

検索質問の緩和と情報統合に基づく クロスメディア・メタサーチ機構

桑原 昭裕[†] 田中 浩也[†]
角谷 和俊[†] 田中 克己[†]

WWW によって取得可能なコンテンツはその量や種類が年々増大しており、ある事象に関する情報を取得したい場合に、これらの情報が大量の Web サイトに分散し、かつ、種々のメディアによって表現されているため、これらの情報を効果的に検索し統合する機能が重要である。そこで、クロスメディア・メタサーチを実現するために検索質問の緩和という方法を提案した。本論文では、検索質問の緩和により情報を収集し、それを統合するためのシステムを提案し、検索質問の緩和に対する有効性の検証のために、実データを用いた実験を行い、質問の緩和についての考察を行う。

Cross Media Meta-Search System Based on Relaxation of Retrieval Query and Answer Integration

AKIHIRO KUWABARA,[†] HIROYA TANAKA,[†]
KAZUTOSHI SUMIYA[†] and KATSUMI TANAKA[†]

Recently, contents of Web have been increasing and the kinds of the media have been varied. To retrieve information of a topic by search engines, users must search many sites in several aspects, because the contents are scattered in several sites and the contents have varied media types. It is important to integrate and arrange various contents distributed in different sites. We have proposed a integration and arrangement method based on query generation for cross media meta-search engines and an extraction method of contents. However, there are some cases which the extractions are too strict to retrieve the required contents. In this paper, we propose a gradual relaxation method of the retrieval queries and verify the level of validity. We show experiments of several relaxation examples and estimate them. We also discuss the result of the method and future work.

1. はじめに

インターネット環境がますます普及してきたため Web ページの数は増大する一方であり、またブロードバンドやデジタルカメラ等の普及により Web ページのコンテンツは画像等が多く取り入れられてますます多種多様になってきている。このように、Web 空間には様々な情報が氾濫しているため、ユーザが有益な情報だけを収集してくることは非常に困難になっている。

情報を効果的に検索し統合する手段として、メタサーチエンジンが挙げられる。従来のメタサーチエンジンでは、ユーザが検索キーワード群を入力し検索を実行すると、各サーチエンジンに入力されたキーワー

ド群を渡し、各々のサーチエンジンがキーワード検索し Web ページを収集してくる。そしてメタサーチエンジンは、各々のサーチエンジンが収集した Web ページの重複を除去したり、自動的に分類等の操作を行い、検索結果を出力する。ここで、メタサーチエンジンに共通する点として3つが挙げられる。

- (1) メタサーチで利用する各々のサーチエンジンは同一タイプ。
- (2) どのサーチエンジンに対しても同一の検索質問が実行される。
- (3) 統合した検索結果として Web ページへのリンクが示される。

このように、従来の検索システムではユーザにとって有益な情報を得ることは容易とはいえない。ユーザにとって重要なことは、一つのサーチエンジンで検索キーワードを入力しただけで、テキスト、画像、動画などの様々な情報が得られることである。またユーザ

[†] 京都大学大学院情報学研究所
School of Informatics, Kyoto University

にとって有益な情報だけを閲覧しやすい状態に表示することである。

本研究の目的は、様々な形の情報を、Web から収集してきて、その情報を従来のように一つの Web ページを独立に存在させるのではなく、様々な Web ページのコンテンツを統合することである。本研究では様々な形の情報を得るために、既存の様々なメディアに対するサーチエンジンを使うことで、それを解決する。また、その統合したコンテンツは様々な形の検索結果を組み合わせて、Web ページを使用した百科事典となるべく、ユーザの検索キーワードに関する情報が分かりやすく記述されているような Web ページのコンテンツを新たに生成することを目標とする。百科事典の構成要素としては、調べたい語句の説明、画像、用例、関連する話題などがあげられる。

本論文では、問題点の一つである、各サーチエンジンには同一の検索質問がただ And 検索として入力されていると言う点に注目し、サーチエンジンにそのエンジンに適した検索キーワードを与えるために、ユーザの検索質問を変換する方法を提案し、その方法の有効性を検証するものとする。この方法を用いることによって、今回は主に検索キーワードに対する、画像とテキスト文書の収集を目的としている。以降、第2章で関連研究、第3章でクロスメディア・メタサーチの概要、第4章で検索質問の緩和について、第5章で検索質問の緩和についての実験、第6章で検索結果の統合について、第7章でまとめと今後の課題について述べる。

2. 関連研究

2.1 NAVER

検索サイトの NAVER の検索サービス⁵⁾ は Web 上にある HTML ページを始め、動画、イメージ、サウンド、文書などを同時に検索し、検索語別にユーザー



図 1 検索サービス Naver

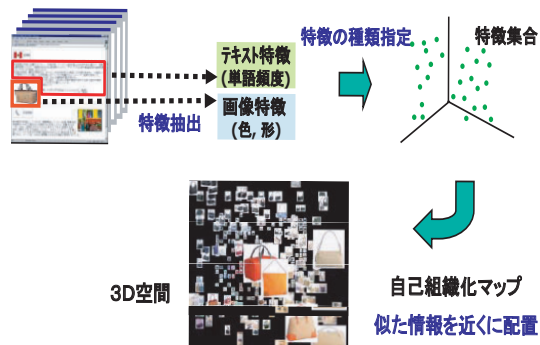


図 2 MIRADOR-Search

の検索意図を予想して検索結果を提供する。統合検索では、Web ページ、動画、イメージ、サウンドの検索を同時に行い、検索結果を一画面にまとめて表示するという、機能を持っている。検索質問は従来のサーチエンジンのような入力方式であり、また統合といっても、画像は画像の領域、テキスト検索の出力結果としての URL のリストを表示する領域は分割されている。様々なメディアを扱ってはいるが、この点で本研究とは異なっている。

2.2 MIRADOR-Search

MIRADOR-Search⁶⁾ とは、富士通によって開発された、テキストによる意味的検索と画像による視覚的検索とを統合した「情報を眺めて選ぶ」クロスメディア検索機能を提供するサービスである。これは、ユーザが検索キーワードを入力、その情報を基に Web Crawler がインターネット上の Web ページを解析し、画像と説明テキストを自動収集する。そして検索結果として、収集してきた画像を画像の色や形などの画像特徴に基づいて、3次元上にマッピングして表示するものである。しかしクロスメディア検索といってもその結果として出力されるのは画像のみであり、テキスト情報は画像の特徴量を出すために使用される。画像に特化した検索だといえる点で、様々なメディアを扱うことを提案している本研究とは異なっている。

2.3 質問の階層構造を用いた検索手法

小山ら⁷⁾によって提案された、検索エンジンに複数のキーワードを入力した場合における、検索質問の階層的構造化を用いた Web 検索手法である。同じキーワードからなる検索に対し、主題を表すキーワード、主題に関する内容を表すキーワードのようにキーワードごとに役割を区別することで検索精度を向上させるといったものである。また検索結果の表示方法として同じキーワードでも役割を変えて検索していった検索結果を比較して、大きな相違があった検索結果集合を

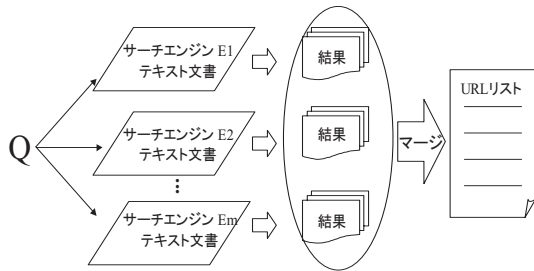


図3 従来のメタサーチ

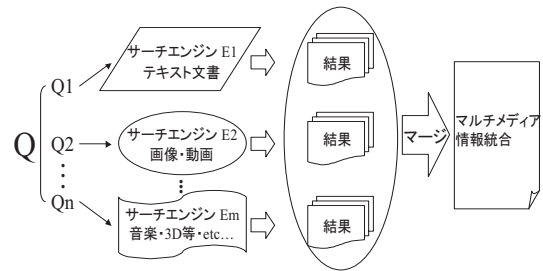


図4 クロスメディア・メタサーチ

並列に提示している．本研究ではこの手法を利用し，検索キーワードを，画像検索に使う質問，テキスト検索に使う質問のように変換させて利用していることで，検索キーワードに役割を持たせている．

3. クロスメディア・メタサーチの概要

複数のキーワード $K_1, K_2, \dots, K_n (n \geq 2)$ からなる conjunctive query Q が与えられたとする．すなわち， $Q = K_1 \wedge K_2 \wedge \dots \wedge K_n$ である．また，種々の利用可能な検索エンジンを， $E_1, E_2, \dots, E_m (m \geq 2)$ とする．従来のメタサーチでは，この複数のキーワードからなるクエリ Q をそのままいくつかの検索エンジンに利用している．またこの時，利用する検索エンジンはほとんどがテキスト検索の検索エンジンである．そしてその後検索結果として様々な検索エンジンの結果から重複などを除去して解の Web ページへのリンクを示している（図3）．

クロスメディア・メタサーチでは，利用可能な検索エンジンは異種のを許している，すなわち，例えば， E_1 は通常の Google⁸⁾， E_2 は AltaVista⁹⁾， E_3 は Google 画像検索エンジン¹⁰⁾， E_4 は音楽検索エンジンなどのように，タイプの異なる検索エンジンの混在を許している点が特徴的である．さらに，与えられた質問 Q ，および，タイプの異なる検索エンジン E_1, E_2, \dots, E_m に対して，質問 Q を変換して各検索エンジンに送り，その結果を統合しようというものである．本論文ではこの変換する方法として，検索質問の緩和という方法を提案する．また，検索結果として従来のような Web ページへのリンクではなく，解の Web ページから検索キーワードに関する情報だけ抽出してそれを統合して，検索結果としてユーザに提示するものである（図4）．

4. 検索質問の緩和

既存の画像検索では，テキスト検索よりもかなりヒット件数が少ない．しかしその分，画像検索でヒッ

トするページは，画像が使用されているだけあって，通常のテキスト検索よりも Web ページに書かれている内容は精度が高いと考えられる．また，テキスト検索の検索結果から検索キーワードに関連する画像を収集するのが困難であるのに対して，画像検索では，検索キーワードに対する画像が与えられるために，画像の収集が容易にできる．このようなことから画像検索をうまく利用すれば，百科事典的な要素である，検索キーワードに対する画像，テキスト文書が容易に収集できるのではないかと考える．

4.1 検索キーワードの役割変換

ユーザが複数のキーワード $K_1, K_2, \dots, K_n (n \geq 2)$ からなる conjunctive query Q を入力したとする．すなわち， $Q = K_1 \wedge K_2 \wedge \dots \wedge K_n$ である．また種々の利用可能な検索エンジンを， $E_1, E_2, \dots, E_m (m \geq 2)$ とする．質問 Q に対する解である Web ページ集合を， $Ans(Q)$ とする．また，質問 Q を検索エンジン $E_i (1 \leq i \leq m)$ に対して行って得られる解集合を $Ans(Q, E_i)$ と表すものとする．但し，解集合は，Web ページ，画像，音楽などのファイル集合である．

本論文では，画像検索を主として取り扱う方法を論じているために， E_1 として Google 画像検索エンジン， E_2 として画像周辺のテキスト検索を用いる．まず，質問 Q を検索エンジンの数に合わせて，部分集合に分割する．すなわち，ここでは検索エンジンは2つであるので

$$\begin{aligned} & \{K_1, \dots, K_n\} \\ & \{K_1\}, \{K_2, \dots, K_n\} \\ & \{K_2\}, \{K_1, K_3, \dots, K_n\} \\ & \vdots \\ & \{K_1, K_2\}, \{K_3, \dots, K_n\} \\ & \{K_1, K_3\}, \{K_2, K_4, \dots, K_n\} \\ & \vdots \\ & \{K_1, \dots, K_n\}, \end{aligned}$$

という部分集合に分解する．ここで，部分集合の前者

の要素には画像検索を、後者の要素にはテキスト検索にかけるという役割を持たせる。

このような役割を割り当てる理由は次のとおりである。テキスト検索では、検索キーワードが1つのページ内に書かれていればヒットするが、画像検索の場合は、ファイル名や、画像へのアンカーテキストに検索キーワードが含まれているものがヒットする。よって、テキスト検索よりも検索キーワードに対するヒット率が低く、検索の条件としては厳しいものになっている。そこで、従来では検索キーワードをすべて And 検索で画像検索にかけていたものを、部分集合に分けていくつかのキーワードをテキスト検索のキーワードとして使用するものである。これは、画像検索するよりもテキスト検索したほうがヒットしやすことをふまえて、検索質問を緩和していくものであると考えられる。第5章では有効性を検証するための実験について述べる。

4.2 Web ページの収集

部分集合の各要素 $\{K_1\}, \{K_2\}, \dots, \{K_1, K_2\}, \{K_1, K_3\}, \dots, \{K_1, \dots, K_n\}$ をそれぞれ E_1 である Google 画像検索にかける。これによって $Ans(K_1, E_1), Ans(K_2, E_1), \dots, Ans(K_1 \wedge K_2, E_1), \dots, Ans(K_1 \wedge \dots \wedge K_n, E_1)$ を得ることができる。これは各要素を Google 画像検索にかけた解集合である。解集合は、画像検索の画像とその画像の参照元の Web ページへの URL によって構成される。

次に、検索結果として出力された画像の参照元の Web ページを収集する。 $Ans(K_1, E_1)$ の Web ページに対しては、まだ使用していない部分集合の要素 $\{K_2, \dots, K_n\}$ が画像の参照元の Web ページにすべて含まれているかを調べる。すべて含まれている場合はこの Web ページを解として収集する。この操作をすべての部分集合に対して行う。ここで解として収集した Web ページは、 K_1 で画像検索をし、 K_2, \dots, K_n でテキスト検索をし、両方の検索結果として出力されたページだけを収集することと変わりはないはずであ

る。つまり、 $Ans(K_1, E_1) \cap Ans(K_2 \wedge \dots \wedge K_n, E_2)$ である。これをすべての部分集合に対し繰り返し行うことによって、 $Ans(Q)$ として

$$\begin{aligned}
 Ans(Q) = & (Ans(K_1, E_1) \\
 & \cap Ans(K_2 \wedge \dots \wedge K_n, E_2)) \\
 \cup & (Ans(K_2, E_1) \\
 & \cap Ans(K_1 \wedge K_3 \wedge \dots \wedge K_n, E_2)) \\
 \cup & \dots \\
 \cup & (Ans(K_1 \wedge K_2, E_1) \\
 & \cap Ans(K_3 \wedge \dots \wedge K_n, E_2)) \\
 \cup & (Ans(K_1 \wedge K_3, E_1) \\
 & \cap Ans(K_2 \wedge K_4 \wedge \dots \wedge K_n, E_2)) \\
 \cup & \dots \\
 \cup & (Ans(K_1 \wedge \dots \wedge K_n, E_1))
 \end{aligned}$$

を得る。

以下簡略化のため、画像検索に使ったキーワードが K_1 である解の集合を $P(K_1)$ と表す。つまり、 $Ans(K_1, E_1) \cap Ans(K_2 \wedge \dots \wedge K_n, E_2)$ を $P(K_1)$ として表し、 $Ans(K_1 \wedge K_2, E_1) \cap Ans(K_3 \wedge \dots \wedge K_n, E_2)$ を $P(K_1 \wedge K_2)$ と表す。

ここで例として、ユーザの入力した検索キーワードが「富士山」「雪」という二つであった場合の収集してくるページを考える。図5はその時のベン図である。図の塗りつぶした部分が収集してくる Web ページと対応している。従来の検索では、テキスト検索エンジンで「富士山」「雪」を検索してきた Web ページ、画像検索で「富士山」「雪」を検索した結果の Web ページだけが解である。本研究では、質問の役割を交換することによって、画像検索で「富士山」を検索した結果の Web ページから「雪」というテキストが記述されているページ、また画像検索で「雪」を検索した結果の Web ページから「富士山」というテキストが記述されているページも、従来の方法に加えて解とするものである。つまりは図6に示したような Web ページも収集対象のページとするものである。

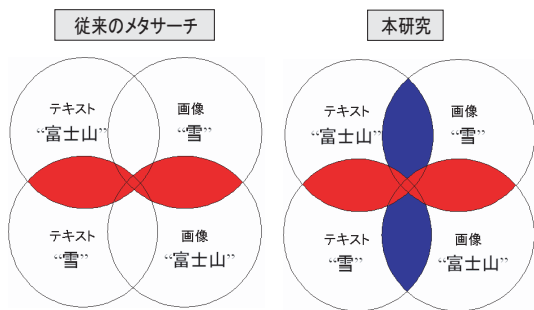


図5 質問の役割



図6 新たな収集対象のページ

検索キーワード 「京都」∧「紅葉」∧「高台寺」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	京都、紅葉、高台寺		6件	6件
1	京都、紅葉	高台寺	38件	13件
	京都、高台寺	紅葉	24件	23件
	紅葉、高台寺	京都	10件	8件
2	京都	紅葉、高台寺	7件	1件
	紅葉	京都、高台寺	6件	2件
	高台寺	京都、紅葉	54件	11件

(a)

検索キーワード 「サッカー」∧「中田」∧「日本代表」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	サッカー、中田、日本代表		19件	17件
1	サッカー、中田	日本代表	32件	22件
	サッカー、日本代表	中田	63件	15件
	日本代表、中田	サッカー	31件	13件
2	サッカー	中田、日本代表	48件	6件
	中田	サッカー、日本代表	40件	26件
	日本代表	サッカー、中田	13件	0件

(b)

検索キーワード 「国民」∧「年金」∧「制度」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	国民、年金、制度		9件	7件
1	国民、年金	制度	17件	10件
	国民、制度	年金	18件	5件
	年金、制度	国民	203件	126件
2	国民	年金、制度	22件	2件
	年金	国民、制度	144件	57件
	制度	国民、年金	24件	8件

(c)

検索キーワード 「台風」∧「大雨」∧「洪水」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	台風、大雨、洪水		14件	4件
1	台風、大雨	洪水	33件	20件
	台風、洪水	大雨	22件	14件
	洪水、大雨	台風	34件	14件
2	台風	大雨、洪水	18件	7件
	大雨	台風、洪水	43件	19件
	洪水	台風、大雨	22件	5件

(d)

検索キーワード 「富士山」∧「夕日」∧「雪」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	富士山、夕日、雪		0件	0件
1	富士山、夕日	雪	8件	6件
	富士山、雪	夕日	12件	12件
	夕日、雪	富士山	3件	3件
2	富士山	夕日、雪	4件	3件
	雪	富士山、夕日	1件	1件
	夕日	富士山、雪	0件	0件

(e)

検索キーワード 「SARS」∧「肺炎」∧「症状」				
緩和度	画像検索	テキスト検索	ヒット数	有効ページ
0	SARS、肺炎、症状		0件	0件
1	SARS、肺炎	症状	0件	0件
	SARS、症状	肺炎	0件	0件
	肺炎、症状	SARS	0件	0件
2	SARS	肺炎、症状	0件	0件
	肺炎	SARS、症状	1件	0件
	症状	SARS、肺炎	0件	0件

(f)

図 7 実験結果

5. 質問の緩和に対する評価

5.1 簡易実験の概要

実際に4章の方法を用いることによって、どのような結果が得られるかということを実証するために、様々なキーワードを入力した場合における、検索のヒット件数、有効ページ数を簡易実験により調べた。実験に使用するキーワードは主にあるニュースページの記事

のタイトルになっていたものをさまざまな分野から選んだものである。

まず、ユーザが入力する検索キーワードは3つとする。ここで質問の緩和度という尺度を定義する。緩和度とはいくつかのキーワードを画像検索からテキスト検索に緩和させたかを表すものである。つまりは検索キーワードの3つを And 検索で画像検索にかけた場合は緩和度0とし、検索キーワードの2つを And 検

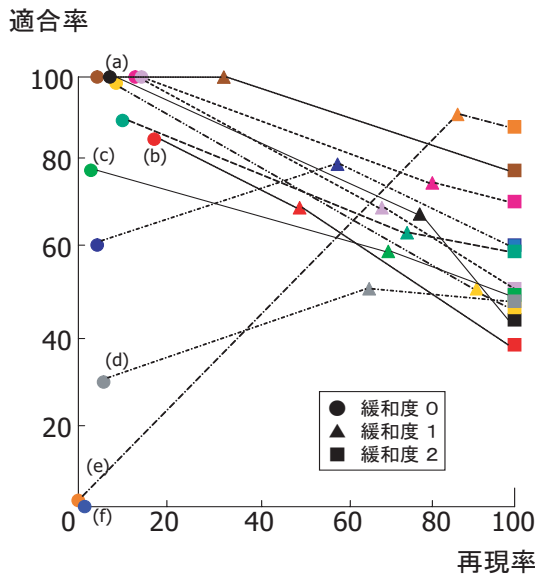


図 8 適合率 再現率グラフ

索で画像検索にかけて残りの 1 つのキーワードをテキスト検索に使用した場合は緩和度 1 とするものである。今回は、緩和度 0、緩和度 1、緩和度 2 の検索結果を比較する。ここで、緩和度 1 の結果は、緩和度 0 の結果も含むこととし、緩和度 2 の結果は緩和度 0、および緩和度 1 の結果を含む。

5.2 実験結果

図 7 に実験結果の一部を示す。ヒット件数は画像検索のキーワードを入れた時の画像検索結果の中からテキスト検索のキーワードが含まれたページの数を示している。有効ページとは、ヒットしたページの中で検索キーワードの内容を適切に記述しているページであると人が判断したものである。また実験の各検索キーワードごとの再現率と適合率についてのグラフを図 8 に示す。適合率、再現率とも緩和度ごとに表示している。ここで、適合率とは各緩和度ごとの結果に対する、平均の適合率である。また、再現率とは、Web 上での解集合全体は分からないため、各実験ごとの有効なページの総和を解集合全体とするものとする。よって緩和度 2 の時の再現率は 100% であり、それと相対的に見た評価であることに注意を要する。また図中のアルファベットは図 7 と対応している。

5.3 考察

実験に対する考察を行う。まず、結果の再現率に注目すると緩和度 0 から緩和度 1 にすることによって非常に再現率が上がることが分かる。これは検索キーワードをテキスト検索に利用することが、検索の条件を緩めて網羅的に収集していることの証明となってい

る。また緩和度 0 から緩和度 1 への再現率の変化に比べて、緩和度 1 から緩和度 2 への変化は小さかった。

またグラフより、適合率 - 再現率グラフには、おまかに二つのパターンの形が存在することが読み取れる。一つ目は (a) (b) (c) のパターンで、緩和度を増やしていくことによって、再現率が上がり適合率が下がっていく一般的な形である。二つ目には (d) (e) のパターンで、緩和度 0 の時よりも、緩和度 1 の時の方が適合率、再現率ともに高くなっている場合である。特に (e) の場合、従来のような緩和度 0 だけを検索しても結果は得られないが、緩和することによって検索結果を得ることができるのである。これは質問を緩和したことによる有効性が顕著に示されるパターンだと考える。特に緩和度 0 の結果が少ない時にこのパターンが現れることが多い。このことから緩和度 0 の結果が少なければ、検索質問を緩和したら有効であることが分かる。また画像検索には少ないワードでもヒット数が少ない、不得意な分野も存在することが分かる (f) の実験結果のようにどのキーワードも具体的な画像が表せないような語句の場合には、緩和度を上げてヒット件数が非常に少ない。これと同様に、実験では時事問題についても画像検索でヒットするページが非常に少ないか場合が多かった。

図 7 から読み取れる事として、(c) では緩和するキーワードによってヒット数、有効ページにかなりの差がついていることが分かる。またキーワードによっては緩和度 2 の結果のヒット数が緩和度 1 の時より減っている場合がある。このような結果より、具体的なキーワードを画像検索に残した場合はヒット数が多くなるのではないかと思われる。しかし、それだけではなく検索キーワード同士の関連度や、共起度によって左右されたり、キーワードが画像のメタデータの内容として使われる頻度なども関係してくると思われる。このようなことから、検索キーワードによって、質問の緩和の仕方を変化させたり、画像検索とテキスト検索を使いわけることが重要となると考えられる。

6. 検索結果の統合

4 章により検索キーワードは部分集合に分けられ、各部分集合の検索結果は $P(K_1), \dots, P(K_n), P(K_1 \wedge K_2), P(K_1 \wedge K_3), \dots, P(K_1 \wedge \dots \wedge K_n)$ として各クラスタに入っている。ここで各クラスタ内に入っている検索結果は Web ページ単位で入っているわけではなく、我々の論文¹⁾ の内容を使い、tf 値等を用いて各 Web ページから検索キーワードに関連した部分だけを抽出してきたものである。このようにして検索結果



図 9 プロトタイプ

を Web ページ単位でなく、Web ページから抽出した部分とする。ここで、各クラスタの中の Web ページから抽出した情報を自動的に統合して新しいコンテンツを生成し、これを検索結果としてユーザに提示する。従来のメタサーチエンジンとは違い、検索結果は URL リストの表示ではなく、検索結果の画像、テキスト文書が新しいコンテンツとしてまとめて表示されることが特徴である。検索結果を統合した例のプロトタイプを図 9 に示す。

現在の所のプロトタイプはこの様であるが、以下に考察段階のインターフェースについて記述する。

質問の緩和と対応するための情報の見せ方としては、各クラスタごとに表示していく方法である。まず、条件が一番厳しい緩和度 0 の $P(K_1 \wedge \dots \wedge K_n)$ を表示する。ユーザが満足いく結果が得られなかった場合は、緩和度を 1 ずつ増やしていく。ここでユーザはどのキーワードをテキストに緩和するかを指定することによって、表示していくクラスタを指定していくものである。これと同時に違うキーワードを緩和した場合の結果の一部を比較して表示する事によって、違うキーワードを緩和したらこのような結果が表示されるということを示す。インタラクションをすることによってユーザ本意で情報を見せれることが利点であるが、各クラスタは内容の意味的なもので分類されたものではないのが問題点である。

また Web を使用した百科事典を目標としていることで以下のようなインターフェースも考えている。情報の見せ方として、検索キーワードの定義や、意味、場所、画像、関連する話題などに分類してそれを目次として表示して見せるという方法である。つまりは収集してきた情報を意味的に類似しているもので再クラスタリングをする方法である。こうすることによって、ユーザはシステムとインタラクションしながら、自分の欲しい情報の目次を選択することによって必要な情

報を得ることができると思われる。

7. まとめと今後の課題

本論文では、以前に我々が提案した¹⁾ 検索質問の緩和の有効性の検証に重点をおき、クロスメディア・メタサーチについて提案した。

今後の課題としては、検索質問の緩和だけでなく、ユーザが入力した検索キーワードをどのように変化させていったら、既存の検索エンジンを有効に活用できるかを考え、既存の検索エンジンでできる検索の限界を調べていきたい。特に、現在の段階での質問の緩和は画像検索を主にして考えているため、テキスト検索のみの検索はほとんど考慮にいれていない。画像というものはユーザが欲しい情報の一部分でしかない。またテキスト検索は画像検索よりも適合率は低いが、再現率は高い。よって様々なメディアのサーチエンジンをうまく利用して、統合させていくのが重要であると考えられる。また、Web ページからの情報の抽出の方法を考え、検索結果をどのように統合させていったらユーザが容易に情報を取得できるのかを模索していきたい。

またこのようなシステムの応用例としては以下のような物を考えていきたい。従来の検索や本論文では、ユーザが検索エンジンのサイトへ移動して、検索キーワードを入力してそれについて結果を得られるような方法で検索を行っている。そこで、Web ページを閲覧しながらその背後で本論文のようなシステムを動かす、ユーザが単語や文章でわからないものがあったならば、その部分を選択することによって、吹き出しのようなものや別 Window によって、わからない情報について統合した結果が表示されるシステムを考えていきたい。こうすることによって、Web ブラウジングを楽しみながらいちいち検索サイトに移動することなく、順次分からない言葉の意味を知ることができ、快適なブラウジングができると考えられる。

謝辞 本研究の一部は、平成 15 年度科研費特定領域研究 (2) 「Web の意味構造に基づく新しい Web 検索サービス方式に関する研究」(課題番号: 15017249, 代表: 田中克己), 平成 15 年度科研費基盤研究 (A)(2) 「モバイル環境におけるコンテンツのマルチモーダル検索・提示と放送コンテンツ生成」(15 年度課題番号: 14208036, 代表: 田中克己) による。ここに記して謝意を表します。

参 考 文 献

- 1) 桑原 昭裕, 小山 聡, 角谷 和俊, 田中 克己 : マルチメディア・メタサーチのための質問変換と検索結果の統合, DBSJ Letters Vol.2, No.1
- 2) M.C. Schraefel, Yuxiang Zhu, David Modjeska, Daniel Wigdor, Shengdong Zhao : Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections, WWW2002, pp.130-131(2002)
- 3) Corin R. Anderson, Eric Horvitz : Web Montage: A Dynamic Personalized Start Page, WWW2002, pp.468-469(2002).
- 4) 奈良先端科学技術大学松本研究室茶筌ホームページ : <http://chasen.aist-nara.ac.jp/index.html.ja>
- 5) NAVER Japan : <http://www.naver.co.jp/>
- 6) 富士通 : MIRADOR-Search, <http://www.labs.fujitsu.com/News/1999/Dec/9-2.html>
- 7) 小山 聡, 田中 克己 : 質問の階層的構造化を用いた Web 検索手法の提案, DBSJ Letters Vol.1, No.1
- 8) Google : <http://www.google.co.jp/>
- 9) Altavista : <http://altavista.com/>
- 10) Google image : <http://images.google.co.jp/>