データベースシステム 104−29
（1995. 7. 19）

ハイパーメディアデータベースの段階的構造化と多重ビュー

上浦 真樹† 森下 淳也‡ 上島 紳一* 大月 一弘* 田中 克己†

†神戸大学大学院自然科学研究科
‡姫路獨協大学外国語学部
*関西大学総合情報学部
*神戸大学国際文化学部

本稿では、ハイパーテキスト文書に対して、視点に応じてリンク設定等が切り替えられる「多重ビュー機構」、および、部分的にオーサリングされた文書群（「不完全web」）から完全なハイパーテキスト文書群を生成していく機構を提案する。前者は、ハイパーリンク設定情報を質問対で表現することで1種の仮想リンクを実現するもので、ここでは、既存のWorld Wide Web上で実際に行った実装についても報告する。後者は、アンカー設定、キーワード付与、リンクの設定などを部分的に行ったハイパーテキスト文書群から集合演算、フィルタリング演算、合成演算を通じて完全なハイパーテキスト文書群を生成するもので、本稿では、この機構の概念やモデルについて主に報告する。

Incremental Organization and Multiple Views for Hypermedia Databases

Masaki KAMIURA† Jun-ya MORISHITA‡ Shinichi UESHIMA*
Kazuhiro OHTSUKI* Katsumi TANAKA†

†Graduate School of Science and Technology, Kobe University.
‡Faculty of Foreign Languages, Himeji Dokkyo University.
*Faculty of Informatics, Kansai University.
*Faculty of Cross-Cultural Studies,Kobe University.

In this paper, we will introduce a multiple-view mechanism and an incremental data organization mechanism for hypertext documents. The multiple-view mechanism makes it possible to give several views (especially, virtual hyperlinks) to hypertext documents. Its implementation over conventional WWW is also described. The proposed incremental organization mechanism is based on *partial webs*, which is a collection of hypertext documents with partial authoring information. By applying set operations, keyword filtering operations, and composition operations to a partial web, users can generate a complete web from his own viewpoint. In this paper, we mainly focus on the concept and the model of the proposed partial webs.

1 まえがき

現在、ネットワーク技術の進展により、インターネットが情報を提供する環境を巨大な分散型のデータベースとしてみる試みが起こっている。その一つが情報を分散型ハイパーテキストとして提供する仕組であるWorld Wide Web（以下WWW）である[1]。WWWによって、ユーザはインターネット上の様々な形式の情報をリンクで結合し、容易にアクセスすることが可能となり、幅広いユーザが様々な情報を共有することができる。

しかし、現在のWWWでは文書データをハイパーテキスト化する場合、次のような問題があると考えられる。

- 全てのデータに対して明示的かつ静的にリンク情報を付けていかなければならない。これはデータが多量である場合、オーサーは多大な労力を必要とすることになる。
- 文書情報の追加、変更、更新、削除といったことが頻繁に起こった場合、それに応じたリンクの更新といった作業が煩雑、複雑となるという問題がある。
- 現在のHTML文書のような形式では、1つのHTML文書を共有しながら、多様な視点から異なるリンク情報を設定するということが困難である。特に、著者以外の利用者が独自の視点から利用者固有のリンク情報を元のHTML文書に付加することが困難である。

この問題に対して本論文では、

- 仮想リンクによる多重ビュー機構
- 不完全webからの段階的なweb生成方式

を提案する。

前者は、我々が以前に提案した、データベース質問対による動的リンク機構（質問対リンク）のWWW環境への導入により、文書データとリンク情報を切り離し、元の文書に対して種々の仮想リンクを設定する機構である。

次に、後者は、多量の未構造化状態の文書データ群に対してオーサリングを複数のオーサーが協調的に行なう場合、各オーサーはどこにどんな文書があるかといった全体の情報が完全に把握できないため部分的にしかオーサリングが行えないという問題に対処しようというものである。この様な場合オーサーは、この文書からリンクを張りたいが、リンク先となる文書が分からないと行った状況が起こりうる。そこで、この文書とあの文書を明示的にリンクするという完全なリンクではなく、リンクの片方の文書が分からないがキーワード等によって部分的に指定できるといった場合があり、そのような不完全なリンクを含むハイパーテキスト文書群をここでは、**不完全web**と呼ぶ。不完全web群に対して、集合演算、キーワードによるフィルタリング操作などを施した後、キーワードマッチングなどにより合成し、リンクが完全に明示的になったハイパーテキスト文書群を生成するのが後者の機構である。

2 多重ビュー

2.1 質問対リンク

質問対リンク[2][3]はハイパーテキストにおけるアンカー間のリンクをデータベースへの2つの質問文によって表すことにより仮想的にリンクをはる、一種の動的リンクを実現している。

ここでいう動的リンクとは、静的リンクがあらかじめリンク先が決まっているのに対して、データベースへの問い合わせによってリンク元のアンカーとリンク先が決定するものである。

HTML[4]がドキュメント内にリンク情報を埋め込んで記述するのに対して、質問対リンクではドキュメントとリンク情報を分割してもっている。リンク情報は2つの質問文からなる質問の対で構成され片方の質問文はアンカーの指定を、もう一方はリンク先の指定を行なう（図1参照）。

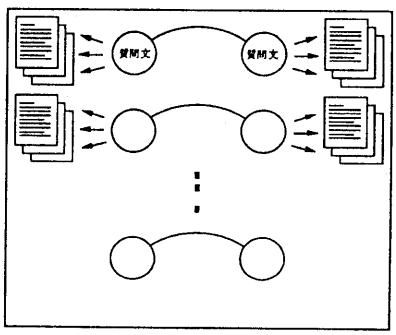

図1: 質問対リンク

2.2 質問対リンクを用いた多重ビューの実現

質問対リンクという動的リンク機構は組織的なリンク設定に有用であるが，これを仮想リンク機構と考えると現在のWWWの次の課題を解決するのにも役立つものと考えられる。

システム構築の留意点としては、次のような事項があげられる。

1. サーバ側にこの機構を装備するだけで、既存のWWWブラウザが多重ビュー機能を使用することができなければならない。
2. 現在のWWWの静的リンク機構と共存する。
3. 元のHTML文書の変更を強いることがない。
4. 仮想リンクで生成されるアンカーが、元のHTML文書に埋め込まれているアンカーと衝突した場合、元のアンカーの情報を失ってはならない。

本研究ではこれらの問題に対して以下のように対処した。まずWWWサーバのCGI(The Common Gateway Interface) 機能を用いて外部プログラムとして構築することによって1目つの問題を解決した。2、3については、元のHTML文書を一度外部に格納し、このデータに対して仮想リンク機構によって生成されたアンカーを埋め込むことで解決した。また、4については、アンカーの衝突の際は、リンク先のURLを引数としてとっておき、元のリンク先と仮想リンクによるリンク先の双方を任意で選択できるようにした。

2.3 システムの概要

本システムの利用者は元のHTML文書群とは別に質問対による仮想リンク情報ファイルを作成する。リンク情報ファイルはアンカーとなる単語を示す質問文とリンク先となるドキュメントを指定する質問文のペアの集合である。WWWサーバに対してあるクライアントからそのHTML文書を参照する要求が来た際、外部プログラムが作動しリンク情報ファイルより得られるアンカーを探し、該当するものに一時的にリンク情報のタグを埋め込んだドキュメントをクライアントに返す。

クライアントがあるアンカーを選ぶと外部プログラムはリンク先を検索し結果をアンカーの埋め込まれたURLのリストとして返す。

クライアントはリストの一つを選択することでリンク先へとたどり着く。

これによって、ユーザは元のHTML文書に埋め込まれているリンクと、質問対リンクによる仮想リンクを意識せず同時に見ることができる。この様な機能によりドキュメント製作者はこのドキュメントに対する様々な視点にもとづく仮想リンクが設定されてもリンク情報の更新を行なう必要がなくなる。

質問対リンクはドキュメント内の単語と他のドキュメント間のリンク関係を表し、形式的には以下のように表す。

$$< uid, query1, query2 >$$

ここで、

- uid:識別子
- query1:リンクの始点となる単語をを表す質問文
- query2:リンク先となるドキュメントを表す質問文

query1、query2の質問文は以下のように構成される。

$$< file, dir, title, word, access >$$

ここで、

- file:ファイル名の指定。指定されたファイルに対して検索を行なう。
- dir:ディレクトリ名の指定。そのディレクトリ内のファイルに対して検索を行なう。
- title:指定した単語が文書のタイトルに含まれる。
- word:指定した単語がドキュメント内に含まれる。
- access:これまでのサーバ内のドキュメントに対するアクセス回数、次のアクセスまでの時間の記録より、ドキュメントに対してランク付けを行なっておき、そのランクを指定する。

質問文の構成要素はすべてが値を持つ必要はない。検索は値を持つ属性それぞれに対して真となるものとなるドキュメントまたは単語を捜し出すことによって行なわれる。通常、file属性とdir属性はどちらかが値を持ち、ドキュメントを限定するためのものであるが、2つがともに値を持つ場合はfile属性のみを有効とする。title属性とword属性はドキュメントのテキストサーチにより行なわれ、これによりドキュメントの内容に対する限定を主な目的とする。ここで、リンクの始点を表す質問文では単語を指すためのword属性が必ず必要となる。access属性はドキュメントに対するアクセスの回数、次のアクセスまでの時間によっ

てそのドキュメントの重要度、または主要度を指定するためのものである。ドキュメントへのアクセスの回数はそのドキュメントへ張られているリンクの数に大きく左右され、一概にアクセス数の多少でドキュメントの価値を判断することはできない。そこで、あるドキュメントがアクセスされ後に次のアクセスが起こるまでの時間をそのドキュメントの参照時間と考え、アクセス回数とその参照時間の2つの要素からこれまでのアクセスの記録から見たドキュメントの価値を決定する。

2.4 システムの実現

多重ビュー機構を実現する為の外部プログラムをプログラミング言語 Perl によって構築した。質問対リンク群は Linkfile に格納される。外部プログラムはアンカー生成とリンク検索の2つからなり、前者は Linkfile よりドキュメント内のアンカーを作成し、後者はリンク先となるドキュメントを検索するものである（図2参照)。

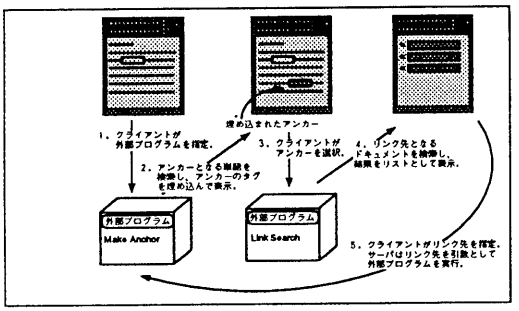

図 2: プログラムの動作の概要

2.4.1 データ構造

- **HTML 文書**
  本来、質問検索を行なう際、HTML 文書はデータベースに格納しておくことが望ましいが、現在の WWW の静的リンク機構と共存させる為、現行のまま UNIX のファイルとして置いておき、WWW の対象となるディレクトリ以下の HTML 文書に対して UNIX の find コマンドと Perl による正規表現のパターンマッチを用いて検索を行っている。

- **Linkfile**
  質問対リンク群は、file1 , dir1 , title1 , word1 , access1 , file2 , dir2 , title2 , word2 , access2 の１０項目のデータ配列に格納され、一つの配列値が1つの質問対リンクとなる。

2.4.2 アンカー生成プログラム

指定されたドキュメントに対し質問対リンクによるアンカーを埋め込む。引数は対象とするドキュメントの URL、出力は与えられたドキュメントのコピーに新しくアンカーを埋め込んだ HTML 文書である。処理の流れは以下の通り。

1. **クライアントから URL で指定され起動する。**
   URL がドキュメント URL を引数としたアンカー生成プログラムの指定となったアンカーをクライアントがクリックすることで起動する。

2. **引数を与えられ、ドキュメントを決定する。**
   引数として与えられた URL をディレクトリパスに変換し、ドキュメントへのパイプをオープンする。

3. **ドキュメントのソースをコピーし配列として取り込む。**

4. **HTML 内の相対パスを絶対パスに変換する。**
   URL はドキュメントの指定を相対パスで記述することができる。しかし、ドキュメントのコピーを外部プログラムで出力すると、カレントディレクトリが変化し相対パスでソースを指定できなくなるため相対パスを絶対パスに変換する。

5. **Linkfike よりそのドキュメントに対して条件を満たす質問対を検索する。**
   Linkfike のすべての query1 に対して検索を行なう。5つの属性は word 属性を除いて値を持つ場合と持たない場合がある。属性が値を持たない場合その属性に対する質問結果はすべてのドキュメントが真となる。5つの属性に対する検索は file,dir,access,title,word の順で行ない、1つ偽を返した時点でその query の結果は偽となる。

6. **質問文が指す単語にアンカーが埋められているか調べる。**
   アンカーのタグがある単語のリンク先を別に保存しておく。その後に、検索の結果、真となった query1 が指す単語に対してアンカー記述を行なうが、その単語が元々アンカーのタグがあるかどうかを判別する。

7. **リンク先、引数を決定し、コピーした配列にアンカーを埋め込む。**

まず、検索の結果、真となったquery1が指す単語をタグで囲む。記述はコピーした配列に対して行なう。タグは、

<A HREF="http://サーバ名/cgi-bin/ リンク探索プログラム?query=配列番号&URL=保存したURL">単語</A>

となり、リンク探索プログラムに配列番号と保存したURLを引数として与えたURLをアンカーとする。

つぎに、リンク先がリンク探索プログラムでないアンカーに対してリンク先をアンカー生成プログラム、引数を元々あったリンク先のドキュメントのURLとして、常に外部プログラムが呼び出されるようにする。

8. クライアントにコピーした配列を返し、終了する。

2.4.3 リンク探索プログラム

指定された質問文よりリンク先を検索し、そのリストを表示する。引数はLinkfileの属性値番号とURL。出力はアンカーを埋め込んだURLのリストのHTML文書。処理の流れは以下の通り。

1. クライアントからアンカーとなる単語を選択され起動する。
   リンク探索プログラムの起動はクライアントがアンカーとなる単語を選択することにより行なわれる。

2. 引数に質問文と元々のリンク先のURLを与えられる。

3. 質問文より該当するドキュメントを検索する。
   指定されたquery2の条件を満たすドキュメントを検索する。属性は、Anchor Make Programと同じくfile,dir,access,title,wordの順で検索し、file,dir,accessを満たすドキュメントを決定した後そのドキュメントに対してtitle,wordでテキストサーチする。

4. 結果をURLのリストとして表示する。
   検索の結果、真となったドキュメントのディレクトリパスをURLに変換し、それをリストとして表示する。その際リストにはリンク先をアンカー生成プログラム、引数をそのURLとしてアンカー記述を行なう。

5. 元々のリンク先のURLがあればリストに加える。
   プログラム起動時に引数URLを与えられていた場合、そのURLもそれと明示してリストに加える。

6. クライアントにリストを表示し、終了する。

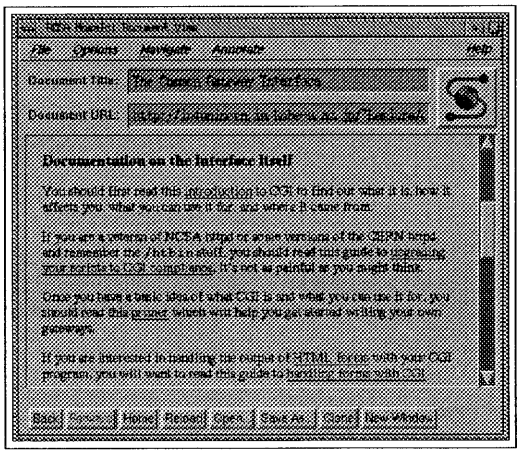

図3: プログラムによるアンカー生成

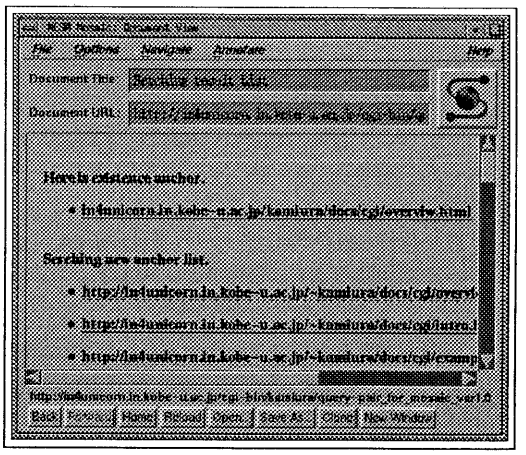

図4: プログラムによるリンク探索の結果

3 Web 文書の段階的構造化

3.1 概要

本節では不完全 Web(Partial Web, 以下 PW と略す) という概念を導入し複数のオーサーが不完全にオーサリングを行なった Web 文書群から、段階的に完全な Web を (仮想的な形で) 構成する機構を提案する。PW は多量の未構造データに対し複数の人間がキーワード付与やリンク付けによるオーサリングを行なう状況において、個々の視点からオーサリングした不完全な Web 文書の集合である。

ここでオーサリングとは、多量のデータに対して複数の人間がリンクによる結合作業を行なう際、オーサーは他人の作業領域を意識せず、個々が自由に自分の作業領域のデータにキーワードやリンクを付けることである。

また不完全なリンクとは二つのデータ間を明示的に結ぶ完全なリンクに対して、二つのアンカー、それを結ぶリンク、のそれぞれが明示的に決定しているのではなく、単なるアンカーのみの指定や、リンク元やリンク先のアンカーが有すべき条件をキーワードなどで指定したものである。これによりリンクを作成するオーサーがこのアンカーからリンクを張りたいがリンク先のアンカーを知らない、しかしリンク先となるアンカーはこの様なキーワードを持つものであって欲しい、と不完全ではあるが、リンクを記述することができる。

この様な不完全で部分的なリンクをそれぞれのオーサーが自由に作成し、そのリンク情報によってハイパーテキストを生成する機構を提供する。

しかし、不完全なリンク情報から生成されたハイパーテキストではオーサーの予期せぬリンクも多数出現し、また冗長なものとなることがある。

そこで、オーサーは自由に部分的なリンクを作成した後、現段階のリンク情報で生成されるハイパーテキストでリンクの張られかたを確認し、必要なリンク不必要なリンクを選別することにより不完全なリンクを確実なリンクへとしていく。この作業を繰り返すことにより多量データのハイパーテキスト化を徐々に行なうという段階的な構造化機構を提案する。

図 5 はいくつかの不完全 Web からハイパーテキストを作成するまでに至る段階を表している。その流れは以下の様となる。

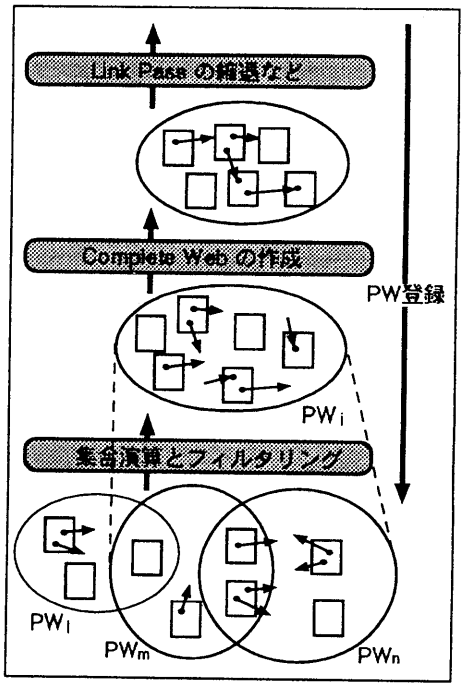

図 5: 段階的構造化の概要

1. オーサーの作業領域の選択とオーサリング。

   図 5 における最下層は不完全 Web 群の全体であり不完全にオーサリングされたドキュメント集合の集合である。オーサリングの作業は主に以下の作業からなる。

   - アンカーの設定。
   - ドキュメントとアンカーのキーワード付け。
   - 不完全リンクの作成。

   リンクの作成はリンクの始点、終点、リンクのキーワード付けによって行なう。始点、終点の指定はアンカー、ドキュメント、キーワード集合のいずれかによって行なう。キーワードで指定した場合、そのキーワード集合を含んでいるアンカーまたはドキュメントが対象となる。

2. 不完全 Web の集合演算及びフィルタリング。

   いくつかの不完全 Web、$PW_1, \ldots, PW_m$ を選び集合和、積、差演算により新たな不完全 Web を生成する。またこの際、アンカーや不完全リンクに付与されたキーワードによって対象となる文書群を選択する。

3. 完全な Web の作成。

不完全リンクから完全なリンクを作成する。一つのアンカーのリンク先が複数となる場合、オーサーは一つのリンク先を選択し、決定することができる。またリンクの縮退などを行ない、リンクの構造を修正する。

3.2 モデル

不完全 Web,PW_iは作業領域内のドキュメントの集合、作業領域内のアンカーの集合、これが持つリンク情報の集合から成り以下のように表す。

$$PW_i = (D_i\ ,\ A_i\ ,\ L_i)$$

- $D_i \subseteq D = \{d_1, d_2, \ldots\}$ はドキュメント全体 D の部分集合。
- $A_i \subseteq A = \{a_1, a_2, \ldots\}$ はドキュメントの一部となるアンカー全体 A の部分集合。
- $L_i = \{l_{i_1}, l_{i_2}, \ldots\}$ は PW_iの持つリンクの集合。

D_iの要素 d_j,A_iの要素 a_jは、それぞれ次のように表す。

$$d_j = (url, keys, anchors)$$
$$a_j = (url, keys)$$

- url

  ソースの指定を行なう識別子。そのアンカーまたはドキュメントとソースをユニークに対応させる。

- keys

  アンカーまたはドキュメントの特徴付けを行なうキーワードの集合。

- anchors

  ドキュメントが含むアンカーの集合。

L_iの要素 l_{i_j}はリンクを表す。リンクは他のアンカやドキュメントの間の関係を結び、ドキュメント間のナビゲーションを可能にする。

そして l_{i_j}はリンクの始点となるアンカーないしドキュメントを指す B_{i_j}、リンクの終点となるアンカーないしドキュメントを指す N_{i_j}、リンクの属性となるキーワードの集合 K_{i_j}を持ち、次のように表される。

$$l_{i_j} = (B_{i_j}, N_{i_j}, K_{i_j})$$

- $B_{i_j} \in D \cup A \cup 2^K$

  B_{i_j}はリンクの始点を指し、その値は D の要素、A の要素、キーワード集合のいずれかとなる。

- $N_{i_j} \in D \cup A \cup 2^K$

  N_{i_j}はリンクの終点を指す。その値は D の要素、A の要素、キーワード集合のいずれかとなる。

- $K_{i_j} \subseteq K$

  K_{i_j}はキーワードの全集合 Kの部分集合でリンク l_{i_j}自身のキーワードである。

3.3 Web 化のための操作

ここでは不完全 Web から完全な Web を作成するまでの操作について述べる。

3.3.1 不完全 Web の集合演算

不完全 Web の集合からまず使用する不完全 Web 群を選択する。複数選択される場合、集合演算によって一つの新しい不完全 Web を生成する。

- 不完全 Web の積は以下のように定義する。

  $$PW_i \cap PW_j = (D_i \cap D_j, A_i \cap A_j, L_i \cap L_j)$$

- 不完全 Web の和は以下のように定義する。

  $$PW_i \cup PW_j = (D_i \cup D_j, A_i \cup A_j, L_i \cup L_j)$$

3.3.2 Partial Web のフィルタリング

集合演算によって得た不完全 Web 内の必要な情報を選択するためにフィルタリングを行なう。これはキーワードの指定によって行なう。キーワードはドキュメント、アンカー、リンクのそれぞれが持っておりその個々について指定する。

3.3.3 不完全 Web の Web 化

集合演算、フィルタリングされた結果得られた不完全 Web のリンク情報からキーワードマッチングでアンカー、ドキュメント間をつなぎ、完全な Web を合成する。このためにリンクの指定がキーワード集合となっている場合、アンカー a、ドキュメント d のキーワードとのマッチングによってリンクの結合を行なう。マッチングが真となるのは以下の場合である（図 6 参照）。

- キーワードとドキュメントのマッチング

  キーワード集合 $\subseteq d.keys$

- キーワードとアンカーのマッチング

  $\{key, \ldots\} \subseteq a.keys$

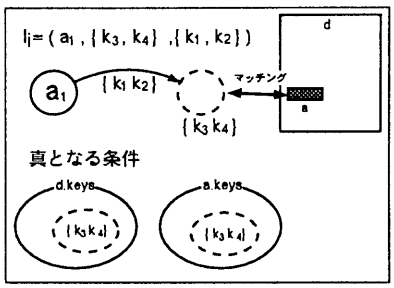

図 6: キーワードのマッチング

3.3.4 リンクパスの縮退

ある不完全リンクのリンク先がキーワードによって指定されており、それにマッチするアンカーまたはドキュメントが存在しないが、別のリンクよりリンク元が同じキーワードで指定されている時、そのリンク先へとリンクを縮退してつなぐ（図7参照)。

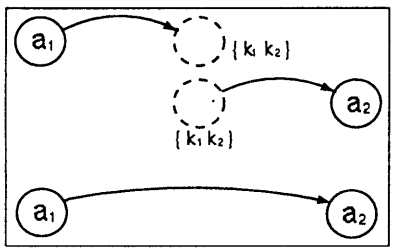

図 7: リンクパスの縮退

3.3.5 条件検索

リンクの結合状態に応じたキーワードによる検索を行なう。図8においてA、B、Cは検索を行なう為のキーワード集合で、それぞれアンカーまたはドキュメントのに対する条件であり、図の右側のような様々なリンク状態を取り出すことができる。

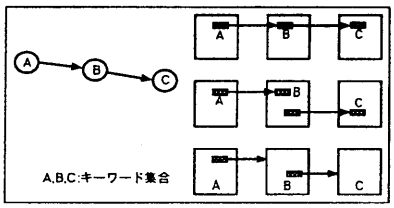

図 8: 条件検索

4 これからの課題

本稿では質問対リンクによるWWWにおける多重ビュー機能の実現、不完全Webモデルによる段階的なオーサリングからのハイパーテキスト生成について述べてきた。今後の課題としては以下の事項が考えられる。

- 仮想リンク機構のproxyによる実現。
- 仮想リンク機構における日本語の扱い。
- ネットワーク上でのURL情報資源に対する質問言語機能の改良。
- 不完全Webの集合演算、フィルタリング、合成機能の実装。

参考文献

[1] 木庭袋 圭祐 益岡 竜介:「World Wide Webによるメディア統合」1994年10月 第100回データベースシステム研究会

[2] K.Tanaka et.al, Query Pairs As Hypertext Links, Proc. of the 7th IEEE Data Engineering Conference, pp.456-463,Kobe,Japan,April 1991

[3] Q.Qian, M.Tanizaki, and K.Tanaka, Abstraction and Inheritance of Hyperlinks in an Object-Oriented Hypertext Database System TextLink/Gem, ADTI'94, Proc. of International Symposium on Advanced Database Techmologies and Their Integration, Nara, Japan, pp.306-313, October 1994.

[4] http://www.ncsa.uiuc.edu/General/Internet/WWW/HTMLPrimer.html