Table 7 Average of E-measure.

	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

表 8 BMIR-match（全被験者の平均）
Table 8 BMIR-match.

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

(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 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，提案手法（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分かる．

表 9 検索終了時における適合率（全被験者の平均）
Table 9 Average of precision at the termination of retrieval.

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

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）が提案手法である 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 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やすい，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．したがって，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つね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 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分かった．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，提案手法（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見ると，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．

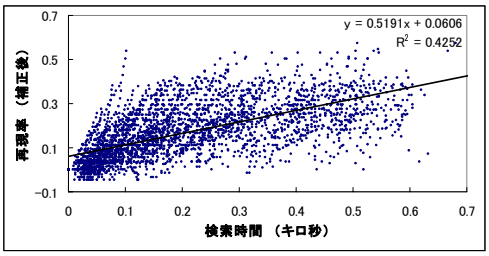


図 6 検索時間と再現率（補正後）
Fig. 6 Search time and adjusted recall.

表 10 検索時間（キロ秒）・再現率間の回帰直線
Table 10 Regressive analysis.

	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

(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（キロ秒）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 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 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（$Recall$）回帰直線を

$$Recall = a \times Time + b \qquad (7)$$

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 回帰直線を

$$Recall = a_q \times Time + b_q \qquad (8)$$

とする．以下の式に従って，補正後の再現率 $Recall'$ を求める：

$$Recall' = \frac{a}{a_q}(Recall - b_q) + b \ . \qquad (9)$$

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は多くなり，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 6）．さらに，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いえる．

表 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）．

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

表 11 アンケート結果（全被験者の平均）
Table 11 Ranking by the subjects.

	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

なっている．このことから，提案手法（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) の結果から，以下の考察が得られる：提案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，提案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え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．

(e) 事後アンケートより（使いやすさ）

《4 システムの比較》

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（一番使いやすかったシステムが 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 11** に示す．

表 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，続いて提案手法の 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提案手法は，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，提案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章でも述べたとおり，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，提案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，提案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 回以上分類したとき（2.2 節ステップ (6)），B は前回とはまった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 の方が使いやすい．

(f) 明確さとエントロピーについての考察

(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（A）とエントロピー（B）のどちらが優れているとはいい難いが，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．

5. ま と め

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提案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提案手法がキーワード手法よりも優れていることがいえた．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，提案手法が特に「難しい」検索課題に対して有効であることが確認できた．さらに提案手法の使いやすさに対する被験者の主観的評価は，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い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

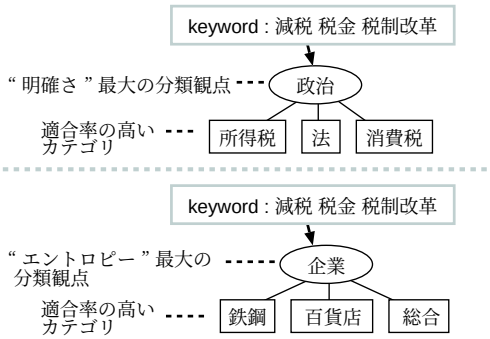


図 7 「減税」に対するカテゴリ構造
Fig. 7 Category structure for the topic “tax reduction.”

かどうかは重要である．3，4 章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，明確さを用いた方が検索課題に対して妥当と思われるものが多かった（**図 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．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まった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分かれた（4.(e) 節）．

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．さらに提案手法の改善として，2 つの評価基準をユーザの好みやその時々の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提案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，提案手法を WWW 文書データベースに適応した際の結果については議論していない．しかし，検索の目的に応じて有用なタイプの文書（たとえばカタログや個人の日記など）を取り出すような前処理と組み合せ，文書の質が揃ったデータベースに対して提案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．関連研究として，文書タイプの分類を前処理として行い，その結果を従来の全文検索技術と組み合せる WWW 検索システム[9]などの研究がある．

また提案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．提案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良い検索を提供できると考えている．

謝辞 分類観点の評価基準について有益なご助言・ご討論をいただきました奈良先端科学技術大学院大学石井信教授に深く感謝いたします．また試作システムの実装に利用させていただいた検索システム『Freya☆』，および，日本語形態素解析システム『茶筌☆☆』の作者の方々に深く敬意を表します．最後に，本研究の全般，特に 3，4 章の実験に数々のご助力をいただきました岩崎正秀氏に深く御礼申し上げます．

参 考 文 献

1) Anick, P.G. and Tipirneni, S.: The Paraphrase Search Assistant: Terminological Feedback for Iterative Information Seeking, *Proc. SIGIR 99*, pp.153–159 (1999).
2) Dreilinger, D. and Howe, A. E.: Experiences with Selecting Search Engines Using Metasearch, *ACM Trans. Inf. Syst.*, Vol.15, No.3, pp.195–222 (1997).
3) Fishkin, K. and Stone, M.C.: Enhanced Dynamic Queries via Movable Filters, *Proc. CHI ’95*, pp.415–420 (1995).
4) Golovchinsky, G.: Queries? Links? Is There a Difference?, *Proc. CHI 97*, pp.407–414 (1997).
5) 原田昌紀：*Freya version 0.92* (1998). http://odin.ingrid.org/freya/
6) 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，DBS120, pp.1–8 (2000).
7) 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).
8) 木谷 強ほか：日本語情報検索システム評価用

☆ http://odin.ingrid.org/freya/
☆☆ http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html

テストコレクション BMIR-J2, 情報処理学会研究報告，DBS114, pp.15–22 (1998).
9) 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，FI53-2, pp.9–16 (1999).
10) 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).
11) 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).
12) 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，HI82, pp.59–64 (1999).
13) 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).
14) Pollitt, A.S.: 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
15) Preece, J., Rogers, Y., Sharp, H., Benyon, D., Holland, S. and Carey, T.: *Human-Computer Interaction*, Addison-Wesley (1994).
16) Robertson, G.G., Card, S.K. and Mackinlay, J.D.: Information Visualization using 3D Interactive Animation, *Comm. ACM*, Vol.36, No.4, pp.57–71 (1993).
17) 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).
18) 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).
19) Tou, J.T. and Gonzalez, R.C.: *Pattern Recognition Principles*, pp.89–97, Addison-Wesley (1974).
20) van Rijsbergen, C.: *Information Retrieval*, 2nd edition, Butterworths (1979).
21) Voorhees, E.M. and Harman, D.K.: Evaluation Techniques and Measures, *The Seventh Text REtrieval Conference* (*TREC 7*), p.A-1, National Institute of Standards and Technology (NIST) (1998).

(平成 12 年 12 月 20 日受付)
(平成 13 年 9 月 12 日採録)

仲川こころ（学生会員）

平成 9 年関西学院大学理学部物理学科卒業．平成 11 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．現在，同大学院大学情報科学研究科博士後期課程在学中．情報検索，ヒューマンインタフェースに興味を持つ．

高田 喜朗（正会員）

平成 4 年大阪大学基礎工学部情報工学科卒業．平成 9 年同大学院博士後期課程修了．同年奈良先端科学技術大学院大学情報科学研究科助手．現在に至る．博士（工学）．ユーザインタフェース，情報検索に関する研究に従事．

関 浩之（正会員）

昭和 62 年大阪大学大学院基礎工学研究科博士後期課程修了．工学博士．同年大阪大学基礎工学部情報工学科助手．同講師，助教授を経て，平成 6 年奈良先端科学技術大学院大学情報科学研究科助教授．平成 8 年同教授，現在に至る．形式言語理論，ソフトウェアの基礎理論に関する研究に従事．平成 9 年度情報処理学会論文賞．