

## 質問キーワードの時間依存性を用いた検索支援方式

小野田 透<sup>†</sup> 角谷 和俊<sup>††</sup>

† 兵庫県立大学大学院環境人間学研究所 〒 670-0092 兵庫県姫路市新在家本町 1 丁目 1-12

E-mail: †nd07o007@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

**あらまし** 現在, Web に対するキーワードの検索結果である URL 集合を, 様々な観点から分類しユーザに提示することで検索を支援する技術が多く提供されている. しかし, 従来の分類手法では検索結果として提示される URL の「現在の内容」を解析するものがほとんどで, URL の過去の更新に注目して解析するものは少ない. また, 過去のページの中にはユーザにとって有用な情報を含むものが数多く存在すると考えられ, これらを利用することでユーザに対し, より有用な検索結果を提示できると考えられる. 本研究では, ユーザによって入力された複数の質問キーワードが, 検索結果の URL においてどのような時間的關係で出現しているかを解析し, キーワード間の時間的關係に基づいて URL の分類および補完を行う手法を提案する. 質問キーワード間の時間的關係の解析は, Web アーカイブに蓄積されている過去の Web ページ情報を用いることで行う. 解析の結果, 判定された質問キーワード間の時間關係によって URL の分類を行う. さらに, 本研究では提示する Web ページに対し, その過去のページを抽出して現在のページと共に提示することで情報の補完を行う. 本稿では, 提案するこれらの Web 検索支援手法について述べる.

**キーワード** Web アーカイブ, 情報検索, クラスタリング

## A Support Method Based on the Temporal Relation of Query Keywords for Web Search

Toru ONODA<sup>†</sup> and Kazutoshi SUMIYA<sup>††</sup>

†

1-1-12 Shinzaike-honcho, Himeji, Hyogo 670-0092, Japan

E-mail: †nd07o007@stshse.u-hyogo.ac.jp, ††sumiya@shse.u-hyogo.ac.jp

**Abstract** Classification and clustering search engines analyzing of Web page's content and returning search results by topic have been provided. However, most conventionally classification and clustering methods analyzed "current contents" of URL shown as search results. And there are few what kind of things which analyzed update of the past. We analyze what kind of relations the plural query keywords which a user input appears with in this study in terms of time in a Web page shown as search results. We suggest method classifying Web pages in based on time relations between query keywords. We perform analysis of time relations between query keywords by using past Web page information accumulated to a Web archive. Furthermore, We supplement information in this study for a Web page to show by We extract a page of the past, and showing it with a current page. In this paper, We explain these Web search support methods.

**Key words** Web Archive, Information Retrieval, Clustering

### 1. はじめに

Web ページ検索手法として, ユーザが自分の知りたいこと, あるいはそれに関するキーワード入力によって検索を行うキーワード検索が一般的なものとなっている. キーワード入力による検索では, 入力キーワードを含む Web ページの URL がランキングされて提示され, ユーザは提示された Web ページを閲

覧することで情報を取得することができる. しかしながら, 検索の結果として提示される URL は多くの場合において膨大な数となる. また, 各 URL の内容も多岐にわたっており, ユーザは多くの URL の中から自分の検索意図と合致する内容のものを探し出す必要がある.

これらの問題を解決するアプリケーションとして, 検索結果をユーザにとって利用しやすく分類して提示するための研究

が数多く行われており、検索エンジンとしても提供されている [1] [2]。Web の検索結果をある観点のもとに分類して提示することは、ユーザの検索を支援する手法として有用であると考えられる。

我々は、検索結果の分類において、提示される URL の現在のページだけでなく、過去のページを含めた解析を行うことで、各 URL において質問キーワードが時間的にどのような関係で出現しているかという観点に基づいて検索結果を分類することができると考えた [3] [4] (注1)。例えばユーザが { お中元, お歳暮 } というキーワードで検索を行うとき、検索結果にはお中元やお歳暮用商品の情報を掲載している URL が多く提示される。提示された検索結果の URL の中には、時期に合わせてお中元やお歳暮用商品の情報を掲載している URL もあれば、時期に関係なく常にお中元, お歳暮用商品の情報を掲載している URL もある。これらの URL は従来のトピックに着目する手法では、分類することは難しい。しかし、それぞれの URL について「お中元」、「お歳暮」というキーワードの時間的な関係を考えてとき、前者の URL であれば特定の時期にだけキーワードが出現するのに対し、後者の URL では常にキーワードが出現していると考えられる。このようなキーワードの時間的な関係を抽出することで、トピックに依存しない分類を行うことが出来る。また、我々はキーワードの時間的な関係を利用して過去のページから有用なページを抽出し、現在のページと共に提示する手法を提案している [5]。分類した検索結果に対し、過去のページによる補完を行うことで、より効果的にユーザの検索を支援することが出来ると考えられる。

本研究では複数の質問キーワードで Web 検索を行った際に、検索結果として提示された各 URL 中にキーワードがどのような時間的な関係をもって出現しているかを解析することで、検索結果を分類する。同時に解析によって得られたキーワードの時間的な関係を用いて、過去のページによる現在のページの補完を行うことでユーザの検索を支援する方式を提案する。

本稿では、2 章で本研究のアプローチについて述べ、3 章では時間依存性の定義と判定手法について述べる。4 章でキーワードの時間的な関係に基づいた分類と補完について述べ、5 章でプロトタイプシステムの構成について述べる。最後に 6 章でまとめと今後の課題について述べる。

## 2. 本研究のアプローチ

### 2.1 本研究の概要

本方式では以下の二つのことを行う。

- 質問キーワードの検索結果として提示された現在のページの分類
- 現在のページを補完する過去のページの提示

まず本方式では、Web アーカイブに蓄積されている過去の Web ページを用いることで、ある一つの URL においてユーザの入力した質問キーワードがどのような時間的な関係を持って出現し

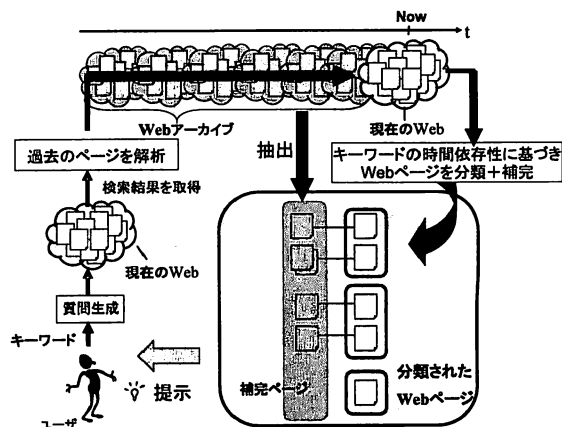


図1 概念図

ているかを解析し、分類を行う。キーワード間の時間的な関係として、「遍在」、「共起」、「前後」、「反復」、「非依存」の5つを定義し、これらを時間依存性と呼ぶ。検索結果として提示されている URL において、質問キーワード間の時間依存性の判定を行うことで、検索結果の分類を行うことが出来る。

分類方式では、検索結果を分類するために過去のページの解析を行うが、ユーザに提示されるのは現在のページのみである。しかし、提示されるのが現在のページのみでは、情報がユーザに十分に提示できない場合がある。よって、同一 URL の Web ページにおいて、過去のページの中から質問キーワードに関する情報を持つページを抽出し、現在のページと共にユーザに提示することで情報の補完を行う。過去のページの抽出はキーワードの時間依存性に基づいて行う。また補完により、これまでの検索では検索結果に現れなかったページを取得することが可能になる。以下に本研究の概要を述べる。

本方式の概念図を図1に示す。まず、ユーザはキーワード検索エンジンに質問キーワードを入力し、現在の Web に対して検索を行う。このとき、ユーザが複数 (2 語以上) のキーワードを入力することを前提とする。ここで、質問キーワードが現在のページと過去のページに分散して出現している URL も検索の対象とするため、検索質問の拡張を行う。ユーザの入力したキーワードが  $a, b$  であった場合、従来のキーワード検索では  $\{a \wedge b\}$  という質問を生成して検索を行うが、本手法では  $\{a \wedge b\}$  に加え  $\{a\}, \{b\}$  でも検索を行う。次に、質問キーワードの検索結果として提示された各 URL 毎に、Web アーカイブに蓄積されている過去のページを調べる。結果、過去のページにおいて質問キーワードが全て出現している URL を抽出し、質問キーワード間の時間的な関係の判定を行う。質問キーワードは過去の一時点で全ての語が出現している必要は無く、蓄積されている過去のページ全ての中で質問キーワードが出現していれば判定を行う。判定された時間依存性によって検索結果の URL を分類し、また、過去のページから検索結果として提示されている現在のページを補完するページを抽出して共に提示する。

(注1) : 過去の Web ページは Web アーカイブを利用して取得することを想定している

## 2.2 関連研究

検索結果として出力された Web ページを解析することで分類を行う手法として、成田ら [6] や大田ら [7] の研究があげられる。これらの研究はどちらも検索結果として提示される URL の現在の内容に対して、そのテキストを解析することで分類を行う。また、Web ページのリンク構造を用いて分類を行う方式として大野ら [8] や Wang ら [9] の研究があげられる。本研究では過去のページを解析し、キーワードの時間的な関係を利用するという点でこれらの研究とはアプローチが異なる。

過去の Web を利用して Web 検索を行う研究として、Adam らは Web アーカイブを用いて検索結果として提示される URL の過去のページを解析し、その更新履歴によって URL の再ランキングを行う方式を提案している [10] [11]。過去の Web 情報を用いて検索結果を変化させるという点で本研究と類似しているが、Adam らの研究では過去の Web ページを用いて、同一 URL の異なる時点における内容の差分を抽出して利用するのに対し、本研究では質問キーワードの時間的な関係を抽出するために過去の Web ページを利用するという点で異なる。

Web ページに対して情報の補完を行う研究として、検索結果として提示された複数の URL を組み合わせてユーザに対して提示する方式が湯本らによって提案されている [12] [13]。これらは複数の Web ページを用いて情報の補完を行うという点で本研究と類似しているが、湯本らの研究では異なる URL を組み合わせて提示するのに対し、本研究では同一 URL の異なる時点のページを組み合わせて提示するという点で異なる。

## 3. 質問キーワードの時間依存性

### 3.1 時間依存性の定義

(1) **遍在** ある URL の Web ページの時系列において、キーワード  $a, b$  がほぼ全ての時区間に出現する場合は「遍在」とする。全ての時区間とは、ある URL に関してそのページ作成時点を開始とし、現在のページが存在する時点、すなわちユーザが検索を行った時点を終端とする区間である。一度ページ中に掲載した内容を消去せずに蓄積していく形で更新を行う URL などに出現するキーワードが、この関係に当てはまる。

(2) **共起** ある URL の Web ページの時系列において、キーワード  $a, b$  が同時期に共に出現する関係を「共起」とする。一般的な「共起」は同一文書の中にキーワードが共に出現することのみを指すが、本稿では  $a, b$  が出現しているのが同一時点の文書でなくても、時間的に近い距離で出現していれば共起とみなす。例えば、ある URL の 2006 年 4 月 10 日時点のページにキーワード  $a$  が出現し、2006 年 4 月 12 日時点の内容にキーワード  $b$  が出現するとき、 $a$  と  $b$  は同一文書内には出現していないが、時間的には近い距離で出現しているといえる。実際の例として、ある URL において桜、花見というキーワードが春先のお花見の時期にのみ出現する場合、「桜」と「花見」は共に同じ時期に依存して出現していると言える、この関係が当てはまる。

(3) **前後** ある URL の Web ページの時系列において、キーワード  $a, b$  の一方が時間的に先に出現し、もう一方が時間的に

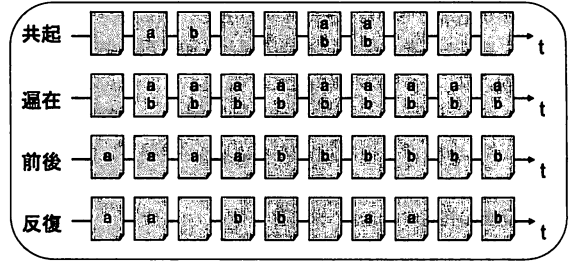


図 2 時間依存性

後に出現する場合、すなわちキーワードの時間的な出現順序が一意に定まる場合を「前後」とする。例えば、小泉、安倍のようなキーワードの場合、小泉の首相在任中はキーワード「小泉」が出現し、安倍に代わった後は「安倍」というキーワードが出現すると考えられる。

(4) **反復** ある URL の Web ページの時系列において、キーワード  $a, b$  がそれぞれ異なる時期に交互に出現し、それがある程度の周期性を持って繰り返される場合を「反復」とする。例えば、お中元、お歳暮というキーワードは、百貨店やオンラインショップのサイトにおいては実際にお中元を贈る時期、お歳暮を贈る時期にあわせてキーワードが交互に出現する。

(5) **非依存** ある URL の Web ページの時系列において、キーワード  $a, b$  が時間に依存せず出現している場合を「非依存」とする。上記の 4 つの時間依存性のいずれにもあてはまらない関係である。

共起、遍在、前後、反復の概念図を図 2 に示す。図中の長方形は Web ページを表し、中の文字はページ中に存在するキーワードである。矢印で連結されたページは同じ URL の異なる時点のページであり、右側のページほど新しく、左のページほど古いものとなっている。

### 3.2 時間依存性の判定

まず、ユーザによって入力された質問キーワードから 2 語を選択して、キーワードの組を作る。仮に質問キーワードとして  $a, b, c$  が入力された場合、 $\{a, b\}, \{b, c\}, \{c, a\}$  の 3 つの組が生成される。時間依存性の判定は生成された全てのキーワードの組について行われる。

#### ● 遍在の判定

判定を行う URL の Web ページの時系列において、全区間に占めるキーワード  $a, b$  の共起区間の割合が閾値以上であれば「遍在」と判定する。共起区間とは、同一 URL の Web ページの時系列において、質問キーワード  $a$  が出現しているページと  $b$  が出現しているページの時間的距離が閾値以下であるとき発生する時区間である。

#### ● 共起の判定

判定を行う URL の Web ページの時系列において、質問キーワードの共起区間と非共起区間を算出し、それらが交互に出現していれば「共起」と判定する。

#### ● 前後の判定

判定を行う URL の Web ページの時系列において、質問キー

ワード  $a$  が出現しているページを始端とし、質問キーワード  $b$  が出現するページを終端とする時区間と、逆に質問キーワード  $b$  が出現しているページを始端とし、質問キーワード  $a$  が出現しているページを終端とする時区間を抽出する。それぞれの時区間の総和の偏りを求め、偏りが閾値以上であれば「前後」と判定する。

● 反復の判定

判定を行う URL の Web ページの時系列において、質問キーワード  $a$  が出現している時区間と  $b$  が出現している時区間が交互に出現している場合、「反復」と判定される。質問キーワード  $a$  が出現しているページを始端とし、質問キーワード  $b$  が出現するページを終端とする時区間と、逆に質問キーワード  $b$  が出現しているページを始端とし、質問キーワード  $a$  が出現しているページを終端とする時区間が周期的に出現している場合、「反復」と判定する。ただし、ある程度一定の時間間隔で質問キーワードが出現していなければ「反復」とは判定されない。ここで「反復」と判定されなかった URL は、キーワードに時間依存性が無いと判断し、「非依存」と判定される。

4. 時間依存性を用いた検索支援

本節では、質問キーワードの時間依存性に基づき検索結果の分類を行う手法について述べる。さらに、分類された検索結果中の各 URL について、ユーザに対して提示される「現在の Web ページ」に不足している情報を同一 URL の過去のページから抽出する手法について述べる。

4.1 時間依存性による Web ページ分類

質問キーワードの組に対して判定された時間依存性を用いて、検索結果として提示される URL の分類を行う。URL の分類では、同一のキーワードの組について判定された時間依存性が同じである URL を同じクラスタとする。仮に質問キーワードが 2 語の場合、生成されるキーワードの組は 1 組となり、時間依存性は遍在、共起、前後、反復、非依存の 5 つであるので、検索結果は最大で 5 つのクラスタに分類される。

質問キーワードが 3 語  $\{a, b, c\}$  の場合、キーワードの組は  $\{a, b\}, \{b, c\}, \{c, a\}$  の 3 つが生成されるので、1 URL につき 3 つのキーワードの組の時間依存性が判定されることになる。この場合、異なる URL において  $\{a, b\}, \{b, c\}, \{c, a\}$  の時間依存性が URL 間で全て一致する場合、この URL は同じクラスタとして分類される。3 通りのキーワードの組に対して、判定される時間依存性がそれぞれ 5 つであるので、最大で 125 のクラスタが生成される。

例として、 $\{\text{お中元, お歳暮, マナー}\}$  というキーワードで検索を行った場合を考える。お中元とお歳暮に関する情報を同時に掲載している URL であれば、 $\{\text{お中元, お歳暮}\}$  の時間依存性は共起または遍在と判定されると考えられる。その上で、マナーについての情報がお中元、お歳暮と同時期に掲載されていれば、「お中元」、「お歳暮」と「マナー」の間の時間依存性も共起または遍在となる。しかし、マナーについての情報がお中元、お歳暮とは関係の無い時期に掲載されていた場合、「お中元」、「お歳暮」と「マナー」の間の時間依存性は非依存となる可能

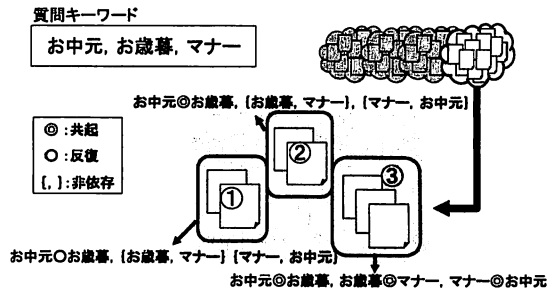


図 3 分類のイメージ

性が高い。このように、各キーワード間の時間依存性によって分類を行うことで、キーワード間の関係をより反映した細かな分類が可能となると考えられる。

4.2 時間依存性によるページ補完

質問キーワードの時間依存性に基づき、検索結果中の URL の現在のページに対して同一 URL の過去のページからページを抽出し共に提示することで補完を行う。本稿では、現在のページの補完を目的として抽出された過去のページを「補完ページ」と呼ぶ。

補完を行うことで、よりユーザの質問意図に適合する結果を提示することが可能になると考えられる。補完ページの抽出は、キーワードの組が持つ時間依存性に基づいてそれぞれ独立して行う。よって、仮に質問キーワードとして  $\{a, b, c\}$  が入力され、生成された 3 つのキーワードの組  $\{a, b\}, \{b, c\}, \{c, a\}$  に時間依存性が判定された場合、補完ページは  $\{a, b\}, \{b, c\}, \{c, a\}$  のそれぞれの時間依存性に基づいて別々に抽出される。

● 共起

質問キーワード  $a, b$  の組が共起と判定されたとき、その URL の Web ページの時系列において、キーワード  $a, b$  が過去において同時期に出現していたと言える。補完ページとして、過去のページから同時期に出現している  $a, b$  の出現ページを抽出することで、ユーザは  $a, b$  についての情報と  $a, b$  の関連を知ることが出来る可能性がある。よって、時間的に距離が近い  $a$  の出現ページと  $b$  の出現ページほどキーワード間の関連が強いと仮定し、補完ページとして抽出する。  $a$  と  $b$  が共に出現しているページの場合、ページ間の時間的な距離は 0 とする。

● 遍在

質問キーワード  $a, b$  の組が遍在と判定されたとき、その URL の Web ページの時系列において、キーワード  $a, b$  がほとんどのページに出現していると言える。そのような URL は、ページ内容の更新において過去の内容を削除するのではなく蓄積していく形で更新を行うページであるか、ニュース記事の URL のように作成時から更新が行われていないページであると予想することが出来る。そのような URL の場合、過去のページに含まれる情報は現在のページにも含まれている可能性が高い。よって、本手法では遍在と判定されたキーワードの組に関しては、過去のページによる補完は不必要であるとし、補完を行わない。

● 前後

質問キーワード  $a, b$  の組が前後と判定されたとき、キーワード  $a, b$  は因果関係を持つと仮定する。キーワード  $a, b$  が前後関係であり、 $a$  が  $b$  よりも時間的に前に出現している場合、まず  $a, b$  の出現区間をそれぞれ抽出する。このとき、例外的なパターンとしてキーワード  $a$  の出現区間とキーワード  $b$  の出現区間に重複が生じる場合、重複区間に存在するページは抽出対象から除く。その上でキーワード  $a$  の出現区間中で  $a$  が最後に出現しているページと、キーワード  $b$  の出現区間中で  $b$  が最初に出現しているページの抽出を行い、補完ページとしてユーザーに提示する。

● 反復

質問キーワード  $a, b$  の組が反復と判定されたとき、その URL の Web ページの時系列において、キーワード  $a$  が出現しているページとキーワード  $b$  が出現しているページが交互に出現している。まず、URL の現在のページについてキーワード  $a, b$  が出現しているかを調べる。仮に、現在のページにおいてキーワード  $b$  のみが出現している場合、補完ページとしてキーワード  $a$  に関する情報を持つページが必要であると、抽出対象区間としてキーワード  $a$  の出現区間を抽出する。抽出した区間の中から、補完ページを抽出するのに適した区間を選択するため、スコアリングを行う。区間のスコアは次の式で求められる。

$$score(i_n) = 1 - \frac{\theta - d(i_n)}{\theta} \quad (1)$$

$i_n$  は抽出区間を表し、 $d(i_n)$  は各区間の長さ、 $\theta$  は全ての抽出区間の長さの平均値を表す。長さが極端に長い、または短い区間はノイズが含まれている可能性が高いため、全ての区間の平均の長さ  $\theta$  を算出し、その長さに近い区間ほどスコアを高くしている。さらに、各区間中のページに対してスコアリングを行う。各ページのスコアは次の式で求められる。

$$score(P_m) = \frac{t(P_m) - t(P_{start})}{t(P_{end}) - t(P_{start})} \quad (2)$$

$P_m$  は区間内のページ、 $t(P_{start}), t(P_{end})$  はそれぞれ区間の始端、終端のページの存在した時点の時間情報である。この式では、同一区間内であれば新しいページほどスコアが高くなる。最終的に区間のスコアとページのスコアを積算し、スコアの高いページから順に抽出する。

**非依存** 質問キーワード  $a, b$  の組が遍在と判定されたとき、そのキーワードの組には時間的な関係性がないと判断し、補完は行わない。

## 5. プロトタイプシステム

プロトタイプシステムは未実装であるが、想定しているシステム構成について述べる。システムは5つのユニットからなる。システム構成図を図4に示す。各ユニットの概要を以下に述べる。

(1) **URL 収集部** ユーザーの入力した複数の質問キーワードによって Web ページの検索を行い、検索結果として提示された URL を収集する。検索の際、ユーザーの入力したキーワードから検索質問を生成する。仮に質問キーワードが  $\{a, b\}$  のとき、

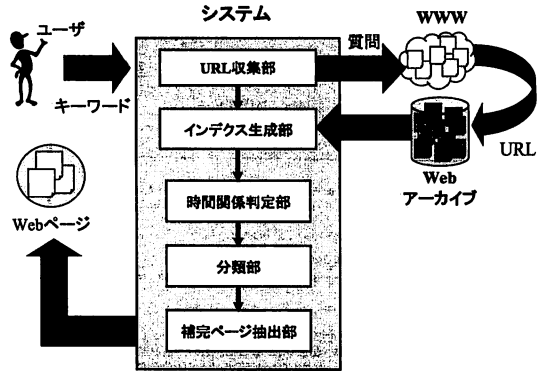


図4 システム構成図

検索質問として  $a, b, a \wedge b$  が生成される。収集された URL をキーとして、Web アーカイブから過去のページを自動的に取得する。過去のページの取得はアメリカの InternetArchive<sup>(注2)</sup> を用いて行う。InternetArchive ではページの URL を入力することで、過去のページを取得することが出来る。このとき、InternetArchive 内の情報から、ページが InternetArchive に収集された年月日をページの時間情報として取得する。

(2) **インデックス生成部** 取得した全ての URL の過去のページから以下を抽出しインデックスを作成する。

- ページの URL
- ページがアーカイブに収集された時間
- ページ内に含まれるテキスト

(3) **時間依存性判定部** 質問キーワードから2語を選択してキーワードの組を生成し、各 URL における時間依存性の判定を行う。判定は生成したキーワードの組すべてについて行う。

(4) **URL 分類部** 判定された各 URL におけるキーワードの時間依存性に基づいて検索結果を分類する。分類は同一のキーワードの組について判定された時間依存性が同じである URL を同じクラスタとする。

(5) **補完ページ抽出部** 判定された各 URL におけるキーワードの時間依存性と、現在のページの質問キーワード出現状況に基づいて同一 URL の過去のページから補完ページを抽出する。補完ページの抽出は、キーワードの組が持つ時間依存性に基づいてそれぞれ行う。

検索結果の出力画面イメージを図5、図6に示す。ユーザーは図5でキーワードを入力し、検索を行う。各キーワードの組ごとに判定された時間依存性によって URL が分類され、分類された URL 数の多いクラスタの順に表示される。ユーザーがクラスタを選択することで、そのクラスタに属する URL が提示される。また、ユーザーが提示された URL を選択した場合、図6のようにその URL の現在の内容と補完ページが提示される。

## 6. おわりに

本稿では、質問キーワード間の時間的な関係に基づく検索結

(注2) : <http://www.archive.org/>

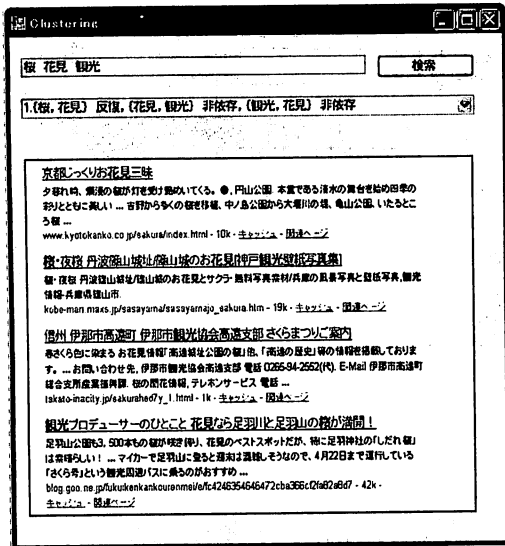


図 5 分類結果表示画面

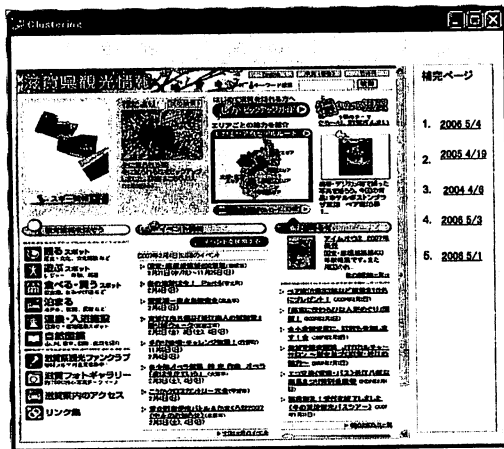


図 6 Web ページ表示画面

果の分類手法, および補完ページの抽出手法を提案した. 検索結果の分類において各 URL の過去の内容を取得し, 質問キーワードの時間的な関係の抽出を行うことで, より各 URL 中の質問キーワードの性質に基づいた分類が可能になると考えられる. また, 過去のページを抽出し, 現在のページと共に提示することで情報の補完を行う手法を同時に提案した. 今後は実際の検索結果の分類実験を行い, キーワードの時間的な関係と実際のページの内容がどのように関連しているのかを検証する.

#### 文 献

[1] Clusty.com: <http://clusty.com/>.  
 [2] Clusty.jp: <http://clusty.jp/>.  
 [3] Kage, T. and Sumiya, K.: A Web Search Method Based on the Temporal Relation of Query Keywords, *Proceedings of The 7th International Conference on Web Information Systems Engineering (WISE 2006)* (2006).  
 [4] 賀家智代, 角谷和俊: Web 検索のための質問キーワードの時間依存性に基づくクラスタリング手法, *情報処理学会研究報告*, Vol. 2006, No. 78, pp. 519-526 (2006).

[5] 小野田透, 賀家智代, 角谷和俊: Web アーカイブを用いた時系列パターンに基づく検索支援方式, *電子情報通信学会 第 18 回データ工学ワークショップ (DEWS2007) 論文集*, 電子情報通信学会 (2007).  
 [6] 成田宏和, 大田学, 片山薫, 石川博: Web 文書検索のための非排他的クラスタリング手法の提案~NOCTURNE(New Overlapping Clustering Tool Using Ranking information of search Engine), 第 14 回データ工学ワークショップ DEWS '03, 電子情報通信学会 (2003).  
 [7] Ohta, M., Narita, H. and Ohno, S.: Overlapping Clustering Method Using Local and Global Importance of Feature Terms at NTCIR-4 Web Task, *Proceedings of the Fifth International Conference on Web Information Systems Engineering (WISE2005)*, pp. 43-52 (2005).  
 [8] 大野成義, 渡辺匡, 片山薫, 石川博, 大田学: Max Flow アルゴリズムによる Web ページのクラスタリング方法, *情報処理学会研究報告*, 2005-DBS-137(1), Jul. 2005, 電子情報通信学会, 情報処理学会 (2005).  
 [9] Wang, Y. and Kitsuregawa, M.: Use Link-based Clustering to Improve Search Results, *Proceedings of the 2nd International Conference on Web Information System Engineering, IEEE Computer Society* (2001).  
 [10] Adam, J., Kawai, Y. and Tanaka, K.: Temporal Ranking of Search Engine Results, *Proceedings of the The Fifth International Conference on Web Information Systems Engineering (WISE2005)*, pp. 43 - 52 (2005).  
 [11] Adam, J., Kawai, Y. and Tanaka, K.: Using Web Archive for Improving Search Engine Results, *Proceedings of The Eighth Asia Pacific Web Conference (APWeb2006)*, pp. 893 - 898 (2006).  
 [12] Yumoto, T. and Tanaka, K.: Finding Pertinent Page-Pairs from Web Search Results, *Proceedings of The 8th International Conference on Asian Digital Libraries (ICADL2005)*, pp. 301 - 310 (2005).  
 [13] Yumoto, T. and Tanaka, K.: Page Sets as Web Search Answers, *Proceedings of The 9th International Conference on Asian Digital Libraries (ICADL2006)*, pp. 244 - 253 (2006).