

Web ページの重ね合わせを利用した効率的な情報アクセス手法

富永 一成[†] 牛尼 剛聡[‡]

[†]九州大学芸術工学部 〒815-8504 福岡県福岡市南区塩原 4-9-1

[‡]九州大学大学院芸術工学研究院 〒815-8504 福岡県福岡市南区塩原 4-9-1

E-mail: [†]kazu9su@gmail.com, [‡]ushiana@design.kyushu-u.ac.jp

あらまし 近年、iPhone, iPad をはじめとするタッチパネルを有する小型のモバイル端末が普及し、それらを利用して Web ページを閲覧する機会が増大している。現在、Web 上で必要な情報にアクセスするためには、検索エンジンを利用することが多い。しかし、このようなモバイル端末では文字入力用のキーボードが存在せず、文字入力に関するユーザの負担が大きいため、効率的に検索を行うことができない。未知の情報を検索する際には、複数の Web ページを参照しながら情報アクセスを行うことが多い。本研究では、ある Web ページを他の Web ページに重ね合わせることで、重ね合わせられる Web ページがリンクしている Web ページの中から、重ね合わせる Web ページに関連性の高いものを提示する。これにより、少ないキーワード入力で、より効率的で容易な情報アクセスが可能になる。

キーワード モバイル端末, 情報アクセス, Web ページ, 重ね合わせ

An Approach for Effective Information Access with Overlapping Operation on Webpages

Kazushige TOMINAGA[†] Taketoshi USHIAMA[‡]

[†] School of Design, Kyushu University 4-9-1 Shiobaru, Minami-ku, Fukuoka-shi, Fukuoka, 815-8504 Japan

[‡] Faculty of Design, Kyushu University 4-9-1 Shiobaru, Minami-ku, Fukuoka-shi, Fukuoka, 815-8504 Japan

E-mail: [†]kazu9su@gmail.com, [‡]ushiana@design.kyushu-u.ac.jp

Abstract

Keyword

1. はじめに

現在、Web 上には膨大な数の Web ページが存在し、Web 上から必要な情報を取得するために Web 検索は必要不可欠な機能となっている。Web 検索は、既知情報の検索と未知情報の検索の 2 種類に大別できる。既知情報の検索では、ユーザはあらかじめ欲しい情報に対する明確なイメージを有している。したがって、検索エンジンに入力するための適切なクエリを思い出すことができるので、欲しい情報を含む Web ページを容易に検索できる。しかし、未知情報の検索では、ユーザは欲しい情報の具体的かつ明確なイメージを有していないため、様々なクエリによって得られる複数のページに含まれる情報から、自分が欲しい情報を段階的に明確化していく必要がある。現在の検索エンジンは、既知情報を検索したり、検索対象が明確である場合には、効率的な情報アクセスを提供する。しかし、ユーザが検索結果をあらかじめ想定できない対象に関しては、効率的な情報取得が行えないことがある。

例えば、登山に関する専門的な知識を持たない福岡

市に住む人物が、日帰りで登山を計画したいという要求を持つ場合、ユーザは福岡市から日帰りで行くことができる山に関する情報を取得し、交通手段や難易度などの様々な情報に基づいて総合的に判断して、登山の計画を立てる必要がある。

このような未知情報の検索では、1 つのクエリの検索結果となる Web ページの内容に基づいて別のクエリを実行したり、リンクをたどることにより複数の情報を比較しながら検討するなどの段階的な操作が踏まれるため、ユーザの作業に要する時間が長くなってしまふことが多い。さらに、最悪の場合、Web ページから適切なクエリを作り出すことができず、欲しい情報までたどり着けないこともある。

また、ユーザが具体的な検索対象が明確化していない段階で一般的な概念を表す単語をクエリとして検索を行ったとき、「ポータルサイト」や「まとめサイト」等、複数の Web ページにリンクを持つ Web サイトの Top ページにたどり着くことが多い。そのような場合、そのページからアクセス可能なページに掲載されている情報を比較して、ユーザの目的を満足する対象を決

定することがある。このような比較検討を行う際に、サイト A で紹介されている x に関する情報を、サイト B で紹介されている情報で確認したい場合が存在する。こうした場合、ユーザは、サイト B のトップページからリンクを辿ったり、サイト内検索等を利用して、サイト B において x に関する情報が掲載されているページとその部分を探す必要がある。しかし、比較する対象が多くなるにつれて、これらの作業はユーザの負担になる。また、場合によっては、対応する情報を別のサイトで見つけることができない場合もある。

このように、明確な検索対象が存在せず、未知の対象を比較検討する作業は、ユーザにとって困難な場合が多い。この問題を解決するために、本論文では、ユーザの欲しい情報に関連する二つの Web ページを重ね合わせることにより、ユーザがクエリを入力する手間を省いて、ユーザが複数の対象の比較検討を行う作業を支援する手法を提案する。

本手法は、クエリの入力を必要としない直観的で分かり易い操作で、スムーズに関連する情報にアクセスすることができる。これは、近年増加しているタッチパネル、小型画面の文字入力に困難さを伴うモバイル端末での情報アクセスにおいて非常に有効であると考えられる。

論文の構成は以降 2. でアプローチについて述べる。3. で本研究の手法を紹介する。4. でプロトタイプの説明をする。5. で考察を行い、6. でまとめと今後の展望を示す。

2. 情報検索の現状と問題点

2.1 未知情報の検索に関する予備実験

本研究を始めるに当たり、被験者 3 人にタスクを課し、未知情報の検索に関する予備実験を行った。タスクの内容と実験結果を以下に示す。

- ・ 福岡近郊で登山をするために必要な情報をすべて調べる
- ・ 登山に関する知識はない
- ・ 出発地は福岡県福岡市中央区天神とする
- ・ 移動には公共交通機関を用いる
- ・ 検索エンジンは Google を使用する

表 1：未知情報の検索に関する予備実験の結果

被験者	所要時間	タスク達成度
A	10 分	100%
B	20 分	90%
C	30 分	――

一人あたりの検索所要時間の平均は約 20 分であり、3 人ともに例外なく目的地(登山する山)はすぐに決めることができていたが、登山口、初心者向けの登山道などの目的地までの詳細な情報、またそれに伴う交通機関(電車、バスの乗り継ぎ)の検索に、多くの時間を要していた。また被験者 C はタスクを最後まで達成することができなかった。

この実験結果から、具体的な検索対象が明確化していない未知情報の検索は、ユーザの作業に要する時間が長くなってしまうこと、さらに、最悪の場合、Web ページから適切なクエリを作り出すことができず、欲しい情報までたどり着けないこともあることが判った。

2.2 モバイル端末での情報検索における問題点

近年、Apple[4]社の iPhone や SAMSUNG[5]社の GALAXY[6]のようなスマートフォンが普及している。スマートフォンとは携帯電話にパソコンや PDA (携帯情報端末)の機能が備わったもので、電子メール機能や Web ブラウザなどを内蔵し、インターネットに接続できる携帯電話である。スマートフォンはインターネットを使いやすくするために、端末本体をできるだけ小型化し、画面を最大限に大きくしている。そのため、キーワード入力にタッチパネル方式を採用しているものが主流である。しかし、タッチパネル方式のキーワード入力には、次のような欠点がある。

- ・ 汚れで画面が見えにくくなることがある。
- ・ 指先での細かな操作を苦手とし、意図しない選択を検出することがある。判定を厳しくすると、強く押さないと検知されないものがある。特に煙草の灰・煙に含まれる脂(やに)などで汚れると、検知できる精度が低下する。
- ・ 一般的にキーボードほどの素早く正確な入力は難しい。
- ・ 押しボタンと異なりクリック感がないので、入力動作がごちこちになる場合がある。

また、現在のスマートフォンの最大通信速度は、iPhone, GALAXY ともに約 5700Kbps、である。携帯電話向けインターネットサービスを開始したころの通信速度が最大 2.4Kbps であったところに比べれば、およそ 30 倍の速度であるが、PC の平均通信速度 10000Kbps と比較すると未だに速いとは言い難い。さらに Web ページの大多数は PC 用に作られている。しかし、PC 用の情報量の多いコンテンツは高速通信下での閲覧を前

提としているため、スマートフォンなどで閲覧する際には PC での閲覧に比べストレスを感じるが多々ある。

表 2：PC とモバイル端末の通信速度の比較

通信機器	通信方式	速度
iPhone4, GALAXY S	W-CDMA	～5700Kbps
PC	光ファイバー	～100000Kbps

未知情報の検索を行う場合、1 つのクエリの検索結果となる Web ページの内容に基づいて別のクエリを実行したり、リンクをたどることにより複数の情報を比較することが多いため、上記で述べた様なタッチパネルでのキーボード入力や通信速度の遅さがユーザの負担になり得る。そこで、本研究では、ユーザの欲しい情報に関連する二つの Web ページを重ね合わせることで、ユーザがクエリを入力したり、同じ Web ページを何度も閲覧するというような作業を省き、ユーザの負担を減らす手法を提案する。

3. アプローチ

Web 上には多種多様な情報が存在しており、その中から必要な情報を効率的に取得するために検索エンジンが広く利用されている。検索エンジンは、検索対象の特徴が明確である場合には、効率的な情報提供を実現可能である。しかし、ユーザは検索対象の特徴が明確でなく、十分な知識を持たない分野において、複数の候補を比較検討しながら、目的を達成したい場合がある。そうした場合には、検索エンジンを利用した検索のみでは、ユーザの目的を直接的に満足することが困難である。

例えば、登山についての専門的知識を有さない福岡県在住のユーザが、次の休日に登山に行きたい状況を考える。これは、未知の情報を取得する代表的な例の一つと考えることができる。このような場合、ユーザの代表的な振る舞いとして以下のような例を考えることができる。

ユーザが、Google で「福岡 登山」というキーワードで Web ページを検索すると、「九州の山」(<http://www.yado.co.jp/yama/yama.htm>)という Web サイトのトップページが最上位にランク付けされた。このサイトは、九州に存在する個々の山について、個別のページとして紹介している。このサイトのトップページには、図 1 に示すように、個々の山の簡単な特徴と

紹介ページへのリンクが表形式でまとめられている。

九州の山一覧

《別画面に表示されます》

県名	山名	標高	
福岡県	宝満山	869m	登山者人気No.1
福岡県	三郡山	936m	若杉〜宝満山
福岡県	油山	597m	市民の森として
福岡県	福智山	901m	筑豊のシンボル
福岡県	皿倉山	622m	夜は100億ドル
福岡県／大分県	英彦山	1200m	日本三大修験
福岡県	古処山	889m	山頂に巨岩の
福岡県	馬見山	978m	古処山からの
福岡県／大分県	御前岳・釈迦岳	1231m	自然めぐり縦走
福岡県	貫山	712m	平尾台カルスト
福岡県	湯川山	471m	湯川山・孔大
福岡県	香春岳	511m	田川のシンボル
福岡県	障子ヶ岳	427m	山頂は広い草

図 1：個々の山を紹介したページへのリンク一覧

一方、上記の検索結果の 2 番目にランキングされたページは「福岡県ガイド情報(登山・ハイキング) - goo 旅行」

(<http://guide.travel.goo.ne.jp/e/goo/jp/kyushu/fukuoka/hiking/>)というページであった。このページでは、福岡県の登山やハイキングに関する書き込み記事へのリンクが一覧としてまとめられている。



図 2：登山に関する書き込みへのリンク一覧

ユーザは、図 2 の一番上の書き込み記事のリンクをクリックし、図 3 に示す「難所ヶ滝の大つらら」の書き

込み記事を閲覧する。

ユーザはこのようなクエリに関する書き込み記事を閲覧し、それを見るための登山の情報を知りたいが、図3の記事は個人の書き込みであるため一般的な情報が掲載されていない。そこで、図1のページに紹介されている山のページの中で、「難所ヶ滝」に関連するページを表示したいと考える。

このように、未知の対象の比較検討を行う際には、複数のページに存在する情報を組み合わせ、総合的に判断することが行われることが多い。しかし、その作業はユーザにとって冗長な作業を含むことが多い。この例では、ユーザは、「難所ヶ滝」が「三郡山」に存在することを調べ、図1の一覧の上から2番目の「三郡山」のリンクをクリックする必要がある。その結果図4の「三郡山」に関するページを閲覧できる。

このように、同一の対象に関する複数のページに存在する情報を利用して、総合的に比較検討する作業を支援することが本研究の目的である。



図3: 書き込み記事の例

図4: 書き込みに関連する Web ページの例

本研究では、ユーザの欲しい情報に関連する複数の Web ページを重ね合わせることにより、ユーザがクエリを入力等の冗長な操作を必要とせずに欲しい情報を効率的に取得できる検索方式を提案する。

具体的には、ユーザが興味を持つ Web ページ A を、関連情報へのリンクを含む Web ページ B に重ね合わせたとする。このとき、Web ページ B から数クリックで到達可能な Web ページの中に含まれる Web ページ A に関連が強い Web ページ C を提示する (図5)。

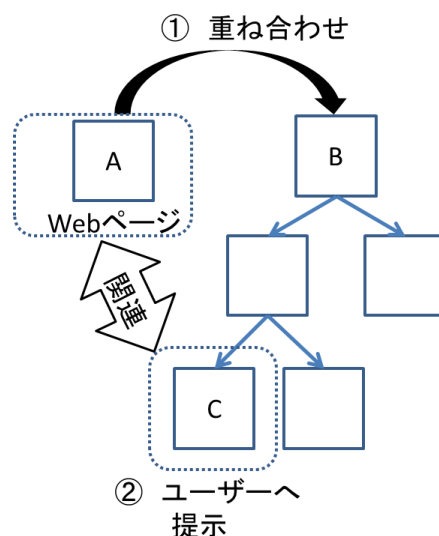


図5: 重ね合わせによる Web ページの提示

例えば、上記の例では、図3に示す「難所ヶ滝」ページを、図1に示す「九州の山」のページに重ね合わせることにより、図4に示す「三郡山」のページが表示されるようにする。

4. 処理手順

上記の提案を実現するために、以下の手法を利用する。

まず、重ね合わせられる Web ページの URL を抽出する。URL の抽出には正規表現を利用する。具体的には重ね合わせられる Web ページのソースコード中で、リンクが宣言してある部分の URL を取得する。(今回はリンクによって同じ Web ページをループする可能性を考慮して2階層下の URL まで取得することとする。) 取得した URL はデータベースに格納される。

次に、重ね合わせる Web ページと、重ね合わせられる Web ページ、抽出した URL の Web ページ全てに関して形態素解析を行う。形態素解析を利用してそれぞ



れの Web ページの文章を品詞に分けて名詞を取得する。更に、取得した名詞の中から特徴的な単語のみを抽出する。

そして、TF・IDF 値を利用して、ユーザの求める情報を含む Web ページのキーワードを取得する。TF・IDF は、文書中を代表する単語を抽出する代表的な特徴量である。この TF・IDF を用いて、抽出した特徴的な単語の TF・IDF 値を求める。

TF・IDF は以下の式で表される。

$$TF \cdot IDF = TF \times \log \frac{N}{DF}$$

ここで、

TF、キーワード抽出対象テキスト中の代表キーワード候補出現数を、N は全てのドキュメント数を、D は代表キーワード候補が含まれるドキュメントの数を表す。

この計算式に基づいて各 Web ページ内での単語の重みを決定する。

次に、重ね合わせる Web ページと重ね合わせられる Web サイトの Top ページがリンクしている Web ページの類似度を求める。類似度の計算にはコサイン相関値を利用する。ここでは重ね合わせる Web ページに含まれる単語の重みベクトルを $\mathbf{a}$ 、重ね合わせられる Web サイトの Top ページがリンクしている Web ページのうち、一つに含まれる単語の重みベクトルを $\mathbf{b}$ とする。

コサイン相関値は次の式で表される。

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{\|\mathbf{a}\| \|\mathbf{b}\|} = \frac{\sum_{i=1}^n a_i \cdot b_i}{\sqrt{\sum_{i=1}^n a_i^2} \times \sqrt{\sum_{i=1}^n b_i^2}}$$

この計算式によって重ね合わせる Web ページと、重ね合わせられる Web サイトの Top ページがリンクしている Web ページとのそれぞれの類似度を求める。

そして、得られた類似度を元に、重ね合わせる Web ページに関連の高い順に重ね合わせられる Web ページがリンクしている Web ページを表示する。

4. 実験

4.1 プロトタイプシステム

上記の手法の有効性を評価するために、プロトタイプシステムを実装した。システムの実装にあたって、プログラミング言語として perl を用い、形態素解析には MeCab を利用した。システムはコマンドライン上で対象となる Web ページの URL を指定することにより動作する。

4.2 実験

今回の実験は、以下のような前提で行った。

- 重ね合わせる Web ページを「福岡県ガイド情報（登山・ハイキング） - goo 旅行 難所ヶ滝」（http://4travel.jp/traveler/omtarou/album/10424558/?cid=goo_oem）とする。
- 重ね合わせられる Web ページを「九州の山」（<http://www.yado.co.jp/yama/yama.htm>）とする
- 対象とする Web ページは、「九州の山」サイト内の Web ページとし、外部リンクは対象外とする。
- 重ね合わせる Web ページに関係がある Web ページ（正解とする Web ページ）は、「難所ヶ滝」の場所、さらに登山に関連のある Web ページかどうかを考慮し、表 3 に示す 4 つの Web ページとする。
- 得られた結果のうち上位 10 件を対象に再現率、適合率を求め、考察を行う。

表 3: 関連のある Web ページ

宝満山（九州の山）	http://www.yado.co.jp/yama/f_houman/houman.htm
三郡山（九州の山）	http://www.yado.co.jp/yama/f_sangun/sangun.htm
太宰府天満宮	http://yado.co.jp/kankou/fukuoka/dazaifu/tenmangu/tenmangu.htm
旅に役立つリンク集	http://yado.co.jp/link/link1/link1.htm

4.3 実験結果

対象とした Web ページは 327 ページであった。実験によって得られた結果より、重ね合わせる Web ページとのコサイン相関度の高い上位 10 件の Web ページとコサイン相関度を表 4 に示す。なお、検索結果の Web ページのスナップショットは付録 A に示す。この場合、実験結果の再現率は、0.75、適合率は 0.3 であった。

表 4: 実験結果

順位	サイト名	類似度	正解
1	写真で見る九州の船	632.9	
2	太宰府天満宮	265.8	○
3	雲仙温泉	259.7	
4	古里温泉	223.9	
5	旅に役立つリンク集	207.1	○
6	古湯温泉	155.9	
7	宝満山(九州の山)	150.4	○
8	ハウステンボス	133.4	
9	普賢岳(九州の山)	129.6	

10	大江天主堂	123.1	
----	-------	-------	--

5. 考察

今回の実験では、関連のある Web ページの内、3 つの Web ページが検索結果上位 10 件に含まれた。しかし、関連のある Web ページよりも、関係のない Web ページのコサイン相関度の方が高かった。

得られた結果のうち、まず 2 位にランクインされた「太宰府天満宮(福岡県太宰府市)・写真満載九州観光」と、7 位にランクインされた「宝満山(九州の山)」に注目する。上記で述べた 2 つの Web ページは、重ね合わせる Web ページとして用いた「福岡県ガイド情報(登山・ハイキング) - goo 旅行 難所ヶ滝」と関連のある Web ページである。難所ヶ滝は、宝満山と三郡山の間にある河原谷の上部に位置している(図)。また、難所ヶ滝へのルートは、福岡市宇美町からの昭和の森ルート、または太宰府市の竈神社から宝満山を通り難所ヶ滝に至るルートが有名である。「太宰府天満宮(福岡県太宰府市)・写真満載九州観光」と、「宝満山(九州の山)」における単語のうち、「宇美」、「太宰府」、「難所」など条件にあてはまる語の TF・IDF 値が高くなっていたことによって上位にランクインされていた。

ここで注目すべきは、「福岡県ガイド情報(登山・ハイキング) - goo 旅行 難所ヶ滝」というページ単体では、「太宰府」、そして「宝満山」という情報を見つけ出すのは困難であるという点である。ブログの内容には、「太宰府」、「宝満山」という単語は存在しておらず、「太宰府」はリンクとして見つけることができるが「宝満山」という情報は全く含まれていない。これは、興味深い結果である。

次に、1 位にランクインされた「写真で見る九州の船」、は、「福岡県ガイド情報(登山・ハイキング) - goo 旅行 難所ヶ滝」に含まれる「賃貸」、「マガジン」、「アトラクション」、「イルミネーション」などの単語の TF・IDF 値が高くなり、関連する Web ページとして上位にランクインされている。これは、広告として旅行特集の雑誌の広告が載っていること、またそれらの情報は「九州の山」の Web サイト全体のうち「写真で見る九州の船」だけに関連する広告が載っていることが原因であると考えられる。

3 位にランクインされた「雲仙温泉(長崎県雲仙市小浜町)・写真満載九州観光」、4 位にランクインされた「古里温泉(鹿児島市)・写真満載九州観光」、6 位にランクインされた「古湯温泉(佐賀市富士町)・写真満載九州観光」、8 位にランクインされた「ハウステンボス(長崎県佐世保市)・写真満載九州観光」は、「予

約」などの旅行に関する単語や「痛」など、温泉に関連する単語の TF・IDF 値が高くなっていることがわかる。これは「福岡県ガイド情報(登山・ハイキング) - goo 旅行 難所ヶ滝」という Web ページはブログであり、旅行記という分類をされているので、旅行に関する広告が多いためであると考えられる。

また 5 位にランクインされた「旅に役立つリンク集」は、登山をしようとしていると仮定したので関連のある Web ページとしたが、「カード」や「カー」、「カメラ」など、広告に関連する単語の TF・IDF 値が高くなったことにより上位にランクインされたと考えられる。全く予想していなかった方法で、関連する Web ページを抽出したという結果になった。

最後に、もう一つの関連のあるページとした「三郡山(九州の山)」がなぜ上位にランクインしなかったのかについて考察する。三郡山に関する二つの Web ページの TF・IDF 値上位 10 件を表 5 及び表 6 に示す。「三郡山」という単語の TF・IDF 値が、「福岡県ガイド情報(登山・ハイキング) - goo 旅行 難所ヶ滝」においてあまり高くないことがわかる。「三郡山」という単語は、ブログ中の本文にでてきているが、他の広告に関連する単語のように何回もでてくることはなく、1、2 回である。



図 5：難所ヶ滝

しかしブログ本文中に出てきているということは重要な単語である。今後はこのような単語にどのように重要度をつけていくかが課題となる。また三郡山(九州の山)は、21 位にランクインされており、コサイン相関度は 83.1 であった。

表 5: 「難所ヶ滝」に含まれ、且つ「三郡山」に含まれる単語の「三郡山」における TF・IDF 値

三郡山	56.7
飯塚	12.6
宇美	6.3
安心	6.3
航空	3.1
三	2.0
郡	1.4
山	1.0
指定	1.0
登山	0.9

表 6: 「難所ヶ滝」に含まれ、且つ「三郡山」に含まれる単語の「難所ヶ滝」における TF・IDF 値

三郡山	10.7
飯塚	8.0
安心	4.5
宇美	3.9
航空	3.7
さん	2.8
三	1.0
お	0.8
人気	0.8
滝	0.8

6. 関連研究

効率的な情報アクセスを可能にするための研究は多数存在する。

近藤ら[1]は Wikipedia などの不特定参加型の百科事典サイトにおいて内容的に関連する項目を探し出す場合において、それまで考慮されていなかった項目内の文書構造にも着目し、セクション単位でリンク解析を行っている。これによって利用者の検索意図に合った、よりの確な関連項目を検索する手法を提案している。この手法は百科事典サイトの中での検索を想定しているため、百科事典サイト内での検索での使用に限定される。本手法では、情報アクセスに用いる Web ページは特定の Web サイトに限定されないため、Web ページ閲覧における様々な状況でユーザの情報アクセスを促進することができる。

田ら[2]は質問キーワードの数が二つの場合を対象に、質問キーワード間の意味的関連を分類した上で、文書内における質問キーワードの近接性を用いて、ウェブ検索結果を改善する手法を提案している。しかし

この手法ではユーザが二つのクエリを明記する必要がある。本手法ではWebページを重ね合わせるという直観的な操作で、新たなクエリを入力することなく情報アクセスを行うことができる。

武田[3]らは実空間上のコンテンツを携帯電話のカメラによって拾い上げ (Pick)、他の実空間上のコンテンツに重ね合わせる (Lap) ことにより、複数の実空間上のコンテンツを意味的な関係性をふまえて仮想的に融合し、ユーザの目的に適した形式で情報を提示するという Pick and Lap 手法を提案している。この研究では、写真と地図など、実空間に存在するコンテンツを重ね合わせることで、それぞれ単体では得られなかった新たな情報をユーザに提供している。この手法は、Pickされるコンテンツと、Lapされるコンテンツがともに重ね合わせることで融合が可能なメタデータを持つことを前提としているので、メタデータを持っていないコンテンツ同士を重ね合わせることはできない。本研究では、二つのWebページを重ね合わせることでユーザに情報を提供する。すべてのWebページは情報を持っているので二つのWebページに関連性を見つけ、考慮することができれば、あらゆるWebページにおいて有効である。しかし武田らのように、あらかじめコンテンツに情報を与え、重ね合わせて提供する情報を想定することで、よりユーザの目的に適した形式で情報を提示できる可能性もある。

旭[4]らは、既知の2つのオブジェクトをクエリとして受け取り、ある観点でオブジェクトを順序づけた場合に、その2つのオブジェクトの間に存在するようなオブジェクトをユーザに提示することで、明確な検索意図を持っているにも関わらず、的確なクエリを思い出すことができないために求める情報にたどり着くことができないユーザへ情報を提供している。この手法は、ユーザが明確な検索意図を持っていることを前提としている。例えば“桶狭間の戦い”と“本能寺の変”の間にあった出来事を知りたい、といった要求をユーザが持っているとき有効である。本研究とは、2つのクエリを用いてユーザに新しい情報を提供するという点で目的は同じだが、本手法では明確な検索意図を持っていないユーザに、より広い意味で有効な情報を提供する。

7. まとめ

本論文では、Web ページの代表キーワードの類似度を求めることによって、キーワードレスで、気軽な情報アクセスを、二つの Web ページを重ね合わせるという操作を行うことで実現する手法を提案した。今後の課題として、今回のプロトタイプシステムを用いて評

価実験を行っていく。さらにシステムの精度改善を行うとともに、よりユーザの視点に立って Web ページ間の関連度を求めるための手法、そしてより汎用性の高いシステムの考察を行っていく。

また、本手法の発展として、例えば以下のような応用が考えられる。

- ・ 登山をしたいが登山する山への交通アクセスがわからないとき、登山する山の Web ページをバス会社の Web ページに重ね合わせて最寄のバス停や時刻表を表示する。
- ・ 自転車などの製品の Web ページを Amazon [5] の Web ページに重ね合わせて Amazon ホームページ内にある製品のページを表示する。
- ・ 旅館や観光名所などの Web ページを GoogleMap[6] の Web ページに重ね合わせて周辺の地図を表示する。

参 考 文 献

- [1] 近藤直樹, 渡辺陽介, 横田治夫: “Wikipedia のセクションを考慮したリンク解析による関連項目検索の実現”, DEIM Forum 2010
- [2] 田 馳, 手塚 太郎, 小山 聡, 田島 敬史, 田中 克己: “質問キーワードの意味的関連と近接性に着目したウェブ検索の精度改善”, DEWS2006
- [3] 旭 直人, 山本 岳洋, 中村 聡史, 田中 克己: “挟み込む検索: 2 オブジェクトからの補完オブジェクト発見”, IPSJ 2009
- [4] 武田 十季, 鶴野 玲治, 牛尼 剛聡: “Pick and Lap: カメラ付き携帯電話を用いた実空間コンテンツの重ね合わせによる効果的な情報閲覧”, DEIM Forum 2010
- [5] Amazon :
<http://www.amazon.co.jp>
- [6] GoogleMap :
<http://maps.google.co.jp>