ウェブコンテンツとテレビ番組の統合型サーチエンジン

宮森　恒[†]　ゾラン・ステイチ[‡]　荒木 禎史[‡]　水口　充[†]　馬　強[†]　田中 克己[†*]

† 情報通信研究機構メディアインタラクショングループ　〒619-0289 京都府相楽郡精華町光台 3-5
‡ （株）リコー　ソフトウェア研究開発本部　〒112-0002 東京都文京区小石川 1-1-17
* 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: † {miya, mmina, qiang}@nict.go.jp, ‡ {araki-t, zoran.stejic}@nts.ricoh.co.jp , * ktanaka@i.kyoto-u.ac.jp

あらまし　本稿では，録画されたテレビ番組のコンテンツとウェブコンテンツを統合的に検索するサーチエンジンを提案する．従来のサーチエンジンは，ウェブコンテンツやデスクトップのデータを検索対象としてきた．しかし，検索対象となるデータは，HD/DVD レコーダやデジカメ，携帯端末，実空間など今後さまざまな場所に格納されると考えられ，それら異種リソースのデータを網羅的・横断的に検索するサーチエンジンが不可欠になると予想される．本稿では，そのような次世代サーチエンジン実現に向けた第一歩として，ハイパーリンクをもつコンテンツであるウェブと，ハイパーリンクをもたないコンテンツであるテレビ番組を対象として，これらを統合的に検索し，さらに，その結果の各項目の関連コンテンツを連鎖的に検索することができるサーチエンジンを提案する．統合検索は，テレビ番組とウェブコンテンツのキーワードに基づく統合インデックスを作成しランキングすることで実現される．また，関連コンテンツの連鎖検索は，統合インデックス内の各コンテンツ間での類似度計算によるランキングを行なうことで実現される．また，ズーミングを利用した結果の表示・閲覧を実現することで，効率よい情報取得が可能になる．本稿では，統合型サーチエンジンのプロトタイプを実装することで，提案する方式が有効に機能することを示す．

キーワード　サーチエンジン，情報検索，情報統合，ウェブコンテンツ，テレビ番組，統合検索，連鎖検索

Integrated Search Engine of Web Content and TV Programs

Hisashi Miyamori[†]　Zoran Stejic[‡]　Tadashi Araki[‡]　Mitsuru Minakuchi[†]　Qiang Ma[†]　Katsumi Tanaka[†*]

† Interactive Communications Media and Contents Group, NICT,
3-5 Hikari-dai Seika-cho Souraku-gun, Kyoto, 619-0289 Japan
‡ Software R&D Group, Ricoh Co., Ltd,　1-1-17 Koishikawa, Bunkyo-ku, Tokyo, 112-0002 Japan
* Graduate School of Informatics, Kyoto University　Yoshida Honmachi, Sakyo, Kyoto, 606-8501 Japan
E-mail: † {miya, mmina, qiang}@nict.go.jp, ‡ {araki-t, zoran.stejic}@nts.ricoh.co.jp , * ktanaka@i.kyoto-u.ac.jp

Abstract　A search engine that can handle TV programs and Web content in an integrated way is proposed. Conventional search engines have been able to handle Web content and/or data stored in a PC desktop as targeted information. However, in the future, the targeted information is expected to be stored in various places such as in hard-disk (HD)/DVD recorders, digital cameras, mobile devices, and even in real space in the future, and a search engine that can search across such heterogeneous resources will become essential. As a first step towards developing such next-generation search engine, a prototype search system is proposed in this paper that can handle Web content as an information source with hyperlinks and TV programs as another without them, that performs integrated search of those content, and that can sequentially search for content related to each ranking result, if required. The integrated search is achieved by generating integrated indices based on keywords obtained from TV programs and Web content and by ranking them, and the chain search for related content is done by calculating similarities and ranking between the contents described by the integrated indices. Also, the zoom-based display of the results enables to acquire information efficiently. In this paper, testing of a prototype of the integrated search engine validated the approach taken by the proposed method.

Keyword　Search engine, information retrieval, information integration, Web content, TV programs, integrated search, chain search

1. はじめに

近年のウェブ関連技術の発展に伴い，さまざまなウェブ検索が実現されている．最近では，ウェブ上の静止画といった特定のメディアや，ブログなどの速報性の高いウェブコンテンツに限定した検索も普及しつつある[1]-[4]．従来，サーチエンジンといえばこれら膨大なウェブコンテンツを検索対象としてきた．

また，コンピュータ内に保存されたさまざまな種類のファイルを全文検索するデスクトップ検索が実現されている[5][6]．電子メールやその添付ファイル，さまざまな形式の文書ファイル，チャットなど，ファイル名だけでなくその中身を考慮した検索が可能であり，サーチエンジンの検索対象は，ウェブだけでなく個々のコンピュータのデータへ徐々に拡大しつつある．

一方，近年のさまざまな家電製品の性能向上に伴い，一般利用者が記録・利用できるコンテンツは身の回りのさまざまな場所で格納できるようになってきている．例えば，現在，1TB の容量をもつハードディスク(HD)レコーダが登場し，画質モードによっては 1700 時間以上の録画が可能である．また， 1300 万画素を超える画素数での撮影や 2GB 以上のメモリを使った写真記録が可能なデジタルカメラや，60GB の容量をもち音楽・写真・動画といったさまざまな種類のコンテンツ再生が可能な携帯プレイヤが次々と実用化されつつある．将来的には，バスの時刻表や駅の広告など身の回りの実空間に各種コンテンツが埋め込まれることも期待されており（ユビキタスコンテンツ），それらとのインタラクションによって利用されるコンテンツ種別も必然的に多様化していくものと予想される．

しかし，格納場所やデータ種別が多様化したとしても，人間が限られた時間で情報取得するには物理的限界があり，有効に情報取得するには，メディアの種別や格納場所を越えた横断的な検索を行えるサーチエンジンが今後不可欠になると考えられる．特に，従来のサーチエンジンはリンク解析や全文検索に基づいているため，それらだけに依存しない様々なコンテンツを場面に応じて統合的にランキングし検索するための手法が重要となる．

本稿では，そのような次世代サーチエンジンの実現に向けた第一歩として，ハイパーリンクをもつコンテンツであるウェブと，ハイパーリンクをもたないコンテンツであるテレビ番組を対象として，これらを統合的に検索し，検索結果の各項目の関連コンテンツを必要に応じて連鎖的に検索できるサーチエンジンを提案する．統合検索は，テレビ番組とウェブコンテンツから得られるキーワードに基づく統合インデックスを作成しランキングすることで実現される．また，関連コンテンツの連鎖検索は，統合インデックス内の各コンテンツ間での類似度計算によるランキングを行なうことで実現される．これにより，例えば，

- キーワードクエリにマッチしたウェブコンテンツと録画番組を同時に検索する，
- 今閲覧中のウェブコンテンツの内容に関連した内容の録画番組を探す，
- 今視聴中のテレビ番組に関連した内容のウェブコンテンツや別の番組を検索する，

といったことが実現できる．本稿では，統合型サーチエンジンのプロトタイプを実装することで，提案する方式が有効に機能することを示す．

本稿の構成は以下の通りである．2 章では，統合型サーチエンジンの概要について述べる．3 章では統合型サーチエンジンの処理手順と実装について詳しく説明する．4 章では考察を示し，5 章で関連研究を整理する．最後に 6 章でまとめを述べる．

2. ウェブコンテンツとテレビ番組の統合型サーチエンジンの概要

図 1 に，統合型サーチエンジンの概要を示す．

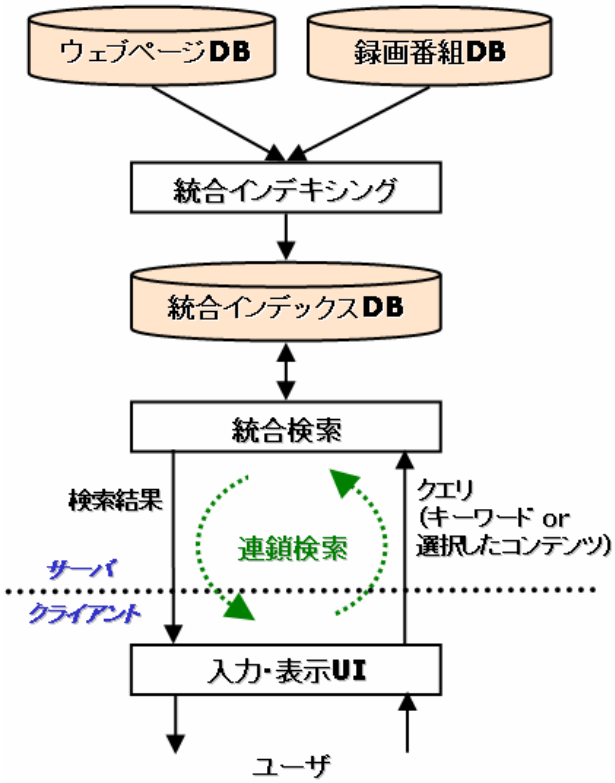

図 1．ウェブコンテンツとテレビ番組の統合型サーチエンジンの概要

まず，サーバ側では，ウェブページを保持するデータベース（ウェブページ DB）と録画番組を保持するデータベース（録画番組 DB）の各コンテンツに対し，ベクトル空間モデルに基づく統合インデックスが予め生成され，データベース（統合インデックス DB）に格納される（統合インデキシング）．

次に，クライアント側のユーザによって入力された

検索要求（クエリ）をサーバが受け付けると，統合インデックスが参照されることで，ウェブページと録画番組を混在させた検索が実行される（統合検索）．

さらに，必要に応じて，統合検索の結果の各ウェブページあるいは各録画番組を新たなクエリとして，それぞれに関連するウェブページまたは録画番組を連鎖的に検索することができる（連鎖検索）．

検索結果の表示方法としては，初期検索結果として統合検索の結果だけをまず一覧表示し，ユーザにより特定のコンテンツが選択されると，そのコンテンツの内容とともに，関連したコンテンツ（連鎖検索で得られる上位結果のコンテンツ）がズーミング操作を介して,随時,一覧表示されるようにシステムを構成した．このようにするのは，関連したコンテンツ同士が適宜まとまった形でユーザに提示されるため，ユーザの興味に合った内容をより的確に提供できる可能性が高くなり，効率のよい情報取得が実現できると考えられるためである．

3. ウェブコンテンツとテレビ番組の統合型サーチエンジンの処理手順と実装

2 節で述べた統合型サーチエンジンの各処理手順とそのプロトタイプの実装について述べる．

3.1. 統合インデキシング

ここでは，ウェブページ DB と録画番組 DB の各コンテンツに対し，統合インデックスを生成する．

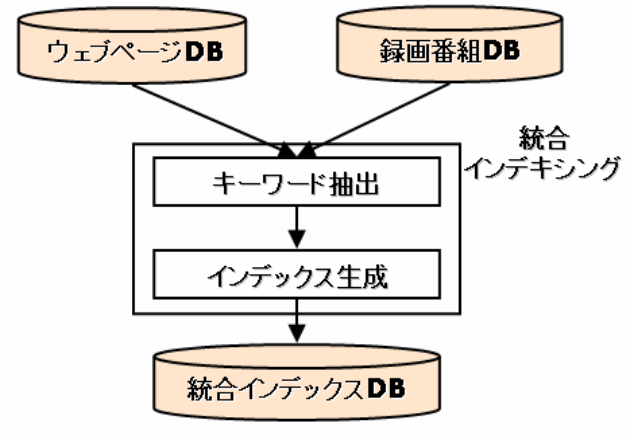

図２．統合インデキシング

まず，図２に示すように，ウェブページのテキストおよび録画番組のクローズドキャプションのテキスト（以下，字幕テキスト）の両方から代表的なキーワードを抽出する（キーワード抽出）．ここでは，形態素解析で得られた名詞語彙のうち出現頻度の高い単語をキーワード集合 K として抽出することとした．

$$K = [K[1], K[2], \cdots, K[N_K]]$$

ここで，N_K はキーワード総数，$K[k]$ は k 番目のキーワードを表す．

次に，抽出したキーワード集合 K に基づき，各ウェブページおよび録画番組を TF-IDF 値によって重み付けした文書／キーワード行列を計算する．ここで，文書とはウェブページのテキストまたは録画番組の字幕テキストに相当し，文書集合 D は次のように表される．

$$D = [D[1], D[2], \cdots, D[N_D]]$$

なお，N_D は文書総数，$D[d]$ は d 番目の文書を表す．

ここで，$tfreq(k,d)$ を文書 $D[d]$ におけるキーワード $K[k]$ の出現頻度，$dfreq(k)$ をキーワード $K[k]$ が出現する文書数とし，関数 idf を次のように定義する．

$$idf(k) = \log_2\left(\frac{N_D}{dfreq(k)}\right) + 1$$

文書 $D[d]$ におけるキーワード $K[k]$ の重みを

$$KeywordWeight(k,d) = tfreq(k,d) \times idf(k)$$

で表すと，文書／キーワード行列 DK は以下のように計算される．

$$DK = \begin{bmatrix} DK[1,1] & DK[1,2] & \cdots & DK[1,N_K] \\ DK[2,1] & DK[2,2] & \cdots & DK[2,N_K] \\ \vdots & \vdots & \ddots & \vdots \\ DK[N_D,1] & DK[N_D,2] & \cdots & DK[N_D,N_K] \end{bmatrix}$$

ただし，$DK[d,k] = KeywordWeight(k,d)$ である．得られた文書／キーワード行列 DK が統合インデックスに相当する．

3.2. 統合検索

ここでは，ユーザによって入力されたキーワード集合のクエリに対し，ウェブページと録画番組の統合検索を行う．

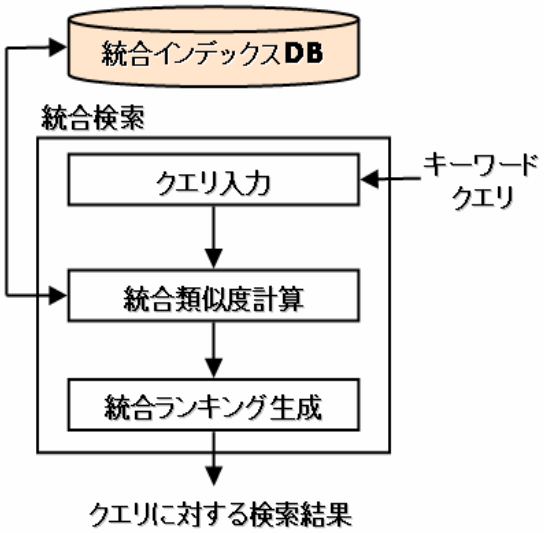

図３．統合検索

まず，図３に示すように，１つ以上のキーワードからなるクエリを受け付け,それを統合インデックスDB内のインデックスと共通の形式に変換する.

クエリをQとすると,

$$Q = [Q[1], Q[2], \cdots, Q[N_Q]]$$

ここで，N_Qはクエリにおけるキーワード総数，$Q[q]$はq番目のキーワードを表す．これを次のようなクエリベクトルQVに変換する.

$$QV = [QV[1], QV[2], \cdots, QV[N_K]]$$

ここで,

$$QV[k] = \begin{cases} idf(k), if(K[k] \in Q) \\ 0, otherwise \end{cases}$$

QVの要素をIDF値としているのは，TF-IDF値のTF値=1としたものであり,各キーワードの全文書に対する重要度を考慮に入れて類似度計算を行うためである.

次に，クエリと統合インデックスDB内の各コンテンツとの類似度を計算する．文書ベクトルDVを以下のように定義すると,

$$DV[d] = [DK[d,1], DK[d,2], \cdots, DK[d, N_K]]$$

統合類似度$DocQuerySim(d, QV)$は以下のように計算される.

$$DocQuerySim(d, QV) = \frac{\sum_{k=1}^{N_k}(DK[d,k] \times QV[k])}{\sqrt{\sum_{k=1}^{N_k}(DK[d,k])^2 \times \sum_{k=1}^{N_k}(QV[k])^2}}$$

最後に，計算された類似度に基づいて，その値が大きいものから順にウェブページと録画番組を統合的にランキングする（統合ランキング).

3.3. 連鎖検索

統合検索結果の各ウェブページおよび各録画番組を閲覧している際に，その閲覧コンテンツ自体を新たなクエリとして関連したウェブページまたは録画番組を検索したい場合がある.

この場合，クエリとして選択された文書$D[d]$と統合インデックスDB内の他の文書$D[e]$間の統合類似度$DocDocSim(d, e)$を次のように計算すること統合検索と同様の検索を行なう.

$$DocDocSim(d, e) = \frac{\sum_{k=1}^{N_k}(DK[d,k] \times DK[e,k])}{\sqrt{\sum_{k=1}^{N_k}(DK[d,k])^2 \times \sum_{k=1}^{N_k}(DK[e,k])^2}}$$

これにより，利用者は，閲覧中のコンテンツを逐次新たなクエリとして連鎖的に検索を繰り返すことができる.

3.4. 検索結果の閲覧

検索結果の表示については，本稿では，初期検索結果として統合検索の結果だけをまず一覧表示し，ユーザにより特定のコンテンツが選択されると，そのコンテンツの内容とともに，関連したコンテンツ（連鎖検索で得られる上位結果のコンテンツ）がズームによって一覧表示されるようにシステムを構成した（図４).

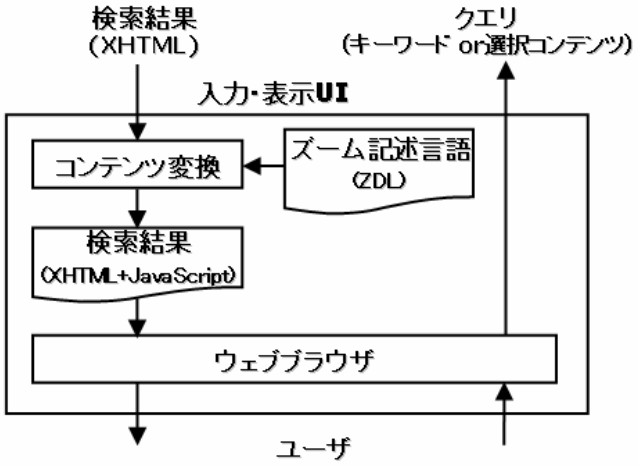

図４．検索結果の表示手順

まず与えられたキーワードクエリによる統合検索を行い，結果として得られたコンテンツごとに連鎖検索を実行する．統合検索および連鎖検索の結果（XHTML）を，詳細度や種類の異なるコンテンツの表示状態をズームによって制御できる記述言語（ZDL[7]; ズーム記述言語）を用いて変換し，ユーザの興味に合わせて必要な詳細度で適宜コンテンツを表示できるようなデータ（XHTML+JavaScript）を出力する．出力されたデータは汎用ブラウザで適宜ズーム操作を行ないながら閲覧することができる.

例えば，統合検索結果の一覧表示においては，ウェブページの場合，タイトルと一致キーワードを含む2,3行の周辺テキストが表示され，録画番組の場合は，一致キーワードを含む周辺の字幕テキストやタイトルが表示される（図５).

また，ユーザによって特定のウェブページあるいは録画番組の項目が選択されズーム操作によって拡大されると，そのコンテンツの内容が表示されるだけでなく（図６の例では，左上の番組映像)，それに関連したコンテンツ（連鎖検索の上位結果のコンテンツ）の一覧がズーム操作に連動して拡大されながら表示される（図６の例では，中央下部に表示されている３つの関連コンテンツがプレビュー表示されている).ユーザがあるコンテンツに着目してそこにズームインすると関

連コンテンツが常にまとまった形で提示されるため，ユーザの興味が別の内容に移った場合も，新たなクエリを考えて再検索をする必要性が少なく，効率のよい情報取得が可能になると考えられる．

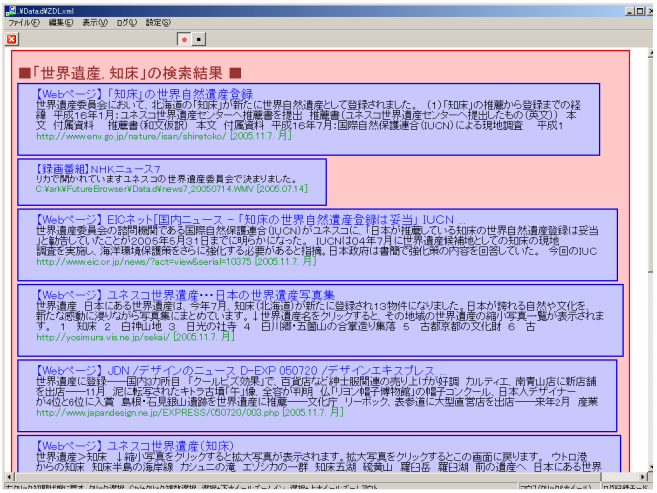

図５．統合検索結果の表示例

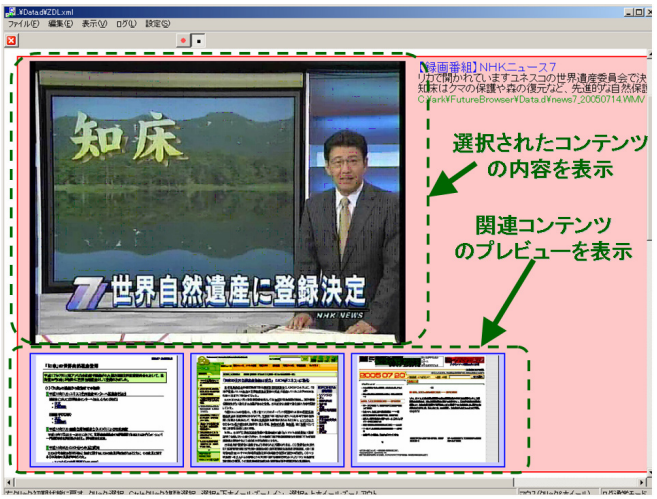

図６．連鎖検索結果の表示例

3.5. 統合型サーチエンジンのプロトタイプの実装

以上で述べた統合型サーチエンジンのプロトタイプを実装した．

NHK のニュースおよび健康番組，計 20 本をテストデータとして録画し，各番組から得られる字幕テキストの出現頻度の高い単語を代表的なキーワードとして抽出した．ウェブページからのキーワード抽出については，ウェブ全体を対象とすると非現実的となるため，録画番組から得られるキーワードをもとに，予め既存の検索エンジンを用いて対象となるウェブページ候補をある程度限定し，各ウェブページ候補から別途キーワードを抽出するという方法をとることとした．得られたキーワードをもとに統合インデックスを作成した．

１つ以上のキーワードを入力クエリとして受けつけ，統合検索および連鎖検索を行ない，検索結果を返すモジュールおよびインタフェースを開発した．図７に，実装したプロトタイプの，ズーミングによる検索結果の操作例を示す．初期検索結果は統合検索表示として，ウェブページの場合はタイトルと短いテキスト，録画番組の場合は字幕テキストの一覧が表示されている（図７の左側）．この状態から，ウェブページ結果をズームインするとウェブページ縮小画像が出現し，さらにズームインすると拡大して，ある時点で真のウェブページに置き換わる．そこで，関連番組縮小画像と字幕テキストが出現し，これをズームインすると関連番組画像が拡大して再生される．また，録画番組結果をズームインすると，番組の縮小画像が出現し，さらにズームインすると関連ウェブページの縮小画像が出現する．そこで，番組縮小画像をズームインすると，それが拡大されて再生され，関連ウェブページ縮小画像をズームインすると，それが拡大されて，ある時点で真のウェブページに置き換わる．図８，９に実装したサーチエンジンでの画面遷移例を示す．

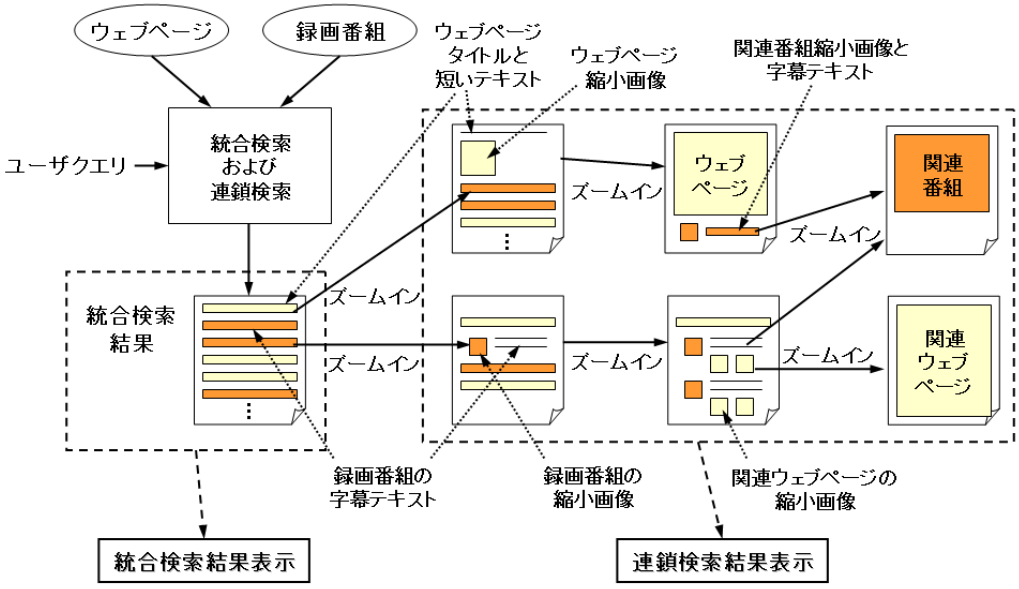

図７．検索結果のズーミングによる閲覧操作例

図８．実装したサーチエンジンの画面遷移
（ウェブ→関連番組）

図９．実装したサーチエンジンの画面遷移
（録画番組→関連ウェブ）

このように，実装したプロトタイプによって，もともとデータ構造や格納場所が異なるウェブコンテンツと録画番組を統合的に処理し，検索可能であることを確認した．また，検索結果一覧から詳細情報までを１つの画面の中で連続的に表示することにより，一覧画面とリンク先を往復する必要がなく，閲覧の文脈を保持しつつ，何度も検索する手間を省いて関連した情報を効率的に閲覧できることも確認した．

4. 考察

本稿で提案する統合型サーチエンジンは，従来，ウェブとローカルなコンピュータ内に限られていた検索対象の範囲を，HD/DVD レコーダなど今後大容量化が予想される家電製品などへも広げていくことができ，異なる特性をもつ多様なコンテンツを横断的に検索可能なシステムが実現できる可能性を示したという点で重要である．

本稿では録画番組がクローズドキャプションと呼ばれる時刻印付の字幕テキストをもつことを利用し，そこに表れる「キーワード」をベースとして，もともと異なるデータ構造をもつ番組映像とウェブコンテンツの統合的なインデキシングができることを示した．このアプローチは，今後さまざまなコンテンツに応用することができると考えられる．例えば，デジカメについては，今後は簡単な音声認識機能などが標準搭載され，コンテンツに付随する「キーワード」を含むメタデータがより手軽に生成・利用できるようになる可能性がある．また，現在，音楽などの携帯プレイヤにメタデータを提供しているデータベースは，今後は，タイトル情報などの書誌的情報だけでなく，歌詞と音楽とのシンクロが可能なレベルのメタデータを提供できるデータベースとして発展していく可能性が考えられる．このように，データ構造や格納場所が異なるコンテンツであっても，今後は何らかの形で内容に関する「キーワード」が付随している可能性はますます高まっていくものと予想される．よって，本稿で示したキーワードをベースとした統合検索のアプローチは，今後録画番組だけでなく多様なコンテンツにも広く応用できる可能性があると考えられる．

なお，スポーツ番組のように，アナウンサーの話す内容とその時刻に映像内で起こっている内容とが異なる場合や，クローズドキャプション自体がほとんど提供されない番組も存在すると考えられる．この場合，本方式では有効な検索結果が必ずしも得られるとは限らない．よって，映像のカット検出や，映像内に表示されるキャプション認識を利用したテキスト抽出，ドメイン知識を利用したイベント認識技術などの映像解析技術を利用することで番組内容に関するメタデータを多角的に補うアプローチが重要となる[8][9]．

また，キーワード以外の情報をベースとした統合検索へのアプローチも今後重要になると考えられる．例えば，ウェブ上には，ブログや掲示板などといったテレビ番組に対する視聴者の反応が多種多様な観点から表現されるコンテンツが存在する一方，テレビ番組は公共放送という性質上，限られた時間のなかである一定以上の質の高い情報を提供することが常に求められる．例えば，番組チャットの解析技術[10]を利用することで，各番組に対する関連情報を，社会的にどのように受け入れられているのかという観点から提供することができると考えられる．これは本稿のような，コンテンツ間の類似という観点からの関連情報の提供とは異なるアプローチである．このように，ウェブコンテンツがもつ情報の多様性と，録画番組がもつ情報の質の高さを相互にうまく補完しながら情報を捉えるアプローチが今後重要になってくると考えられる．

多様なコンテンツの統合検索を考える上では，その検索結果の閲覧インタフェースも重要な要素と考えられる．例えば，Google や Yahoo などのウェブ検索における検索結果の閲覧インタフェースは，リスト形式のため一覧性は高いものの，その内容を確認するには，

一つ一つのアンカーをクリックして該当するウェブコンテンツを画面上に表示させる必要がある．その際，検索結果の一覧画面が該当ウェブコンテンツに置き換わって表示されたり，そのコンテンツを表示した新しいウィンドウが次々に起動する，といったことがしばしば起きる．特に，多角的な観点から情報を収集したい場合は,アンカーを何度もクリックする必要がある．よって，多種多様なコンテンツの統合検索を考える際には従来のこのアプローチは最適とはいえない．本稿では，このような観点からズーミングを利用した検索結果の閲覧インタフェースを実現した．ズーミングを利用することにより，ユーザは着目部分だけを拡大して閲覧でき，新しいウィンドウがいくつも起動するということがなくなる．また，１つのウィンドウ内に常に文脈を保存することができるため，連鎖的に検索を進める際にもユーザが多くのウィンドウの中に埋もれて迷子になるという危険性が極めて少ないと考えられる．今後，ユーザ実験を通して閲覧インタフェースの有効性評価と向上を図っていく必要がある．

5. 関連研究

従来，異種リソースの情報を統合的に扱う研究は活発に行なわれてきた．例えば，ラッパーにより情報源固有のインタフェースから共通情報モデルへ変換することによる異種情報源の統合手法はよく知られている[11]．また，複数データベース間の意味的な相互互換性を確保するために，意味の数学モデルとそれを応用した意味的連想検索方式が提案されている[12][13]．

一方，テレビ番組とウェブの各コンテンツの性格の違いに着目した情報検索の研究も行なわれている．例えば，テレビ番組のクローズドキャプションから話題構造と呼ばれるデータを抽出し，これに基づきその番組内容を補完すると期待されるウェブコンテンツを検索する補完情報検索が提案されている[14]．また，インターネット上の番組チャットの書き込み内容の頻度や表現解析により視聴者の反応度合い，喜びや落胆などの感情の度合いを抽出し，テレビ番組の内容を補完する方式が提案されている[10]．

本稿で実装した統合検索は，多種多様なコンテンツのいずれにも共通して付随する可能性が高いと考えられるキーワードを利用したベクトル空間モデルに基づいたものである．今後，検索対象のコンテンツのそれぞれの特性を活かしたキーワード抽出，関連情報抽出を考えていく必要がある．

また，検索に関連する閲覧インタフェースについても数多くの研究がなされている．例えば，技能訓練に必要な例示映像とメタデータを入力し，図書館などの様々な情報源を検索して得られる関連情報を例示映像と関連付けたインタフェースを提供するシステムが提案されている[15]．このインタフェースでは，例示映像の再生画面右横に最新の関連情報が常に表示されるが，画面配置は一定のままである．一方，視聴中のテレビ番組を一時停止させ，そのシーンに関連するウェブコンテンツを逐次提示する方式が提案されている（寄り道視聴）[16]．このインタフェースでは，テレビ番組が全画面に表示されている画面，一時停止したテレビ画面に対して関連情報を表示した画面,および，選択した関連情報を全画面に表示した画面の３つの画面構成を持っており，ユーザはそれぞれの画面間をズームインタフェースによって動的に遷移させることができる．さらに，テレビ番組内の各シーンを縦方向に並べたストーリーボード形式のウェブコンテンツに変換し，関連した情報へのリンクを各シーンの代表アイコンの右横に並べて配置したインタフェースが提案されている（ウェブ化ビデオ）[17]．このインタフェースでは，番組全体の一覧性が高く，関連情報へのアクセスも素早く行なえるという利点がある．関連情報として埋め込まれる内容は，シーン分割されたクローズドキャプションをアンカーとした該当シーンへのリンク,補完検索で得られたウェブコンテンツへのリンク,番組チャット解析で得られた各シーンの視聴者反応の大きさなどである．このインタフェースも，番組全体をいくつかの異なる詳細度で分割したウェブコンテンツとしての画面を持ち，利用者はそれぞれの画面間をズームインタフェースによって滑らかに切り替えることができる．

本稿で実装した閲覧インタフェースは，検索結果の一覧をテキスト表示する画面，検索結果の各項目のプレビュー画面と説明テキストを表示する画面，該当コンテンツそのものを表示・再生し，より詳しい説明テキストを表示する画面といった画面構成を持ち，ユーザはそれぞれの画面間を，検索結果の全項目あるいは選択した特定項目についてズームによって動的に遷移することができる．さらに，ズームに連動して，着目コンテンツに関連したコンテンツを，着目コンテンツの直下に表示させることができる．今後，ユーザ実験を通して閲覧インタフェースの評価と改良を進めていく必要がある．

6. まとめ

本稿では，異なるメディアや格納場所を越えた横断的な検索可能な次世代サーチエンジンの実現に向けた第一歩として，ウェブコンテンツとテレビ番組を対象とした統合型サーチエンジンを提案した．ハイパーリンクをもつウェブとそれをもたない録画番組に対して，リンク解析や全文検索だけに依存しない統合的な検索を実現するため，キーワードに基づくベクトル空間モデルを考えることで統合検索を可能とした．また，統合検索結果の各コンテンツに関連したコンテンツをズーム操作によって逐次まとめて表示・非表示できるようにインタフェースを工夫した．ズーミングを利用してこれら検索結果を表示・閲覧できるようにすること

で，今閲覧中のウェブコンテンツに関連した内容の録画番組を探すなどの，効率よい情報取得が可能になることを確認した．

今後は，より多くの実験データを用いたサーチエンジンの性能の検証，および，ユーザ評価実験等を進めていく予定である．

文　献

[1] Google Image Search. http://www.google.com/imghp?hl=en&tab=wi&q=

[2] Technorati. http://www.technorati.com/

[3] Ask.jp. http://ask.jp/

[4] Goo ブログ. http://blog.goo.ne.jp/

[5] Google デスクトップ.http://desktop.google.co.jp/ja/

[6] MSN Japan. http://www.msn.co.jp/home.armx

[7] Araki, T., Miyamori, H., Minakuchi, M., Kato, A., Stejic, Z., Ogawa, Y., Tanaka, K.: Zooming Cross-Media: A Zooming Description Language Coding LOD Control and Media Transition. Proc. of the 16th International Conference on Database and Expert Systems Applications (DEXA2005), LNCS3588, pp.260-269, 2005.

[8] Nakamura, Y., Kanade, T.: Semantic analysis for video contents extraction - spotting by association in news video. ACM Multimedia, pp.393-401, 1997.

[9] Miyamori, H.: Automatic annotation of tennis action for content-based retrieval by integrated audio and visual information. CIVR2003, LNCS2728, Springer-Verlag, pp.331-341, 2003.

[10] Miyamori, H., Nakamura, H., Tanaka, K.: Generation of Views of TV Content Using TV Viewers' Perspectives Expressed in Live Chats on the Web, Proc. of the 13th Annual ACM International Conference on Multimedia (ACM Multimedia2005), pp.853-861, 2005.

[11] 北村泰彦: インターネット上での新しい情報共有技術，システム制御情報学会, Vol.42, No.9, 1998.

[12] Kitagawa, T. and Kiyoki, Y.: The mathematical model of meaning and its application to multidatabase systems, Proc. of 3rd IEEE International Workshop on Research Issues on Data Engineering, pp.130-135(1993).

[13] 清木康，金子昌史，北川高嗣: 意味の数学モデルによる画像データベース探索方式とその学習機構，電子情報通信学会論文誌,D-II, Vol.J79-D-II, No.4, pp.509-519, 1996.

[14] 馬 強，田中 克己: 話題構造に基づく放送と Web コンテンツの統合のための検索機構，情処論文誌: データベース Vol.45 No.SIG 10 (TOD23), pp.18--36, 2004.

[15] Gordon, Andrew S.: Using Annotated Video as an Information Retrieval Interface. 2000 International Conference on Intelligent User Interfaces, 2000.

[16] 宮森恒，馬強，田中克己: 放送番組とウェブコンテンツの融合による新しい視聴方式，電子情報通信学会モバイルマルチメディア通信研究会，MoMuC2005-46，pp.89-94, 2005.

[17] Miyamori, H., Tanaka, K.: Webified Video: Media Conversion from TV Programs to Web Content for Cross-Media Information Integration, Proc. of the 16th International Conference on Database and Expert Systems Applications (DEXA2005), LNCS3588, pp.176-185, 2005.