

部分構造の主題の相互関係を考慮した文献検索

大石貴治 吉川正俊 渡辺正裕

奈良先端科学技術大学院大学 情報科学研究科

本研究ではベクトル検索に基づく文献検索の手法を提案する。文献が一つの話題 (main topic) だけから成っているのではなく複数の話題も含んでいることを考え、一つの文献をいくつかの部分的な話題 (subtopic) に分割する。文献全体からベクトルを作成するだけではなく、subtopic ごとにベクトルを作成することにより、話題の相互関係を考慮した文献検索を行う。

Document Retrieval Based on Relationship among Topics of Substructures

Takaharu OISHI, Masatoshi YOSHIKAWA and Masahiro WATANABE

Graduate School of Information Science
Nara Institute of Science and Technology (NAIST)

In this paper, a new method for document retrieval is proposed. Subtopics are considered as well as their main topics. Documents are divided into some parts, and vectors are produced for documents as a whole and their parts. We have carried out experiments using master thesis of our institute, and the results show the effectiveness of our method.

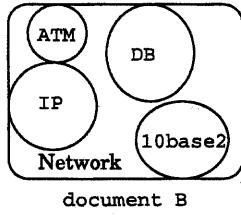
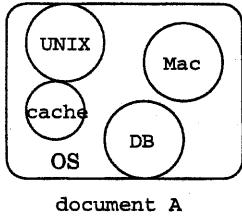


図 1 文献の例

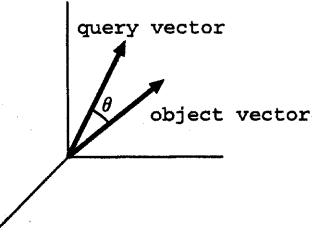


図 2 vector space model

1. はじめに

文献集合からある内容の文献を検索する際に、文献内の内容の構造に基づいて検索ができるれば便利である。例えば図 1において DB に関する話題は document A で一部でしか述べられておらず、OS の話題が広範囲に述べられている。本研究ではこの広範囲に述べられている話題を main topic と呼び、一部で述べられている話題を subtopic と呼ぶ。一方 document B においても DB は一部でしか述べられおらず、ネットワークについては全体で話題になっている。これらののような文献があった場合にどちらか一方だけを検索したいという要求がある。これに対しこまでの方法は文献・単語行列をつくる際に単語の分散、集中の違いによって重み付けを行うことをしてきた。従来のではこの要求を満たすことは難しい。

したがって、document A, B を区別して検索できるようにするために、本研究では一つの文献をさらに細かい segment に分けることによりその segment のベクトルを作り、文献の内部構造を考慮にいれた問合せを行う方法を提案する。

さらに日本語の文献を扱うということで奈良先端科学技術大学院大学の松本研究室の茶筅 [Mat97] を使用して語幹の切り出しをする。

2. 背景と目的

2.1 研究の背景

一般に文献検索システムに共通して指摘されている問題として、適合文献の収集の困難さと適合文献の選別の困難さの二つが挙げられる。これらはそれぞれ再現率 (recall) と適合率 (precision) で表される。またこれらの両方をともに高く保つことは、一般に困難であることが知られている。

適合文献の収集が困難であるといった状況は適合文献をみつけるための情報が少ない場合に生じる。各文献に対して割り当てられたキーワード、概要などの情報が個々の文献を正確に表すことができるとは一概には言えないであろう。これらの情報のみを頼りにして検索式を構成し問合せを実行したとしても、その検索式が利用者の意図を十分に反映したものであるかは判定できない。さらにその結果の文献集合が適合しているかは定かではない。

適合文献の選別が困難であるといった状況は検索式の構成の際の表現力の低さによって生じる。全文情報を持つ文献検索システムで検索式の構成にブーリアンモデルを用いるものなどがこれに相当する。この場合検索式として自由語または統制語を “and” や “or” で連結した式が用いられる。しかしそのような

検索の結果として返される文献集合はランキングによるフィルタリングはなされるものの、自由語リストを含むかどうかで選別された結果に過ぎず、これも利用者の意図を十分に反映したものであるかは定かではない [森村 96]。

2.2 研究の目的と特徴

既存の文献検索システムの問題点を考える。その原因の一部として、(1) 検索システムの検索対象として、文献全体を一つの話題 (main topic) として考え、いくつかの subtopic が含まれていることを考慮していないこと、(2) 検索式の構成における表現力が低いこと、といった点が考えられる。

本研究では、これらを鑑み、文献が一つの main topic だから成っているのではなく複数の subtopic も含んでいることを考え、従来の検索システムより検索精度を高めることのできるような文献検索システムを提案することを目的とする。

本研究は、このように文献をいくつかの大きさに区切りベクトルをつくることによって、文献が一つの話題 (main topic) だから成っているのではなく複数の subtopic も含んでいることを考慮することを特徴とする。

3. 関連研究

3.1 ベクトル検索モデル

オブジェクトや検索要求をベクトルで表現する検索モデルをベクトル検索モデル (図 2) という。とくにオブジェクトが文書である場合には、出現する単語から文書のベクトルを生成する様々な方法が研究されている。検索要求ベクトル (query vector) とオブジェクトベクトルの類似度を余弦などを用いて計算し、検索要求ベクトルと類似度の大きいものから順にランク付けして検索結果とする [小川 96][Oga96]。

3.1.1 ベクトルの作り方

文献に出現する単語一つを一つの次元として捉え、出現する単語の種類数の次元の多次元ベクトルを考える。文献からベクトルを構成する方法はいろいろある。

- 文献中にある単語が出現すれば 1、しなければ 0 にする方法
- tf*idf を使い単語の出現数と文献の長さから正規化する方法
- エントロピーを考える方法

$$X_0 = T_0 \times S_0 \times D_0^T$$

$m = \text{rank}(X) = \text{数万} \sim \text{数十万}$

図 3 SVD for X

$$\hat{X} = T \times S \times D^T$$

$k = \text{rank}(\hat{X}) = 100 \sim 200$

図 4 SVD for \hat{X}

3.2 Latent Semantic Indexing

Latent Semantic Indexing(LSI)[FO95][DDF+90][PTVF88][BDL95]はSaltonのSMARTシステムで用いられた伝統的なベクトル空間技術を上回る性能を発揮する、ベクトル空間情報検索手法である。

完全な文献集合から単語-文献行列が生成される。この行列の各成分はどの文献にどの単語が出現しているかを表すものである。この行列の特異値分解(Singular Value Decomposition(SVD))が計算され、小さい値は削除される。結果の単価ベクトルと単価行列は文献と問合せの単語頻度ベクトルを、単語-文献行列からの意味関係が保存される一方で、使用される単語の使用変化が少なく抑えられているような部分空間に射影する。文献と問合せのベクトルの正規化された内積で余弦類似尺度を計算する。そして文献はこの計算された値である関連度(類似度)の順に並べられる。

3.2.1 記法

単語-文献行列 X は t 行(文献集合に出現する各単語に対応)と d 列(文献集合中の各文献に対応)から構成される。SVD $X = T_0 S_0 D_0^T$ は、

- $t \times m$ 行列 T_0 : 各列は left singular vector と呼ばれ直交する
- $m \times m$ 行列 S_0 : 降順に並べられた正の singular values

- $d \times m$ 行列 D_0 : 各列は right singular vector と呼ばれ直交する

m は行列 X の rank である。

に帰着する。ここで図3に X の SVD を示す。

T_0, S_0, D_0 によって X は正確に再現される。LSIにおける主な革新は、 S_0 行列において大きいものから k 個の singular values だけを保存し、他を 0 にすることである。 k は設計によって決まるパラメータである。100~200 の値が一般に用いられる。はじめの行列 X は $\hat{X} = T S D^T$ で近似される。ここで

- $t \times k$ 行列 T : 各列は直交する
- $k \times k$ 行列 S : 降順に並べられた正の対角行列
- $d \times k$ 行列 D : 各列は直交する

である。ここで図4に \hat{X} の SVD を示す。

3.2.2 文献マッチング

LSI の有効性は文献集合における単語頻度から主要な特徴を抽出する能力に依存している。これを理解するためには、まず SVD を構成する三つの行列の操作上の解釈をすることが必要である。もともとのベクトル空間で、 $X^T X$ は文献どうしの内積で構成される $d \times d$ 対称行列である。ここで各文献はベクトルは単語頻度で表現される。 $X^T X$ 行列は各列は X 行列の対応する列の文献ベクトルと、文献集合中の各文献の間の内積の集合である。

文献 i と j のコサイン相関値は次のように計算される。

$$\frac{(\hat{X}^T \hat{X})_{(i,j)}}{(\hat{X}^T \hat{X})_{(i,i)} (\hat{X}^T \hat{X})_{(j,j)}}$$

LSIによって導入された唯一のSVDからの変更点は、 S_0 から小さい値の singular values を削除することである。このことは、小さい値に関する概念は実際には文献類似度を計算するときには不適当であり、それらを含めると適合性の判定の正確さが減少するということである。保持される特徴は m -次元空間内の文献ベクトルの位置にもっとも大きな影響を持つものである。

文献と文献の間の相関を調べる際に、相関行列 $X^T X$ を求めることが一般に行われるが、これは X の次元数を減らした行列 \hat{X} の相関行列

$$\hat{X}^T \hat{X} = DS^2 D^T$$

を求めて近似される。

\hat{X} の次元数が小さいことにより、 X_0 について計算する場合にくらべ効率的に処理することができる。問合せに対する文献のランキングや文献集合のクラスタリングなどにおいてもコサイン相関値や文献ベクトル間の距離の計算が一般に求められるが、次元数を減らした行列 \hat{X} を利用することに、少ない計算量で近似的な値を求めることが可能である [渡辺 96][Davi94][Pete95]。

3.2.3 單語マッチング

XX^T 行列は文献の単語頻度ベクトルの内積で構成されている。これは単語ベクトルと呼ばれ、これまでの文献ベクトルと区別される。 XX^T の各列は対応する X^T の列中の単語ベクトルと文献集合中の各単語のための単語ベクトルとの内積のベクトルである。

文献と文献の比較と同じように単語と単語の間の相関を調べる際に、相関行列 XX^T を求めることが一般に行われるが、これは X の次元数を減らした行列 \hat{X} の相関行列

$$\hat{X} \hat{X}^T = TS^2 T^T$$

を求めて近似される。

4. 部分構造の主題の相互関係を考慮に入れた文献検索

一つの文献をいくつかに分解して情報を得ることで、文献全体どうしで比較を行うよりも、部分構造の主題の相互関係を考慮にいれた検索の方が精度の高い検索ができるのではないかと考える。

そのため、一つの文献から一つのベクトルを作るだけでなく、一つの文献をいくつかの sentence, word ごとの塊に分け、その塊からベクトルを作る。この塊のことを segment 呼ぶ。文献がベクトルを一つだけもつよりも、たくさんのベクトルで表した方がその文献の特徴をとらえやすいのではないかという観点からこのアプローチを提案する。

また、そのベクトルの組合せと問合せのベクトルの組合せから検索を行う。その際、文献全体で話題になっている main topic と一部でしか話題にならない subtopic の両方から検索を行うために二つの問合せベクトルを作り検索を行う。検索を行う際には LSI を使用し、ベクトル検索における効率をはかる。

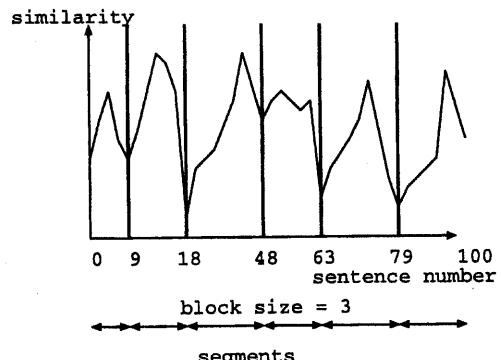


図 5 関連度曲線の例

4.1 segment を作る方法

一つの文献をいくつかの segment に分けなければならぬが、ここで二種類の方法を述べる。

- segment を作る方法 その 1
ヒューリスティックに構成する二つの方法を以下に述べる。

- あらかじめ著者が章、段落などに文献を分けておく。その章、段落を一つの segment にする。
- segment を作る人が文献、文章の内容を理解してそれに基づいて segment を作成する。この場合 segment は連続していないてもよいし、segment 同士がオーバーラップしてもよい。

- segment を作る方法 その 2

自動的に segment を作成する方法

- まず一つの文献 (n 文で構成) に対して Text-Tiling [A.H93][AP93] でブロックの大きさを 1 から n までのすべてのパターンを行い、 n 個の関連度曲線を作る。(ブロックの大きさが 2 というのは一つのブロックが 2 文からできていることを示す。)

- ii) 山と谷の関連度の差の大きさに閾値をもたせ n 個の関連度曲線からいくつかを pick up する。どのくらいの閾値をもたせるか、どのくらいの数の関連度曲線を選ぶのかは課題である。

- iii) 取り出した関連度曲線 (図 5) から山の部分をみつけ、それを一つの segment とする。

4.2 問合せ

4.2.1 問合せの方法

問合せの方法は main topic と subtopic の問合せベクトルを作り、そのベクトルの組合せを用いて検索を行う。

main topic を一つの文献全体と考えることもできるし、一つの文献の中の一部を main topic と考え、

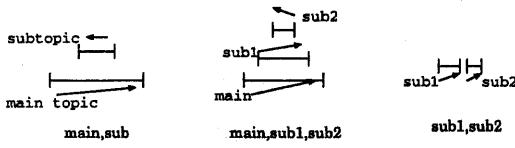


図 6 問合せの関係

さらにその中の一部を subtopic と考えることもできる。 $(main\ topic \sqsubset subtopic)$

さらに main topic と subtopic と二つに分けるだけでなく、main topic, subtopic1, subtopic2 のように三つに考えて三つの問合せベクトルの組合せであることもできる $(main\ topic \sqsubset subtopic1 \sqsubset subtopic2)$ 。

また、包含関係だけでなく並列な関係、subtopic1, subtopic2 の二つのベクトルの組合せでの問合せも可能である(図 6)。

具体的に問合せベクトルを作る方法は main topic, subtopic ともにユーザの知っている文献を入力してその文献からベクトルを作る。その際に 1 つの文献からではなく 2 つ以上の文献からもベクトルを構成できる。

またキーワードから問合せベクトルを作る。キーワードは 1 つだけでなく、いくつかを組み合わせることもできる。

また文献とキーワードの両方から問合せベクトルを構成することもできる。

したがって、main topic に対する問合せベクトルを文献の名前を入力して作り、subtopic に対する問合せベクトルをキーワードから作るよう、一方の問合せベクトルが文献名から、もう一つの問合せベクトルがキーワードからのように文献名とキーワードを組み合わせる問合せも行える。もちろん main topic に対する問合せベクトルを文献の名前とキーワードの組み合わせから作り、subtopic に対する問合せベクトルも文献名とキーワードから作るような検索も行える。

以上をまとめると

- main topic 問合せベクトル
 - 文献の組み合わせから構成
 - キーワードの組み合わせから構成
 - 文献とキーワードの組み合わせから構成
- subtopic 問合せベクトル
 - 文献の組み合わせから構成
 - キーワードの組み合わせから構成
 - 文献とキーワードの組み合わせから構成

例えば main topic が OS で subtopic が UNIX であるような文献を検索したい場合には main topic にオペレーティングシステム(文献名)と入力し、subtopic には The UNIX Super Text(文献名)と入力する。

また同じように main topic が OS で subtopic が UNIX であるような文献を検索したい場合に main topic にオペレーティングシステム(文献名), OS(キーワード)と入力し、subtopic には system V(キーワード), solaris(キーワード)と入力することによっても検索できる(図 7)。

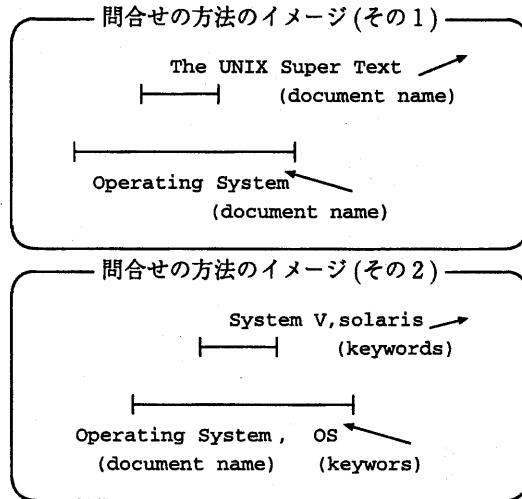


図 7 問合せのイメージ

しかし The UNIX Super Text(文献名)から作るベクトルと system V(キーワード), solaris(キーワード)から作るベクトルが同じものではないため検索結果は同じになるとは限らない、むしろキーワードの選び方にもよるが、違う検索結果になる可能性が高い。

4.3 LSI を使用したベクトルの比較

前章で述べたように LSIにおいて文献を比較するには文献ベクトルどうしの内積で構成される $d \times d$ 対称行列 $X^T X$ を考える。文献 i と j のコサイン相関値は次のように計算される。

$$\frac{(\hat{X}^T \hat{X})_{(i,j)}}{(\hat{X}^T \hat{X})_{(i,i)} (\hat{X}^T \hat{X})_{(j,j)}}$$

さらに一つの文献を表すベクトルから列ベクトル X_q からコサイン相関値として利用できる列ベクトルの内積への一次関数として X^T 行列を考えることができる。SVD を用いて X^T を拡張すると、 $X^T X_q = D_0 S_0 T_0^T X_q$ となる。 $D_0 S_0^{\frac{1}{2}}$ と $S_0^{\frac{1}{2}} T_0^T$ に分けて考える。まず $S_0^{\frac{1}{2}} T_0^T X_q$ について考える。この演算は問合せベクトル X_q を m -次元空間に射影する。基本的に T_0^T 行列は文献ベクトルを t -次元“文献ベクトル空間”から m -次元“文献特徴空間”に射影する。 $S_0^{\frac{1}{2}}$ は実対角行列である。したがって、 $S_0^{\frac{1}{2}}$ 行列は各特徴ごとに縮小することによって、文献ベクトルを文献特徴空間内の文献ベクトルに縮小して設計し直す。 $m \times t$ 行列 $S_0^{\frac{1}{2}} T_0^T$ は文献ベクトル空間から文献特徴空間への射影である。この特徴空間は文献類似度を計算する際には、いくつかの特徴が他より重要であるという考えを導入している。

文献特徴ベクトルが利用できるようになると、 $d \times m$ 行列 $D_0 S_0^{\frac{1}{2}}$ は我々が求める内積の計算をすること

に使用できる。これは文献特徴空間内の m -次元の $S_0^{\frac{1}{2}}T_0^TX_q$ ベクトルと, $D_0S_0^{\frac{1}{2}}$ の各行の内積を計算することである。 $D_0S_0^{\frac{1}{2}}$ の各行は、文献ベクトルとして解釈され、文献特徴空間に射影され、 X_q と同様にして設計し直される。

LSI では S_0 から小さい値を削除して k 次元にする。保持される特徴は m -次元空間内の文献ベクトルの位置にもっとも大きく影響を与えるものである。Deer-Wester らはこの選択が単語-文献内の概念をとらえ、単語の使用の変化によるノイズを排除すると述べている [DDF+90]。これは m -次元“文献特徴空間”から k -次元“文献概念空間”への射影である。

S_0 から小さい値を削除することで SVD を $\hat{X} = TSD^T$ に減少させる。ここで文献の内積を $\hat{X}X_q = DST^T$ で計算することができる。 $S^{\frac{1}{2}}T^T$ 行列が \hat{X} 行列のランクを m から k に減少させることによって単語の利用変化による影響を抑圧する。

この解析により $S^{\frac{1}{2}}T^T$ を単語ベクトルから概念ベクトルへの一次関数として考え、 $DS^{\frac{1}{2}}$ の各行を対応する文献集合中の関連する概念ベクトルとして考えることができる [FO95]。

5. 実験システムについて

奈良先端科学技術大学院大学の 1994, 1995, 1996 年度の修了生のうち、373 人の修士論文を対象に、本論文で提案している手法を適用した実験システム(図 8)を構築した。以下にこのシステムの処理の流れについて述べる。

- i) 一つの修士論文全体を main topic のセグメントとし、それぞれの修士論文を章ごとに分けたものを subtopic のセグメントとする。373 個の main topic 用のセグメントと 2394 個の subtopic 用のセグメントを合わせた 2767 個のセグメントを集合全体とする。
- ii) 各セグメントに語幹の切り出し処理(stemming)を施し、不要語(stop words)を除去する [FBY92]。
- iii) 文献集合で用いられている異なり単語(different term)の数を数え、文献(セグメント)-単語行列を作る。
- iv) 行列の特異値分解を行う。
- v) 今回の実験では main と sub の二階層の包含関係を用いた検索を行う。まずははじめに main topic と subtopic それぞれについての検索を行う。main topic の場合にはそれぞれの関連度と関連度の一一番高い main topic(文献)の関連度との割合を計算する。この値を m とする。sub の場合にもそれぞれの関連度と関連度の一一番高い subtopic(章)の関連度との割合を計算する。この値を s とする。

それぞれの文献の関連度

$$m = \frac{\text{それぞれの文献の関連度}}{\text{関連度の一番大きい文献の関連度}}$$

それぞれの章の関連度

$$s = \frac{\text{それぞれの章の関連度}}{\text{関連度の一番大きい章の関連度}}$$

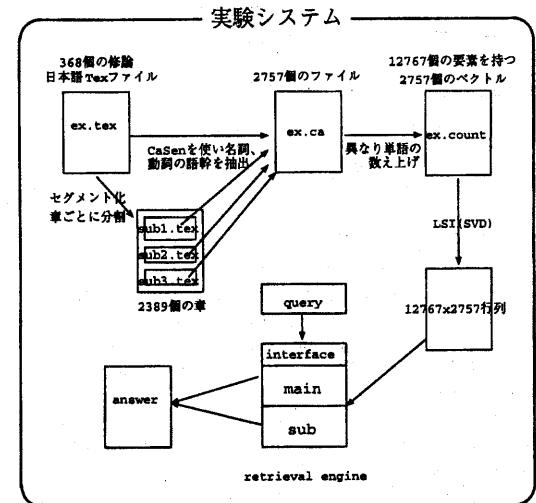


図 8 実験システム

ここで

$$sum = \alpha m + \beta s \quad (\alpha + \beta = 1)$$

という式を使い絞り込む。この α, β の値はユーザが指定する。さらにユーザは sum の上位何個の結果を得たいということも指定する。結果として文献の名前とその章番号と章のタイトルを返す。

5.1 問合せ例と期待する検索結果

maintopic が“映像データベースのための論理データモデルとその実装”(堀内優希さんの奈良先端大修論)、subtopic が“時間”(キーワード)、“軸”(キーワード)、“映像”(キーワード)、“演算”(キーワード)、“合成”(キーワード)、“場面”(キーワード)、“問合せ”(キーワード)で検索を行った場合には“映像データベースのための演算体系と圧縮情報構造に関する研究”(小川政行さんの奈良先端大修論)の第 3 章(映像の論理構造)が本システムにおいて期待する検索結果(図 9)である。

5.2 実験結果

上述の問合せを $\alpha = 0.7, \beta = 0.3$ とした場合の上位 3 個の結果を図 10 に示す。

5.3 考察

main topic(文献)に対する問合せの中で問合せの文献自身よりも関連度の高い文献があるがこれは LSI をしたときの近似によるものであると考えられる。またベクトルを構成する際の重み付けの仕方で文献の長さを考慮していない点も原因の一つであると考えられる。

6. 結論

6.1 まとめ

本研究では、文献が一つの main topic だけから成っているのではなく複数の subtopic も含んでいること

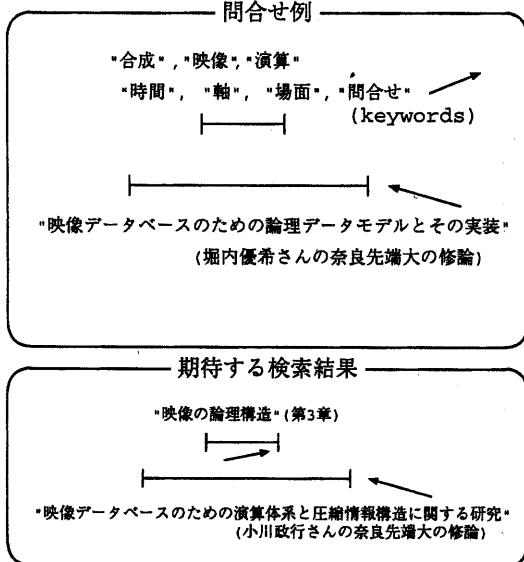


図 9 検索したい問合せと期待する検索結果

を考え、従来の検索システムより検索精度を高めることのできるような文献検索システムを提案した。

6.2 今後の課題

現在、実装した実験システムの評価、改良、評価の方法は難しいと考えられるが、いくつかのサンプルの問合せに対して、何人かの人間が検索を行いそれを答えにし、そのこたえをもとに精度と再現率を算出したいと考えている。

謝辞

本研究を進めるにあたって日頃から有意義な御指導、御討論をいただいた植村研究室の皆様に感謝いたします。

参考文献

- [A.H93] Marti A. Hearst. Textiling: a quantitative approach to discourse segmentation. 1993.
- [AP93] Marti A. Hearst and Christian Plunt. Subtopic structuring for full-length document access. In *SIGIR'93*, June 1993.
- [BDL95] M.W. Berry, S.T. Dumais, and T.A. Letsche. Computational methods for intelligent information access. In *Proceedings of Supercomputing*, 1995.
- [Davi94] David Ellis(原著), 細野公男(監訳), 斎藤泰則(訳), 鈴木志元(訳), 村上泰子(訳). 情報検索(認知的アプローチへの展望). 丸善, 1994.
- [DDF+90] Scott Deerwester, Susan T. Dumais, George W. Furnas, Thomas K. Landauer, and Richard Harshman. Indexing by latent semantic analysis. In *Journal of the American Society for Information Science*, Vol. 41-6, pp. 391-407, 1990.
- [FBY92] William B. Frakes and Ricardo Baeza-Yates, editors. *Information Retrieval - Data Structures & Algorithms* -. Prentice-Hall, 1992.

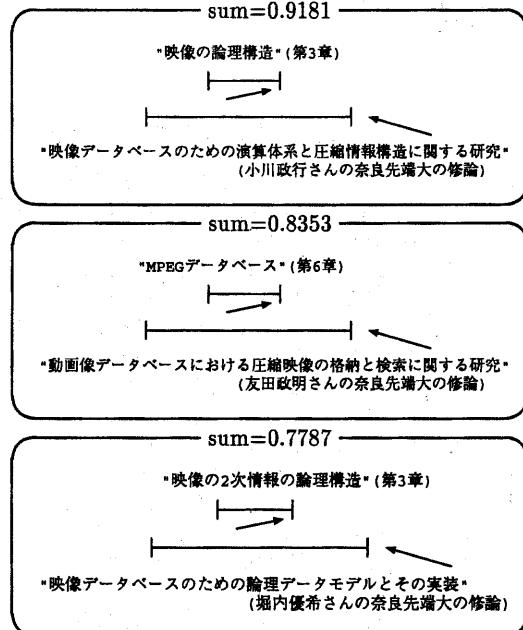


図 10 実験結果

- [FO95] Christos Faloutsos and Douglas W. Oard. A survey of information retrieval and filtering methods. Technical Report CS-TR-3514, University of Maryland, August 1995.
- [Mat97] Matsumoto lab. NAIST. Chashitu. <http://cactus.aist-nara.ac.jp/lab/nlt/chasen.html>, july 1997.
- [Oga96] Yasushi Ogawa. Effective & efficient document ranking without using a large lexicon. In *Proc. of the 22nd International Conference on Very Large Data Bases (VLDB)*, pp. 192-202, Bombay, September 1996.
- [Pete95] Peter Ingwersen(原著), 藤原鎮男(監訳), 細野公男(訳), 後藤智範(訳), 岸田和明(訳). 情報検索研究(認知的アプローチ). 凸版, 1995.
- [PTVF88] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C*. Cambridge University Press, 1988.
- [小川 96] 小川泰嗣. 情報検索の最近の動向. 日本ソフトウェア科学会 インターネットと情報検索 講習会資料, pp. 1-16, 1996.
- [森村 96] 森村一雄. 構造化文書からの概念構造の抽出とそれに基づく意味的検索に関する研究. Master's thesis, 奈良先端科学技術大学院大学, 1996. NAIST-IS-MT9451206.
- [渡辺 96] 渡辺正裕, 石川佳治, 吉川正俊, 植村俊亮. 多次元ベクトルの視覚的探索機能を有する情報検索. 情報処理学会データベースシステム研究会研究報告 96-DBS-109, pp. 7-12, July 1996.

付録

main topic(文献)に対する関連度

- 1, 1.000, 158581.06, 9351078
- 2, 0.887, 140670.90, 9351103
- 3, 0.883, 140080.62, 9451024
- 4, 0.735, 116480.11, 9451002
- 5, 0.725, 114908.00, 9351101
- 6, 0.601, 95290.48, 9551076
- 7, 0.585, 92746.14, 9451112
- 8, 0.557, 88269.25, 9451099
- 9, 0.556, 88187.35, 9551033
- 10, 0.556, 88097.77, 9351106
- 11, 0.555, 87945.91, 9551060
- 12, 0.509, 80674.30, 9451036
- 13, 0.500, 79219.63, 9451082
- 14, 0.480, 76188.19, 9351062
- 15, 0.465, 73708.18, 9351051
- 16, 0.457, 72401.03, 9451206
- 17, 0.449, 71145.83, 9551120
- 18, 0.439, 69578.25, 9451069
- 19, 0.434, 68768.68, 9451097
- 20, 0.433, 68618.28, 9351065

subtopic(章)に対する関連度

- 1, 1.000, 2938.70, 94510243
- 2, 0.576, 1692.10, 95510332
- 3, 0.526, 1545.10, 93511033
- 4, 0.468, 1375.00, 93511064
- 5, 0.451, 1326.70, 93510786
- 6, 0.413, 1213.00, 94510025
- 7, 0.407, 1195.50, 93511065
- 8, 0.353, 1036.50, 93511035
- 9, 0.331, 973.60, 93510785
- 10, 0.288, 846.80, 94511012
- 11, 0.261, 767.50, 95511204
- 12, 0.246, 723.30, 93511032
- 13, 0.238, 700.70, 94510024
- 14, 0.238, 698.10, 94510022
- 15, 0.230, 677.30, 94510246
- 16, 0.225, 662.40, 95511203
- 17, 0.222, 651.30, 94510242
- 18, 0.220, 646.30, 95510602
- 19, 0.220, 645.80, 94511013
- 20, 0.218, 640.00, 95510854