

利用者の検索意図に応じた検索エンジン選択手法

水野 広大[†] 鈴木 優^{††} 川越 恭二^{†††}

[†] 〒 525-8577 滋賀県草津市野路東 1-1-1

立命館大学大学院理工学研究科

E-mail: [†]mizuno@coms.ics.ritsumei.ac.jp, ^{††}suzuki@ics.ritsumei.ac.jp, ^{†††}kawagoe@is.ritsumei.ac.jp

あらまし 本稿では、利用者の検索意図に応じて動的に検索エンジンを選択するための手法を提案する。利用者が検索を行うとき、その目的、意図に応じて適切な検索エンジンを選択することができる。ところが、利用者が全ての検索エンジンの特性や特徴を把握することは困難である。そのため、適切な検索エンジンを選択することができず、適切な検索結果が得られない場合がある。本研究では、ある検索意図に適していない検索エンジンを選択した場合、適している場合と比較して検索結果数が少なくなることに着目し、検索意図に対する検索エンジンの適合度を算出する。利用者はシステムが算出した適合度を利用することによって適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となる。

キーワード 情報検索, 検索エンジン, 検索エンジン選択, 検索結果数

A Search Engine Selection Method for User's Information Need

Kodai MIZUNO[†], Yu SUZUKI^{††}, and Kyoji KAWAGOE^{†††}

[†] Graduate School of Science and Engineering, Ritsumeikan University

Nojihigashi 1-1-1, Kusatsu, Shiga, 525-8577 Japan

E-mail: [†]mizuno@coms.ics.ritsumei.ac.jp, ^{††}suzuki@ics.ritsumei.ac.jp, ^{†††}kawagoe@is.ritsumei.ac.jp

Abstract In this paper, we propose a dynamic search engine selection method for user's information need. When users retrieve on the Internet, the users can select the best search engine for their purpose or information need. However, a problem is that the users cannot understand characteristics and features of all search engines. Consequently, because the users cannot select the best search engine, the users cannot get the best retrieval results. In this paper, we focus that the search engine not matching user's information need outputs less retrieval result than the best one, and we calculate search engine's matching score for the user's information need. As a result, users can select the best search engine using the score that our system calculated, and users can get the best retrieval result.

Key words information retrieval, Search Engine, Search Engine selection, number of retrieval result

1. はじめに

利用者が Web ページや画像といったコンテンツを検索するために、様々な検索エンジンが存在する。利用者が検索を行うとき、利用者はその目的、意図に応じて適切な検索エンジンを選択することができる。適切な検索エンジンとは、全ての検索エンジンと比較して利用者の検索意図を満たす検索エンジンである。例えば、あるアーティストの画像を必要とする利用者が適切な検索エンジンを選択するとき、適切な検索結果はあるアーティストの画像であることが考えられる。そのため、他の検索エンジンと比較して画像に特化した検索結果を出力する点において、利用者にとって適切な検索エンジンは画像検索エンジンであることが考えられる。このことより、利用者はあるアーティストの画像に対して適切な検索エンジンとして、画像

検索エンジンを選択することができる。また、検索エンジン選択において、ある検索エンジンが何の検索を行うのかといった検索エンジンの特性把握、問合せに対してある検索エンジンは適切なのかといった特徴把握を用いて、利用者は適切な検索エンジンを選択することが考えられる。

ところが、利用者が全ての検索エンジンの特性や特徴を把握することは困難である。そのため、適切な検索エンジンを選択することができず、適切な検索結果が得られない場合がある。例えば、Java のアプレットに関するコンテンツを必要とする利用者が適切な検索エンジンを選択するとき、適切な検索結果は Java のアプレットに関するコンテンツであることが考えられる。ところが、Web ページ、プログラムソースや画像といったコンテンツは、Java のアプレットに関するコンテンツを含んでいることが考えられる。この場合、利用者は適切な検索エン

ジンの選択を行うために、Java のアプレットに関するコンテンツの検索を行う検索エンジン全ての特性および特徴を把握する必要があることが考えられる。ところが、利用者が検索エンジン全ての特性および特徴を把握することは困難である。そのため、利用者は Java のアプレットに関する適切な検索エンジンを選択することができず、適切な検索結果を得られない場合が考えられる。

そこで本研究では、利用者が適切な検索結果を得るために、利用者の検索意図に応じて動的に検索エンジンを選択する手法を提案する。提案手法では、二つの事項に着目し、各検索エンジンが問合せに対して適している度合いを示す適合度を算出する。一つは、ある検索意図に適していない検索エンジンを選択した場合、適している別の検索エンジンを選択した場合と比較して検索結果数が少なくなることである。この事項を用いて、対象検索エンジン全体における平均検索結果数と各検索エンジンの検索結果数との割合算出を行う。もう一つの事項は、ある検索エンジンが検索意図に適していない場合、ある検索エンジンが検索意図に適している場合と比較して検索結果数が少なくなることである。この事項を用いて、各検索エンジンにおける平均検索結果数と各検索エンジンの検索結果数との割合算出を行う。算出した二つの割合を用いて、問合せに対する各検索エンジンの適合度を算出する。利用者は算出した適合度を利用することによって、適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となる。

2. 関連研究

本研究に関連する研究として、複数のデータベースから問合せに適切なデータベースの検索を行う研究がある [1] [2]. Gravano ら [3] [4] は、問合せに適切な文書データベースの検索を行うために、問合せに対する文書データベースの適合度を全ての文書データベースに対して算出している。算出した適合度を用いて問合せに適切な文書データベースの検索を行うことから、問合せに対する文書データベースの適合度を算出する点において、本研究と類似している。ただし、文書データベースの検索を行うために、問合せを構成する各キーワードの出現頻度を用いていることから、問合せキーワード全体に対する文書データベースの特徴を考慮していない。そのため、問合せキーワード全体に対する文書データベースの特徴を把握することは困難であると考えられる。この点において、問合せに対する検索エンジンの適合度を算出するために、検索結果数を用いて問合せキーワード全体に対する検索エンジンの特徴を把握している本研究と異なっている。

一方で、複数の検索エンジンの検索結果を統合することにより、利用者の問合せに適切な検索結果を与える研究がある [5]. Liu ら [6] は、ニュースメタ検索エンジンにおいて、各検索エンジンの検索結果の統合を行い、利用者の問合せに対して適切な検索結果の出力を行っている。利用者の問合せを用いて、各検索エンジンの検索結果が問合せに適しているかの判断を行う点において、本研究と類似している。ただし、利用者へ出力する検索結果として、各検索エンジンの検索結果を統合した検

索結果を用いている。そのため、各検索エンジンが問合せに適しているのかを利用者が判断することを考慮していない。この点において、問合せに対して各検索エンジンの適合度算出を行い、適切な検索エンジンの選択を利用者が行う本研究と異なっている。

3. 検索意図に応じた検索エンジン選択手法

利用者が検索を行うとき、利用者は様々な検索エンジンから、利用者の検索意図を満たす検索エンジンの選択を行う。ところが、検索エンジンは様々な特性を持っており、各検索エンジンの特徴は異なっていることが考えられる。このことに対し、利用者が全ての検索エンジンの特性や特徴を把握することは困難である。そのため、利用者が適切な検索エンジンを選択することができず、適切な検索結果を得られない問題が考えられる。

例として、Java のアプレットに関する検索結果を必要とする利用者が適切な検索エンジンを選択する場合を考える。ここで、利用者の選択対象となる検索エンジンは Google, Excite, GoogleCodeSearch および GoogleImageSearch とする。この場合、Web ページ、プログラムソースおよび画像といったコンテンツは、全て Java のアプレットに関するコンテンツを含んでいることが考えられる。そのため、利用者は適切な検索エンジン選択のために、4 個の検索エンジンの特性および特徴を把握する必要があることが考えられる。ところが、Web ページ、プログラムソース、画像といったコンテンツは、全て Java のアプレットに関するコンテンツを含んでいるため、Google, GoogleCodeSearch, GoogleImageSearch のうち、どの検索エンジンが適切な特徴をもつのかを利用者が把握することは困難である。また、Web ページ検索エンジンとしての特性が同じであるため、Google と Excite との特徴の差異を利用者が把握することは困難である。そのため、利用者が適切な検索エンジンを手動選択した場合では、利用者は Java のアプレットという問合せに対して、適切な検索エンジンを選択することができない場合が考えられる。

そこで本研究では、利用者が適切な検索エンジンを選択するために、検索意図に対する検索エンジンの適合度を算出する。提案手法を用いたシステムが算出した適合度を利用することによって、利用者は検索意図に対して適切な検索エンジンの選択を行う。例として、前段落で述べた例を用いる。利用者が「Java アプレット」という問合せに適していない GoogleImageSearch を選択した場合、適している Google を選択した場合と比較して、検索結果数は少なくなることが考えられる。このことにより、問合せ「Java アプレット」に対して、4 個の検索エンジンにおける平均検索結果数と GoogleCodeSearch の検索結果数との割合算出を行う。また、GoogleCodeSearch が「北京オリンピック 結果」という問合せに適していない場合、GoogleCodeSearch が「Java アプレット」という問合せに適している場合と比較して、検索結果数は少なくなることが考えられる。このことにより、GoogleCodeSearch における平均検索結果数と問合せ「Java アプレット」に対する GoogleCodeSearch の検索結果数との割合算出を行う。算出した二つの割合を用いて、問合せ「Java アプ

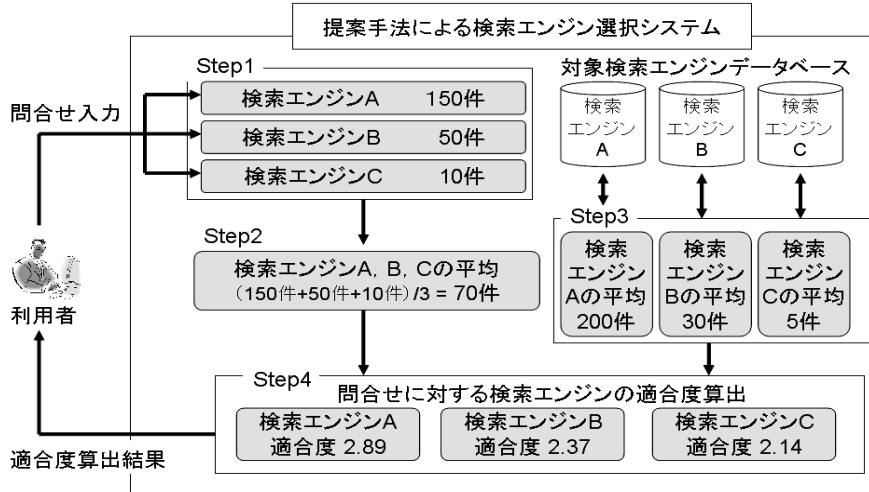


図1 提案手法による検索エンジン選択システムの概要

レット」に対する GoogleCodeSearch の適合度を算出する。同様の割合および適合度算出により、問合せ「Java アプレット」に対する Google、Excite および GoogleImageSearch の適合度を算出する。利用者は算出した適合度を利用することによって、適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となる。

3.1 処理の概要

提案手法による検索エンジン選択システムの概要を図1に示す。本システムでは、問合せに対する対象検索エンジン全体の平均検索結果数、各検索エンジンにおける平均検索結果数を用いて、検索エンジンの適合度を算出する。利用者は、システムが算出した適合度を利用することによって、適切な検索エンジンの選択を行う。

適合度算出のために、利用者が過去に行った全ての問合せに対する各検索エンジンの検索結果数を前処理によって取得する。そして、対象検索エンジンデータベースの構築を行う。対象検索エンジンデータベースには、利用者が過去に行った全ての問合せに対する各検索エンジンの検索結果数を格納する。

利用者は、検索意図を表すキーワードを問合せとして入力する。システムは、Step1として各検索エンジンが問合せに対して出力する検索結果から、各検索エンジンの検索結果数を抽出する。そして、Step2では問合せにおける対象検索エンジン全体の平均検索結果数を算出する。一方、Step3として、対象検索エンジンデータベースに格納された各検索エンジンの検索結果数により、各検索エンジンにおける平均検索結果数を算出する。問合せに対する対象検索エンジン全体の平均検索結果数および各検索エンジンにおける平均検索結果数を取得すると、システムはStep4として問合せに対する各検索エンジンの適合度を算出する。最後に、システムは利用者に対して適合度算出結果を出力する。

3.2 問合せに対する検索エンジンの適合度算出

提案手法では、問合せに対する検索エンジンの適合度を算出する際に、検索エンジンの検索結果数から得られる二つの値を用いる。一つは、問合せに対する対象検索エンジン全体の平均検索結果数である。この値によって、対象検索エンジン全体の平均検索結果数と問合せに対する各検索エンジンの検索結果数との割合を算出する。問合せ i_h ($h=1,2,\dots,n$) に対して対象検索エンジン $E_1, E_2, \dots, E_N$ は $E_1(i_h), E_2(i_h), \dots, E_N(i_h)$ 件の検索結果数を得る。このとき、対象検索エンジン全体の平均検索結果数と、検索エンジン E_H ($H=1,2,\dots,N$) との検索結果数の割合 $R_H(i_h)$ の算出式を (1) 式に示す。

$$R_H(i_h) = \frac{E_H(i_h)}{\sum_{K=1}^N E_K(i_h)} \times N \quad (1)$$

もう一つの値は、対象となる各検索エンジンの平均検索結果数である。この値によって、各検索エンジンの平均検索結果数と問合せにおける各検索エンジンの検索結果数との割合を算出する。検索エンジン E_H は、問合せ $i_1, i_2, \dots, i_n$ に対して $E_H(i_1), E_H(i_2), \dots, E_H(i_n)$ 件の検索結果数を得る。このとき、問合せ $i_1, i_2, \dots, i_n$ に対する検索エンジン E_H の平均検索結果数と、問合せ i_h に対する検索エンジン E_H の検索結果数との割合 $r_H(i_h)$ の算出式を (2) 式に示す。

$$r_H(i_h) = \frac{E_H(i_h)}{\sum_{k=1}^n E_H(i_k)} \times n \quad (2)$$

算出された二つの割合を用いて、システムは問合せに対する各検索エンジンの適合度を算出する。問合せ i_h に対する検索エンジン E_H の適合度 $A_H(i_h)$ の算出式を (3) 式に示す。

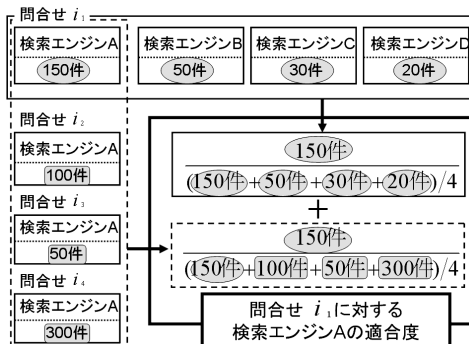


図 2 検索結果数による検索エンジンの適合度算出手法

$$A_H(i_h) = R_H(i_h) + r_H(i_h) \quad (3)$$

本研究における検索エンジンの適合度算出の具体例を図 2 に示す。図 2 では、問合せ i_1 に対する検索エンジン A の適合度を算出している。まず、システムは問合せ i_1 における対象検索エンジン A, B, C, D の平均検索結果数を算出する。図 2 により、対象検索エンジン A, B, C, D の平均検索結果数は、 $\frac{150+50+30+20}{4} = 62.5$ となる。そして、この値と問合せ i_1 における対象検索エンジン A が出力した検索結果数との割合を算出する。図 2 により、問合せ i_1 における対象検索エンジン A が出力した検索結果数の割合は、 $\frac{150}{62.5} = 2.4$ となる。次に、システムは問合せ i_1, i_2, i_3, i_4 における検索エンジン A の平均検索結果数を算出する。図 2 により、問合せ i_1, i_2, i_3, i_4 における検索エンジン A の平均検索結果数は、 $\frac{150+100+50+300}{4} = 150$ となる。そして、この値と問合せキーワード i_1 における対象検索エンジン A が出力した検索結果数との割合を算出する。図 2 により、問合せキーワード i_1 における対象検索エンジン A が出力した検索結果数の割合は、 $\frac{150}{150} = 1$ となる。最後に、算出した二つの値を用いて、システムは問合せキーワード i_1 に対する検索エンジン A の適合度を算出する。図 2 により、問合せキーワード i_1 における対象検索エンジン A が出力した検索結果数の割合は 2.4、問合せキーワード i_1 における対象検索エンジン A が出力した検索結果数の割合は 1 となる。この二つの値によって、問合せキーワード i_1 に対する検索エンジン A の適合度は、 $2.4 + 1 = 3.4$ となる。

図 2 と同様の適合度算出によって、システムは問合せキーワード i_1 に対する検索エンジン B, C, D の適合度をそれぞれ算出する。算出した適合度を利用することにより、利用者は問合せキーワード i_1 に対して適切な検索エンジンの選択を行うことが可能となる。

4. 予備実験

評価実験のために、対象となる各検索エンジンの平均検索結果数を算出する予備実験を行った。

まず、予備実験では研究室の学生 11 名から 105 件の問合せを取得した。次に、取得した問合せを本システムに入力した。

そして、入力した問合せに対する各検索エンジンの検索結果から各検索エンジンの検索結果数の抽出を行い、検索結果数を対象検索エンジンデータベースに格納した。以上の手順を、取得した 105 件の問合せ全てに対して行った。

対象検索エンジンデータベースに格納された各検索エンジンの検索結果数を用いて、各検索エンジンの平均検索結果数の算出を行った。まず、4.1 節では、評価実験の対象となる検索エンジンについての詳細を述べる。最後に、4.2 節では、予備実験の結果についての詳細を述べる。

4.1 評価実験の対象となる検索エンジンの決定

評価実験の対象となる検索エンジンと検索対象となるコンテンツを表 1 に示す。評価実験では、Google、Infoseek などの Web ページ検索エンジン 4 個を用いる。また、画像、プログラムソース、動画、百科事典の記事、掲示板の記事の検索を行う検索エンジンをそれぞれ 1 個ずつ用いる。本研究では、この 9 個の検索エンジンを対象として、予備実験および評価実験を行う。

4.2 各検索エンジンの平均検索結果数の決定

予備実験における、各検索エンジンの平均検索結果数を表 2 に示す。予備実験では、対象となる全ての検索エンジンにおいて、Excite の平均検索結果数が 25,624,773 件となり、最も平均検索結果数の多い検索エンジンであることを示した。一方で、対象となる全ての検索エンジンの中で、Wikipedia の平均検索結果数は 76 件となり、最も平均検索結果数の少ない検索エンジンであることを示した。また、対象となる全ての Web ページ検索エンジンにおいて、最も平均検索結果数の多い検索エンジンである Excite と最も平均検索結果数の少ない検索エンジンである Mooter との間では、平均検索結果数に 25,507,337 件

表 1 評価実験の対象となる検索エンジン

検索エンジン	検索対象となるコンテンツ
Google	Web ページ
Infoseek	Web ページ
Excite	Web ページ
Mooter	Web ページ
GoogleImageSearch	画像
GoogleCodeSearch	プログラムソース
YouTube	動画
Wikipedia	百科事典の記事
教えて!goo	掲示板の記事

表 2 各検索エンジンの平均検索結果数

検索エンジン	平均検索結果数 (件)
Google	5548651
Infoseek	1002864
Excite	25624773
Mooter	117436
GoogleImageSearch	876610
GoogleCodeSearch	65200
YouTube	4565
Wikipedia	76
教えて!goo	383

表 3 問合せに対する各検索エンジンの正解数

検索エンジン	問合せキーワード				
	「Java アプレット」	「Java HTML」	「C 入門」	「料理 レシピ」	「北京オリンピック 結果」
Google	20	8	16	24	17
Infoseek	22	9	15	24	19
Excite	19	12	12	27	21
Mooter	18	13	13	26	24
GoogleImageSearch	16	7	6	23	2
GoogleCodeSearch	24	11	2	0	0
YouTube	5	2	1	25	4
Wikipedia	6	10	2	17	10
教えて!goo	23	11	18	19	1

の差異が生じた。ここで算出した平均検索結果数を用いて、本研究は提案手法の評価実験を行う。

5. 評価実験

5.1 実験方法

提案手法によってシステムが算出した適合度 1 位の検索エンジンにおける検索結果上位 30 件の再現率および精度と、既存手法における検索結果上位 30 件の再現率および精度を比較することによって、提案手法の評価を行う。また、利用者が検索を行うときに、様々な検索エンジンの特性および特徴を利用者が把握することによって、検索エンジンの選択を行うことが考えられる。そのため、既存手法として、問合せに対して適切な検索エンジンを利用者が手動選択する手法を用いる。本実験では、既存手法の再現率および精度と比較して、提案手法によってシステムが選択した検索エンジンの再現率および精度が高い値を示すことによって、提案手法の有用性を示す。

本実験では、予備実験と同じ検索エンジン 9 個を、本実験の対象となる検索エンジンとして用いた。対象となる検索エンジンの平均検索結果数に、予備実験で算出した平均検索結果数を用いた。問合せキーワードには、Java のアプレットに関する問合せである「Java アプレット」、Java の HTML に関する問合せである「Java HTML」、C の入門に関する問合せである「C 入門」、料理のレシピに関する問合せである「料理 レシピ」、そして、北京オリンピックの結果に関する問合せである「北京オリンピック 結果」を用いた。また、既存手法によって、利用者がこの 5 件の問合せに対して適切であると選択した検索エンジンは Google である。

本実験の問合せ 5 件に対する各検索エンジンの正解数を表 3 に示す。本実験の正解集合は 3 人の人手により作成した。問合せに対する正解の基準として、3 人の被験者のうち 3 人が問合せに対して適切であると判断していることを正解の基準とした。

5.2 実験結果

提案手法および既存手法による再現率精度曲線について、図 3 に示す。ここで、提案手法および既存手法の各曲線において 11 点平均精度を算出すると、それぞれ 0.779、0.730 といった値を得られた。これらの値により、提案手法では、既存手法の 11 点平均精度と比較して 0.049 の向上を得ることができ、提案手法は既存手法と比較して有用性が高いといえる。

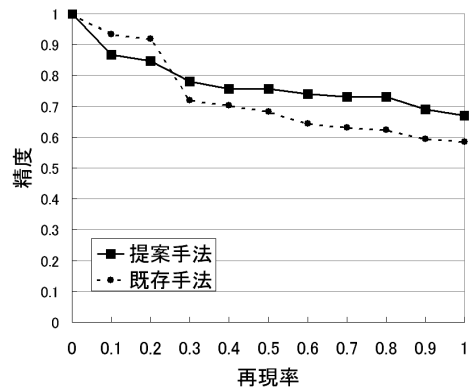


図 3 提案手法および既存手法による再現率精度曲線

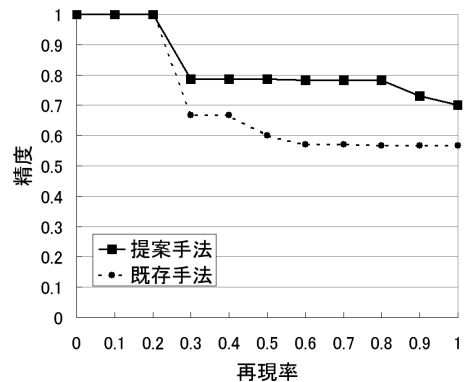


図 4 問合せ「北京オリンピック 結果」に対する提案手法および既存手法による再現率精度曲線

問合せ「北京オリンピック 結果」に対する提案手法および既存手法の再現率精度曲線を図 4 に示す。ここで、提案手法を用いたシステムが算出した適合度 1 位の検索エンジンは Excite である。問合せ「北京オリンピック 結果」において、提案手法が既存手法の 11 点平均精度と比較して 0.124 の向上を得ることができ、本実験で用いた 5 件の問合せにおいて最大の値を示す。

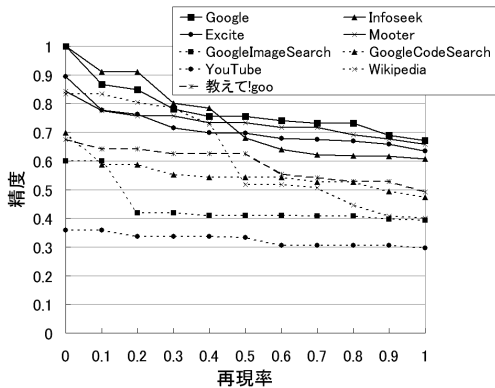


図5 5件の問合せに対する検索エンジン9個の再現率精度曲線

した。この結果は、提案手法による検索結果数の比較によって、本システムが各検索エンジンの特性および特徴の把握を行うことができた結果であると考えられる。このことより、提案手法では、既存手法では困難であったWebページ検索エンジンとしての特性が同じGoogleおよびExciteの特徴把握が可能となることが考えられる。

5件の問合せに対する検索エンジン9個の再現率精度曲線を図5に示す。Webページ検索エンジン4個の再現率精度曲線は実線を用いて示す。図5より、本実験で用いた5件の問合せに対してWebページ検索エンジン4個が、それ以外の検索エンジン5個と比較して平均精度が高いという結果が得られた。しかし、問合せ「Java HTML」に対する適切な検索エンジンとして、提案手法を用いたシステムが1位に算出した検索エンジンはGoogleCodeSearchであった。また、問合せ「Java HTML」において、提案手法が既存手法の11点平均精度と比較して0.108の向上を得ることができた。このことにより、既存手法と比較して提案手法では、問合せに対してある検索エンジンが適切なのかといった特徴把握が可能となることが考えられる。

6. おわりに

本研究では、利用者が適切な検索結果を得るために、利用者の検索意図に応じて動的に検索エンジンを選択する手法の提案を行った。検索エンジンを選択するために、二つの事項に着目し、各検索エンジンの問合せに対する適合度を算出した。一つは、ある検索意図に適していない検索エンジンを選択した場合、適している別の検索エンジンを選択した場合と比較して検索結果数が少なくなることである。この事項により、提案手法では、問合せに対して対象検索エンジン全体における平均検索結果数と各検索エンジンの検索結果数との割合算出を行った。もう一つの事項は、ある検索エンジンが検索意図に適していない場合、ある検索エンジンが検索意図に適している場合と比較して検索結果数が少なくなることである。この事項により、提案手法では、各検索エンジンにおける平均検索結果数と各検索エンジン

の検索結果数との割合算出を行った。算出した二つの割合を用いて、検索意図に対する各検索エンジンの適合度を算出した。システムが算出した適合度を利用することによって、適切な検索エンジンを選択することが可能となり、適切な検索結果を得ることが可能となることを評価実験によって示した。また、提案手法では、既存手法では困難であった検索エンジンの特性および特徴把握に対して有用な手法であることを評価実験によって示した。

今後の課題として、二つの事項について検討する。一つは、検索エンジンの適合度における精度向上である。本研究では、問合せに対する検索エンジンの適合度算出を行うために、対象検索エンジン全体における平均検索結果数および各検索エンジンにおける平均検索結果数と各検索エンジンの検索結果数との割合の算出を行い、割合を用いて適合度を算出した。評価実験によって、提案手法での適合度算出は、既存手法と比較して有用であることを示した。今後は、さらなる適合度の精度向上のために、各検索エンジンの適合度算出手法について改善を図ることを考える。また、本研究において用いた検索結果数だけでは、検索エンジンにおける特性や特徴の全てを把握することはできていないと考えられる。そこで、検索エンジン選択における、適切な特性、特徴の把握についても検討する。

もう一つの課題は、複数のキーワードによって構成された問合せに対する適合度の精度向上である。現段階では、問合せに対する各検索エンジンの検索結果数によって適合度算出を行うのみで、問合せを構成するキーワードのうち、どのキーワードが利用者の検索意図にあたるのかを考慮できていない。この課題に対して、今後は複数のキーワードによって構成された問合せにおける各キーワードの関係性に着目し、どのキーワードが他のキーワードと比較して利用者の検索意図を表現しているのかを考慮することを行う。

文献

- [1] Milad Shokouhi, Justin Zobel and Yaniv Bernstein: "Distributed text retrieval from overlapping collections", ADC '07: Proceedings of the eighteenth conference on Australasian database - Volume 63, pp.141-150, 2007.
- [2] Jack G. Conrad and Joanne R. S. Claussen: "Client-system collaboration for legal corpus selection in an online production", ICAIL '03, pp.262-273, 2003.
- [3] L. Gravano, H. Garcia-Molina and A. Tomasic: "The effectiveness of GLOSS for the text database discovery problem", ACM SIGMOD Record, pp.126-137, 1994.
- [4] L. Gravano, H. Garcia-Molina and A. Tomasic: "GLOSS: text-source discovery over the Internet", ACM Transactions on Database Systems (TODS), v.24 n.2, pp.229-264, 1999.
- [5] Yves Rasolofo, Faiza Abbaci and Jacques Savoy: "Approaches to collection selection and results merging for distributed information retrieval", CIKM '01, pp.191-198, 2001.
- [6] King-Lup Liu, Weiyi Meng, Jing Qiu, Clement Yu, Vijay Raghavan, Zonghuan Wu, Yiyao Lu, Hai He and Hongkun Zhao: "AllInOneNews: development and evaluation of a large-scale news metasearch engine", SIGMOD '07, pp.1017-1028, 2007.