

大規模テストコレクションNTCIR-1における 検索課題の難易度に関する分析

江口 浩二[†] 栗山 和子[†] 神門 典子[†]

[†]国立情報学研究所

〒101-8430 東京都千代田区一ツ橋 2-1-2

e-mail: {eguchi,kuriyama,kando}@nii.ac.jp

あらまし テストコレクションの信頼性という観点から、検索課題の難易度あるいは検索課題セットの難易度分布の予測に関する必要性が高まっている。それを目的として、本稿では、NTCIR-1 (NACSIS Test Collection for Information Retrieval System-1) における検索課題、文書データベース、正解文書リストに関する種々の特徴を計量し、これらの相関性の分析に基づいたNTCIR-1のテストコレクションとしての特徴を報告する。

Analysis of the Topic Difficulty for NTCIR-1: NACSIS Test Collection for Information Retrieval System-1

Koji EGUCHI[†] Kazuko KURIYAMA[†] Noriko KANDO[†]

[†]National Institute of Informatics

2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430, Japan

e-mail: {eguchi,kuriyama,kando}@nii.ac.jp

Abstract Test collections are required, from the point of view of the reliance, to predict the topic difficulty or its distribution for a set of topics. For that object, this paper measures the various features on the search topics, the document database and the list of relevant documents, of which NTCIR-1 (NACSIS Test Collection for Information Retrieval System-1) is composed, and analyzes the correlations among them.

1 まえがき

近年、いくつかの評価ワークショップが実施され、注目を集めつつある [1, 2, 3]. 評価ワークショップとは、複数の参加者による複数のシステムを用いて、ある問題を解決する情報技術の性能を共通の基盤上で評価することにより、相互の特徴比較を行なうことを目指すものである。NTCIR は、情報検索システムに関する評価ワークショップで、共通のテストコレクションを用いて各システムの検索有効性 (retrieval effectiveness) を比較する。なお、テストコレクションとは、(1) 文書データベース、(2) 検索課題の集合、(3) 各検索課題に対す

る正解文書のリストからなる検索実験用データセットのことである [3, 4, 5].

本研究では、情報検索システム評価用ツールとしてのテストコレクションにおける検索課題の性質について考察する。検索課題に望ましい性質として、「自然さ」と「難易度のバランス」が挙げられる。検索課題の内容は、現実の検索過程においてシステムに与えられる検索要求と同様に自然なものであることが望ましい。NTCIR-1では、検索課題を自然なものとするを目指し、検索課題の作成を各領域の専門家 (大学院生以上の研究者) から収集した。ところで、TREC-6では、人間による検索課題の難易度の分類と検索結果の評価による数値的な

難易度に相関があるとは言えないことが報告されている [6]. しかしながら、検索課題が易し過ぎるものや難し過ぎるものに偏ることにより、テストコレクションがシステムの特定の振舞いに関する評価にしか利用できないことを避けることが必要とされ、難易度にバランスがとれていることが望ましい。本稿では、NTCIR-1における検索課題、文書データベース、正解文書セットに関する種々の特徴量を求め、それらと検索課題の難易度との相関性を分析することにより、検索課題の難易度の予測可能性について検討する。また、以上のような観点から NTCIR-1 のテストコレクションとしての特徴を報告する。

2 テストコレクション NTCIR-1 の概要

本章では、テストコレクション NTCIR-1 を構成する (1) 検索課題、(2) 文書データベース、(3) 正解文書リストについて概要を述べる¹[3, 4, 5].

2.1 文書データベース

国立情報学研究所が日本国内の 65 学協会の協力を得て、全国大会や研究会等の発表論文の要旨を集めた学会発表データベースから、約 33 万件の文書を選択し、各文書ごとに特定の項目を抽出したものが用いられる [3, 4, 5]. 約半数の文書は日英対訳であり、各レコードは、表題、著者名、会議録名、学会名、発表年月日、要旨、著者キーワードから成る。

2.2 検索課題

検索課題 (search topics) は、利用者の検索要求を一定の書式の自然言語で明文化したものであり、訓練用 30 課題、評価用 53 課題がある。これらは、各領域の専門家 (大学院生以上の研究者) から収集した。検索課題は主に、検索要求文 (description)、検索要求説明 (narrative)、タイトル (title)、概念リスト (concept) から成る。検索要求は、利用者の検索要求を記述した自然言語文であり、検索要求説明は、背景説明・検索の目的・正解判定基準・用語の定義などを含み、検索要求の背景情報を提供する。タイトルは、検索課題を簡潔に表現したものであり、概念リストは検索課題を表す重要な概念、キーワードのリストである。検索実験では以上のいずれの項目を使用しても良く、検索課題の記述を参照

¹NTCIR-1 には、(1),(2),(3)に加えて、タグ付きコーパスが含まれている。文書データの一部に、語構成要素まで考慮した詳細な形態素タグを付与したものである。(1),(2),(3)が情報検索システムの評価を目的としたものであるのに対して、タグ付きコーパスは自然言語処理の基礎的データを提供することを目的としている [7] が、詳細については割愛する。

しながら対話的にクエリの作成と入力を行なっても良いが、結果提出に際しては検索課題のどの項目を使用したか、対話型検索であるかどうかを報告する必要がある。次に、検索課題の例を示す。

```

<TOPIC q=0035>
<TITLE>
電子図書館
</TITLE>
<DESCRIPTION>
分散環境における電子図書館についての研究はないか。
</DESCRIPTION>
<NARRATIVE>
様々な人がネットワークを利用するようになり、ネットワークを介した情報提供サービスも数多く実現してきている。電子図書館もその一つでネットワークを通じて遠くにある電子化された出版物や画像を検索したり閲覧するというサービスが行なわれてきている。ネットワーク上の利用者や資源は基本的に分散して存在するものであり、電子図書館に保存される資料も複数の場所に分散してすることも考えられる。このように、電子図書館を分散環境で利用するために必要な技術について述べている論文が欲しい。ネットワークを通じての電子図書館の利用について知りたいので、所蔵品を電子化して検索できるシステムを設置しましたという論文は要求を満たさない。新しい研究を始めるにあたり、このトピックの現状を知りたい。
</NARRATIVE>
<CONCEPT>
<J.CONCEPT>
a. 電子図書館,
b. 分散環境, ネットワーク
</J.CONCEPT>
<E.CONCEPT>
a. Digital Library, Electronic Library, Virtual Library,
b. Distributed System, Distributed Environment, Network
</E.CONCEPT>
<A.CONCEPT>
c. Z39.50
</A.CONCEPT>
</CONCEPT>
<FIELD>
1. 電子・情報・制御
</FIELD>
</TOPIC>

```

2.3 正解文書

日本語の検索要求に対して日本語および英語の正解文書を検索する「随時検索」と日本語の検索要求に対して英語の正解文書を検索する「言語横断検索」の 2 つのタスクに対して、参加者が各自のシステムによる検索結果を提出し、それらに基づく評価が実施された。検索課題ごとに、各システムの検索結果文書セット (以下、提出結果と呼ぶ) の上位一定数の和集合に加えて、再現率重視の対話型検索の結果に対して正解判定を実施することで、網羅的な正解文書セットを収集する [8, 9, 10]. 正解判定に際しては、2 名のクロスチェックに基づく最終判定を行なった。また、判定は検索要求に「適合」、「部分的適合」、「不適合」の 3 段階で実施した。

3 検索課題の分析

実際の検索課題の難易度を特定するために、3.1 項に後述する通り、提出結果ごとの検索有効性 (retrieval effectiveness) を示す非補間平均精度 (non-interpolated

average precision) の中央値に基づいて検索課題をクラス分けする。このとき、対話型検索か非対話型検索か、あるいは、検索課題中のどの項目を使用したかによって、検索有効性の分布の傾向が異なることを避けるため、本稿では、「随時検索」の際に検索課題中の検索要求文のみを用いた²非対話型システムの提出結果を用いて、評価用検索課題³の分析を実施する。本稿では、実際の提出結果に基づいた難易度と、人間が判定した難易度、検索課題の各種特徴量との相関性を分析する。

3.1 検索課題ごとの提出結果に関する統計量

検索課題ごとの提出結果セットに関して、各種統計量を求める。

- 正解文書の総数 ($|REL|$),
- 提出結果セットにおける各提出結果の非補間平均精度の平均値 (*ave*), 標準偏差 (*stdev*), 中央値 (*med*), 歪度 (*skew*), 尖度 (*kurt*).

ところで、上記の中央値を検索課題の難易度の指標と見なし、中央値に基づいて検索課題を3つのクラスに等分割した。中央値の小さいクラスから順に *hard*, *middle*, *easy* とし、これらを難易度 (*diff*) とした。

以上の結果を表1に示す。

3.2 機能分類

検索課題を BMIR-J2 の機能分類に準拠し、以下の6種の機能を設定した。ただし、BMIR-J2[14] の「F1:基本機能」を「F0:基本機能」と「F1:シソーラス機能」に細分した[11]。

- F0. 基本機能: キーワードの存在確認、あるいは、それらの語の存在に関する論理式 (AND や OR など) の充足判定など。
- F1. シソーラス機能: キーワードのシソーラスによる拡張語の存在確認。および、それらの語の存在に関する論理式の充足判定。
- F2. 数値・レンジ機能: 数の数え上げや、数値などの範囲を正しく解釈する。数値の大小比較や単位の理解・変換なども含む。
- F3. 構文解析機能: 複数のキーワードの間の係受け関係を判断する (構文解析を行なう)。
- F4. 内容解析機能: 通常の構文解析に必要とされるよりも深い言語知識を利用する。文脈を理解することや、言葉の深い意味を理解することを含む。
- F5. 知識処理機能: 世界知識を利用する。常識的な判断や、蓄積された事実からの推論などを含む。

機能分類に関する判定は2名の図書館情報学を専攻する大学院生が実施した。判定の結果、該当する機能の有無をそれぞれ「○」あるいは「×」で、例えば、(F0, F1, F2, F3, F4, F5) = ○○○○○○ のように表現し、判定のパター

²このとき検索課題における「概念リスト」を用いた提出結果は除外した。

³訓練用検索課題の分析については[11]を参照されたい。

ンによって検索課題を次のような6グループに分類した。このとき、A,B,C,D,E,Fの順に必要とされる処理が多くなることから、一般的にはその順に検索要求に適切な検索の実行が困難であると考えられることができる。すなわち、3.1項で述べた実際の提出結果に基づく難易度とは異なり、人間により判定された難易度に関する一指標になり得る。機能分類の結果を表2中に *func* として示す。

A. 基本機能のみ:	○×××××
B. シソーラス機能のみ:	○×××××
C. 構文解析機能のみ:	○×××××
D. シソーラス機能と構文解析機能:	○×××××
E. シソーラス機能と内容解析機能:	○×××××
F. シソーラス機能と構文解析機能と内容分析機能:	○×××××

3.3 検索課題の特徴量

検索課題の特徴を示すであろう、以下の検索課題文の各種特徴量に着目する。

- 検索課題文の語数 (*word*), 文字数 (*char*),
- 検索課題文中の語に関する文書データベースにおける語頻度,
- 検索課題文中の語に関する文書データベースにおける文書頻度。

検索課題文に対して形態素解析⁴を実行して求めた形態素群をもとに、いくつかのルール[13]を用いて複合語を求め、名詞あるいは未知語と判定された形態素と複合語の合計数をもって語数とした⁵。なお、3.4項に述べる検索課題を構成する語 *tm* は、上記の形態素および複合語を示す。

以上のようにして求めた検索課題文の語数 (*word*), 文字数 (*char*) を表1に示す。検索課題文中の語に関する文書データベースにおける語頻度および文書頻度については、3.4, 3.5項にて述べる。

3.4 検索課題文中の語に関する文書データベースにおける語頻度

検索課題 *tp* に対して、以下のように $tf_rel(tp)$, $tf_db(tp)$, $tf_rat(tp)$ を定義し、それらの全検索課題に関する総和平均を求め、表2に示す。ただし、*TT* は検索課題文を構成する語集合、*tm* は *TT* の要素すなわち語である。*REL*, *DB* は、それぞれ正解文書セット、文書データベースを示す。また、 $tf(tm, A) =$ 文書セット *A* における語 *tm* の出現頻度、である。

⁴日本語形態素解析システムとして、『茶筌』[12]を用いた。

⁵ただし、後ほど式(12)に示す *idf* 等の計算のため、文書データベース中において検索課題文中の語が出現する頻度が0である場合、その語あるいは複合語は語数に含めなかった。

$$tf_rel(tp) = \frac{1}{|TT|} \sum_{tm \in TT} tf(tm, REL) \quad (1)$$

$$tf_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} tf(tm, DB) \quad (2)$$

$$tf_rat(tp) = \frac{1}{|TT|} \sum_{tm \in TT} \frac{tf(tm, REL)}{tf(tm, DB)} \quad (3)$$

3.5 検索課題文中の語に関する文書データベースにおける文書頻度

以下のように $df_rel(tp)$, $df_db(tp)$, $df_rat(tp)$ を定義し、それらの全検索課題に関する総和平均を求め、表 2 に示す。ただし、文書セット A 中における語 tm を含む文書の出現頻度を $df(tm, A)$ で示す。

$$df_rel(tp) = \frac{1}{|TT|} \sum_{tm \in TT} df(tm, REL) \quad (4)$$

$$df_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} df(tm, DB) \quad (5)$$

$$df_rat(tp) = \frac{1}{|TT|} \sum_{tm \in TT} \frac{df(tm, REL)}{df(tm, DB)} \quad (6)$$

3.6 TF-IDF

情報検索研究の成果の一つとして、TF-IDF なる語の重み付け法がしばしば用いられる [15, 16] が、これを検索課題の特徴量の計算に適用する。以下のように $ltf_db(tp)$, $idf_db(tp)$ を定義し、それらおよび $tfidf_db(tp)$, $ltfidf_db(tp)$ の全検索課題に関する総和平均を求め、表 2 に示す。

$$ltf_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} ltf(tm, DB) \quad (7)$$

$$idf_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} idf(tm, DB) \quad (8)$$

$$tfidf_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} tf(tm, DB) \cdot idf(tm, DB) \quad (9)$$

$$ltfidf_db(tp) = \frac{1}{|TT|} \sum_{tm \in TT} ltf(tm, DB) \cdot idf(tm, DB) \quad (10)$$

ただし、

$$ltf(tm, A) = \log(tf(tm, A)) + 1.0 \quad (11)$$

$$idf(tm, A) = \log(N/df(tm, A)) \quad (12)$$

4 分析結果

実際の提出結果の各種統計量とそれに基づいた難易度、機能分類、検索課題の各種特徴量との相関性を分析する。相関係数として順位に基づく Kendall の τ を用いた。順位相関係数とその両側有意水準を表 3 に示す。表 3 から以下の事実が観察される。

- 提出結果セットにおける平均精度の中間値 med の分布に関する歪度および尖度は、検索課題の難易度と明らかな正の相関がある。また、標準偏差は、検索課題の難易度と明らかな負の相関がある。このことから、検索課題の難易度が高くなるほど、検索結果の平均精度の分布は、低平均精度領域に偏りかつ尖ったものとなることが確認された。

- 文書データベースにおいて検索課題文を構成する語が出現する頻度 tf_db とそれらの語を含む文書の頻度 df_db とは、順位相関係数が約 0.80 と大きく、統計的検定の結果からも明らかな正の相関があった。また、それぞれの変形である ltf_db と idf_db については明らかな負の相関があり、同じく相関の度合は大きかった。従って、 tf_db と df_db (あるいは ltf_db と idf_db) は、独立な特徴量の対であるとはいいがたい。以下、 df_db をもって、これら文書データベースに対する頻度情報を代表する特徴量とする。

ところで、 df_db と検索課題の難易度とは、順位相関係数が約 0.33 と、度合はそれほど大きくないものの、統計的検定の結果から明らかな正の相関が認められる。このことは、文書データベース中に検索課題中の語を含む文書が多いほど、検索が難しいことを示唆する。しかしながら、このような検索課題文を構成する語に関する文書データベース中の頻度情報のみに基づいて、検索課題の難易度の分布を予測するのは容易でないと思われる。

- 実際の提出結果に基づく難易度 $diff$ と人間により判定された難易度を示す機能分類 $func$ とは明らかな相関が見られなかった。また、各機能ごとの検索課題の難易度と他の各種特徴量の相関、各難易度のクラスごとの機能分類と他の各種特徴量の相関を求めたが、特に明らかな事実は発見できなかった。

- 正解文書セットにおいて検索課題文を構成する語が出現する頻度 tf_rel と文書頻度 df_rel は、検索結果の難易度とは明らかな相関が認められなかった。一方、正解文書セットと文書データベースに関する、検索課題文を構成する語が出現する頻度の比率 tf_rat と文書頻度の比率 df_rat について

は、いずれも検索結果の難易度と明らかな相関が認められた。両者はいずれも正解文書セットにおける頻度が高く、文書データベースにおける頻度が低ければ、大きい値をとる特徴量である。しかしながら、*tf_rat*あるいは*df_rat*を求めるには、正解文書セットを必要とするため、検索課題の難易度の分布を予測するという目的にはそぐわない。

5 むすび

テストコレクションの信頼性という観点から、検索課題の難易度あるいは検索課題セットの難易度の分布を計測、予測することを目的として、NTCIR-1における検索課題、文書データベース、正解文書リストに関する種々の特徴量を求め、それらの相関性に関する分析を行った。分析結果により、NTCIR-1の一側面からの特徴が明らかになった。以下の通りである。

- 検索課題の難易度が高くなるほど、多様な情報検索手法による検索結果の平均精度分布は、低平均精度領域に偏りかつ尖ったものとなることが観察された。
- 実際の提出結果に基づく難易度と人間により判定された難易度を示す機能分類とは明らかな相関が見られなかった。
- 検索課題の難易度と、文書データベースにおける検索課題文中の語の頻度情報には、度合は大きくはないものの明らかな相関が認められた。このことは、現在までに提案されている多様な情報検索手法が、検索要求を記述する語の頻度情報に依拠する傾向にあることを示唆する。
- 実用的に、検索課題の難易度あるいは検索課題セットの難易度の分布を予測するには、検索課題の各種特徴量およびその組合せに関するなお一層の検討が必要であると考えらる。

謝辞

本研究は、日本学術振興会未来開拓学術研究推進事業「高度分散情報資源活用のためのユービキタス情報システム」(課題番号 JSPS-RFTF96P00602)による。

参考文献

- [1] Voorhees, E. and Harman, D. K.: Overview of the Seventh Text REtrieval Conference (TREC-7), *Proc. of 7th Text REtrieval Conference (TREC-7)*, NIST Special Publication 500-242, pp. 1-23 (1998).

- [2] 関根, 井佐原: IREX プロジェクト概要, IREX ワークショップ予稿集, pp. 1-5 (1999).
- [3] Kando, N., Kuriyama, K., Nozue, T., Eguchi, K., Kato, H. and Hidaka, S.: Overview of IR tasks at the First NTCIR Workshop, *Proc. of 1st NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*, Tokyo, Japan, pp. 11-44 (1999).
- [4] Kando, N., Kuriyama, K., Nozue, T., Eguchi, K., Kato, H., Hidaka, S. and Adachi, J.: The NTCIR Workshop: The First Evaluation Workshop on Japanese Text Retrieval and Cross-lingual Information Retrieval, *Proc. of 4th International Workshop on Information Retrieval with Asian Languages (IRAL'99)*, Taipei, Taiwan, pp. Inv-1-7 (1999).
- [5] 神門: 情報検索システムの評価プロジェクト: NTCIR ワークショップ, 情報処理, Vol. 41, No. 6, pp. 689-697 (2000).
- [6] Voorhees, E. and Harman, D. K.: Overview of the Sixth Text REtrieval Conference (TREC-6), *Proc. of 6th Text REtrieval Conference (TREC-6)*, NIST Special Publication 500-240, pp. 1-24 (1997).
- [7] Kageura, K., Yoshioka, M., Takeuchi, K., Koyama, T., Tsuji, K., Yoshikane, F. and Okada, M.: Overview of TMREC Tasks, *Proc. of First NTCIR Workshop on Research in Japanese Text Retrieval and Term Recognition*, Tokyo, Japan, pp. 415-416 (1999).
- [8] Gilbert, H. and Sparck Jones, K.: Statistical Bases of Relevance Assessment for the 'Ideal' Information Retrieval Test Collection, *BL R&D Report 5481* (1979).
- [9] 栗山, 江口, 野末, 神門: 大規模テストコレクション NTCIR-1 の構築 (1): プーリングと正解判定の分析, 情報処理学会第 59 回全国大会講演論文集 (3), No. 4P-1, pp. 105-106 (1999).
- [10] 栗山, 神門, 野末, 江口: 大規模テストコレクション構築のためのプーリングについて: NTCIR-1 の分析, 学術情報センター紀要, No. 12 (2000).
- [11] 栗山, 神門: 大規模テストコレクション構築について: NTCIR-1 の訓練用検索課題の分析, 情報処理学会研究報告, No. 99-FI-55, pp. 41-48 (1999).
- [12] 松本, 北内, 山下, 今一, 今村: 日本語形態素解析システム『茶筌』version 1.0 使用説明書 (1997).
- [13] Kando, N., Kageura, K., Yoshioka, M. and Oyama, K.: Phrase Processing Methods for Japanese Text Retrieval, *SIGIR Forum*, Vol. 32, No. 2, pp. 23-28 (1998).
- [14] 木谷, 小川, 石川, 木本, 中渡瀬, 芥子, 豊浦, 福島, 松井, 上田, 酒井, 徳永, 鶴岡, 安形: 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, 情報研報, DBS114-1, pp. 15-22 (1998).
- [15] Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley (1989).
- [16] Buckley, C., Salton, G. and Allan, J.: The Effect of Adding Relevance Information in a Relevance Feedback Environment, *Proc. of ACM-SIGIR'94*, pp. 292-298 (1994).

表 1: 検索課題ごとの提出結果に関する統計量と難易度

tp	REL	ave	stdev	med	diff	skew	kurt
0031	21	0.1984	0.1063	0.2099	middle	-0.2093	-0.4495
0032	23	0.0237	0.0103	0.0227	hard	-0.8221	0.3783
0033	162	0.1497	0.1351	0.1031	hard	1.9867	3.8642
0034	15	0.1473	0.1554	0.1159	hard	1.4436	1.6541
0035	32	0.3278	0.1825	0.4252	easy	-0.8424	-0.9062
0036	14	0.5473	0.2831	0.6177	easy	-0.9108	-0.4180
0037	65	0.0443	0.0483	0.0198	hard	1.4814	2.0974
0038	39	0.2161	0.1719	0.1602	middle	0.5720	-0.9874
0039	16	0.2866	0.1110	0.2951	middle	-0.9806	1.1246
0040	47	0.2426	0.1190	0.2437	middle	-0.0588	-0.4148
0041	16	0.2420	0.1396	0.3000	easy	-0.5114	-1.0984
0042	22	0.0905	0.0738	0.0794	hard	0.5330	-0.8500
0043	35	0.5810	0.2732	0.6863	easy	-1.1314	-0.0481
0044	15	0.1555	0.1357	0.0833	hard	0.7788	-0.4147
0045	18	0.0178	0.0163	0.0114	hard	1.5280	1.2156
0046	37	0.1668	0.1352	0.1346	middle	0.9170	-0.3253
0047	30	0.0298	0.0284	0.0198	hard	0.8456	-0.3362
0048	34	0.0092	0.0113	0.0044	hard	1.5827	2.0953
0049	20	0.0297	0.0346	0.0149	hard	1.5632	2.4572
0050	37	0.3173	0.1958	0.3356	easy	-0.0799	-0.1780
0051	20	0.0400	0.0408	0.0398	hard	0.9295	-0.1811
0052	9	0.2333	0.1410	0.2602	middle	-0.1126	-0.7624
0053	84	0.1472	0.1037	0.1612	middle	0.4551	0.2571
0054	584	0.0565	0.0429	0.0573	hard	0.4213	-1.0014
0055	40	0.2987	0.1327	0.3360	easy	-1.0173	0.1926
0056	68	0.1515	0.0792	0.1703	middle	-0.7704	-0.5874
0057	187	0.3286	0.1660	0.3915	easy	-0.9158	-0.3216
0058	10	0.6777	0.3116	0.8038	easy	-1.3600	0.5010
0059	61	0.2944	0.1569	0.3218	easy	-0.4605	-0.8614
0060	10	0.3022	0.1554	0.2958	easy	-0.3061	-0.1251
0061	24	0.1812	0.1117	0.2213	middle	-0.5002	-1.1493
0062	22	0.1398	0.0695	0.1329	middle	-0.3816	0.5757
0063	43	0.1064	0.0687	0.1044	hard	0.0987	-0.4297
0064	59	0.3481	0.1698	0.3492	easy	-0.4925	0.3238
0065	10	0.5128	0.3227	0.6149	easy	-0.6812	-1.2188
0066	33	0.5661	0.2822	0.6216	easy	-0.6189	-0.4104
0067	23	0.4341	0.2224	0.5315	easy	-1.0703	-0.1337
0068	52	0.1518	0.1089	0.1291	middle	0.6651	-0.6654
0069	12	0.0182	0.0244	0.0134	hard	4.0900	19.0359
0070	111	0.2934	0.1518	0.2805	middle	-0.0341	-0.7638
0071	13	0.1858	0.0915	0.1862	middle	0.2619	0.5090
0072	21	0.1341	0.0745	0.1645	middle	-0.7121	-0.6104
0073	11	0.1009	0.0934	0.0666	hard	0.2647	-1.7578
0074	17	0.2275	0.1417	0.2348	middle	-0.0480	-0.4838
0075	14	0.1682	0.1149	0.2046	middle	0.5540	1.8577
0076	17	0.1357	0.1022	0.1235	hard	0.7600	-0.3330
0077	6	0.2950	0.1665	0.3278	easy	-0.4446	-0.6095
0078	6	0.5964	0.2938	0.6778	easy	-1.0385	0.6047
0079	10	0.5542	0.2903	0.4721	easy	-0.2619	-0.8745
0080	16	0.1611	0.1262	0.1687	middle	0.8830	1.3761
0081	9	0.0885	0.1133	0.0655	hard	2.7200	9.6553
0082	9	0.4942	0.2193	0.5741	easy	-1.0186	-0.1256
0083	36	0.0270	0.0460	0.0111	hard	3.4642	13.9048

表 2: 検索課題に関する各種特微量

tp	diff	func	word	char	tf_rel	df_rel	tf_db	df_db	if_rat	df_rat	ltf_db	idf_db	tfidf_db	ltfidf_db
0031	middle	D	12	43	32.08	9.75	41266.50	21327.83	0.02	0.02	9.58	4.74	86688.23	29.63
0032	hard	D	4	37	22.75	8.50	17097.00	12177.25	0.01	0.01	9.90	4.30	52851.65	34.12
0033	hard	F	9	39	232.78	59.44	27244.00	13523.89	0.12	0.12	9.30	5.02	76560.18	28.40
0034	hard	F	9	32	19.00	5.44	50525.22	29366.33	0.00	0.00	10.52	3.75	95838.99	30.17
0035	easy	D	7	25	63.14	18.43	55997.86	37209.43	0.07	0.05	10.27	4.13	86537.69	31.92
0036	easy	G	7	27	35.29	8.71	14050.71	7321.71	0.04	0.03	9.02	5.29	43451.48	36.91
0037	hard	D	6	32	68.83	19.00	19546.00	10766.50	0.02	0.01	9.85	4.39	52411.51	36.82
0038	middle	D	10	32	56.70	16.90	34654.30	19616.60	0.05	0.05	9.27	5.10	66800.05	33.84
0039	middle	D	6	26	22.33	6.17	13438.83	9081.50	0.12	0.13	9.02	5.33	41619.40	34.39
0040	middle	B	8	27	74.00	21.63	46904.88	27674.00	0.03	0.03	10.27	4.04	89282.46	32.09
0041	easy	F	9	26	34.22	8.33	23039.33	12294.56	0.02	0.01	9.65	4.70	65702.75	34.99
0042	hard	H	12	34	53.08	10.00	96469.67	47679.00	0.08	1.00	10.52	3.74	139419.45	23.28
0043	easy	D	7	31	71.00	20.71	86900.00	40290.57	0.07	0.05	10.40	4.01	117330.96	30.40
0044	hard	H	6	26	22.17	7.00	49118.17	29268.17	0.00	0.00	11.08	3.23	107161.02	29.85
0045	hard	F	7	25	21.71	7.14	44954.00	20453.71	0.00	0.00	10.26	4.20	99888.88	32.97
0046	middle	F	7	28	28.29	9.86	19712.00	12712.86	0.07	0.07	9.00	5.41	47890.68	36.47
0047	hard	F	16	49	22.50	9.19	46085.81	27935.44	0.01	0.01	9.80	4.49	84361.19	29.71
0048	hard	F	6	19	24.00	9.17	25049.67	14458.50	0.01	0.01	9.75	4.49	64802.64	34.80
0049	hard	F	6	22	23.67	9.17	49860.67	33005.67	0.01	0.00	10.59	3.51	78412.69	30.44
0050	easy	F	5	15	50.60	21.80	10732.40	8053.40	0.13	0.11	9.47	4.73	35400.65	36.70
0051	hard	B	7	41	11.29	6.86	46930.14	31138.14	0.01	0.01	9.68	4.53	61134.21	34.37
0052	middle	D	2	20	9.00	3.00	1547.00	890.00	0.22	0.33	5.99	8.44	8148.04	32.42
0053	middle	F	8	32	68.13	26.88	14945.50	9639.25	0.02	0.01	9.42	4.81	40800.16	36.62
0054	hard	I	6	20	1021.17	284.33	63091.83	27354.67	0.07	0.07	10.14	4.39	117905.97	30.09
0055	easy	F	17	72	54.76	15.41	24148.94	15465.53	0.04	0.03	9.20	5.27	51879.85	35.55
0056	middle	D	16	60	57.50	18.25	26860.75	17667.13	0.04	0.03	9.02	5.21	48321.05	35.40
0057	easy	D	12	53	160.83	50.58	16286.92	9259.92	0.16	0.15	8.75	5.45	46103.55	34.41
0058	easy	F	5	32	12.60	4.60	9619.60	7547.80	0.19	0.17	8.72	5.46	30277.47	35.88
0059	easy	F	7	37	50.14	16.29	17584.57	11973.86	0.04	0.03	8.46	5.60	39166.70	33.35
0060	easy	G	9	27	13.44	3.22	19852.67	10916.56	0.11	0.08	6.99	7.18	44941.64	27.38
0061	middle	F	12	35	44.00	10.75	22383.83	12464.33	0.03	0.03	9.53	4.74	62973.11	33.12
0062	middle	G	5	26	34.80	8.40	7222.60	2529.40	0.05	0.03	8.39	6.13	28081.44	42.61
0063	hard	H	12	30	43.00	12.75	53521.42	32324.00	0.10	0.10	9.32	5.00	81586.37	28.99
0064	easy	F	8	30	116.13	26.13	8743.50	3674.75	0.08	0.05	9.05	5.39	35523.81	39.56
0065	easy	F	8	26	30.75	6.88	22844.75	9662.38	0.01	0.00	10.29	4.20	72801.62	36.95
0066	easy	D	12	32	73.17	20.92	28912.00	18020.25	0.05	0.04	9.51	4.90	54529.16	36.55
0067	easy	D	13	51	43.31	10.46	10834.08	7107.77	0.04	0.03	8.81	5.51	35013.91	35.72
0068	middle	D	11	41	15.45	9.45	14724.09	10380.64	0.08	0.10	7.60	6.51	34116.54	31.62
0069	hard	D	7	32	7.71	3.71	13431.29	8597.71	0.00	0.00	8.23	5.92	34977.90	32.36
0070	middle	D	7	19	229.43	61.14	30662.14	16855.29	0.05	0.04	10.27	4.19	72236.83	35.44
0071	middle	D	8	31	8.63	3.88	18985.38	15273.00	0.04	0.03	6.82	7.27	30235.24	28.68
0072	middle	D	7	30	24.86	7.71	46703.00	27518.43	0.02	0.01	9.25	4.99	76286.70	29.81
0073	hard	F	9	31	9.78	4.00	25863.56	13339.56	0.01	0.02	8.87	5.46	57625.64	35.68
0074	middle	D	12	46	26.00	9.75	27772.25	11098.92	0.02	0.02	9.36	5.09	72015.11	35.83
0075	middle	F	10	37	15.10	5.50	13006.10	7729.40	0.02	0.01	9.30	4.95	44081.26	35.59
0076	hard	D	6	22	19.50	6.17	17362.50	9451.17	0.02	0.02	8.91	5.43	48805.08	35.16
0077	easy	D	6	27	9.67	2.67	23281.00	11567.33	0.02	0.03	8.45	6.09	44834.43	38.27
0078	easy	F	12	44	15.25	2.75	21206.00	8815.83	0.06	0.04	8.57	5.87	54500.90	37.41
0079	easy	F	10	37	22.50	5.90	19448.90	8981.40	0.12	0.10	7.82	6.46	43803.49	31.40
0080	middle	D	11	43	20.18	5.64	27180.73	12746.36	0.01	0.03	9.18	5.11	174740.95	43.10
0081	hard	F	11	37	6.18	2.45	23931.00	10511.82	0.01	0.02	8.96	5.26	147850.13	41.49
0082	easy	D	5	20	28.20	5.80	20401.40	7559.60	0.19	0.13	8.58	6.16	61804.75	37.05
0083	hard	F	11	38	73.45	9.55	29643.18	13601.64	0.04	-0.02	9.16	5.10	72090.62	31.98

表 3: 検索課題に関する各種特徴量および難易度の間の順位相関

	diff	func	REL	ave	stdev	med	skew	kurt	word	char	tf_rel	df_rel	tf_db	df_db	tf_rat	df_rat	tfidf	ltfidf	
diff		0.094 0.087 0.443 0.424	-0.798 -0.688 0.000 0.000	-0.824 0.655 0.000 0.000	0.227 -0.068 0.035 0.548	-0.014 -0.142 0.189 0.562	0.063 0.296 0.006 0.002	0.333 -0.421 0.000 0.001	-0.362 0.312 0.004 0.007	-0.291									
func	0.094 0.443		-0.032 -0.090 0.771 0.401	-0.023 -0.114 0.829 0.287	-0.114 0.110 0.307 0.954	0.110 0.006 0.954 0.795	0.029 -0.133 0.224 0.746	0.035 -0.024 0.822 0.888	0.015 -0.026 0.810 0.802	0.028 -0.002 0.987 0.449	0.081 -0.058 0.589 0.589	0.170 0.139 0.065 -0.116	0.139 0.065 -0.116 -0.222	0.086 0.163 0.494 0.222	0.389 -0.203 0.196 0.038	0.330 -0.134 0.221 0.019	0.330 -0.134 0.221 0.019	0.156 0.019	
REL	0.087 0.424	-0.032 0.771		-0.119 0.211	-0.167 0.080	-0.119 0.214	0.064 0.504	-0.062 0.514	0.113 0.258	0.091 0.348	0.639 0.000	0.780 0.000	0.122 0.200	0.195 0.040	0.170 0.086	0.139 0.163	0.065 0.163	-0.116 0.494	-0.222 0.222
ave	-0.798 0.000	-0.090 0.401	-0.119 0.211		0.795 0.000	0.901 0.000	-0.592 0.000	-0.181 0.055	0.060 0.541	0.004 0.969	0.110 0.247	0.047 0.618	-0.193 0.041	-0.266 0.005	0.424 0.000	0.389 0.000	-0.203 0.032	0.196 0.038	
stdev	-0.688 0.000	-0.023 0.829	-0.167 0.000	0.795 0.000	0.736 0.000	-0.430 0.000	-0.182 0.054	0.045 0.648	-0.007 0.939	0.071 0.452	0.012 0.902	-0.147 0.119	-0.246 0.009	0.336 0.001	0.330 0.001	-0.134 0.001	0.221 0.156	0.221 0.019	
med	-0.824 0.000	-0.114 0.287	-0.119 0.214	0.901 0.000	0.736 0.000		-0.669 0.000	-0.200 0.035	0.058 0.556	0.009 0.926	0.115 0.225	0.041 0.667	-0.201 0.034	-0.274 0.004	0.389 0.000	0.358 0.000	-0.214 0.024	0.221 0.019	
skew	0.655 0.000	0.110 0.307	0.064 0.504	-0.592 0.000	-0.430 0.000	-0.669 0.000		0.244 0.010	-0.015 0.883	-0.019 0.847	-0.114 0.228	-0.050 0.597	0.174 0.066	0.186 0.050	-0.339 0.001	-0.288 0.003	0.199 0.036	-0.160 0.091	
kurt	0.227 0.035	0.006 0.954	-0.062 0.514	-0.181 0.055	-0.182 0.054	-0.200 0.035	0.244 0.010		-0.080 0.416	0.082 0.393	-0.127 0.179	-0.140 0.139	-0.126 0.182	-0.112 0.237	-0.116 0.237	-0.161 0.102	-0.052 0.581	0.094 0.319	
word	-0.068 0.548	0.029 0.795	0.113 0.258	0.060 0.541	0.045 0.648	0.058 0.556	-0.015 0.883	-0.080 0.416		0.609 0.000	0.121 0.902	0.125 0.667	0.171 0.902	0.148 0.667	-0.022 0.000	0.018 0.000	0.148 0.000	-0.093 0.000	
char	-0.014 0.902	0.133 0.224	0.091 0.348	0.004 0.969	-0.007 0.939	0.009 0.926	-0.019 0.847	0.082 0.393	0.609 0.000		0.012 0.902	0.042 0.667	-0.010 0.92	0.017 0.860	-0.061 0.545	-0.021 0.834	-0.005 0.957	-0.020 0.835	
tf_rel	-0.142 0.189	0.035 0.746	0.639 0.000	0.110 0.247	0.071 0.452	0.115 0.225	-0.114 0.228	-0.127 0.179	0.121 0.221	0.012 0.902		0.797 0.000	0.160 0.090	0.140 0.139	0.264 0.007	0.196 0.047	0.144 0.127	-0.024 0.800	
df_rel	-0.063 0.562	-0.024 0.822	0.780 0.000	0.047 0.618	0.012 0.902	0.041 0.667	-0.050 0.597	-0.140 0.139	0.125 0.207	0.042 0.667	0.797 0.000		0.149 0.116	0.190 0.045	0.231 0.019	0.189 0.056	0.116 0.223	-0.075 0.429	
tf_db	0.296 0.006	0.015 0.888	0.122 0.200	-0.193 0.041	-0.147 0.119	-0.201 0.034	0.174 0.066	-0.126 0.182	0.171 0.084	-0.010 0.92	0.160 0.090	0.149 0.116		0.803 0.000	-0.209 0.034	-0.152 0.123	0.714 0.000	-0.338 0.000	
df_db	0.333 0.002	-0.026 0.810	0.195 0.040	-0.266 0.005	-0.246 0.009	-0.274 0.004	0.186 0.050	-0.112 0.237	0.148 0.135	0.017 0.860	0.140 0.139	0.190 0.045	0.803 0.000		-0.207 0.035	-0.158 0.109	0.569 0.000	-0.417 0.000	
tf_rat	-0.421 0.000	0.028 0.802	0.170 0.086	0.424 0.000	0.336 0.001	0.389 0.000	-0.339 0.001	-0.116 0.237	0.022 0.827	-0.061 0.545	0.264 0.007	0.231 0.019	-0.209 0.034	-0.207 0.035		0.871 0.000	-0.274 0.005	-0.033 0.740	
df_rat	-0.362 0.001	-0.002 0.987	0.139 0.163	0.389 0.000	0.330 0.001	0.358 0.000	-0.288 0.003	-0.161 0.102	0.018 0.863	-0.021 0.834	0.196 0.047	0.189 0.056	-0.152 0.123	-0.158 0.109	0.871 0.000		-0.195 0.048	-0.025 0.798	
tfidf_db	0.312 0.004	0.081 0.449	0.065 0.494	-0.203 0.032	-0.134 0.156	-0.214 0.024	0.199 0.036	-0.052 0.581	0.148 0.135	-0.005 0.957	0.144 0.127	0.116 0.223	0.714 0.000	0.569 0.000	-0.274 0.005	-0.195 0.048		-0.232 0.014	
ltfidf_db	-0.291 0.007	-0.058 0.589	-0.116 0.222	0.196 0.038	0.221 0.019	0.221 0.019	-0.160 0.091	0.094 0.319	-0.093 0.349	-0.020 0.835	-0.024 0.800	-0.075 0.429	0.338 0.000	-0.417 0.000	-0.033 0.740	-0.025 0.798	-0.232 0.014		

各セルの上段:Kendall の τ , 下段:両側有意水準, 強調文字:相関係数が1%水準で有意(両側), 下線:相関係数が5%水準で有意(両側).