

Web サーチエンジンの新しい評価手法

大塚 崇志[†] 山名 早人[‡]

[†] 早稲田大学大学院理工学研究科 〒169-855 東京都新宿区大久保 3-4-1

[‡] 早稲田大学 〒169-855 東京都新宿区大久保 3-4-1

E-mail: [†] t-ohtsuka@toki.waseda.jp, [‡] yamana@yama.info.waseda.ac.jp

あらまし 現在, Web 空間は増加の一途を辿り, 目的の情報を得ることは容易ではない. そのため検索エンジンの必要性が高まり, 性能の向上が求められている. 既存の方法では適合文書による単純な得点の評価を主な性能評価の基準としていたが, 検索エンジンの性能を評価するための方法として最善のものであるとは言い難い. 本報告では, ユーザの立場に立った文書の連続性に基づく新たな評価手法 SS を提案する. また, NTCIR-3 で使用された DCG や MWRR と SS という評価法を比較する. 更に総検索結果数と不適合文書の割合との相関を考慮し, 実際的な性能評価法を提案する.

キーワード 情報検索, 評価方法, Web 検索エンジン

New Evaluation Methods for Web Search Engine

Takashi Ohtsuka[†] Hayato Yamana[‡]

[†] Graduate school of Science and Engineering, Waseda University 3-4-1 Okubo Shinjuku-ku Tokyo, 169-8555 JAPAN

[‡] School of Science and Engineering, Waseda University 3-4-1 Okubo Shinjuku-ku Tokyo, 169-8555 JAPAN

E-mail: [†] t-ohtsuka@toki.waseda.jp, [‡] yamana@yama.info.waseda.ac.jp

Abstract Nowadays, it is not easy to obtain the target information because the Web space is increasing. Therefore, the search engine is necessary and its performance improvement is required. However, in the conventional methods, the performance evaluation is mainly based on simple score of relevant documents, conventional evaluation methods for search engine performance are not best ones. In this paper, we propose new evaluation method SS that is based on documents sequence from a user's viewpoint. Moreover, as a we compare SS with DCG and MWRR that evaluation methods are used in NTCIR-3. Moreover, we propose the practical performance evaluation methods by considering the correlation between total search results number and irrelevant documents rate.

Keyword Information Retrieval, Evaluation Method, Web Search Engine

1 はじめに

現在, Web 空間は増加の一途を辿り, 目的の情報を得ることは容易ではない. そのため検索エンジンの必要性が高まり, 性能の向上が求められている. 既存の方法では適合文書による単純な得点の評価を主な性能評価の基準としていたが, 検索エン

ジンの性能を評価するための方法は最善のものであるとは言い難い. 本報告では, ユーザの立場に立った文書の連続性に基づく新たな評価手法 SS(Sequence Score)を提案する. また, NTCIR-3[1]で使用された DCG (Discounted Cumulative Gain) [2]や MWRR (Mean Weighted Reciprocal Rank) [2]

と SS という評価法を比較する。更に総検索結果数と不適合文書の割合との相関を考慮し、実際的な性能評価法を提案する。

本報告では、以下第 2 節において評価尺度の概要と NTCIR での試みについての概要を紹介する。第 3 節において新たな評価手法 SS を提案し、第 4 節で検証を行う。第 5 節で今後の課題を述べる。

2 NTCIR[1]

情報検索の研究・開発は 1950 年代から 60 年代にかけて本格化し、近年ではインターネットの検索エンジンの開発・普及を経て今日に至っている。この発展は検索実験に基づいていると言われている。その中でも 1990 年代前半に米国で開始された TREC (Text REtrieval Conference) [3]が有名である。このプロジェクトによって近年の情報検索技術が大きく発展した。更に現在では、インターネットの爆発的な普及・発展によって情報検索への関心が高まっている。

このような背景を基にし、日本において同様の検索実験プロジェクトとして NTCIR (NII/NACSIS Test Collection for Information Retrieval) が開始された。NTCIR の特徴として、検索対象が日本語テキストであるということが挙げられる。これは日本語に対する情報検索には未解決な問題が多く、更なる技術発展が必要とされているという点で意義のある試みとなっている。

NTCIR は TREC における検索実験を参考にし、1998 年に開始され、現在までに 3 回の実験が実施されている。各回の実験では、いくつかの個別的な研究課題が設定されていて、最新の NTCIR-3 において研究課題として Web 検索が初めて導入された。

2.1 Web 検索の特徴

World Wide Web の普及に伴い、Web 検索エンジンの重要性は非常に高まってきた。Web 上の情報の単位となるものが Web 文書であり、主に Web 文書を対象とした情報検索が Web 検索と呼ばれる。Web 文書には従来の情報検索が扱ってきたような新聞記事や論文などとは異なる特徴がある。具体例としては、ジャンルの多様性 (論文, カタログ, 日記などが混在), 表現の多様性 (タグを用いたレイアウト, 表や画像), リンクによる参照 (ハイパーリンクによるページ間の参照) などがある。

Web 検索システムではこれらの特徴に対応する様々な手法が提案され、検索エンジンに応用されているが、研究課題がまだ多く残されている。

Web 検索技術を実際に研究する上では様々な問題が存在する。まず、Web 検索システム全体の検索性能を決定する要因が検索手法以外に、収集ページ数、ハードウェア性能、ページデザインなど多く存在するため、比較評価を客観的に行うことが困難である。また、検索対象の Web 空間が巨大であるため、計算機資源が足りない、実用化を見据えた研究が行いづらいといったことも挙げることができる。更に利用者の検索要求と検索結果に対する評価のモデルが十分に研究されていないため、客観的評価のための基準が欠如している。具体例として、検索結果の

適合判定ではテキストだけを見るのか画像なども評価するのか、リンク先のページも見るとか、内容の信憑性や重要性をどのように評価するのかというように、従来の情報検索の評価尺度では評価できない事項が多く存在する。

2.2 NTCIR-3 での Web 検索

Web 検索は NTCIR で初めて取り上げる研究課題であるため、従来の情報検索システムの評価法の考え方を継承しつつ、可能な限り Web 文書や Web 検索の特徴に対応するという方針で実施された。

次に NTCIR-3 での Web 検索において用いられた評価尺度を説明する。

2.3 評価尺度

本節では評価尺度の概要と共に、NTCIR-3 で実際に使用された評価尺度を説明する。

2.3.1 精度と再現率

検索が成功したかどうかの伝統的な評価法として、精度と再現率が存在する。

検索課題を 1 つ固定し、文書集合に含まれる検索課題に対する適合文書の総数を R とする。ここで、あるシステムを使用して検索を実行し、 n 件の文書を出したとき、その中に含まれる適合文書の数が r 件であったとする。このとき精度は r/n 、再現率は r/R と定義される。

上述の評価尺度では適合度順出力評価を行うことができないため、適合度順出力の評価のために以下のような指標が存在する。

- 精度 (λ): 上位 λ 件の文書集合における精度
- 再現率 (λ): 上位 λ 件の文書集合における再現率
- R 精度: 上位 R 件の文書集合における精度 (R は適合文書総数)
- 平均精度: 最上位の文書から順に調べ、適合文書が出現した時点でそれぞれ精度を計算し、最後にそれらを平均化したもの

上記の評価尺度に加え、NTCIR-3 では次の評価尺度が使用された。

2.3.2 Discounted Cumulative Gain[2]

DCG (Discounted Cumulative Gain) は多段階適合性に適した評価基準であり、 $d(i)$ は i 番目にランクされた文書を示し、 $g(i)$ は文書 $d(i)$ の得点、 $dcg(i)$ は上位 i 番目までにランクされているページの累積得点を示し、以下のように定義される。

$$dcg(i) = \begin{cases} g(1) & \text{if } i=1 \\ dcg(i-1) + g(i) / \log_c(i) & \text{otherwise} \end{cases} \quad (1)$$

$$g(i) = \begin{cases} h & \text{if } d(i) \in H \\ a & \text{if } d(i) \in A \\ b & \text{if } d(i) \in B \end{cases} \quad (2)$$

ここで、 H , A , B はそれぞれ高適合文書、適合文書、部分適合文書の集合であり、多段階の適合判定が可能である。 h , a , b

はそれぞれに割り当てられている重みであり、等式(2)で示される得点の重要性が次の2適合レベルにおいて定められた。

適合レベル1: $(h, a, b)=(3, 2, 0)$

適合レベル2: $(h, a, b)=(3, 2, 1)$

NTCIRでの実行では、等式(1)で $c=2$ とした。

2.3.3 Mean Weighted Reciprocal Rank[2]

検索質問ごとに、最初に出現した正解の順位の逆数を求め、それらを全質問にわたって平均することで MRR(Mean Reciprocal Rank)を定義した。次の wrr (weighted reciprocal rank) を全検索質問にわたって平均することで MWRR(Mean Weighted Reciprocal Rank)を求める。

$$wrr(m) = \max(r(m))$$

$$r(m) = \begin{cases} \delta_h / (i-1 / \beta_h) & \text{if } (d(i) \in H \wedge 1 \leq i \leq m) \\ \delta_a / (i-1 / \beta_a) & \text{if } (d(i) \in A \wedge 1 \leq i \leq m) \\ \delta_b / (i-1 / \beta_b) & \text{if } (d(i) \in B \wedge 1 \leq i \leq m) \\ 0 & \text{otherwise} \end{cases}$$

ここで $d(i)$ は2.3.2と同様に i 番目にランクされた文書を示し、 m は評価に考慮する実行結果のランクを示す。また

$$\delta_h \in \{1, 0\}, \delta_a \in \{1, 0\}, \delta_b \in \{1, 0\}, \beta_b \geq \beta_a \geq \beta_h > 1$$

はそれぞれが満たす重み係数である。 δ_x と β_x の組み合わせを次のように設定し、2つの適合レベルをサポートした。簡単のため、 β_x の値は十分に大きなものと仮定した。

レベル1:

$$(\delta_h, \delta_a, \delta_b) = (1, 1, 0), (\beta_h, \beta_a, \beta_b) = (\infty, \infty, \infty)$$

レベル2:

$$(\delta_h, \delta_a, \delta_b) = (1, 1, 1), (\beta_h, \beta_a, \beta_b) = (\infty, \infty, \infty)$$

NTCIRでの評価では、MWRRを算出する際に、 m を5, 10, 15, 20と設定した。

NTCIR-3では上述の評価尺度を用いることによって、情報検索システムの性能評価を行った。

3 提案する新しい評価手法

先に述べたとおり、現在の検索システムの性能を評価する指標としては、主に精度と再現率、そして NTCIR-3 で導入された DCG と MWRR が使用されている。しかしこの評価指標は、文書集合が異なる特徴を持つ実際の Web 文書に対して本当に有効な評価指標であるとは言い難い。また、実際の非常に巨大な Web 空間での性能評価が可能となるような評価指標が必要であると考えられる。そこで実際の Web 空間への検索要求に対して有用な検索ができたかどうかの判定を行う新たな評価手法として次に示す手法を提案する。

3.1 提案

ユーザが実際に Web 空間での検索を行う場合、通常は多くても検索結果の上位 50 件程度までしか検索結果を参照しないと考えられる[4]。つまり上位 50 件以下に検索された結果は意味を持たないと考えることができる。更に Web 文書の膨大さを考慮すれば、全適合文書数が不明なので、再現率の考慮は意味を成さないと考えられる。

そこで検索結果の評価としては、上位にランキングされた文書の精度を評価対象とする。またユーザが検索結果を参照する場合、ランキングの最上位から降順に結果を見ていくのが普通である[5]。そのためユーザがより参照し易い、ランキング上位部分に適合文書が並ぶことが望ましい。また検索結果の適合文書と不適合文書が混在してランキングされている場合よりも、適合文書は適合文書同士、不適合文書は不適合文書どうし隣り合っていたほうが、ユーザとしては検索結果の満足度が高くなると考えられる。これは例えば適合文書が隣り合っていれば、ユーザの入力した検索クエリが有効であると判断可能であるし、不適合文書が隣り合っていれば、入力した検索クエリを直ぐに変更することができると考えられるためである。

以上のような考えから、次のような評価尺度を提案する。

3.2 新しい評価尺度 Sequence Score

新しい評価尺度として SS(Sequence Score)を次のように定義する。

$s(i)$ を i 番目にランクされた文書の得点とし、 $s(1)=1$ とする。

2.3.2 で用いた H, A, B において、以下のように得点を定義する。

$$H \cup A \cup B = X,$$

$$s(i), s(i+1) \in X \Rightarrow s(i+1) = a \cdot s(i),$$

$$s(i), s(i+1) \notin X \Rightarrow s(i+1) = a \cdot s(i),$$

$$s(i) \in X, s(i+1) \notin X \Rightarrow s(i+1) = 1,$$

$$s(i+1) \in X, s(i) \notin X \Rightarrow s(i+1) = 1$$

ここで a は $a > 1$ である重み係数とする。つまり X を適合文書とみなしたとき、検索結果のランキングの順番で、適合文書・不適合文書が連続して出現した場合に重みを与える。最終的に $s(i)$ の和を求める。この得点と DCG, MWRR それぞれとの比較を行うことによって、評価を行う。

4 テスト実験

ここでは NTCIR-3 での検索実験において使用された 47 個の検索課題中 38 個の検索課題を用いて、実際の Web 検索エンジンと実際の Web 空間を用いた実験を行った。実験に使用した検索エンジンは Google, Alltheweb, Altavista, Goo の 4 種類であり、日本語ページを検索対象ページとした。

4.1 テストデータ

検索語として各検索課題に設定されている TITLE の 3 単語を使用し、検索結果の上位 50 件までを人間の手によって判定し

た. また検索には AND 検索を使用した. 以下の図 1 において, 検索課題例を示す. 図 1 中の標準的な各タグの意味は以下の通りである.

- TOPIC : 検索課題の境界を表す
- NUM : 検索課題の識別番号を示す
- TITLE : 検索課題の内容を数語で表現したもの
- DESC : 一文で要求する情報を表現したもの
- NARR : 背景の説明, 単語の定義, 適合判定の基準を説明する

```
<TOPIC>
<NUM>0008</NUM>
<TITLE CASE="b">サルサ, 学ぶ, 方法</TITLE>
<DESC>サルサを踊れるようになる方法を知りたい</DESC>
<NARR>
<BACK>
最近はやっているサルサという踊りを学ぶためにどうすればよいか具体的な方法を知りたい. 例えば教室に通うという場合には, その場所や授業形態など, 具体的な内容を必要とする.
</BACK>
<RELE>
具体的な方法の表記のない, 流行であることのみを扱った文書は不適合とする.
</RELE>
<NARR>
<CONC>サルサ, 習う, 方法, 場所, カリキュラム</CONC>
<RDOC>NW011992774, NW011992731, NW011992734</RDOC>
<USER>大学院修士 1 年, 女性, 検索歴 2.5 年</USER>
</TOPIC>
```

図 1 検索課題例

判断基準は検索課題中の NARR に従い, 高適合, 適合, 部分適合, 不適合の段階で判定を行った.

4.2 テスト結果

以下に実験結果を示す. 表 1 で DCG, MWRR, 提案手法 SS の各検索エンジンにおける得点, 表 2 で総検索結果数と不適合文書の割合を示す. ここで DCG の計算には 2.3.2 における適合レベル 2 かつ, $c=2$ とした. また MWRR の計算は 2.3.3 では wzr の平均としたが, ここでは各検索質問に対して, 2.3.3 における $r(m)$ の累積を MWRR とした. また δ_x と β_x の組み合わせは 2.3.3 におけるレベル 2 とした. 提案手法 SS の計算には 3.2 において $a=1.1$ を使用した.

4.3 考察

DCG, MWRR と提案手法 SS との比較から, 提案手法の得点が特に高い場合には DCG, MWRR 共に得点が落ち込んでいることが分かる (表 1 において, 各検索エンジンの SS の上位 5 つのセルを網掛け表示.). 適合文書に対して得点が与えられる

DCG, MWRR の得点が低下することから, 提案手法による連続性への得点の付与は主に不適合文書に対するものであることが分かる. このことから, 提案手法 SS の得点が高い, つまり連続性が高い場合とは主に不適合文書が連続的に並んでいる場合であり, あまりよい検索結果とならなかったことを示している. しかし, ユーザにとっては直ぐに次の検索語に変更しやすいということも示す.

全ての検索エンジンにおける NUM040 の DCG, MWRR, SS の全ての得点が低くなっているが, これは検索結果における総検索結果数自体が少ないためによるものである.

また検索エンジン Google における NUM018 と NUM037 において特に提案手法の得点, DCG が共に高くなっている (表 1 において, 太字で表示.). このことから NUM018, NUM037 においては適合文書が並び, ユーザにとって満足のできる検索結果であったと考えることができる.

更に, MWRR は適合文書がランキングのどの辺りに現れているかの指標となることから, MWRR と SS を組み合わせて評価することにより, ランキングのどの辺りにどの程度の連続性があるかを評価することも可能となる. つまり, SS の得点が高く, MWRR の得点も高ければ, ランキングの上位に適合文書が並び, 逆に MWRR の得点が低ければ, ランキング上位には不適合文書が並んでいると判断が可能となる.

提案手法 SS を導入し, 文書の連続性と MWRR を組み合わせることで, ユーザにとっての検索エンジンの使い心地が評価可能であると考えられる.

次に連続性, 総検索結果数, 不適合文書の割合との考察である. これは各検索課題における総検索結果数が多すぎる場合の考慮を目的とした. 総検索結果数が非常に多い場合, より一般的な意味を持った検索語であると判断でき, 不適合文書の割合が多くなると考えられ, ランキングの上位 50 件以内に適合文書が上手く並ぶとは考えにくいという仮定による. そこで, 総検索結果数と不適合文書の割合の相関を求めることにより, 不適合文書に適合文書以上の評価を与えることを考えた. つまり総検索結果数が非常に多い状況での上位 50 件における適合文書はまぐれあたりであり, むしろ不適合文書をランキング上位に並べ, ユーザに検索語の変更を促した方が良いということである.

しかし, 実際には総検索結果数と不適合文書の割合には主だった相関は見受けられず, 上で述べたような仮定は成り立たなかった.

5 まとめと課題

本稿では Web 検索エンジンの新しい評価手法を提案した. NTCIR-3 で実際に使用された検索課題を用いてテスト実験を行い, 文書の連続性を考慮した新手法 SS と, 既存の評価手法である DCG, MWRR との比較を行った. 更に, 評価された文書の連続性を応用するための, 不適合文書の割合と総検索結果数との相関を調査したが, 特別な相関は見られなかった.

今回のテスト実験では評価の対象としなかったが、ランキングのより上位に一度出現したページもしくは、同一サイトのページだと判断されるページがある程度の割合で出現していた（フレームを使用したサイトなどに多数みられた）。これらはユーザにとっては、同じ情報が与えられたことになり、たとえ適合文書であったとしても2度目以降は価値のないものとなってしまふ。このようなユーザにとって価値のないページの扱いを更に掘り下げて評価する必要がある。

参考文献

- [1] :- “NTCIR Workshop 3 Meeting Web Site ,”
<http://research.nii.ac.jp/ntcir-ws3/>
- [2] 岸田和明, 岩山真, 江口浩二, “検索実験の方法と実際 : NTCIR ワークショップでの試み,” Pre-meeting Lecture at the NTCIR-3 Workshop, Oct.2002.
- [3] :- “Text REtrieval Conference (TREC) Home Page,”
<http://trec.nist.gov/>
- [4] Amanda Spink, B. J. Jansen, D. Wolfram & T. Saracevic, “From E-Sex to E-Commerce: Web Search Changes,” IEEE Computer, 35(3), pp.107-109, Mar.2002.
- [5] :- “Japan internet.com デイリーリサーチ - 検索結果は上から順に。上位表示サイトのクリック率高まる、87% -,”
<http://japan.internet.com/research/20020320/1.html>
- [6] Koji Eguchi, Keizo Oyama, Emi Ishida, Noriko Kando and Kazuko Kuriyama, “Evaluation Methods for Web Retrieval Tasks,” NTCIR-3 Workshop, Tokyo, Dec.2002.

表1 DCG, MWRR, 提案手法 SS の各検索エンジンにおける得点

NUM	Google			Alltheweb			Altavista			Goo		
	DCG	MWRR	SS	DCG	MWRR	SS	DCG	MWRR	SS	DCG	MWRR	SS
008	5.49	1.09	59.89	1.99	1.09	121.84	2.90	0.59	167.74	3.95	1.82	75.26
010	13.17	2.99	56.15	6.20	1.24	54.78	6.87	1.52	57.01	7.59	1.73	58.88
011	8.18	1.51	56.03	4.66	1.04	72.69	4.77	0.71	62.20	8.27	2.37	57.57
012	12.25	1.79	55.84	4.09	0.67	65.09	12.86	2.42	59.50	8.75	1.18	54.64
013	11.84	2.60	60.67	9.66	2.75	67.78	6.32	1.89	71.35	10.70	2.82	55.14
014	6.21	1.06	62.23	0.00	0.00	1163.91	4.75	1.07	220.39	7.82	2.68	68.34
015	15.38	2.62	57.72	8.04	1.76	62.18	10.17	2.40	56.99	11.51	2.13	59.25
016	8.20	1.24	59.13	6.19	1.81	64.26	2.57	0.62	189.19	6.78	1.66	68.56
017	5.12	0.86	62.67	3.06	0.75	96.96	2.02	0.44	189.11	4.88	1.05	83.55
018	31.69	4.40	191.74	22.93	3.78	64.08	25.87	4.04	86.27	25.57	3.86	61.41
019	14.06	3.41	63.31	2.20	1.23	85.81	8.95	2.36	55.97	13.81	3.25	56.33
020	11.94	2.97	66.85	10.18	3.05	63.08	11.85	3.04	53.86	9.19	2.93	69.84
022	7.61	0.70	67.40	2.81	0.45	73.15	11.33	2.15	59.58	7.21	1.63	56.40
023	15.41	3.54	61.15	3.63	0.76	91.90	11.83	3.05	60.88	12.97	2.94	53.80
024	10.22	2.77	56.35	2.72	1.32	123.66	5.95	1.38	61.25	5.99	1.49	58.48
027	12.10	3.26	56.10	9.78	2.46	56.81	5.75	1.66	54.97	10.96	3.33	57.89
028	7.43	1.79	57.30	6.57	1.39	66.09	3.51	1.45	89.54	1.61	0.48	174.37
029	5.16	1.10	65.60	1.20	1.03	224.49	3.90	1.51	66.51	1.52	0.75	658.84
030	8.69	1.61	54.13	6.50	2.44	65.22	7.48	1.77	54.62	9.66	2.01	55.96
031	11.06	1.84	65.06	8.56	2.65	59.00	8.31	2.99	57.17	8.55	2.75	61.99
032	13.11	3.30	57.09	4.56	2.20	81.66	5.98	2.16	69.61	7.82	2.00	60.57
033	6.85	1.70	78.04	3.69	1.16	207.62	7.64	2.53	61.56	4.16	1.67	65.85
034	17.71	4.07	74.48	9.48	3.42	57.20	10.62	3.44	54.12	9.67	3.36	59.57
035	16.38	3.50	65.07	5.45	1.57	60.97	12.68	2.75	54.83	12.88	2.82	54.77
036	15.62	3.70	57.73	5.55	1.89	60.79	6.53	1.10	64.54	6.60	1.66	57.27
037	22.14	4.34	143.17	19.13	4.08	59.26	19.09	3.90	60.93	20.80	3.27	69.45
038	13.38	2.96	58.19	1.26	0.44	112.67	6.99	1.43	60.21	6.56	2.15	61.50
039	10.30	1.73	55.01	7.37	1.40	54.30	10.70	2.18	60.30	7.88	1.52	61.83
040	3.83	1.94	26.64	3.13	0.81	8.41	0.43	0.20	5.74	1.80	0.45	25.10
041	9.99	2.02	56.60	7.30	2.38	54.56	4.44	0.65	95.12	8.14	1.86	60.47
042	18.24	4.21	69.93	10.61	2.93	56.72	10.74	3.38	56.19	10.13	3.29	58.91
043	16.52	3.75	57.44	10.14	2.58	66.09	7.75	1.28	55.81	8.09	1.65	62.44
044	5.08	1.67	97.88	4.89	0.76	53.42	4.85	1.53	54.60	4.66	0.81	82.19
046	14.74	2.88	61.23	4.14	1.17	69.38	6.01	1.90	58.42	8.75	2.39	53.74
047	8.88	1.88	57.08	7.96	2.22	100.83	6.78	2.19	72.22	5.17	0.98	138.47
048	4.39	0.55	72.79	6.86	1.72	62.92	5.75	1.44	88.84	6.29	1.63	66.73
049	15.39	3.93	79.88	5.96	1.67	68.68	4.42	1.43	62.03	7.96	2.44	58.16
052	9.58	1.53	57.02	4.33	1.15	69.76	8.53	2.53	103.47	5.46	1.00	56.37

表2 総検索結果数と不適合文書の割合

NUM	Google		Alltheweb		Altavista		Goo	
	総検索結果数	不適合文書の割合	総検索結果数	不適合文書の割合	総検索結果数	不適合文書の割合	総検索結果数	不適合文書の割合
008	925	0.72	260	0.92	364	0.94	295	0.82
010	1680	0.52	1055	0.68	721	0.66	986	0.68
011	651	0.66	271	0.84	335	0.74	402	0.72
012	3750	0.52	2877	0.80	1819	0.56	2484	0.64
013	64400	0.46	70578	0.68	53835	0.74	28804	0.62
014	141000	0.74	101598	1.00	54695	0.86	56022	0.74
015	575	0.30	262	0.68	303	0.60	337	0.60
016	1730	0.50	1158	0.68	821	0.86	713	0.74
017	3030	0.76	1387	0.92	2051	0.90	1943	0.88
018	892	0.08	573	0.40	429	0.22	578	0.30
019	6080	0.56	4233	0.90	9316	0.70	4675	0.54
020	855	0.54	363	0.68	501	0.48	559	0.62
022	5810	0.58	4542	0.80	2968	0.54	3384	0.60
023	2720	0.50	1539	0.88	1290	0.64	1890	0.46
024	9370	0.60	6989	0.90	3626	0.78	5641	0.74
027	455	0.48	188	0.64	257	0.72	365	0.48
028	5600	0.56	2824	0.70	4848	0.84	2875	0.88
029	15700	0.76	19495	0.96	6838	0.82	10155	0.94
030	329	0.50	129	0.72	134	0.50	196	0.48
031	17800	0.36	1637	0.68	1132	0.56	4860	0.66
032	640	0.56	398	0.84	330	0.88	460	0.74
033	1030	0.58	387	0.88	530	0.70	446	0.88
034	7210	0.18	3670	0.42	2811	0.34	5126	0.32
035	2850	0.42	1612	0.74	1184	0.58	2192	0.54
036	132	0.38	45	0.73	78	0.70	70	0.68
037	30800	0.08	28018	0.20	12277	0.24	19965	0.14
038	2180	0.60	1194	0.92	1066	0.72	1227	0.72
039	8780	0.48	3845	0.66	908	0.60	4517	0.68
040	22	0.64	8	0.63	5	0.80	21	0.81
041	4130	0.40	2661	0.62	2054	0.84	2705	0.74
042	2190	0.16	1354	0.58	1149	0.52	1288	0.50
043	6660	0.28	254	0.72	741	0.58	311	0.62
044	137	0.82	37	0.81	30	0.83	52	0.88
046	738	0.36	362	0.86	520	0.72	599	0.56
047	99	0.58	232	0.72	42	0.74	51	0.94
048	1210	0.82	1325	0.80	649	0.82	865	0.82
049	5390	0.36	4389	0.70	3515	0.76	3732	0.62
052	1920	0.62	1002	0.74	1356	0.72	1079	0.60