

グラフィンタフェースにおけるクエリ操作に基づく検索結果予見システム

A Foreseeing System of Search Results based on Query Operations on the Graph Interface

市倉 丈寛[†] 北山 大輔[†]

[†] 工学院大学情報学部コンピュータ科学科 〒163-8677 新宿区西新宿 1-24-2

E-mail: tj110015@ns.kogakuin.ac.jp, ††kitayama@cc.kogakuin.ac.jp

あらまし 多くの Web 検索サービスではクエリの推薦機能が提供されている。ユーザは推薦されたキーワードから自身の検索意図に合うキーワードを選択することで容易に検索結果を改善することができる。しかしそれらの推薦キーワードは検索結果の概要については予想は困難である。本研究では入力クエリに対して、得られる検索結果をクラスタリングし、クラスタの重心となる特徴語をノードとした操作可能なグラフィンタフェースにより次の検索結果の概要を予見する手法を提案する。提案手法では推薦キーワード同士の共起を計算することで、対話的な推薦キーワードを操作にともない、関連ある他の推薦キーワードも連動して動くインタフェースを提案する。その動作によりどのような結果が得られるか予見させる。

キーワード クエリ生成, web 検索, グラフィンタフェース

1. はじめに

現在、世界に存在する Web ページ数が膨大なものとなっている。また、年々 Web ページの数は増加している。2008 年 7 月時点では、世界に 10 兆もの URL が存在していると Google-Blog^(注1)で公表された。また、Web ページには、Blog や日記といった個人的なコンテンツ、コンピュータ言語のサンプルや Tips のような教科書のようなコンテンツ、国や企業のホームページなどの公的なコンテンツなどの様々な種類のコンテンツが混在している。多種多様な情報を持ち、膨大な数が存在する Web ページの中から信頼性の高い Web ページや得たい Web ページを見つけ出すのは困難である。Web ページを見つけたい時には、検索エンジンが利用される。検索エンジンでは、特定のクエリに対して検索結果の上位しか閲覧されていない。そのため、検索結果で上位になるということが、重要視される。近年、検索結果で上位を獲得するために、検索エンジン最適化 (SEO) 技術が競われるようになってきている。SEO とは、Web ページの内容を向上させるのではなく、html ソースの書き方などを変えることで検索上位を取得するというものである。そのため、必ずしもユーザの得たい情報と検索結果の順位がかみ合わなくなってしまう。この様な状況で、Web 上で目的の情報を探索することは、ますます困難になってきている。

一般的な検索エンジンはクエリとなるキーワードを入力するテキストボックスと検索を実行するボタンからなる。ユーザがテキストボックスにクエリを入力し、検索を実行するとクエリを含む Web ページを検索し、ランキングしてユーザに提供する。しかし、一般に検索する際にユーザは、必要な情報に辿り着くために必要なキーワードを事前に想像しなくてはならない。また、複雑なクエリの方が検索の精度は向上するが、ユー

ザが思いつくクエリは短い。そして、上手なクエリを生成したとしても、必要な情報が得られるとは限らない。そのため、数万件の検索結果から検索意図に合致するキーワードを発見し又、必要であれば、クエリの追加・削除などをしたりユーザは満足が出来る情報を得られるまで、様々なクエリの組み合わせを試行錯誤する必要がある。無駄な検索行為を繰り返す。目的の情報が曖昧な場合には適切なクエリを入力することは難しく、この試行錯誤の回数は多くなる。この問題に対して、ユーザの検索意図を推測したり、検索の候補をユーザに提示する検索の支援を行う研究が盛んに行われている。

本研究では、Web を用いた情報探索を支援することを目的とする。Web 情報探索の支援のためには、クエリに対する Web 検索結果の概観の把握が必要であると考えた。通常、Web 検索結果の概観情報としては入力クエリによる Web 検索結果数の表示や、タグクラウドによる頻出キーワードの可視化といったことが行われるが、入力したクエリによって検索結果がどのように変化するかを予め知ることは出来ない。そこで、Web を用いた情報探索支援のために、検索結果が予見できる視覚的なインタフェースの開発を目指す。

2. 関連研究

2.1 Web 検索結果の概観

一般的に Google 等の Web 検索サービスでは検索結果の概観性を高めるために、HITS や PageRank のようなアルゴリズムでランキングし、タイトルとスเปニットといった Web ページの要約を表示する。さらに進んだ取り組みとして Rerank.jp [6] では、検索結果に対してリランキングを行うことができる。このサービスでは、検索結果中の気になった箇所を選択することで、検索結果を自由に並びかえる事ができる。

また、多次元の軸を操作しリランキングを行うシステムとして STAR [10] がある。STAR では、ユーザがレーダーチャート中

(注1) : <http://googleblog.blogspot.com/2008/07/we-knew-web-was-big.html>

に表示された多次元の軸を操作して変更することにより、検索結果のリランキングを行うものである。例えば、宴会場を探す場合は、「リーズナブル」や「まんぶく度」といった軸がある。ユーザは値段の安い宴会場を探しているなら「リーズナブル」の値を増やし、高くてもいいなら値を減らすといった操作を行う。すると、システムは求める条件に合う店を上位に表示するようにリランキングを行う。

Clusty.com [2] は、vivisimo 社の提供するクラスタリングメタサーチエンジンである。クラスタリングとは、検索結果に対し、出現頻度の高いキーワードによる分類を行い、キーワードごとに検索結果の Web ページをまとめることである。

これらの技術は、一つのクエリに対する検索結果を詳細に把握するのに有効である。また、収束型の探索であり、検索結果の中から目的の Web ページに辿りつくことが目的である。本研究では Web 検索結果の概覧のためにこれらの技術を参考に、Web 検索結果のクラスタリングを行い、そのクラスタリングに属するキーワードを可視化することで、クエリ操作と Web 検索結果の概覧表示の両方を実現している。

2.2 グラフインタフェースによる可視化

検索結果から得られる膨大なデータを多面的に理解するために、グラフの可視化および閲覧システムが必要である。このような検索結果から多面的な情報を得ることを目的とした検索を IBM の Andrei Broder は情報志向型探索と呼んでいる [7]。クエリに対する Web の検索結果の全体像をユーザに提示することで、直感的に必要な情報を理解することができ、検索効率向上を目指した研究 [9] がある。Web 検索結果をページの内容によって適切にクラスタリングし、分類結果にラベルを付加した、amazonode [3] は書籍を視覚的に見せることができ、書籍同士の関連度をエッジ、書籍をノードとするグラフ構造で表現しており、書籍同士の関係性を理解することが容易である。納豆ビュー [4] は Web 空間を 3 次元グラフィックスにより可視化するもので、あるノードを持ち上げる操作により芽づる式にそのノードに関係するノードが全て持ちあげることができるため、関係の深いノードを閲覧することが容易である。

これらの技術は検索結果をユーザが様々な視点から理解する際に有効である。また、探索的な検索であり、1 つのクエリを様々な視点から多角的に見ることが目的である。これらの技術は自分が入力したクエリの検索結果内容を一覽的に表示することで検索効率の向上を目指しているが、入力したクエリによる検索結果のみを示している。本研究ではこれらの技術を参考に、予測されるクエリによる結果を含ませて見せることで、キーワード操作による Web 検索結果の予見を目指している。

2.3 クエリ生成

ユーザの入力したキーワードに関連した単語 (関連語) を提案するシステムがいくつか研究されている。Web 検索用のクエリ作成の支援を目的として、関連語を視覚的に表示した研究 [1] がある。まず、入力したクエリに対する話題語とその話題語に対する共起語を Web 上から抽出する。そして、抽出したクエリを相関度ごとに円型の相関グラフ上に配置し、関係のあるものは、線で結ばれる。また、話題語や共起語と入力したクエリ

の距離で AND, OR, NOT の演算子を変更する。

Quintura [5] は、キーワードに大きさ、位置の情報を持たせて、視覚的に提示する。また、視覚化したキーワードを用いて、動的なクエリ生成を提供する。Quintura のキーワードの大きさには 3 種類しかなく、一度に読み取れる情報量が少ない。

VQuery [8] はベン図を用いたインタフェースであり、楕円の中にクエリを割り当てることを可能にしている。もし楕円が互いにオーバーラップするように配置されておりユーザがその共通部分を選択した場合は AND 検索が実行されが、ここでユーザが共通部分を選択しなかった場合、対応する語集合に基づく OR 検索処理が実行される。NOT 検索はディスプレイ上のアクティブな領域で対象となる楕円が選択されていない状態に対応する。

これらの取り組みはユーザが適切なクエリのキーワードを想起できない場合に有効である。また、これらの取り組みは検索結果をユーザが円滑に試行錯誤することが目的である。しかし、これらの技術はキーワードだけがユーザに提示されるのでどのような背景で推薦されたキーワードなのかユーザは把握できない。本研究ではキーワード操作と共に、連動するキーワードもユーザに提示することで、提示されたキーワードのコンテキストの理解とクエリの想起の両方を実現している。

3. 検索結果予見システム

Web においてキーワードとは Web から目的の情報を見つけるための道具として考えられてきた。これは、目的の情報が明確な場合には有効な手段である。しかし的確なキーワードが想起できていない場合の検索ではクエリの入力し、検索結果の閲覧、適切な語をクエリに追加、検索結果の閲覧といった検索行為を繰り返す必要がある。我々は、ある 1 つの推薦キーワードを操作する時に、関連のある他の推薦キーワードが連動することでそのキーワードを追加/否定した際に変化と予測させるトピックを予見させ、またそのキーワード操作によって生成した、仮想クエリによる検索結果から強調/淡化する新たなキーワードをユーザに提示することで、ユーザに次の検索結果の概観を予見しながらクエリを生成ができると考えた。そこで本研究では検索の試行錯誤を効率的に行うことを目的としたインタフェースを提案する。

3.1 推薦キーワードの抽出

Web 検索結果 N 件のスペニット中に出現する単語頻度を基にキーワードを決定する。タグクラウドのように頻出語の上位のみを用いた場合、特定のトピックに関するキーワードが多くを占めてしまい、推薦キーワードとしてはふさわしくないと考えた。提案手法ではスペニット N 件を階層的クラスタリングし、クラスタの重心となる次元を代表的な語として抽出することで、検索結果に含まれるトピックを網羅するキーワード選択を実現する。また、同じクラスタから抽出した代表的な語は関連のある語として定義する。クラスタサイズに応じて、キーワードの大きさを変更することで、検索結果におけるそのトピックの占める割合を表現する。

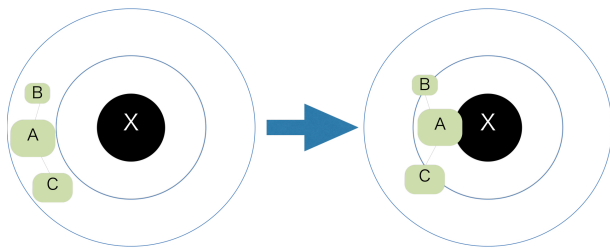


図1 AND 検索時のキーワード連動

3.2 グラフを用いた推薦キーワードの操作

提案インタフェースでは操作したキーワードは円を超えるが、連動するキーワードは円を超えない。円に隣接するキーワードは円の外部にある連動するキーワードに AND 条件で追加されるが、キーワード同士は OR 条件となる。図1はグラフ操作の例を示している。「A」のキーワードを操作して内円の内側に移動させると、AND 検索として検索式を生成する。「B」、「C」のキーワードは円の線上に移動し、OR 検索に加わる。

図2の上部は「X」というクエリで検索した時に出現するキーワードと提案インタフェースを操作してできた「 $X \wedge A \wedge (B \vee C)$ 」というクエリで検索をした時の出現するキーワードの重なりを表した図を示している。図2の下部は、キーワード操作によって強調/淡化するキーワードの様子を表している。「 $X \wedge A \wedge (B \vee C)$ 」というクエリから得られた「G」「H」「I」を強調し、キーワード、「X」というクエリからしか得られないキーワードの「E」「D」「F」は淡化する。「K」「J」はキーワード操作によって得られたクエリの「 $X \wedge A \wedge (B \vee C)$ 」からも出現するキーワードなので変化しない。

4. ForeseeSearch:プロトタイプシステム

図3に今回提案する検索結果予見システム ForeseeSearch のインタフェースのイメージを示す。左にインタフェース、右に検索結果を表示させる。ユーザが操作可能なのはキーワードとして表示されているノード部分のみであり、ノードを操作した結果、生成された検索結果から抽出された新たなキーワードと Web ページが表示される。エッジに関しては表示していると見づらい場合があり、非表示だと動かすまで関係が分からないため、キーワードにマウスオーバーしたときに該当のエッジのみ表示する。検索結果はユーザからの操作が一定時間ない場合に、その時点のキーワードの配置に基づくクエリによる結果を表示する。

図3は「夜景」のキーワードを操作した際に連動して「スポット」「東京」「首都」のキーワードが OR 検索に加わった様子を示している。

本システムは、クエリを入力することから始まる。入力されたクエリに対し、Google Custom Search API から Web 検索結果を取得し、取得した Web 検索結果を Mecab で形態素解析する。ページ毎に単語の TF をカウントし、階層的クラスタリングす

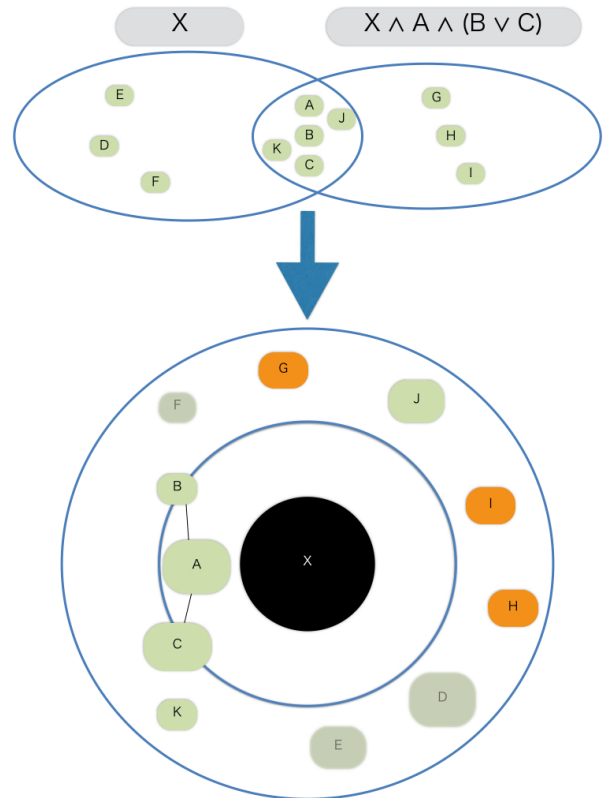


図2 AND 検索時のキーワード強調/淡化

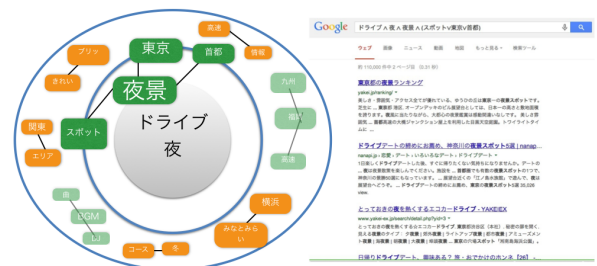


図3 インタフェース

る。代表的な語はクラスタの重心から選出し視覚化する。ノードが描写されてからはノードを操作し、クエリを生成することができる。図4が全体のフロー図である。

5. 実験および考察

5.1 推薦キーワードの抽出に関する実験

抽出したキーワードが検索結果の概要を表しているか検証するため、Web 検索結果 50 件を階層的クラスタリングし、各クラスタから抽出した語に対して、10 人の被験者が適合、不適合を判断した。結果を表1に示す。ドライブのクラスタから抽出したキーワードに関して評価が一致したのが 39 件、評価の不一致は 21 件あった。デートのクラスタから抽出したキーワ

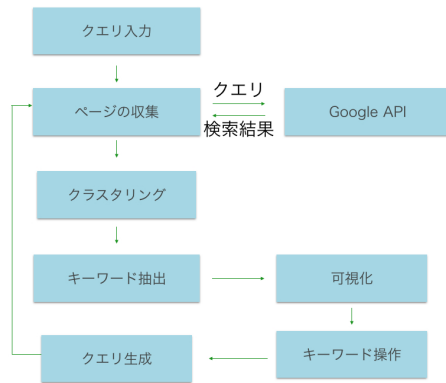


図4 全体のフロー図

表1 実験結果1

	評価が一致	評価が不一致
ドライブ	39	21
デート	40	29
イベント	25	6

表2 実験結果2

	A	B	C
最も参考にならなかった	0%	0%	0%
参考にならなかった	0%	20%	20%
どちらも言えない	10%	20%	10%
参考になった	20%	50%	30%
最も参考になった	70%	10%	40%

表3 実験結果3

	A	B	C
時間がかからなかった	70%	20%	20%
中間	0%	30%	70%
最も時間がかかった	30%	50%	10%

ドに関して評価が一致したのが40件、評価の不一致は29件あった。イベントのクラスタから抽出したキーワードに関して評価が一致したのが25件、評価の不一致は6件あった。各クラスタに関して評価が一致するクラスタの方が多い結果となった。クラスタは必要な単語が連動するため、ユーザが必要な語を操作した際に適切な語が連動すると考えられる。しかしながら、評価が分かれたクラスタは必要な単語と不要な単語が連動するため、改善が必要である。

5.2 提案インタフェースの比較実験

提案するプロトタイプシステムをインタフェースAとし、通常の検索結果をインタフェースB、タグクラウドをインタフェースCとして、2種類の比較実験を被験者10人に行った。実験結果を表2と表3に表す。表2、表3の各値は10人中の回答結果を表している。

実験2の内容は「ルート」をインタフェースの検索ワードに加えた際の出力を被験者が見て、検索ワードとして加えてよいか参考になったかどうかの評価実験を被験者10人に行った。実験3の内容は実験2の判断にかかった時間を同じ被験者10人

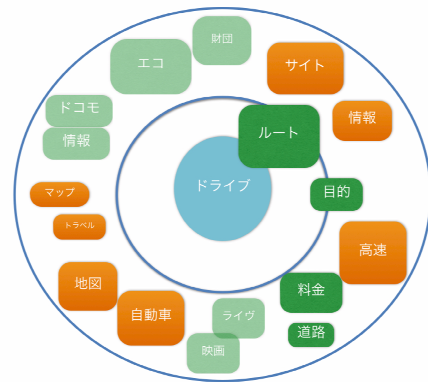


図5 インタフェースA:プロトタイプシステム



図6 インタフェースB:通常の検索結果

に評価実験を行った。表の2では70%の人がインタフェースAが参考になったという結果になった。インタフェースAでは操作したキーワードに対して、連動するキーワードや、淡化するキーワード、新たに出現するキーワードといった視覚情報があるため、インタフェースB、インタフェースCよりもキーワードを加えるべきかどうか決定する際の判断材料が分かりやすいため、最も参考になったと考えられる。

表の3ではインタフェースAが最も判断に時間がかからなく、インタフェースCが2番目に時間がかかり、インタフェースBが最も判断に時間がかかるという結果になった。インタフェースAでは判断材料が他のインタフェースと比較して分かりやすいため、最も時間がかからなかったと考えられる。

5.3 インタフェースの考察

本インタフェースでは推薦キーワードをノード、その関係をエッジとしたグラフとして視覚化させる。内円と外円に対するノードの位置関係を利用してクエリを生成するシステムにすることにより、複数のキーワードの組み合わせによるクエリを直感的に作成、理解できると考えた。

図8は「ドライブ 夜」をクエリとして検索した際のインタフェースの初期位置である。AND検索の例としては図9はユーザが「夜景」のキーワードを操作した際に、連動して「スポッ

ウェイ

エコ 価格 高速 交通

サイト 財団 最新 スポット **情報**

自動車 DVD ディスク どこ 道路 バス 無料

ライブ 料金 **ルート**

図7 インタフェース C: タグクラウド

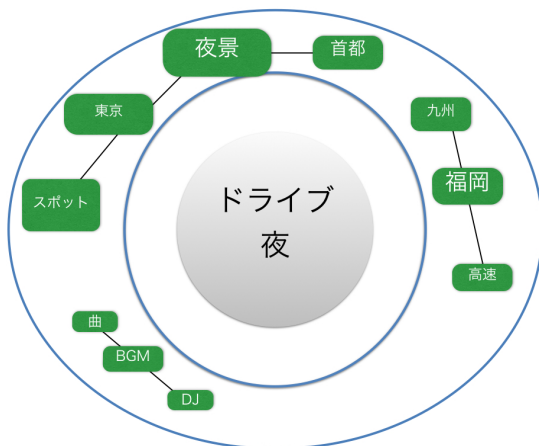


図8 検索結果予見システム

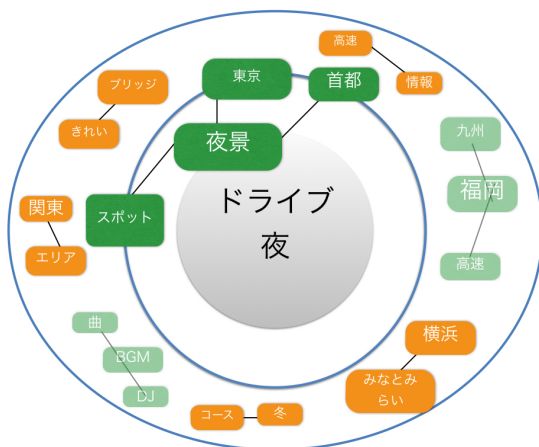


図9 キーワード「夜景」を近づけた場合の挙動

ト「東京」「首都」の3つの推薦キーワードがOR検索に加わり、“ドライブ ∧ 夜 ∧ (夜景 ∧ (スポット ∨ 東京 ∨ 首都))”の検索式が生成される様子を示したものである。新たに追加する語はキーワード操作によって出来た検索式から検索結果を取得し、推薦キーワードを再抽出する。図9は新しい語として「横浜」「みなとみらい」「コース」「冬」「関東」「エリア」「ブリッジ」「きれい」を取得する様子を表している。

NOT検索の例としては図10はユーザが「福岡」のキーワードを操作した際に、連動して「九州」「高速」の2つの推薦キー

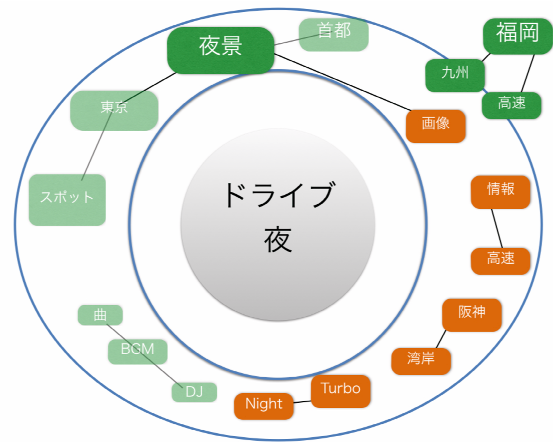


図10 キーワード「福岡」を遠ざけた場合の挙動

ワードがOR検索に加わり、“ドライブ ∧ 夜 ∧ -(福岡 ∧ 九州) ∨ -(福岡 ∧ 高速)”の検索式が生成される様子を示したものである。新たに追加する語はキーワード操作によって出来た検索式から検索結果を取得し、推薦キーワードを再抽出する。図10は新しい語として「画像」「高速」「情報」「阪神」「湾岸」「Turbo」「Night」を取得している様子を表している。

上記で説明したユーザのクエリ生成を支援する検索結果予見システムについて考察する。キーワードを操作すると関連するキーワードが動くことで、いま操作しているキーワードがどんな情報を持っているかユーザは知ることができる。例えば図9は、「夜景」のキーワードを操作すると「東京」「スポット」「首都」が連動する様子を示している。この連動するキーワードから「夜景」のクエリをAND検索に加えると、次の検索結果には「東京」に関する検索結果が増えそうだと予見することができる。

また、図10は「福岡」のキーワードを操作して「九州」「高速」のキーワードが連動する様子を示している。この連動するキーワードから「福岡」のクエリをNOT検索に加えると、次の検索結果には「九州」や「高速」に関する検索結果が減りそうだと予見することができる。このことからユーザは連動するキーワードを参考に検索結果の概要を予見しながら、検索式を生成できると考える。

キーワード操作と同時に新たなキーワードを抽出してくることで、キーワード操作による次の検索結果の概要を予見することができる。図9では「横浜」「みなとみらい」「冬」「コース」「関東」「エリア」「ブリッジ」「きれい」というキーワードが新たに提示された様子を表している。これらの新たに提示されたキーワードから検索結果には関東エリアや、横浜みなとみらいに関する情報が含まれそうだと予見することができる。

また、図10では新たに「画像」「高速」「情報」「阪神」「湾岸」「Turbo」「Night」というキーワードを提示している様子を表している。自分の検索意図とは合わないキーワードを操作した結果、新たなキーワードとして「阪神」「湾岸」を取得した。このことから次の検索結果には「阪神」や「湾岸」に関する情

報が含まれそうだと予見することができる。このことからユーザは新たに提示されたキーワードから、検索式を生成する際に検索結果の概要を予見しながらキーワードの操作が可能だと考える。

6. まとめと今後の課題

本研究ではキーワードを操作し、クエリを生成する手法と、提案インタフェースを組み合わせた検索結果を予見する ForeseeSearch を提案した。具体的には連動するキーワードをユーザに提示し、また生成したクエリの実行結果に含まれるキーワードを強調/淡化することで検索結果の概観を把握可能にした。評価実験としては推薦キーワードの抽出に関する実験を行った。抽出したキーワードは連動してクエリを作るために有用であることを確認したしかしながら、評価が分かれたクラスタについても無視できない量存在し、改善が必要であるとわかった。また、今後の課題はインタフェース部分や Web 検索結果からキーワードを抽出する部分のシステムを実装する予定である。ユーザが試行錯誤した時間および回数を測定し、従来の検索よりも効率がよくなったかを実験する。またインタフェースの感想のアンケートを集計し、提案インタフェースに対する主観評価を測定する。

謝 辞

本研究の一部は、平成 25 年度科研費若手研究 (B)(課題番号：24700098) によるものです。ここに記して謝意を表すものとします。

文 献

- [1] 吉田大我, 小山聡, 中村聡史, 田中克己, : Web 検索結果におけるキーワード出現相関の可視化と対話的な質問変換, 電子情報通信学会第 18 回データ工学ワークショップ第 5 回 DBSJ 年次大会, C7-2, 2007.
- [2] Clusty:<http://clusty.com/>
- [3] Amazonode:<http://amaznode.fladdict.net/>
- [4] 塩澤秀和, 西山晴彦, 松下温, : 「納豆ビュー」の対話的な情報視覚化における位置づけ, 情報処理学会論文誌, Vol38, No11, pp2331-2342, 1997.
- [5] Quintura:<http://www.quintura.com/>
- [6] Rerank:<http://rerank.jp/>
- [7] Broder, A. : A taxonomy of web search, SIGIR Forum, Vol. 36, No. 2, 2002.
- [8] Jones S, VQuery: a graphical user interface for Boolean query Specification and dynamic result preview, 1998.
- [9] 小林拓海, 佐藤大介, 三末和男, 田中二郎, : Web 検索結果の概観提示による情報収集支援, The 19th Annual Conference of the Japanese Society for Artificial Intelligence, 3C3-03, 2005.
- [10] 中谷文彦, 河合由起子, 熊本忠彦, : 柔軟な Web コンテンツ検索のためのレーダーチャート検索システムの提案, 電子情報通信学会 第 19 回データ工学ワークショップ論文集, B5-4, 2008.