

ブラウジングと検索の融合：閲覧履歴からの関連情報の検索・自動提示にもとづくウェブ閲覧

Blend of Browse and Search: Web Browse with Autonomous Retrieve of Related Information by Using Web Browsing History

吉田 大我[◇] 中村 聡史[◇]
田中 克己[◇]

Taiga YOSHIDA Satoshi NAKAMURA
Katsumi TANAKA

本研究では、ウェブ閲覧中の関連情報閲覧を支援するシステムを提案する。システムはユーザのウェブ閲覧に連動して自動的に関連ページを検索し、提示する。その際、ユーザのウェブ閲覧履歴からキーワードを抽出し、それを利用することにより提示する情報の精度を向上させることを試みた。関連ページ中のキーワードは、キーワードマップと名付けたグラフにより可視化を行う。キーワードマップ上には、関連ページ中の語がノードとして表示される。ユーザはキーワードマップ上での対話操作により、関連情報を含んでいるページの閲覧が可能となる。

In this research, we propose a system that supports users to browse related information while browsing the Web. The system automatically retrieves and presents related pages synchronizing with users' Web browsing. In this system, we tried to improve accuracy of information by extracting keywords from users' Web browsing history, and using them. Keywords in related pages are visualized by the graph named a keyword map. On the keyword map, keywords in related pages are displayed as nodes. Users become possible to browse related pages by interactive operations on the keyword map.

1. はじめに

インターネット上において情報を入手しようとする際、起点となる行為は大きく2つに分けられる。1つはポータルサイトやリンク集などからページ中に設置されたハイパーリンクを辿ることにより目的の情報が書かれているページを発見する方法であり、そしてもう1つは検索エンジンを用いて関連するキーワードを入力することでページを見つける方法である。適切なリンク集を知っている場合や、適切なキーワードで検索することができる場合はこれらの方法により

瞬時に情報を得ることができる。しかし、起点となるページやクエリを思いつかなければ目的の情報を得ることは困難である。

また、インターネットを初めとするプル型配信メディアの短所として、ユーザが要求した情報以外の情報にユーザがたどり着くことが困難であり、発見が少ないという点が挙げられる。テレビや新聞などのプッシュ型メディアではユーザが能動的に情報を探していくわけではないため、意図していない情報に触れ、新たな分野に触れる機会が多いが、インターネットの場合は目的とする情報に特化したコンテンツが多く、そうした発見の機会が少ない。

これらの問題を解決するために、トラックバックやリンク元ランキングなどの逆リンクを取り入れることによって話題性のある情報や関連情報の入手を容易にしようと試みているページも存在する。しかし、全てのウェブページにおいてランキングや関連情報へのリンクがあるというわけではなく、ユーザが幅広い情報を得ることができるページの割合はわずかである。

また、関連情報へのリンクを含むページであっても、ある1つのページ中に存在するハイパーリンクは通常1人の作者によって定義されたものであり、話題の網羅性に欠けていたり主観的になり偏っていたりする可能性がある。例えば、過去に書かれたページには新しい情報へのハイパーリンクは存在せず、また企業のホームページ中のある商品に関するページにはライバル企業へのハイパーリンクや、悪い評価を与えるレビューなどはめったに存在しない。

そのため、ユーザはある情報について調べる際、同様の事柄に対して書かれたページを幅広く閲覧し、公正な判断を行うことが求められる。しかし前述の通り、あるページからリンクされているハイパーリンクはページの作者によって選別されたものであり、リンク先のページもまた作者と同じ傾向をもったものだけに限定される可能性がある。現在、ウェブページを閲覧中に関連する情報を幅広く入手したいと思った際、関連する情報を一覧的に提示するシステムは存在しない。

そこで本研究では、任意のページに対してページ中の話題に関連する情報や注目されている情報を提示するシステムを提案する。本システムは、ユーザのウェブ閲覧に連動して閲覧中のページにおいて、メインテーマとなっているキーワードを判別する。そして、そのキーワードだけでは検索結果中の話題がまとまりのないものになってしまうため、統一性のある内容のページを検索できるようにクエリ修正を行い、関連するページを検索する。この際、ユーザの過去の閲覧ページの履歴を考慮することにより、ユーザが興味を持ち、かつ話題性の高いページを提示することを試みた。

システムは検索したページを自動的にスコアリングし、再ランキングを行う。また、再ランキング後に関連ページのタイトルおよびスニペットを順番にユーザに提示することにより、ユーザはウェブブラウジング中における受動的な情報閲覧が可能となる。

2. 関連研究

WebGlimpse[1]は、検索とウェブ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を対象として検索することができる。しかし、このシステムにより検索可能なページは狭い範囲に限定されてしまうという問題点があった。我々が提案するシステムは、ウェブブラウジング中にユーザが情報を検索するのではなく、システムがバックグラウンドで動的に情報を検索することにより、情報の発見を支援することを目的としている。

[◇] 学生会員 京都大学大学院情報学研究科 博士前期課程
yoshida@dl.kuis.kyoto-u.ac.jp

[◇] 正会員 京都大学大学院情報学研究科社会情報学専攻
{nakamura, tanaka}@dl.kuis.kyoto-u.ac.jp

クエリフリー検索とは、ユーザが操作を行うことなしに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。提案するシステムでは、従来のクエリフリー的な情報提示に加え、ユーザとのインタラクションにより、表示される情報の傾向を動的に変更することができる。

受動的な情報提示を目的としたシステムとして、Goromi-TV[4]やPointCast[5]が挙げられる。Goromi-TVは大坪が開発した、ローカルに存在するビデオを受動的に閲覧可能なシステムであり、PointCastはあらかじめチャンネルとして株価や天気などの閲覧したいコンテンツを指定しておくことでスクリーンセーブとして情報を自動的に提示するシステムである。これらのシステムでは、あらかじめユーザがコンテンツを指定することにより、閲覧するコンテンツの傾向を指定することができる。

ウェブ上のコンテンツを受動的に閲覧するためのシステムとして、TVML[6]やfeedTV[7]、Energy Browser[8]やAmbient Browser[9]などがある。TVMLとは、林らによって開発された、ウェブページなどのコンテンツをテレビ番組のような受動的に閲覧できる内容に変換してユーザに提示するシステムである。feedTVとは、RSSリーダーのようにいくつかのページを登録しておくことで、それらのページを一定時間毎に切り替えて表示するサービスである。これらのシステムにより、ユーザは受動的にコンテンツを閲覧することができる。Energy BrowserおよびAmbient Browserは、前者はユーザの運動、後者は照明の状態や気温などの環境に関する情報といった、ユーザの文脈に応じて表示するコンテンツの内容や表示速度などを変更することができるシステムである。これらのシステムでは表示されるコンテンツをユーザが事前に指定する必要があった。本研究では自動的に検索を行うことにより、閲覧できる話題の種類を広げることを試みた。

自動的に検索を行い、ユーザに提示するシステムとして、Memorium[10]やGoromi-Web[11]などがある。渡邊らによって開発されたMemoriumというシステムは、最初にユーザが入力したキーワードをクエリとして自動的に検索を行なう。ユーザはシステムに表示される情報を眺めることにより、受動的にコンテンツを閲覧することができる。Goromi-Webも初期に入力したキーワードをクエリとしてウェブページを検索し、自動的に関連する画像や文章をユーザに提示する。両システムでは、マウス操作によってユーザが閲覧するコンテンツの傾向を変えることができる。これらのシステムはユーザが最初に入力したクエリによる検索結果中のコンテンツに対し、時間変化やユーザ操作によって表示されるコンテンツが変化するシステムであった。それに対し、本研究ではユーザのウェブ閲覧に連動して動的に関連するコンテンツを提示することを目的としている。提案するシステムは、クエリの代わりにユーザが閲覧しているページおよび過去に閲覧したページの履歴を入力として利用する。

3. システムの概要

本研究では、ユーザのウェブ閲覧に連動して、閲覧中の情報に関連する情報を自動的に検索および提示するシステムを提案する。提案システムにより、ユーザは受動的なウェブページの閲覧が可能となる。また、ユーザはシステムに対し能動的にかかわることにより、提示されるコンテンツの内容を変更することができる。

本研究はユーザに対して閲覧中の情報を補完する情報を提示することを目的としているため、システムをブラウザとは独立したアプリケーションとして実装した。ユーザがブラウジングを行う際、システムをサイドバーのように表示して利用することを想定している。

3.1 履歴の利用

システムはユーザが閲覧しているウェブページ中に出現する語を解析する際、ページ中における主な話題が何であるかを分析する。しかし、1つのページにおいて取り扱われる話題は1つであるとは限らず、複数の話題が同時に存在する可能性がある。そして、どの話題に興味を持ちページを閲覧しているかはユーザによって異なると考えられる。

例えば、あるユーザが「京都大学工学部」のページを閲覧しているとき、京都の有名な場所を調べている際にたまたま京都大学の情報を見ているのならば興味の対象は「京都」にあると考えられ、京都大学に存在する学部(医学部・法学部など)を順番に閲覧しているならば「京都大学」自体に興味があると考えられる。また、ユーザがさまざまな大学の工学部のページばかり閲覧しているとすれば、興味の対象は「工学部」にあるといえるだろう。

システムは閲覧ページ中からキーワードを抽出し、それをクエリとして自動的にページの検索を行う。しかし、前述のように同じページの場合でもユーザによって興味を示している内容は異なる場合があるため、それに応じてクエリも変える必要がある。そこで、本研究ではユーザのウェブ閲覧履歴を利用することによって、ユーザ操作により検索に利用するクエリを変更することを試みた。具体的には、現在のページから抽出したキーワード群に対し、閲覧ページおよび閲覧履歴中におけるそれらのキーワードの出現傾向を分析することによりユーザが興味を示していると考えられるキーワード群を抽出する。システムは閲覧ページから抽出した語を主題語候補として定義し、スコアリングを行うことにより検索に利用する主題語を決定する。

抽出した主題語を元に、システムは関連ページの検索を行う。だが、単純に主題語のみをクエリとして検索した場合、検索結果ページ中にさまざまな話題が混在し、主題語とあまり関係のないページやユーザの興味が無いページが多く含まれてしまう可能性がある。一度に提示するコンテンツの話題を統一性のあるものにするため、本研究では検索エンジンのクエリログを利用することによってクエリの拡張を行い、関連ページの検索を行うようにシステムを実装した。クエリログとは、インターネット上のユーザが検索するために入力したクエリを蓄積したログである。クエリログはインターネットユーザ全体の検索履歴であり、これを利用することにより表示するコンテンツの内容を変化させ、また話題性のある情報をユーザに提示することができると考えられる。



図1：システムの全体図
Fig.1 System overview

3.2 受動的なウェブ閲覧のための自動的検索

システムイメージを図1に示す。システムは、キーワードマップおよび主題語やクエリを選択するためのプルダウンメニューから構成される。ブラウザと共にシステムを起動した状態でユーザが新しいページを開くと、それに連動してシステムは自動的に閲覧中のページに関連する情報を提示する。本システムを利用する際のシステムの振る舞いを列挙すると以下ようになる。

1. ユーザがブラウザ上で新しいページを表示する
2. システムは閲覧ページ中から主題語候補を抽出する。そして、閲覧ページおよび履歴中のキーワード出現傾向にしたがって主題語候補をスコアリングし、最もスコアが高いものを主題語とする
3. システムはクエリログを利用して、主題語を含み、かつ頻繁に検索されているクエリのリストを取得する。それらのクエリをユーザのウェブ閲覧履歴中のページに出現するキーワードと比較することによりスコアリングを行い、検索に利用するクエリを決定する
4. システムは作成したクエリで検索を実行する。システムは得られた検索結果から選んだページをスライドショーの形式で表示し、一定時間ごとに表示する内容を切り替えることで、ユーザに受動的なコンテンツ閲覧をさせることができる

詳細については、4, 5, 6章で解説を行う。

3.3 能動的な操作による提示コンテンツの内容変更

システムが自動的に検索し、提示したコンテンツに興味が無い場合、ユーザはシステムに対して操作を行うことにより、表示されるウェブページの傾向を変えることができる。ユーザはキーワードマップ上に表示されたノードをクリックするか、図2のようにプルダウンメニューによってクエリを選択することにより、検索に利用するクエリを変更することができる。主題語も変更したい場合は、主題語を表示しているプルダウンメニューをクリックし、新しい主題語を選択することで変更できる。主題語を変更するとシステムは再びクエリ拡張を行い、新しく検索した結果ページをユーザに提示する。

4. 主題語の抽出方法

4.1 主題語候補の選択

システムはユーザが閲覧中のページ中からテキストを抽出し、形態素解析を行う。形態素解析とは、文章を形態素と呼ばれる単位ごとに分割し、品詞情報を付加する操作である。なお、本研究では形態素解析器としてMeCab[12]を用いた。

形態素解析を行った結果の品詞情報を受け、システムは主題語候補となるキーワードを抽出する。

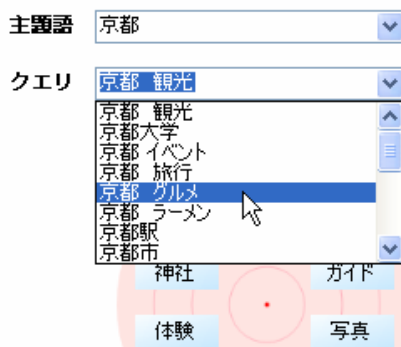


図2 : プルダウンメニューによるクエリの選択
Fig.2 Query selection with pull-down menu

主題語とは、その語について何らかの議論を行うことが可能なオブジェクトの名称として定義した。例を挙げると、「京都」や「ヘッドホン」などといった実際の物体を想像することができるものは主題語候補としてふさわしく、「走る」や「美しい」といった漠然とした語は主題語候補として適切ではない。

主題語候補を抽出するため、形態素解析においてその品詞が名詞と解析されたキーワードを採用した。しかし、単純に名詞のみを抽出しただけでは「私」や「それ」といった代名詞や「積極的」といった形容動詞語幹（「積極」）や接尾語（「的」）も名詞として解析されてしまうため、より詳細にキーワードとしてみなす品詞を分類した。主題語候補として抽出する語の品詞の例を以下に示す。

抽出する名詞

- ・ 名詞, 一般
- ・ 名詞, 固有名詞

抽出しない名詞

- ・ 名詞, サ変接続
- ・ 名詞, 形容動詞語幹
- ・ 名詞, 数
- ・ 名詞, 接尾
- ・ 名詞, 代名詞
- ・ 名詞以外の語

また、「人」や「情報」といった一般的過ぎる語や「ページ」や「トラックバック」といった意味の薄い語を除外するためストップワードリストを定義し、リストに含まれる語は主題語候補から除外した。

4.2 主題語の選択

システムが抽出した主題語候補は、閲覧ページ中における出現頻度や履歴中におけるキーワードを含む文書数に応じてスコアリングを行い、最もスコアが高いキーワードを主題語として採用する。その他の主題語候補は、スコアの低い順にプルダウンメニューに格納される。ユーザはメニューから語を選択することにより、主題語を変更することができる。各主題語候補のスコア S_{T_i} は以下のように決定される。

$$S_{T_i} = tf_i \times \sum_{k=1}^N \left(\frac{x_k}{k} + 1 \right) \quad (1)$$

ただし、 i は主題語候補に振られた通し番号、 N はユーザのウェブページ履歴の総数、 x_k は履歴を最終訪問時刻が最近のものから順に並べたとき、 k ページ目のページタイトルにキーワードが含まれたら1, 含まれなければ0をとる変数である。

このスコアリングにより、現在閲覧中のページに多く含まれ、かつ最近閲覧したページ中に多く含まれるキーワードほど高いスコアが与えられるため、ユーザが興味を示している語を主題語として提示することができる。

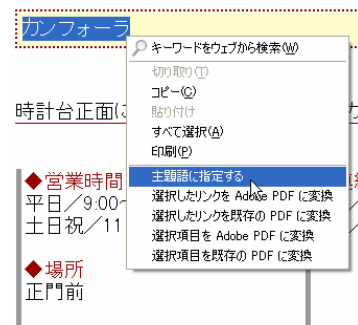


図3 : コンテキストメニューからの主題語指定
Fig.3 Selecting topic term from contextual menu

プルダウンメニュー中に目的の主題語が表示されていない場合は、直接入力するか、もしくは図3のようにブラウザ上で主題語にしたいキーワードを選択した状態で右クリックを行い、コンテキストメニュー中から「主題語に指定する」という項目を選ぶことで直接主題語を指定することができる。

5. クエリ拡張

5.1 クエリ候補の抽出

システムは選択された主題語を元に、関連ページの検索を行うためのクエリを構築する。しかし、1つのキーワードのみをクエリとした場合、検索結果として検索される話題の傾向が定まらず、ユーザにとって有用な情報が検索結果中に含まれない可能性がある。例えば、「京都」というクエリで検索した結果中には観光情報や大学、不動産などの情報が混在している。

そのため、偏った話題やユーザの興味のないページばかりを提示してしまう可能性がある。そこで、本研究では主題語に対してクエリ拡張を行い、それを利用することにより、関連ページを検索することを試みた。

クエリ拡張をする方法として、いくつかの手法を考えることができる。それらの手法のうち1つが、主題語のみを検索クエリとして検索し、その結果中から主題語に関する話題を限定することができると思われる語を抽出する方法である。しかし、先述の通り「主題語」のみをクエリとした検索結果ページ中にはさまざまな話題に関するページが混在しており、単純に語の出現頻度が高い語を抽出すると一般的な語ばかりが抽出されてしまう。それを防ぐため、検索結果ページのクラスタリングを行うことにより、話題ごとにページを分類し、各クラスタから特徴的な語を抽出することも考えられるが、適切にクラスタリングを行うためには大量の検索結果ページが必要となり、解析に時間がかかってしまう。また、検索結果中から抽出可能な話題語候補は、ページの作成によって書かれた文書中においてよく使われるキーワードであり、実際にユーザが求めている情報とはかけ離れている可能性がある。例えば、商品名によって検索した場合には商品を売っているサイトのページばかりが検索され、ユーザが求めているレビューや感想といった情報が見つからない場合などが考えられる。

その他の手法として、クエリログなどの一般ユーザによってよく検索されているクエリを解析し、その中で主題語を含むものをクエリの候補として採用することが考えられる。本研究では、有用な検索結果ページを含む検索結果を取得可能なクエリを見つけるため、クエリログとしてYahoo!の関連検索ワード[13]を利用した。Yahoo!関連検索ワードとは、Yahoo!においてよく検索されているクエリのうち、指定したキーワードを含むものをリストとして取得できるサービスである。関連検索ワードとして取得したキーワードを利用することにより、主題語に関連する各話題を検索することができるクエリを取得することができる。また、関連検索ワードのキーワードは実際に検索を行うためにユーザが作成したクエリを元に作成されているため、ユーザのニーズを反映しており、人気の高い情報を探すことのできるクエリが多く存在すると考えることができる。

主題語の関連検索ワードとして得られるクエリには2種類存在する。1種類目は主題語を含む1語のキーワードのみから構成されるクエリ、そして2種類目は主題語ともう1つの語という2語のキーワードから構成されるクエリである。例えば、主題語が「京都」である場合、前者の例として「京都大学」や「京都駅」、後者の例として「京都観光」や「京都地下鉄」

などといったクエリが得られる。

5.2 クエリの決定

システムは、関連検索ワードによって得られた各クエリ候補に対してスコアリングを行い、最もスコアの高いものをクエリとして採用する。各クエリ候補のスコア S_{Q_i} は、以下の式によって求められる。

$$S_{Q_i} = (100 - r_i) + n_i \quad (2)$$

ただし、 i はクエリ候補に振られた通し番号、 r_i は関連検索ワードリスト中におけるクエリの利用頻度の順位、 n_i はユーザのウェブページ閲覧履歴中においてページタイトルに主題語ではない方のキーワードを含むページ数である。ユーザが興味のある内容を幅広く提示するため、主題語のスコアリングの場合とは異なり履歴中に存在するキーワードに対してページを閲覧した時期に関係なくスコアリングを行っている。このスコアリングにより、他のユーザが頻繁に検索しており、かつユーザが興味のあるキーワードをクエリ候補として提示できると考えられる。

システムは最もスコアの高いクエリを用いて関連ページを検索する。ユーザが主題語に関係する別の話題を閲覧したいと思った場合には、プルダウンメニューから別のクエリを選択することにより、新しい話題に関連するページを閲覧することができる。

6. 自動的な関連情報の提示

6.1 検索結果からの特徴語の抽出

システムは指定されたクエリを元に関連ページの検索を行う。このとき、検索結果を単純に上位から順番に表示するだけでなく、ユーザによって提示する情報のスコアリングを可能とし、より興味のある情報を閲覧することを支援することを試みた。

主題語およびクエリが指定された後、システムはまず特徴語の抽出を行う。特徴語とは、クエリによる検索結果において頻出する単語であり、検索結果中の話題を特徴付けると考えられる語である。いま、生成されたクエリが1語から構成されている場合はクエリをKとおき、2語から構成される場合は主題語以外のキーワードをKとおく。このときシステムは「主題語AND K」(正例)と「主題語NOT K」(負例)という2つのクエリを作る。そして、正例と負例の両クエリに対し、それぞれ100件のページを検索する。正例のクエリにより2つの語を共に含むページのタイトルおよびスニペットを100ページ分、負例のクエリにより主題語は含むけれどももう1つの語を含まないページのタイトルとスニペットを100ページ分取得することができる。

このとき正例と負例における語の出現傾向を分析すると、正例の検索結果中には「主題語のK」についての話題に関係する語が多く含まれ、逆に負例の検索結果中には「主題語のK」についての話題に関係する語の出現回数が少なくなると考えられる。したがって、それぞれの検索結果中における語の出現回数を比較することにより、「主題語におけるK」に関する話題を特徴付ける語を発見できると考えられる。本研究ではこれらの語を特徴語と呼び、以下のようにして抽出した。

1. 「主題語 AND K」と「主題語 NOT K」の各クエリの検索結果を100件ずつ取得する(各結果ページ集合をP, Qとする)
2. PおよびQのタイトルおよびスニペットを形態素解析し、名詞を抽出する
3. 抽出したn個の語 T_i ($i = 1 \sim n$)について、P中のページのうちタイトルもしくはスニペット中に T_i を含むページ数 N_p を求める

4. 同様に T_i ($i = 1 \sim n$) と Q についても語を含むページ数を求め、 N_Q とおく
5. N_P, N_Q について、表 1 のような表を作りカイ 2 乗検定を行う
6. カイ 2 乗検定の値を T_i のスコアとする (ただし、 $N_P < N_Q$ の場合は符号を反転する)

表 1 P と Q における T_i を含む/含まないページ数

Table 1 Number of pages containing/not containing T_i

| | T_i を含む | T_i を含まない | 合計 |
|-----------|-----------|-------------|-----|
| 主題語 AND K | N_P | $100 - N_P$ | 100 |
| 主題語 NOT K | N_Q | $100 - N_Q$ | 100 |

カイ 2 乗検定では、「主題語 AND K」の検索結果における T_i を含む文書数と「主題語 NOT K」の検索結果における T_i を含む文書数が等しいと仮定して、カイ 2 乗値と呼ばれる値を求める。カイ 2 乗値 χ^2 は以下の式により求められる。

$$\chi^2 = \frac{200\{N_P(100 - N_Q) - (100 - N_P)N_Q\}^2}{N_P(100 - N_P)N_Q(100 - N_Q)} \quad (3)$$

カイ 2 乗値は 0 から無限大までの値をとり、各検索結果における T_i を含む文書数が大きく異なるほど大きな値をとる。したがって、カイ 2 乗値を T_i のスコアとして用いると、「主題語 NOT K」の検索結果にはあまり現れず、「主題語 AND K」の検索結果中には多く出現する語に対して高いスコアを与えることができる。つまり、カイ 2 乗値によるスコアが大きい語ほど検索結果中のページの話題を特徴づける語であると考えることができる。

6.2 関連ページの検索と提示

システムは主題語をクエリ拡張することで得られたクエリを実行し、関連ページを取得する。そして、検索したページのタイトルおよびスニペットを検索結果の上位から順にシステムウインドウに提示する。それと同時に、システムは検索結果中から抽出した特徴語をユーザに提示する。

特徴語をユーザに提示することにより、ユーザは検索結果中の話題を特徴付けている語を一覧的に閲覧でき、検索結果の内容を漠然と把握することができる。本研究では、特徴語を単に一覧表示するだけでなく、システムが提示する関連情報の傾向をユーザが動的に変えるためにも特徴語を利用することを試みた。

このとき、特徴語を単純に並べて表示するだけではユーザとのインタラクションによってどのキーワードを重視して検索結果ページのスコアリングを行うのかといった情報を伝えることができない。そこで、本研究ではキーワードマップと名付けた表示方法をとることにより、キーワードを一覧的に表示し、ユーザにとってのキーワードの重要度に応じて対話的な操作を行うことを可能にした。

キーワードマップとは、2次元平面上に特徴語をノードとして描画する表示方式である。キーワードマップ上に表示されたノードはユーザのマウス操作によって移動することができ、閲覧するコンテンツの話題傾向を変えることができる。キーワードマップの例を図 4 に示す。

システムは特徴語のうち、カイ 2 乗値から計算されたスコアの高いものから 10 語を選択し、キーワードマップ上に表示する。ユーザはキーワードマップ上のノードの中で、重視したい語を中心に近づけ、必要の無い語を中心に遠ざけることにより、語の重み付けを変更することができる。



図 4：キーワードマップの例 (初期状態)

Fig. 4 Example of keyword map (initial condition)



図 5：キーワードマップの例 (重み付け後)

Fig. 5 Example of keyword map (after weighting)

システムは変更後の特徴語の重みを元に検索結果の再ランキングを行い、再ランキング後の順位が高いものから改めて関連ページを提示する。特徴語の重み付けを行った後のキーワードマップの例を図 5 に示す。

各特徴語ノードは固有の重みを持ち、ノードの重みは円の半径を r 、円の中心からノードまでの距離を d として $r - d$ という式により計算される。ノードは初期状態として円の中心から均等な距離 r に配置され、スコア 0 をもつ。

ユーザがノードを移動すると、特徴語の重み付けが更新される。次に、システムは特徴語の重み付けを利用し、検索された各結果ページの優先順位を計算する。以下の式によって各検索結果ページのスコア S_{P_i} が計算され、スコアの高いものほど優先順位が高くなる。

$$S_{P_i} = \sum_{k=1}^{10} (T_k) + \sum_{k=1}^{10} (S_k) \quad (4)$$

ただし、 i は検索結果ページに振られた通し番号、 T_k は検索結果のタイトルが k 個目の特徴語を含むならば 1、含まないならば 0 とする変数、 S_k は検索結果のスニペットが k 個目の特徴語を含むならば 1、含まないならば 0 とする変数である。つまり、タイトルおよびスニペットに重みの大きい特徴語を多く含むページほど高いスコアが与えられ、優先的にユーザに提示される。

7. まとめと今後の課題

本研究では、ユーザが閲覧中のウェブページの内容に連動して自動的に関連情報を検索し、提示するシステムを提案した。提案したシステムは閲覧ページ中から主題語を抽出し、主題語を含むクエリを Yahoo! の関連検索ワードから取得することによりクエリの拡張を行う。また、システムは主題語の話題に強く関連する特徴語を検索結果中から取得する。ユーザは特徴語の重み付けを変えることにより、閲覧するコン

テンツの優先順位を変えることができる。

本システムにより、興味のあるウェブページを連続的に閲覧することが可能となった。ユーザはマウスによる簡単な選択行為をしながらシステムに表示される情報を眺めるだけで、閲覧中のページの主題に関する情報を幅広く得ることができる。また、ユーザのウェブ閲覧履歴およびYahoo!の関連検索ワードという2種類の履歴を用いることにより、提示する関連情報の質を向上させることができた。具体的には、前者によってユーザの興味のある話題を含む情報を提示することができるようになり、後者によって話題性のある情報を提示することができるようになった。例えば、ユーザが「iPod」に関するページを閲覧している際、そのユーザが「youtube」のページをよく見ているならば、iPodの関連検索ワードとして「iPod AND youtube」というクエリを取得できるため、システムは「iPodでyoutubeの動画を見るための手法が書かれたページ」をユーザに推薦することが可能となった。キーワードを抽出する手法は他にも考えられるため、今後はその他の手法との精度の比較を行い、最適なキーワード抽出手法を模索していく予定である。

現在のシステムの問題点としては、表示される情報が検索結果のタイトルおよびスニペットのみであり、一つ一つの話題に対して十分な情報量を提示することができないという点が挙げられる。今後の課題として、各話題に関して十分な情報量をユーザに提供するため、複数のページから得られた情報を集約して提示したり、検索結果中のページを整理してブラウザに表示するという機能を実装していきたい。

また、ある話題に関する情報には最適な閲覧順序があると考えられる。例えば、システムがある話題に関する関連ページを提示する際、ユーザがその話題についてあまり詳しくない場合には、局所的な話題に特化したページを表示することは好ましくない。この場合は、幅広く概論的な内容を取り扱ったページを先に表示し、その後により詳細な話題について取り扱ったページへとユーザを誘導すべきである。また、このようなユーザに対し、専門用語を多用したページを表示する際には事前に関連用語について説明の書かれたページを表示し、事前知識を与えた上で専門用語を用いて説明を行っているページへと導くべきである。このように、ページの閲覧順序を考慮した関連情報の提示も今後の課題である。

本システムでは、キーワードマップに対するユーザの操作を許容することにより、ユーザが表示される話題の傾向を変化させることが可能となった。例えば、ある商品に関する関連情報を表示している際、オンラインショップのページを閲覧したいのか、もしくは感想やレビューを読みたいのかといった意図をシステムに伝えることができ、システムはその意図を反映して表示するページの傾向を変えることができる。しかし、現在のシステムでは自動的に抽出された特徴語に関してのみしか話題の傾向を変えることができない。感想やレビューに関するページを閲覧したいが、良い評価を与えている情報しか閲覧したくないなどといった、多様なユーザの要望にこたえるため、より詳細な意図伝播を行うことができるインタフェースを考案し、実装する予定である。

[謝辞]

本研究の一部は、文部科学省研究委託事業「知的資産の電子的な保存・活用を支援するソフトウェア技術基盤の構築」、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）、計画研究「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）、文部科学省科学研究費補助金若手研究(B)No.18700129、および、文部科学省グローバルCOE拠点形成プログラム「知識循環社会のための情報学教育研究拠点」（研究代表者：田中克己、平成19～23年度）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] U. Manber, M. Smith and B. Gopal: "WebGlimpse - Combining Browsing and Searching", Proc. of the Usenix Technical Conference (1997).
- [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] 大坪五郎: "Goromi-TV", WISS2006 論文集, pp. 47-52 (2006).
- [5] "PointCast
<http://www.pointcast.com/>".
- [6] 林正樹: "番組記述言語TVMLを使ったインタラクティブアプリケーション", 第57回情報処理学会全国大会論文集(3), p. 641 (1998).
- [7] "FeedTV
<http://feed-tv.com/>".
- [8] S. Nakamura, M. Minakuchi and K. Tanaka: "Energy Browser: To Make Exercise Enjoyable and Interesting", pp. 258-261 (2005).
- [9] S. Nakamura, M. Minakuchi and K. Tanaka: "AmbientBrowser: Web Browser in Everyday Life.", pp. 83-92 (2005).
- [10] 渡邊恵太: "Memorium: 眺めるインタフェースの提案とその試作", pp. 99-104 (2006).
- [11] G. Otsubo: "Goromi-Web: browsing for unexpected information on the web", pp. 267-268 (2007).
- [12] "MeCab: Yet Another Part-of-Speech and Morphological Analyzer
<http://mecab.sourceforge.net/>".
- [13] "Yahoo!関連検索ワード
<http://developer.yahoo.co.jp/search/webunit/V1/webunitSearch.html>.

吉田 大我 Taiga YOSHIDA

京都大学大学院情報学研究科社会情報学専攻修士課程在学中。主に情報検索の研究に従事。日本データベース学会学生会員。

中村 聡史 Satoshi NAKAMURA

京都大学大学院情報学研究科社会情報学専攻特任助教。2004年大阪大学大学院情報学研究科博士後期課程修了。博士(工学)。主にヒューマンコンピュータインタラクション、ウェブ検索の研究に従事。情報処理学会、日本データベース学会会員。

田中 克己 Katsumi TANAKA

京都大学大学院情報学研究科社会情報学専攻教授。1976年京都大学大学院修士課程修了。博士(工学)。主にデータベース、マルチメディアコンテンツ処理の研究に従事。IEEE Computer Society, ACM, 人工知能学会、日本ソフトウェア科学会、情報処理学会、日本データベース学会等各会員。