

## 語の出現頻度情報に基づく適合度順検索の問題点

相良 佳弘<sup>†</sup>

Some problems on ranking retrieval systems based on term frequencies

Yoshihiro Sagara<sup>††</sup>

### <Abstract>

Ranking retrieval systems based on term frequencies present users a sequence of documents ranked in descending order of similarity between query and document. These systems come into use in online database retrieval or WWW search engines. Most preceding researches pointed out many advantages of ranking retrieval systems. Whether ranking by systems satisfy users or not, however, has not been examined. In this research, an experiment comparing the ranking by system with the ranking made by user on his relevance or utility judgement was carried out. From this result, existing ranking retrieval systems based on term frequencies have some problems in ranking. The conditions to be fulfilled in order that the system ranking should be similar to user ranking were identified. Under following four conditions, system ranking is different from user ranking.

- 1) When various fields are covered by a database
- 2) When record lengths in a database vary
- 3) When many topics are treated in one record
- 4) When vague query or keywords are used

Ranking retrieval systems using only term frequencies are not enough to make ranking similar to ranking by user. These problems of the ranking retrieval systems may be attributed to the fact that the main part of the process is based on keywords that used at conventional Boolean retrieval. Ranking retrieval systems shall be improved by the use of some methods that can reflect user's information needs in addition to term frequencies.

### 1. はじめに

近年、適合度順検索がオンラインデータベース検索やインターネット上のサーチエンジンなどの形で実用化が進みつつある。この適合度順検索は、SaltonのSMARTシステム以来、非ブール型の検索方法として多くの手法が研究されてきているが、それら研究の多くが適合度順検索には以下のような利点があることを指摘している。<sup>1)2)</sup>

- ①ユーザが出力件数を調整するために検索式の修正をすることなく、任意の件数を出力できる
- ②適合すると思われる順に出力されるので、精度を優先させた検索と再現率を優先させた検索とを同じ検索式で行える
- ③利用者、特に検索の経験のあまりない初心者にとってブール演算子を用いる必要がないので直観的でわかりやすい

しかし上記の利点のうち①及び②は、適合度順検索システムが実際にユーザの望む順に検

索結果を出力可能であることを前提としている。適合度順検索では、各レコードに順位を付け出力する点が重要な特徴であり、いかにユーザの望む順位付けを行うかが非常に重要な要素である。適合度順検索システムの検索性能を考える上で、ユーザがより適合していると考えられるレコードがきちんと上位で出力されることが求められている。しかし、これまでに行われてきた適合度順検索システムへの評価実験では、適合と非適合という二値的なレコードへの評価しか行っていないものが多いため、この順位付けという観点を十分に反映しているとは言えない。本研究では、システムの出力する順位付けとユーザによる順位付けとを実験により比較することで、現状でどの程度ユーザの情報要求を満足させる順位付けが可能か、また適合度順検索が先行研究で指摘されている利点を十分に発揮しているかについて明らかにする。

## 2. 実験の概要

実験は DIALOG によって提供されている TARGET を用いて行った。TARGET は SMART システム以来研究がなされ、今日実用化が進んでいる適合度順検索システムの多くが用いている語の出現頻度情報から検索結果の順位付け出力を行う手法を使用している。<sup>3)</sup>実験は、大学院修士課程在籍以上の研究者及び検索に習熟した情報専門家の計 16 名を対象として、以下の手順に従って行った。

- ①被験者が情報要求を示す
- ②被験者がその情報要求から検索に用いるキーワードを抽出
- ③TARGET を用いて検索を行う
- ④出力されたレコードのうち上位 30 件を、無作為順に並べ替える
- ⑤被験者が無作為順となった 30 件のレコードを自らの情報要求と照らし合わせて実際に入手したい文献順に順位を評価する
- ⑥同時に、被験者は各レコードに対して 5 段階で利用価値の高さ、つまり有用度の評価を行う
- ⑦システムの出力と被験者の評価との比較を行う

今回の実験では、それぞれ LISA、PAPERS、PsychINFO、NTIS、MEDLINE といったデータベースを対象とした合計 25 件の情報要求について、語の出現頻度に基づいた適合度順検索システムによる順位付けと、実際に情報を用いるユーザによる順位付けとの比較を行っている。なお、この順位付けの一致度を見るために順位相関係数を用いた。ユーザによる順位付けと全く同じ順位付けで TARGET が 30 件を出力した場合、順位相関係数の値は 1 となり、反対に逆順にならんでいる場合は -1 となる。つまり順位相関係数が正の値を示している場合は、被験者が望んだ順位付けに近く、逆に負の値を示している場合は被験者の望む順位付けとは逆に出力したことになり、TARGET がどれだけユーザが望んだ順位付けに近い順位付けができているかを数値化することが可能である。ただし本研究では、順位相関係数はあくまで順位付けの一致度の目安を示すために用いたものであり、統計的な検定は行っていない。

## 3. 結果

実験で用いた情報要求及び実験で得られた結果は、第 1 表に示す。また、情報要求 No.11 から No.25 による実験の結果得られた順位付けのデータは、第 2 表及び第 3 表に示す。実

験の結果からも明らかなように、語の出現頻度情報に基づいて順位付けを行っている現状の適合度順検索システムでは、十分にユーザの望んだ順位付けを実現できてはいない。しかし、

第1表 情報要求と順位相関係数

No.	質問文	DB	相関係数
1	電子出版物の納本について	LISA	-0.08565
2	図書館サービスの設計	LISA	0.021135
3	図書館における電子情報と公正使用との関係に付いて	LISA	0.291657
4	人間の情報理解、情報利用における情報や知識の概念	LISA	0.298776
5	人間のコミュニケーションにおける情報と知識の関連	LISA	0.260067
6	英国市民憲章における公共図書館サービスのクオリティマネジメント	LISA	0.436707
7	ユーザー志向アプローチを用いた研究例	LISA	-0.03092
8	電子メールのメリット、デメリット。電子メール利用によりコミュニケーション形態は変化し	LISA	0.317464
9	図書館情報学におけるインターネットというメディアの特質についての先行研究	LISA	0.064739
10	インターネット上のサーチエンジンについて	LISA	0.34327
11	アメリカ各地の地ビールについて	PAPERS	0.267631
12	日本の貿易黒字に関する記	PAPERS	0.000222
13	アメリカのペット葬祭業の動向	PAPERS	0.288543
14	インターネットのもたらす弊害とそれに対する規制	PAPERS	-0.15595
15	最近の偽札事情	PAPERS	0.002447
16	中年期女性のストレス	PsychINFO	0.304561
17	父親不在による子どもの発達への影響	PsychINFO	0.609789
18	夫不在の家庭における妻のストレスについて	PsychINFO	0.582202
19	電子メールと通常の郵便との心理的差違	PsychINFO	-0.21824
20	並列処理、コネクショニスト、ニューラルネットワークの手法による情報検索	NTIS	0.05762
21	発展途上国における電気通信分野の専門用語制定	NTIS	0.290323
22	フロン破壊処理技術	NTIS	0.08832
23	QUINAPRILのDNA、mRNA関与の心筋への作用	MEDLINE	-0.00289
24	ニトレンジピンとリファンピシンの相互作用	MEDLINE	0.361513
25	子宮筋腫または子宮内膜症治療に使用のGnRH誘導體製剤の副作用	MEDLINE	-0.15551

全ての事例において有効な順位付けが行えなかったのではなく、ある程度有効な順位付けが行えた事例も存在することから、システムが有効な順位付けを行えるか否かに関して、データベースや情報要求の特性などの要因による影響があると考えられる。

まず、データベースの特性の差違では、抄録や索引語の質が高いと考えられる PsychINFO のようなデータベースの方が、ユーザの情報要求に近い並べ替えが行えていた。PsychINFO では、キーワード付与に統制語彙が用いられ、ほぼすべてのレコードに抄録が付与されている。MEDLINE も統制語彙は用いられているが、情報要求 No.25 には抄録の含まれないレコードが検索されており、これが順位付けの一致度が低かった原因であると考えられる。キーワードの出現頻度に基づいているため、ある程度のレコード長は必要であり、その長さはある一定の範囲内に収まることが望ましいようである。また、PAPERS に見られるように同一レコード内に複数のテーマを扱ったレコードが存在する場合や、NTIS のように単一のデータベース内で様々な分野を対象としている場合は、順位付けが混乱する傾向が見られた。これは対象とする主題が幅広くなった場合、同一のキーワードに対するユーザの捉え方も幅広くなることによるものと思われる。これらは、順位付けをキーワードに対する重みのみで行っていることの問題点を示すと考えられる。情報要求の特性の面では、特定性の高い情

第2表 情報要求No.11からNo.18における順位付けの差異

出力 順位	PAPERS										PsychINFO					
	No.11		No.12		No.13		No.14		No.15		No.16		No.17		No.18	
	順位	評価	順位	評価	順位	評価	順位	評価	順位	評価	順位	評価	順位	評価	順位	評価
1	1	5	7	4	16	2	26	3	20	1	28	1	3	5	4	5
2	25	3	14	3	4	2	27	2	24	1	15	2	5	5	7	5
3	20	3	20	2	13	2	30	1	26	1	7	4	9	4	19	2
4	5	5	4	5	15	2	8	4	28	1	27	1	8	4	3	5
5	4	5	9	4	24	2	21	3	14	4	9	3	2	5	10	5
6	6	5	10	4	1	3	28	2	8	5	5	4	1	5	8	5
7	24	3	15	3	17	2	2	5	16	3	12	2	6	4	11	5
8	29	2	28	1	5	2	6	5	25	1	10	3	7	4	24	1
9	13	4	26	2	2	3	18	4	12	4	1	5	4	5	9	5
10	19	3	19	2	19	2	19	4	6	5	16	1	20	1	13	4
11	9	5	18	3	12	2	4	5	2	5	6	4	12	1	6	5
12	21	3	24	2	21	2	10	4	10	5	14	2	16	1	20	1
13	17	3	5	4	26	2	1	5	3	5	2	5	14	1	16	3
14	16	4	16	3	8	2	24	3	7	5	3	5	18	1	5	5
15	12	4	3	5	6	2	5	5	19	2	23	1	21	1	12	4
16	2	5	22	2	20	2	23	3	18	2	20	1	26	1	2	5
17	7	5	23	2	25	2	13	4	13	4	4	4	28	1	28	1
18	15	4	6	4	22	2	11	4	5	5	29	1	17	1	21	1
19	18	3	17	3	14	2	14	4	21	1	30	1	25	1	1	5
20	10	4	27	2	28	2	29	2	9	5	8	4	29	1	18	2
21	8	5	29	1	3	3	20	4	27	1	25	1	22	1	26	1
22	14	4	30	1	9	2	7	5	4	5	13	2	24	1	15	3
23	28	2	8	4	29	2	3	5	1	5	21	1	30	1	22	1
24	27	2	21	2	30	2	12	4	23	1	11	3	23	1	30	1
25	23	3	12	3	11	2	22	3	29	1	22	1	10	2	25	1
26	11	4	13	3	10	2	25	3	11	4	19	1	27	1	17	3
27	3	5	11	4	7	2	15	4	30	1	24	1	19	1	23	1
28	26	2	1	5	18	2	16	4	17	3	26	1	11	1	27	1
29	22	3	25	2	27	2	17	4	22	1	18	1	15	1	14	3
30	30	1	2	5	23	2	9	4	15	4	17	1	13	1	29	1

報要求の方が情報要求に合致した順位付けが行えている。これは、特定性の低い情報要求に対してシステムはより多数のレコードを対象に順位付けを行うが、実際にはこの様な場合でもユーザ側には立場や嗜好など情報要求として表現されない何らかの判定基準があり、これらを順位付けに反映できないことが大きく影響していると考えられる。適合度順検索では、出力件数が過大になる特定性の低い情報要求に対しても検索可能としている先行研究もあるが、この様な情報要求に合致した順位付けを行うためには、ユーザの立場、嗜好を検索に反映する必要がある。また、検索語群に含めなかった語がユーザの判定に大きな影響を与えるケースが多く見られた。このことからキーワードへの重み付けを行う手法では、不要な語に対してマイナスの重みを付与できるようにする機能が必要になると考える。

今回の実験では、データベースや情報要求の特性が以下の四つの何れかに当てはまる場合、ユーザの順位付けに近い出力が行えない事例が見られた。

- ①データベースが多数の分野のレコードを扱っていた場合
- ②データベース内のレコード長にばらつきがあった場合
- ③1件のレコード内で同時に複数の主題を扱っていた場合

第3表 情報要求No.19からNo.25における順位付けの差異

出力 順位	PsychINFO		NTIS				MEDLINE							
	No.19 順位	評価	No.20 順位	評価	No.21 順位	評価	No.22 順位	評価	No.23 順位	評価	No.24 順位	評価	No.25 順位	評価
1	30	1	15	1	17	1	8	3	6	3	3	2	6	3
2	28	1	17	1	22	1	1	5	4	3	4	2	26	1
3	15	1	14	1	5	2	16	2	30	1	1	3	27	1
4	21	1	23	1	4	2	21	2	1	5	14	1	7	3
5	14	1	1	4	16	1	24	2	26	1	26	1	15	2
6	7	2	18	1	30	1	15	2	10	3	28	1	29	1
7	24	1	2	4	7	2	12	3	2	4	19	1	30	1
8	26	1	26	1	6	2	13	3	16	3	7	1	18	2
9	16	1	25	1	1	3	20	2	25	1	6	1	16	2
10	29	1	24	1	8	2	18	2	29	1	10	1	24	1
11	1	4	22	1	19	1	25	1	14	3	9	1	1	5
12	6	2	16	1	9	2	3	4	7	3	18	1	25	1
13	12	1	4	4	29	1	27	1	24	1	8	1	5	3
14	4	2	3	4	3	2	22	2	23	1	11	1	22	1
15	20	1	21	1	23	1	26	1	27	1	17	1	8	3
16	8	1	20	1	14	1	4	4	18	2	29	1	9	3
17	18	1	19	1	12	2	9	3	20	2	16	1	10	3
18	19	1	7	3	11	2	5	4	22	1	2	2	28	1
19	22	1	8	3	18	1	14	2	9	3	27	1	4	5
20	3	2	6	3	28	1	11	3	13	3	12	1	2	5
21	9	1	30	1	10	2	23	2	12	3	23	1	3	5
22	2	3	9	3	20	1	19	2	15	3	22	1	14	2
23	5	2	13	1	21	1	7	4	21	1	30	1	20	2
24	10	1	12	3	13	2	6	4	19	2	20	1	13	2
25	25	1	11	3	27	1	29	1	5	3	15	1	21	1
26	23	1	5	4	15	1	28	1	11	3	13	1	12	2
27	27	1	29	1	26	1	2	4	28	1	25	1	19	2
28	11	1	28	1	2	2	30	1	3	4	24	1	23	1
29	17	1	27	1	25	1	10	3	8	3	21	1	11	3
30	13	1	10	3	24	1	17	2	17	2	5	2	17	2

④曖昧な検索式、キーワードを用いた場合

これらはブール検索において検索を行う上でも障壁となりうる問題点である。適合度順検索も、語の出現に基づくキーワード検索と言う点でブール検索と同様であり、ブール検索において困難なケースは適合度順検索でも困難なケースになるのである。むしろ問題となるのは、ユーザの考える適合度順と一致していない適合度順検索の方が、ユーザが何件でも出力できてしまうため、検索効率が低下する可能性も十分にあり得る点である。従来の適合度順検索研究では、多くの実験において、ある一定の再現率を達成した時点での精度を比較する研究が多い。そして、再現率が低い段階では非常に精度が高く、また再現率が必要になる場合は出力件数を増やすことによって、高い再現率が得られるとしている。しかし、実際の検索を行うユーザは再現率を知ることはできない。そのため、ユーザがシステムの行った順位付けに対して信頼できない場合、必要以上に出力件数を増やし、本来適合度順検索の利点であるべき精度が低下する事態を招きかねないのである。

確かに適合度順検索システムでは、先行研究の多くで指摘されるように適合レコードは上位の出力に多く、下位の出力では少なくなる傾向が見られる。つまり上位の出力において

は精度が高く、下位まで出力するほど適合文献の数は減少していく。他のデータベースより順位相関係数が高く、ある程度順位付けがうまくいっていると考えられる PsychINFO 及び LISA では、明らかにこの傾向を見ることができる。また PAPERS も下位では適合文献数が減少している。しかし従来の精度を用いた評価方法では、上位 5 位内の適合文献よりも、26 位から 30 位に含まれたレコードの方が有用度は高かったとしても、その点は反映されないのである。NTIS、MEDLINE は全体での適合文献数が少なかったため、下位に有用度の高いレコードが目立つ結果となっているが、順位付けの一致度から見ると、PAPERS が必ずしも MEDLINE より良い結果を得られたデータベースというわけではない。PAPERS においても、下位の出力において有用度の高いレコードが出力されているのである。このことは、上位 5 件の中に 5 段階評価で 5 とされたレコードが含まれた事例と同様に、下位 5 件の中にも 5 段階評価の 5 とされたレコードが存在することからも、明らかである。これまでに行われてきたように、単純に適合文献と非適合文献の数や割合による比較を行うだけでは、適合度順検索システムの評価を行うには不十分であると言える。

今回行った実験では、適合度順検索システムが 26 位以下に出力したレコードを、30 件の中でも有用度が高いと評価した事例は多く見られた。26 位以下に出力したレコードに対して、30 件中 5 位以上の評価をユーザが与えた事例が、25 件中約半数の 12 件あった。特に情報要求 12 は、ユーザが最も高く評価したレコードと二番目に高く評価したレコードが、それぞれ 28 位と 30 位に出力されている。また、今回の実験では相関に負の値が得られた事例が 5 件あった。多くは相関として認められるほどの値は示さなかったが、そのうち 1 件は弱いながらも負の相関が認められている。この場合、ユーザの評価に対して逆順に近い評価を検索システムが行ったことになる。つまり、出力を進めて行くほど有用度の高いレコードが出力されているのである。下位に有用度が高く評価されたレコードが出力されていると、ユーザは出力を止めるべき点を容易に決定できなくなってしまい、結果的に検索効率が悪化することが予想される。

#### 4. 考察

適合度順検索システムは、従来のブール検索と比較し多くの利点を持っていると言われている。しかし、今日実用化が進みつつある語の出現頻度情報に基づいた適合度順検索システムでは、ユーザが欲しいと望んだレコードの順位付けに近い順位付けが十分に行えているとは言い難い。これは現状の適合度順検索がキーワード検索にすぎないことによる限界であると考えられる。適合度順検索が十分に利点を発揮するためには、キーワードのみではなく、よりユーザの情報要求を反映させることのできるシステムを考えていく必要がある。

† 慶應義塾大学大学院文学研究科図書館情報学専攻博士課程

†† Graduate school of Library and Information Science, Keio University.

1) Cleverdon, C. "Optimizing convenient online access to bibliographic databases". Information services & use, Vol.4, p.37-47(1984)

2) Stibic, V. "Influence of unlimited ranking on practical online search strategy". Online review, Vol.4, No.3, p.273-279(1980)

3) Keen, E. M. "How does dialog's target work?", Online & CDROM review. Vol.18, No.5, p.285-288(1994)