

補完情報の自律型検索機能を有する Web ブラウザ

吉田 大我[†] 中村 聡史[†] 小山 聡[†] 田中 克己[†]

† 京都大学大学院情報学研究科社会情報学専攻 〒 606-8501 京都市左京区吉田本町

E-mail: †{yoshida,nakamura,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

あらまし 本稿では、ウェブ閲覧と検索を融合した新しいブラウジング手法を提案する。提案手法では、ウェブ閲覧に連動して関連ページを動的に検索して提示する。また、閲覧中のページと関連ページを、キーワードマップにより可視化する。ユーザがキーワードマップ上で対話操作を行うと、自動的にキーワードの重み付けが変更され、関連ページが再サーチおよび再ランキングされる。この仕組みによりユーザは意図に応じた情報探索が可能となる。

キーワード 情報検索, 可視化, 対話操作, クエリフリー検索

A Web Browser with Autonomous Search for Complementary Information

Taiga YOSHIDA[†], Satoshi NAKAMURA[†], Satoshi OYAMA[†], and Katsumi TANAKA[†]

† Department of Social Informatics, Graduate School of Informatics, Kyoto University

Yoshida-Honmachi, Sakyo, Kyoto 606-8501 Japan

E-mail: †{yoshida,nakamura,oyama,tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract In this paper, we propose a Web browsing method which merges web browsing and search. The system automatically searches and shows related pages when a browsing page is changed. Our system visualizes a currently-browsed page and related pages by a keyword map. When a user operates a keyword map, weights of keywords operated are changed and searched related pages are re-ranked. The user can find a target page and navigate to the page by this interaction.

Key words information retrieval, visualization, interactive operation, query-free search

1. はじめに

近年、情報を収集するために Web が広く利用されている。ユーザは Web ブラウザを利用して Web ページ上にあらかじめ設置されたハイパーリンクを辿ることにより各種情報の閲覧を行うことができる。一方、ユーザはサーチエンジンを利用することにより、ハイパーリンクでは接続されていないページを発見し、閲覧することができる。これまで、ハイパーリンクを辿ることによる Web 閲覧とサーチエンジンを用いた Web ページ検索は互いに独立に用いられてきており、ハイパーリンクによるナビゲーションとサーチは連動していない。

Web 閲覧と Web 検索における問題は次のようになる。

(1) ページからページへのナビゲーションは、Web ページの作者があらかじめ定義したハイパーリンクに限定されている。例えば、ある企業のページを閲覧中にライバル企業の情報を知りたいと思っても、そのようなライバル企業のページへのリンクが設置されていることは非常にまれである。

(2) サーチエンジンにより検索を行う際、ユーザがこれま

でにどのようなページをナビゲーションにより閲覧していたかという情報は考慮されない。

ユーザがあるページを閲覧している際に現在閲覧中のページを補完する情報を知りたいと思った場合、これらの問題はより深刻なものとなる。

本論文では、上述の問題を解決するための新しい Web ブラウザを提案する。提案システムの主な特徴を以下に示す。

- (1) Web 閲覧に連動して自動的に検索
- (2) 補完情報の検索
- (3) キーワードマップによる補完情報の可視化
- (4) 情報補完ページへのハイパーリンクの自動生成

本システムは、ユーザの Web ブラウザ上でのページ閲覧に連動して関連ページの検索を行う。ユーザがブラウザであるページを表示したのに連動し、システムは閲覧しているページを分析し、クエリを生成して補完情報を持つ関連ページを検索する。システムは、閲覧ページ中の語および検索した関連ページ中の語を一種のハイパーリンクとしてキーワードマップ上に可視化する。ユーザは、キーワードマップ上の語を操作することで表

示する関連ページを切り替えることができる。

本稿では、まず研究の概要について説明する。そして、システムのデザインと実装について説明する。その後、システムの評価を行い、関連研究について述べる。最後にまとめを行い、今後の方針について論じる。

2. 研究の概要

2.1 研究の動機

我々の研究班は、Web 検索エンジンの使い方に関する調査を行った [1]。この調査では、我々の研究に関連する質問が 2 つあった。ひとつは、Web 検索を行う状況に関する質問であり、もうひとつは Web 検索を行う理由の質問である。調査により、ユーザがサーチエンジンを用いる主な状況は、ユーザがある特定の情報について調べたいと思ったとき、および Web ブラウジングの最中が多く、続いて、特に目的なく Web 検索を行う場合が多いことが分かった。また、ユーザはある事柄に対して、それに関する説明を知りたい (46.0%)、もしくはより詳しく知りたい (36.8%) ということを経験している理由に Web 検索を行うことが多いことが分かった。これらの調査は、多くの人にとって Web ナビゲーション中の Web 検索が重要であり、求められていることを示唆している。しかし、多くの人にとってブラウジング中に関連するページを検索することは容易ではない。ユーザは関連する Web ページを見つけるためにサーチエンジンにアクセスし、適切なクエリを入力する必要がある。

2.2 研究の目的

本研究の目的は、ユーザの Web 閲覧中に補完情報を持つページを自動的に検索し提示することにより、ユーザの情報収集を支援することである。通常、ユーザは本システムを一般的な Web ブラウザと同様に利用する。システムはユーザの Web ページ閲覧に連動して、自動的に補完情報を持つ関連ページの検索を行う。また、システムはキーワードマップ上に補完情報に相当する語となる話題語を表示する。

本研究では、話題語を以下のように分類する。

- 内在語：閲覧ページ中から抽出した話題語
- 外在語：閲覧中のページに関連するページから抽出した話題語（ただし、閲覧ページ中から抽出した語は外在語の抽出対象から除外する）

キーワードマップのイメージを図 1 に示す。閲覧中のページおよび関連ページから抽出した語はキーワードマップ上に表示される。話題語を配置する位置は以下のように決定する。

- 内在語はキーワードマップ中の円内に配置
- 外在語はキーワードマップ中の円の周囲に配置

ユーザは、キーワードマップ上の語を見ることで閲覧中のページの話題を把握できる。また、キーワードマップはユーザが操作することが可能であり、語のノードをマウス操作することにより理想的な関連ページを見つけることができる。

図 2 はシステムのイメージ図である。ユーザが新しいページを表示した際のシステムの動作を以下に示す。

- (1) ユーザがブラウザ領域に新しいページを表示
- (2) システムは閲覧ページから内在語を抽出

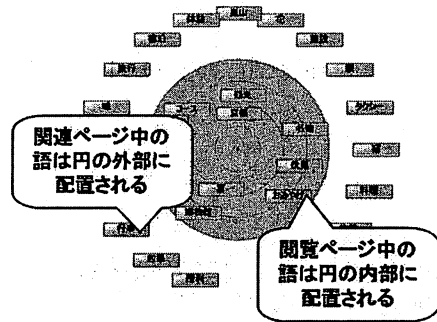


図 1 キーワードマップのイメージ

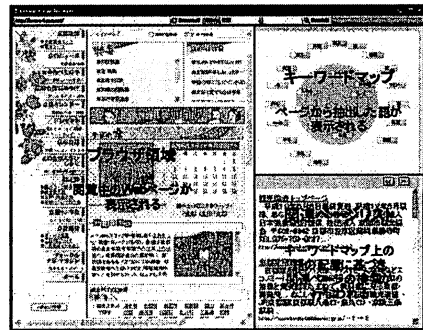


図 2 システムのイメージ図

- (3) 内在語から数語を選び、検索クエリを作成・検索
- (4) 取得した検索結果を分析し、外在語を取得
- (5) 内在語と外在語をキーワードマップ上に配置
- (6) 同時に検索した関連ページのリストを表示

キーワードマップ上に表示された語ノードは、クリックや移動といったマウス操作を行うことが可能である。ユーザはキーワードマップ上に表示された円の中心に語ノードを近づけることで、その語に関連する話題を含む関連ページを閲覧することができる。逆に、ユーザはその語ノードを円の中心から遠ざけることにより、その語を含まない関連ページを閲覧できる。つまり、キーワードマップ上でのユーザ操作が一種のハイパーリンクのように動作する。ユーザは画面上のボタンを押すことにより、次々と関連ページを見ることが可能である。

2.3 システムの利用例

カメラを購入しようとしているユーザを例に、システムの利用シナリオについて説明を行う。この例では、ユーザがカメラを購入しようとしているけれども、どの機種を購入しようか悩んでいる状況を想定している。

- (1) まず、ユーザは amazon のページでデジタルカメラの項目を選ぶ
- (2) ランキングを見ることにより、Canon の IXY が人気であることが分かる
- (3) しかし、ユーザは他のデジカメに関する情報も見たいと感じる
- (4) そこで、ユーザはキーワードマップ上の「比較」とい

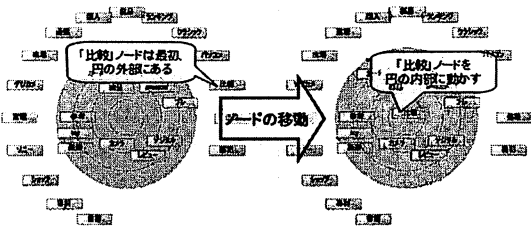


図3 操作前と操作後のキーワードマップ

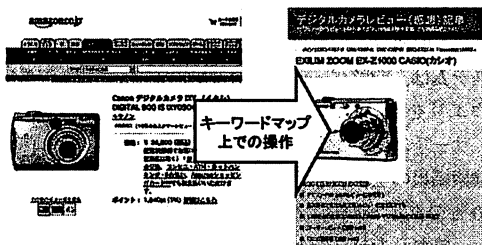


図4 閲覧ページと検索された関連ページ

うノードを円の内部に移動する

(5) システムは、自動的にデジカメの比較を行っているページの検索を行い、ユーザに提示する

(6) ユーザはデジカメの比較を行い、購入するデジカメを決定する

デジカメを比較したページを見つけるためにキーワードマップに対しユーザが語ノードの移動を行った際のキーワードマップの様子を図3に示す。左側の図が操作を行う前のキーワードマップ、右側が操作後のキーワードマップである。ユーザ操作によって「比較」ノードが移動している。

図4は左側のページがユーザが最初に開いたIXYのページ、右側がシステムが提示した関連ページである。検索された関連ページでは、複数のデジカメのレビューが行われており、比較を行うことができる。

3. システムのデザインと実装

システムを実装するにあたり、いくつかの機能が必要となる。要求される機能には以下のようなものがある。

- 閲覧中のページからの話題語（内在語）の抽出
- キーワードマップ上に内在語を配置
- 抽出された内在語からクエリを生成
- 関連ページの検索を行い、関連ページから話題語（外在語）を抽出し、キーワードマップ上へ配置
- キーワードマップ上の配置に基づき、表示する関連ページを変更

本章では、これらの機能について説明する。

3.1 システムのデザイン

システムを利用する際、ユーザは閲覧したいページのURLを入力、もしくは検索クエリを入力を行う。URLを入力し「Connect」ボタンを押すとそのページが表示され、クエリを入力して「Search」ボタンを押すとそのクエリによる検索結果が

表示される。ユーザはシステム画面左側のブラウジング領域で、Webページ閲覧を行うことができる。

ページのロードが完了すると、システムはページ中のテキストから語の抽出を行う。これらの語は、閲覧ページ中の頻出名詞を求めることにより抽出される。ただし、「それ」や「もの」といった意味を持たないと考えられる語は抽出の対象外とした。

システムは、頻出する語から10語を選びキーワードマップ上に配置する。それぞれの語は、語の出現頻度の基づいて配置する位置が決められる。閲覧ページ中における出現頻度の高い語ほど、キーワードマップ内の円の中心付近に配置される。逆に、出現頻度の低い語は円の内側の円周付近に配置される。

そして、システムは現在閲覧中のページに関連するページの検索を自動的に行う。まず、システムはキーワードマップ上の円の中心付近にある語をいくつか選択し、クエリの生成を行う。システムは、サーチエンジンでクエリを検索し、100件の検索ページを取得する。

キーワードマップ上には、閲覧ページにおける出現頻度に応じて語が配置されている。それはつまり、中心付近に配置された語を多く含むページは閲覧ページに類似したページであると考えられることを意味する。システムは、検索した関連ページ一件ずつに対して、キーワードマップ上の語の重み（中心からの距離により計算される）と関連ページ中における語の出現頻度に基づきスコアを計算する。システムは、計算したスコアに基づき関連ページの再ランキングを行い、スコアの高い何件かの関連ページを画面上に表示する。

サーチエンジンでページを検索した際、検索されたページはいくつかの要素で表される。要素とは主に、タイトル、URL、スニペットである。スニペットとは、ページ中からクエリの周辺の文章を抽出することにより作られる要約文のことである。システムは、検索結果として取得した100件のページに対して、どのような語が何回出現したかをカウントする。そして、頻出単語上位20個がキーワードマップ上に配置される。

キーワードマップ上に語が配置される際、閲覧ページ中から抽出した語は円の内部に、関連ページ中から抽出した語は円の外周に配置される。そこで、キーワードマップにおける語の配置はユーザの意図を表しており、キーワードマップの円内はユーザが閲覧したいと考えているページを表していると考えられることができる。システムはユーザが何らかの操作を行うまでユーザの意図を十分に知ることはできないため、初期状態では閲覧中のページに類似したページを関連ページとする。

ユーザは、マウス操作によってキーワードマップ上の語ノードを移動することができる。語ノードを移動することで、語ごとの重みづけが修正される。ある語ノードを中心に近付けたならばその語に対して高いスコアが与えられ、逆に外側に遠ざけたならばその語に対して低いスコアが与えられる。ユーザが語ノードの移動を行うと、システムは自動的に関連ページのスコアを再計算し、関連ページの再ランキングを行ったうえで、ページを提示する。

外在語を抽出する際、すでに閲覧ページ中から抽出した語は関連ページ中からは抽出されない。したがって、関連ページ中

から抽出された語は閲覧ページ中には全くもしくはわずかしが含まれない語であると考えられる。そこで、ユーザがある語をキーワードマップの円外から円内に移動したならば、ユーザは閲覧中のページにはその語に関する情報があまり載っていないと感じており、さらにその情報について知りたいと感じていると考えられる。逆にユーザがある語ノードを円の内部から外側に向けて移動をしたならば、現在閲覧中のページにはその語が含まれてはいるが、その情報は見たくないと感じていると考えられる。システムは、そのような語に対して負のスコアを与える。

ページが表示された際、システムは関連ページのリストを表示する。つまり、ユーザがどのようなページを閲覧したいと考えているかという情報は、現在表示されているページにアクセスしたという情報以外には存在しない。そのため、キーワードマップの初期配置に基づき生成されたクエリはユーザの意図を反映しておらず、検索された関連ページ中にはユーザの求めている情報は存在しない可能性がある。もし、ユーザが検索されたページ中に求めるページがないと感じたならば、ユーザはクエリを変えて再検索を行うことが可能である。語ノードをダブルクリックすることにより、システムはその語を含むクエリを生成し、再検索された関連ページを表示する。

ユーザがキーワードマップに対して操作を行うと、システムはキーワードマップ上の語の配置に基づいて1つの関連ページを画面上に表示する。さらに、関連ページのリストを画面右下に表示する。表示される関連ページのリストは数件であるが、キーワードマップへの操作や次の関連ページを表示するためのボタンを押すことにより、さまざまな関連ページを閲覧することが可能である。

3.2 システムの実装

3.2.1 閲覧中のページからの話題語（内在語）の抽出

システムの実行において、まず、システムは表示されたWebページのHTMLコードを取得し、その中からHTMLタグを除いてテキストを抽出する。そして、システムは閲覧ページのテキストの形態素解析を行い、名詞および未知語を抽出する。なお、形態素解析にはMeCabを用いた。ただし、話題語にはならないと考えられる語はストップワードとして抽出の対象外とした。システムは抽出した語に対し、それぞれ出現回数を数え、出現回数の多い方から10語を話題語として抽出する。

3.2.2 キーワードマップ上への内在語の配置

システムは、閲覧ページ中における出現回数に基づき、抽出した語をキーワードマップ上に配置する。頻繁に閲覧ページ中に出現する語ほど、キーワードマップ内の円の中心付近に来るように配置される。システムは、以下に示す式により、キーワードマップの中心と語ノード間の距離 $d(t)$ を決定する。

$$d(t) = \frac{r}{\sqrt[4]{\text{freq}(t) + 1}} \quad (1)$$

ただし、 r はキーワードマップ内の円の半径、 $\text{freq}(t)$ は閲覧ページ中の語の出現回数を表す。 $d(t)$ は語の出現回数に応じて、 $0 \sim 0.841r$ までの値を取る。システムは、円の中心からの距離が $d(t)$ となるように語をキーワードマップ上に配置する。

3.2.3 抽出された内在語からのクエリ生成

システムは、キーワードマップ上の円の中心に近い語ノード3つを選び、クエリを生成する。これらの語を T_1, T_2, T_3 とおくと、クエリ Q は以下のように表される。

$$Q = (T_1 \text{ OR } T_2) \text{ AND } (T_2 \text{ OR } T_3) \text{ AND } (T_3 \text{ OR } T_1)$$

このクエリによって、3語のうち2語以上を含むページが検索される。3語のAND検索である $Q = T_1 \text{ AND } T_2 \text{ AND } T_3$ というクエリを生成しないのは、 $T_1 \sim T_3$ 中に話題とは関連性の低い語が含まれ、検索結果の件数が非常に少なくなってしまう危険性を極力減らすためである。

3.2.4 関連ページの検索および外在語の抽出と配置

システムは、閲覧ページ中の語の出現頻度に基づいて生成されたクエリを用いて、サーチエンジンで関連ページの検索を行う。サーチエンジンとして、Yahoo! Web Search API を用い、検索結果の上位100件を取得して利用する。検索結果ページ100件のタイトルおよびスニペット中に頻出する語を20語抽出し、これらを外在語とする。外在語は、閲覧ページ中にはほとんど含まれていない語であり、閲覧中のページを補完する話題に関連する語であると考えられる。語の抽出方法は、4.2.1と同じである。抽出した語は、キーワードマップ上の円の周囲に配置される。

3.2.5 キーワードマップの配置に基づく関連ページの表示

キーワードマップ上の語は、システムによりスコアが付与される。語のスコアは、語ノードとキーワードマップ中の円の中心との距離により計算される。検索された関連ページは、このスコアに基づきスコアリングが行われ、再ランキングされる。ページのスコア $S(p)$ の計算式は以下ようになる。

$$S(p) = \sum_{t=1}^N x_{pt}(r - d(t)) \quad (2)$$

ただし、 p は検索結果ページ、 N はキーワードマップ上の語の数、 t, r, d は4.2.2と同様である。 x_{pt} は、ページ p が語 t を含むかどうかを表し、含むならば1、含まないならば0の値をとる。語ノードが円の外側に配置されていたならば、その語のスコアはマイナスとなり、その語を含むページは低く評価されることとなる。ただし、関連ページ中から抽出された話題語である外在語が円の外側にある場合はスコアを0とした。

システムは、検索された関連ページをスコア $S(p)$ が高い順にソートする。そして、 $S(p)$ の高い関連ページ上位数件が関連ページリストとして表示される。なお、ユーザがキーワードマップの操作を行うと、システムはユーザが強調した語に関する話題の補完情報を含むページを見つけるために、 $S(p)$ の再計算を行い、最もスコアが高いページをブラウザ領域に表示する。

4. システムの評価

システムの有効性について調べるため、いくつかのページに対して本システムを用い、適切な関連ページが表示されるかどうかを調べた。実験は、Yahoo!ニュースのヘッドライン (<http://headlines.yahoo.co.jp/hl>) からリンクされているニュース記事15件および、Amazonのエレクトロニクスコーナー

(<http://www.amazon.co.jp/b?ie=UTF8&node=3210981>)で取り扱われている商品ページ15件に対して行った。評価手法としては、システムが検索した関連ページに対し、再ランキング後に1位となったページが閲覧ページ中の話題と同一の話題のページであった場合、類似するページであった場合、関連しない話題のページであった場合の3段階で評価を行なった。

実験の結果は、Yahoo!ニュースのヘッドラインのニュースページにおいて、15件中9件で同一の話題のページ、2件で類似するページを発見でき、4件は関連ページを表示することができなかった。なお、ニュースの話題としては、野球や年金、大雨や地震といったものがあった。また、Amazonのエレクトロニクスコーナーの商品ページでは、15件中7件で同一話題ページ、5件で類似ページを発見でき、関連ページの表示ができなかったのは2件であった。対象とした商品は、電子辞書やヘッドホンといったものがあり、液晶ディスプレイのページを見ている際に液晶テレビが関連ページとして表示された場合などは類似ページと判断した。

Yahoo!ニュースの結果を分析すると、ページ中の話題の規模と自動検索された関連ページの正解・不正解の間に相関性は見られなかった。実験において、システムがユーザに提示するページが閲覧中のページと関連するページである場合は、システムが生成したクエリ中に話題に関連する語が含まれている場合が多く、逆に関係のないページを提示してしまった場合には、生成したクエリが閲覧ページの話題とはほとんど関係のない場合が多かった。したがって、適切にクエリを生成できた場合は関連ページを検索することができるが分かる。また、関連ページの検索において、適切なクエリを生成することが重要であると考えられる。

また、Yahoo!ニュースの記事とAmazonの商品ページの実際結果を比較すると、Amazonの方が高い精度で関連ページの検索ができていることがうかがえる。これは、Amazonの商品ページの方がYahoo!ニュースの記事に比べてページ内の文章が長いためであると考えられる。Yahoo!ニュースの記事において関連ページを検索することができなかったものの中には、ニュース記事自体は数行しかなく、関係のない文章の割合が多いため、話題の中心となる語を抽出できない場合があった。適切な話題語を抽出するためには、単純に頻度だけで語を選ぶのではなく、文書の構造や、語の使われ方を分析する必要があると考えられる。

5. 関連研究

クエリフリー検索とは、ユーザが操作を行うことなくシステムが自動的に関連ページの検索を行う手法のことである。Hartら[2]は、クエリの入力なしにユーザがプリンタのマニュアルを検索することを可能とするシステムを構築した。また、Henzingerら[3]はユーザがテレビを閲覧する際に自動的に字幕キャプションを解析し、関連するニュースページを検索するシステムを構築した。我々が提案したシステムの手法もクエリフリー検索の一種であると考えられる。しかし、我々のシステムはユーザの意図を取得するためにキーワードマップを採用し

た点でこれらの研究とは異なる。提案するシステムでは、いくらかのユーザとのインタラクションにより、ユーザの意図をより深く把握することが可能となる。

ユーザが検索以外の行為を行っている際に関連するページを提示する研究は他にもある。

WebTelop[4]はユーザが見ているテレビ番組に関連するページを検索するシステムである。システムは、現在見ているテレビ番組を補完する情報を自動的に提示する。このシステムでは、話題をどのような視点から見た情報が欲しいかという意味表示を行うことができない。我々のシステムは、キーワードマップを操作することにより、関連ページの話題傾向を変えていくことができる点で異なる。

灘本ら[5]は、ユーザが閲覧中のページと類似ページを比較することができるシステムを構築した。CWBと呼ばれるこのシステムでは、ユーザがあるニュースサイト内でニュースページを閲覧する際、システムが別のニュースサイトから類似するニュースページを取得し、並べて提示する。ユーザは、システムにより類似するニュースページを閲覧することができるが、比較対象とするニュースサイトのURLはユーザが最初に指定する必要がある。

湯本ら[6][7]によるシステムやCWS[8]と呼ばれるシステムでは、二つのオブジェクト名からページを検索し、得られたページを分析することにより二つのオブジェクトを比較することができる。しかし、これらのシステムでは、ユーザは比較するオブジェクト名をあらかじめ知っている必要がある。

我々のシステムは、ユーザが関連ページを閲覧する際、比較対象を入力する必要がない点でこれらの研究とは異なる。

KeyGraph[9]はテキストデータ中の重要な語をノードとし、それらを共起関係によって結んだネットワーク図で可視化することで、文章構成のキーワード抽出を行うソフトである。KeyGraphは1つの文書を解析対象としているだけであり、Web検索結果全体などは解析対象としていない。また、可視化したネットワークを対話的に操作することは想定されていない。

Yahoo! Mindset[10]は、ユーザの操作を検索結果に反映させるシステムである。このシステムは、検索結果リストに表示される各検索結果について買い物(shopping)に適したページなのか、情報収集(researching)に適したページなのかといったスコアリングを行っている。ユーザは、検索対象に応じて、スライドバーを操作し、重みづけを変化させることで検索結果を動的に再ランキングすることができる。しかし、Yahoo! Mindsetではユーザの意図を示すための軸を新たに生成することはできない。閲覧ページ中から語を抽出することにより、話題に合わせた評価軸をキーワードマップ上に自動的に生成する点で本研究は異なる。

WebGlimpse[11]は、検索とWeb閲覧を融合するシステムである。ユーザは目的とするページを現在閲覧中のページからハイパーリンクによって辿ることができるページ群から検索することができる。しかし、このシステムにより検索できるページは現在閲覧中のページからハイパーリンクで辿ることのできるページに限定されるという問題点があった。我々が提案する

システムでは、柔軟にクエリを生成し、サーチエンジンで検索を行なうことにより、表示される関連ページは限定されず、あらゆるページを関連ページとして検索可能である。

6. まとめと今後の課題

本稿では、関連ページの検索機能を持った新しいブラウザを提案した。システムは、ユーザのブラウジング中に自動的に関連ページの検索および提示を行う。本章では、システムの有効性について考察を行い、今後の課題について述べる。

Web ページを閲覧する際、目的のページを見つけることはユーザにとって困難である。

まず、閲覧中のページに自分が求める情報が含まれているかどうかを判断することは、ページ中の情報を一度すべて読む必要があり、ユーザの負担となる。我々の提案したシステムは、閲覧ページ中に頻出する単語をキーワードマップ上に表示することにより、ページ中の話題傾向を一覧的に把握することを可能としている。

次に、もし閲覧中のページが求める情報を含んでいなかった場合、別のページを見つける必要がある。閲覧中のページから張られているハイパーリンクをすべて辿ってページを探すことは困難であるし、目的のページへのリンクがあるとは限らない。サーチエンジンを使ってページを探す場合も、適切なクエリを考えることは一般のユーザにとって難しいことである。提案システムは、関連ページを検索するためのクエリを自動的に作り、関連ページのリストと関連ページ中の話題を提示することによってユーザの支援を行う。ユーザは、キーワードマップから自分の目的に合った語を選んでいくだけでよく、適切な語が何であるかを考える必要がない。

ユーザがサーチエンジンでクエリの検索を行う際、多くのページが検索結果として表示される。しかし、これらのページ中に目的のページが含まれているかどうかを調べるためには一件ずつ結果を見ていく必要があり、全てのページを見ることは非常に困難である。我々のシステムは得られたページの再ランキングを行うことで、結果中にユーザが求める話題のページが含まれているかどうかの判断を支援可能である。

しかし、我々が提案するシステムには問題点もある。ひとつの問題は、関連ページを検索するために時間がかかってしまうことである。ユーザがページを閲覧する際、ユーザはそのページ自体に意識を向けていると考えられる。そこで、ページの表示が完了した後に、関連ページの検索のためにユーザを待たせることは望ましくない。ブラウザを作業不可能な状態にすることなく、関連ページの検索を行うようにすることが重要である。

本システムを用いると、ユーザは閲覧ページと関連ページにおける話題傾向を一覧的に把握することができる。システムは、キーワードマップを用いてどんな話題が閲覧ページや関連ページに含まれているかを提示する。しかし、ユーザはどの話題が一般的な話題で、どの話題がマイナーな話題なのかといった情報はキーワードマップからは得ることができない。キーワードマップ上のノードの大きさを変えるなどして、話題の一般性をユーザに提示できるようにすることは、今後の課題である。

また、システムの利用時にキーワードマップ上に操作を行いたい語のノードが配置されていないことがある。そこで、ユーザにより自由にキーワードマップ上に語の追加を行えるようにすることが必要であると考えられる。ユーザの追加した語に関しても重みを変更し、ユーザが求めるページの検索に利用することができるよう、システムを改良していく予定である。

ユーザが長いページを閲覧する際、ページの一部だけを見られるようにしたいといった、ページのレイアウトをカスタマイズしたいと感じる場合がある。我々はシステムを改良し、キーワードマップによる操作でページ中の段落ごとに文字サイズを変更したり、文字のハイライトを行うなど、ページのレイアウトを変えることができるようにしていきたいと考えている。

謝辞 本研究の一部は、文部科学省研究委託事業「異メディア・アーカイブの横断的検索・統合ソフトウェア開発（研究代表者：田中克己）」、文部科学省科学研究費補助金特定領域研究「情報爆発時代に向けた新しい IT 基盤技術の研究」における、計画研究「情報爆発時代に対応するコンテンツ融合と操作環境融合に関する研究」（研究代表者：田中克己、A01-00-02、課題番号 18049041）、「情報爆発に対応する新 IT 基盤研究支援プラットフォームの構築」（研究代表者：安達淳、Y00-01、課題番号：18049073）によるものです。ここに記して謝意を表するものとします。

文 献

- [1] S. Nakamura, S. Konishi, A. Jatowt, H. Ohshima, H. Kondo, T. Tezuka, S. Oyama and K. Tanaka: "Trustworthiness Analysis of Web Search Results", Proc. of the 11th European Conference on Research and Advanced Technology for Digital Libraries (ECDL) (2007).
- [2] P. Hart and J. Graham.: "Query-free information retrieval", IEEE Expert, pp. 32-37 (1997).
- [3] M. Henzinger, B.-W. Chang, B. Milch and S. Brin: "Query-Free News Search", Proc. of the 12th International World Wide Web Conference (WWW) (2003).
- [4] Q. Ma and K. Tanaka.: "WebTelop: Dynamic tv-content augmentation by using web pages", Proc. of IEEE International Conference on Multimedia and Expo (ICME), pp. 173-176 (2003).
- [5] A. Nadamoto and K. Tanaka: "A Comparable Web Browser (CWB) for Browsing and Comparing Web Pages", Proc. of the 12th International World Wide Web Conference (WWW), pp. 727-735 (2003).
- [6] T. Yumoto and K. Tanaka: "Finding Pertinent Page-Pairs from Web Search Results", Proc. of The 8th International Conference on Asian Digital Libraries (ICADL) (2005).
- [7] T. Yumoto and K. Tanaka: "Page Sets as Web Search Answers", Proc. of the 9th International Conference on Asian Digital Libraries (ICADL), pp. 224-253 (2006).
- [8] J.-T. Sun, X. Wang, D. Shen, H.-J. Zeng and Z. Chen: "CWS: a comparative web search system", Proc. of the 15th international conference on World Wide Web (WWW), pp. 23-26 (2006).
- [9] 大澤幸生, N. E. Benson, 谷内田正彦: "KeyGraph: 語の共起グラフの分割・統合によるキーワード抽出", 電子情報通信学会論文誌 J82-D-1, No.2, pp. 391-400 (1999).
- [10] "Yahoo! mindset.
<http://mindset.research.yahoo.com/>".
- [11] U. Manber, M. Smith and B. Gopal: "WebGlimpse - Combining Browsing and Searching", Proc. of the Usenix Technical Conference (1997).