デジタルアーカイブに対する効率的な検索の提案
—神戸大学電子図書館システムを例として—

依田平[¶], 大月一弘[‡,†], 森下淳也[‡], 清光英成[‡]

¶ 神戸大学大学院総合人間科学研究科, ‡ 神戸大学国際文化学部
† 神戸大学付属図書館研究開発室

デジタルアーカイブ（アーカイブス）は，書籍・逐次刊行物・新聞の抜粋・チラシ・パンフレットなど様々な種類の文献を収集したものである．近年では，紙媒体のデータだけでなく，映像・音声やそれらの複合メディアデータを含むアーカイブスも珍しくない．本研究はアーカイブス中の資料から利用者所望の部分を取り出し，アーカイブスのビューを提示することを目的とする．そのため，各資料を論理構造単位（以下部分資料）に分割し，メタデータを与えている．ここで問題となるのが，同一資料中の異なる部分資料に検索対象となるメタデータが散在することである．本稿では，AND 演算子を拡張することにより，アーカイブスに対する効率的な検索方法を提案する．

An effective method to retrieve from Digital Archive
—By way of example to Kobe University Digital Library Archive Search—

Taira Yoda[¶], Kazuhiro Ohtsuki[‡], Jun-ya Morishita[‡], Hidenari Kiyomitsu[‡]

[¶]Graduate School of Cultural Studies and Human Science, Kobe University
[‡]Department of Cross-Cultural Studies, Kobe University

Digital Archive is a collection of materials that contains books, serials, extracts from newspapers, magazines, journals, leaflets and so on. Recently, it begins to collect non-paper media that are continuous media data such that audio, video and their compositions. Our major objective is to retrieve sub-materials from the archives and provide a view of archives to our users. We divide a material to some sub-materials in its logical structure, and give metadata to each sub-resource. Here, we have to resolve a problem that corresponding metadata are scattered around sub-resources in a material. In this paper, we propose augmented AND operations for providing an effective method to retrieve sub-materials from the archives.

1. はじめに

デジタルアーカイブ（アーカイブス）は，様々な種類の文献や複数メディアのデータを統一的に格納・編集・検索できるシステムである．アーカイブスに対する検索は，一つの資料全体を取り出すだけでなく，利用者所望の部分を取り出す機能も必要である[1]．神戸大学電子図書館プロジェクトでは，文献データを文書の論理構造に基づいて分割して格納する方式を採用している[2],[3]．本論文では，現状の神戸大学電子図書館システムの特徴と問題点を整理し，デジタルアーカイブに対する効率的検索手法を提案する．

提案方法においては，分割された部分情報間の包含関係を階層的な木構造グラフによって関連付ける．資料の集合に対して行われる検索は，キーワードに対する検索を個々の部分情報を対象に行った後，階層木に基づいて再構成し評価することで検索結果を決定する．結果の掲示は，検索結果を元に集約し，適切な部分資料を提示する．

一般的な非分割資料に対する検索結果は複数の文献となるが，文献データを論理構造に基づいて格納しておくと，任意の文献の部分を検索結果として得ることが容易になる．さらに我々は，分割された部分資料に対してメタデータを与え，文献の論理構造とは別に扱っている．これにより，一つの文献を木構造表現したとき，各ノードがデータを保持するという特徴を持つ．この格納方式は，キーワードによる情報検索と構造検索に有効な文献データの格納構造といえる．つまり，メタデータに対する検索，構造に対する検索を単体で行うことと，メタデータと構造に対して複合的な検索式を定義することが可能である．また，文献データの構造をあらかじめ定義できなくてもアーカイブスに格納できる利点も有する．

グラフの形状を利用して部分資料に分割された資料を検索する方法としては，田島[4]らのweb情報に対する検索方法がある．田島らの方法は，グラフがネットワーク構造となっているため，2つの資料間の距離をもとに検索を行っているが，本方式は，もともとの資料単位がわかっておりグラフが木構造であることを利用した検索を行う特徴をもつ．

本論文の構成は以下のとおりである．2.ではデジタルアーカイブの構成を述べ，3.ではアーカイブスに対する検索の要件について考察する．4.では神戸大学電子図書館システムの特徴を紹介し問題点を整理する．5,6.では，提案する格納構造に有効な検索方法を議論する．さらに7.で提案検索方法に対する評価を行う．

2. アーカイブの構成

本研究で対象とするアーカイブは，種別やメディアを問わずに網羅的に資料が収容されているものとする．このようなアーカイブにおいて，利用者の欲する情報の単位は，資料そのものでなく，部分資料であることが多いことが考えられる．これは，一つの資料に様々な情報が入っている資料の場合，利用者はその中から特定の情報のみを欲している場合が多いことである．例えば，神戸大学電子図書館が収容しているアーカイブスである「震災文庫」の場合，

- **「火災」**に関する情報が得たい利用者は，書籍やムービーなど資料の種類にこだわらず**「火災」**に関する資料が網羅的にほしい．

といった要求を持つ利用者が多い．このような要求に対しては，資料そのものを結果として利用者に提示するよりも，該当する部分のみを抽出して掲示することが重要となる．このため，資料を細分化する必要が生じる．

一つの資料内に多くの部分資料を含んでいる資料の場合，資料の位置関係による包含関係にのみ基づいて分割される．資料の細分化の例を図1に示す．資料は原則的に，章節構造に基づいて細分化されているが，写真や図といった個別の資料単位となりうる部分も，章節構造と同等に取り扱う．このような細分化方法を取ることで，どのような資料に対しても簡単に細分化を行うことができる．

一つの元資料はpart_ofの関係に基づいて，階層的な木構造によって関係付けられる．図2は図1の資料の細分化をツリーグラフであらわしたものである．各ノードには子ノードに細分化されなかった部分の情報のみが記述される．例えば，図1の①，②の部分に記載されている情報は，図2におけるルートノードに記述される．すなわち，

図1　資料の細分化

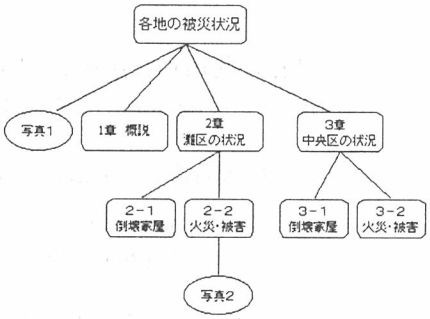

図2　ツリーへの関連付け

リーフノード以外のノードにもそれぞれの部分資料（ルートノードに関しては資料全体）に関する部分情報が記述されている．

ツリーグラフにおいて，部分資料とは，任意のノードを頂点とする部分ツリーで表される．例えば，**「2章」**という部分資料は，図2において**「2章」**を頂点とし，**「2-1」**，**「2-2」**，**「写真2」**からなる部分となる．

アーカイブ資料の構造の特徴として，書物などが定型の構造を持つのに対し，アーカイブスでは多様な資料を取り扱っているために，ツリーの深さも不定であるような構造が非定型な半構造データであることが挙げられる．このため，SGMLやXMLを用いて資料を一つのファイルとして記述することは困難であり，別のデータ記述方法が必要となる．

3. アーカイブス検索

アーカイブに対する部分資料の検索とは，**様々な資料のツリーからなる森から，適切な部分ツリーを選出すること**である，と考えることができる．

本章では，アーカイブをWebなどを用いて一

般に公開する場合の，検索機構に求められる要件について述べる．

3-1. 非定型半構造型資料に対する検索

利用者はアーカイブスから自分の欲する部分資料を抽出する際に，取り扱う資料の多様性から，アーカイブスに含まれている資料の種類にこだわらない，あるいは種類を知らないことが考えられる．また，構造が不定形であるため，XML や SGML のような構造を特定した検索を行うのも困難となる．

このため，部分資料を抽出する問い合わせを行う際に，簡単な問い合わせ指定から利用者の要求を理解し，半構造データの構造を利用して部分資料を選出するための検索手法を確立する必要がある．

3-2. 検索結果の集約

検索結果として得られた部分資料をそのまま提示するのではなく，それら部分資料を集約した形，つまりフィルタリングを行った形で表示したほうが利用者に取って利便を図れる場合がある．これは，一つの資料から多数の部分資料がマッチした場合，すべての部分資料を検索結果一覧として表示すれば検索結果一覧は冗長なものとなるためで，利用者に取っては見やすいように集約した形で表示することが望まれる．例えば，

- 写真集から大量の写真がマッチした場合は，すべての写真ではなく，写真集としてまとめた形で表示する．

3-3. コンテンツの再成

部分資料を利用者に提示するためには，細分化された資料からサブツリーを再構成し，表示コンテンツを生成できる必要がある．また，部分資料から元資料を再構成できることも重要となる．このため，木構造に基づく資料の再構成を行えることが重要となる．

3-4.目標とするシステム

アーカイブスに対する検索の要件をまとめると，**利用者が資料の種別を意識せずに，検索語ならびに簡単な検索条件を指定するだけで，利用者の欲する部分資料を適切な範囲で抽出する**，こととなる．この特徴を実現するためにシステムに必要な条件は，

- 資料の種別を意識せずに検索が行えること．
- 部分資料が取り出せること．
- 適切な範囲を集約した形の部分資料として取り出せること．
- 分割資料を合成することによって，表示コンテンツの生成ができること．

となる．

4. 神戸大学システムの現状と課題

4-1. 神戸大学システムの概略

神戸大学電子図書館ではＷｅｂで公開することを前提にシステム（以下現システム）を作成している．このため，利用者は通常のＷｅｂと同様に，「検索語」と検索語に対する「ブール演算」のみを指定するだけで，

- 部分資料単位で情報を取り出せる．
- フィルタリングにより，部分資料の集約ができる．
- 検出した部分資料から元資料を最構成し，利用者は元資料とその構造がわかる．

といったこと目指したシステムとなっている．これらのことを実現させるため現システムの大きな特徴として，

- 資料をツリーに分割して格納
- 各分割資料に対してメタデータを作成
- 検索機構と表示機構とに分割

していることが挙げられる．

4-2. データ格納形式 [2]

資料の分割は 2 章で述べた part_of の関係に基づいて行い，分割した個別の単位で格納する方法を取っている．格納した単位それぞれに対して DublinCore [5] に準拠したメタデータを作成している．各メタデータは，図 2 におけるノードにそれぞれ対応し，そのノードをルートとする部分資料に対応した情報が書かれている．図 3 に図 2 の**「2-2 火災・被害」**のノードのメタデータを示す．図に示しているように，このノードのメタデータの<TITLE>には，元資料のタイトルではなく，節のタイトルが記載されている．すべてのノードに対してこのように記述することで，すべての部分資料が独立したひとつの資料として取り扱えるようになっている．

```
<KUMETATABL>
  <METAID>00200005</METAID>
  <TOPMID>00200000</TOPMID>
  <LEVEL>00005000_00000003_00000002</LEVEL>
  <TITLE>2-2 火災・被害</TITLE>
  <CREATOR>神戸大学付属図書館</CREATOR>
  <ORGPUBLISHER>神戸大学付属図書館</ORGPUBLISHER>
  <TREE>各地の被害状況@@@2章 灘区の状況@@@2-2 火災
  ・被害</TREE>
  <CONTENTS>～～～～～</CONTENTS>
</KUMETATABL>
```

図 3　メタデータの例（一部省略）

各メタデータには資料間の関係を記述する項目が設けられている．これは図 3 のメタデータに

おける<LEVEL>項目に該当する．図3には，このノードの下位に属する（あるいは，この部分資料に含まれる）「写真 2」に関する記述は記載されていない．しかしながら，「写真 2」の<LEVEL>を用いることで，ノード間の関係を表すツリー構造が再構成でき，部分資料を生成することができる．

このメタデータの構造を一般化すれば次のようになる．

n＝（ID, キーワード$^{+}$, 上位ノード$^{-}$）・・(1)

(1)式における項目は以下に示す．

n：ノード．

ID：各ノード固有の番号．

キーワード$^{+}$：検索語群（タイトル名や著者名などの項目や全文テキストデータを含む）．

上位ノード$^{-}$：同一パス上にある上位ノード．

4-3.検索機構

利用者の問い合わせでは，抽出する部分資料の構造に関する指定はされない．そこで，**検索条件を満たす最小の部分資料を求める結果とする**と定義する．即ち，このできる限り範囲を絞った部分資料を結果として採択する検索は，メタデータに対して行う．検出されたメタデータをルートノードとする部分資料が結果となる．よって検索においては，すべてのメタデータをそれぞれ独立したものとして取り扱い，ツリーに基づく評価は一切行わないものとなっている．

4-4. 表示機構（上位ノード項目の利用）

表示機構では，検索機構によって得られた結果を表示する．図 4(a)，(b)に結果表示の例を示す．結果一覧には，検索に該当したメタデータの内容をもとにリスト形式で表示される．リストから利用者が資料を選択すると，その資料に対する詳細が別画面に表示される．画面の右側にメタデータに記載された詳細情報が示され，左側には，構造をたぐって元資料の構成が表示される．

同一資料内から複数のメタデータが抽出された場合，集約する処理（フィルタリング）を行う．集約は(1)式における「上位ノード」項目を用い，検索機構で得られた結果をツリーに基づいて再評価することで行う．神戸大学では，集約の評価として次のルールを採用している．

集約ルール：　同一検索で得られた 2 つの部分資料が包含関係にあった場合，包含するほうのみを表示する．□

4-5. AND 検索への対応

個々のメタデータを完全に独立したものとして

図 4(a)　検索結果一覧

図 4(b)　データ詳細表示

取り扱い AND 検索を行った場合，資料の分割のために，同一の資料に含まれるキーワードが共起しなくなるという問題が生じる．

神戸大学ではこの問題に対応するため，本来下位の部分資料に記述されていない情報でも，下位の部分資料にとって重要な意味を持つと思われる項目は，下位の部分資料のメタデータに記述する方法を取っている．図 3 における<TREE>項目がこれに該当する．<TREE>には，上位ノードのタイトルがルートノードから順に記述されている．**「灘区の火災」**に関する資料を得るためにキーワードを**‘灘区’and‘火災’**と指定すれば，このメタデータから<TREE>に記述されている**「灘区」**と**「火災」**がマッチするため，この部分資料を AND 検索の結果として得ることができる．

5. AND 検索の検討

5-1. 部分資料に対する AND 検索

現システムの AND 演算は，資料を表すツリーグラフにおいて，「２語が同一パス上に存在し，かつ，上位ノードに付与されているキーワードを下位ノードにも記述している場合」のみしか検索できず，一般に利用者が想定する，「ひとつの資

料内に2語が共出現するものを探す」検索は行えないものとなっている.

そこで，我々は，部分資料に対するAND検索を，一般的なAND検索の概念と同じ意味になるように次のように定義する.

AND検索： 2語が共出現する最小の部分資料を探すこと. □

この検索は，図5に書かれた手順で実現される. 利用者の入力に基づき，まずメタデータに対して個別のキーワード検索を行い，それぞれから得られた結果を，木構造にマッピングし評価を行う. 木構造に照らし合わせて 2 語を含む最小のサブツリーのルートノードを検索機構の結果とし，これを表示機構に渡す.

このツリーに基づいて再評価する方式では，メタデータに記載されている<TREE>項目などの上位ノードの情報のコピーは省いても良い.

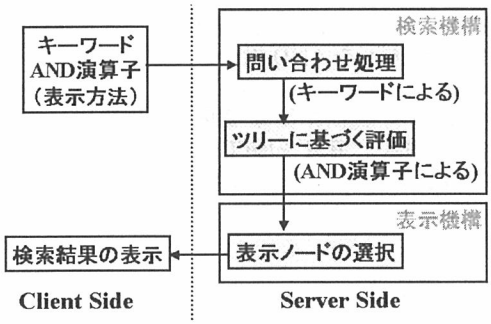

図5 利用者の入力と検索処理

5-2. AND演算の意味の違い

資料に共出現する2語の関係は大きく，

- 2語が関係を持って使われている.（この関係には単語の性質によって差異がある.）
- 2語に関係はなく，偶然一つの資料に使われている.

の2つの場合に分けることができることが指摘されている[6].

例えば，利用者が**‘灘区’and‘写真’，‘倒壊家屋’and‘火災・被害’**という2種類のAND検索を行ったとする. この場合，最初の問い合わせにおける利用者の意図は，**「灘区に関する写真が欲しい」**ことであることが想定でき，**「灘区」**と**「写真」**が関係をもって使われていると考えることができる. これに対し，後者は，**「倒壊家屋と火災・被害に関する資料がみたい」**と利用者が欲していると想定でき，偶然一つの資料に両語が使われている資料を欲していると考えることができる.

通常のAND検索は，ひとつの資料に対して行うものであるため，構文解析や文章の位置関係の解析などを行わない限り，これらの違いは区別されない[6]. しかし，part_of関係によって分割された資料の，ツリーグラフ上の2語間の位置関係を利用すれば，ANDに対する意味の違いをある程度区別することができる.

部分資料間の関係には，ある部分資料に含まれているものと含まれていないもの，言い換えれば同一パス上にあるものとないものが存在するが，2語が関係を持って使われている場合は，2語が同一パス上に存在する可能性が高いと考える. つまり**「灘区の写真」**は，**「灘区で分類された中の写真という項目」**か**「写真という項目の中の灘区という項目」**にある可能性が高い. これに対し，**「倒壊家屋」**と**「火災・被害」**の資料は，図2に示すように同一パス上よりも併記あるいは，資料の別々場所に記載されている可能性が高いと考えられる. 図2の資料には**「中央区」**と**「写真」**という語を含んでいるが，同一パス上には存在していない. **‘中央区’and‘写真’**という検索を行う人の意図が**「中央区の写真」**であった場合，この資料は利用者の欲するで資料ある可能性が低いと想像することができる.

5.3 拡張アンド演算

これらのことから，利用者がキーワードの関係と，その関係はツリーではどのように表現されるのか（ノード間の相対関係はどのようになっているのか）を考慮すれば，自分の意図を反映した検索が行えることとなる. そのため，我々はAND検索を拡張し，5-1.で述べたAND検索（これを*標準AND*と呼ぶ）のほかに，3つの*AND演算子*を用意した.

1) *直列AND*：

キーワードを含むノード間の関係が，包含関係になっている（直列の）組み合わせを選択する. また，直列ANDにはキーワードが同一ノードに共出現する場合も含むものとする.

デジタルアーカイブにおいては，ノード間の包含関係が逆転している場合もよくあると考えられるため，*直列AND*ではキーワードを含むノードの上下関係は問わないものとする. 例えば，資料には**「灘」**で分類された中に**「写真」**が入っている場合と，**「写真」**で分類された中に**「灘」**が入っている場合が考えられる. **「灘の写真」**を欲している利用者にとって，この分類の順序は関係ない. すなわち，ノード間の上下関係に意味はなくなっており，上下関係を問うことは情報を得損なうことにつながるため，*直列AND*ではそれは問わないものとする.

2) *並列 AND*：

キーワードを含むノード間の関係が包含関係になっていない(並列の)組み合わせを選択する.

3) *兄弟 AND*：

並列 AND の特殊例で，2つのノードが共通の親ノードを持つ組み合わせを選択する.

一般に資料に共出現する2語が包含関係になっていないノードの組み合わせに記述されていた場合，それら2語に関係のない場合が多い．そのため，*並列 AND* が意味を持つのは，ツリーにおいて両者の距離が最も近い共通の親ノードを持つ場合であると考え，*並列 AND* については距離を考慮した*兄弟 AND* を用意した.

5-4. 資料内に同一語が複数出現する場合の処理

標準 AND あるいは*並列 AND* 検索において，ツリー中に検索する 2 語を含んだノードが多数ある場合，どのようにサブツリーを決定するのかが問題となる.

原則的には，任意の2語のペアにすべてに対して，それぞれ結果を得ることになる．ここで，図6のように，キーワードを含むノードが2箇所に分かれて部分的に集中している場合を考える．すべてのサブツリーを検索結果とすると，距離の離れた 2 語をペアとして関連性の低い組み合わせも取ることとなる．この関連性の低い組み合わせを検索結果とした場合，◎のついたノードが検索結果となる.

利用者に部分資料を提供するためには，可能な限り小さなサブツリーを選択したほうが良いため，本システムでは，この関連性の低い組み合わせを取ることは有用でないと考える．そこで，キーワードを含むノードが多数存在している場合は，自分に一番近いノードを自分の相手とし，可能な限り小さなサブツリーを選択することとする．図6の場合，色のついたノードをルートとするサブツリーが結果となる.

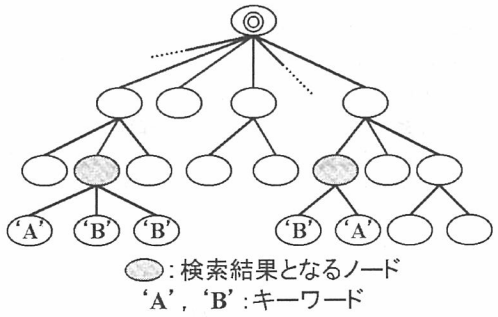

図6　同一語が複数出現する場合の処理

5-5. 絞込み検索

一般の検索においては，絞込み検索と AND 検索は同じ結果を得るが，部分資料に対する絞込み検索では，意味が異なる．部分資料に対する検索において，絞込みとは，「ある検索語（'A' とする）で検索した部分資料に対し，さらに別の検索語（'B' とする）を含む部分を探す.」という意味になる．即ち，部分資料に対してさらに部分の範囲を絞り込む場合に使われる．図7ように，絞込みを行うと，結果の件数が増える場合もある.（この場合でも，資料の範囲は絞り込まれていることに注意して欲しい）.

また，絞りこみを行う順番（'A' で検索した後 'B' で絞り込むのか，その逆か)で結果が異なる.

絞込み検索は，直列 AND 検索において 2 語の上下関係を指定し，かつ，同一パス上にある 2 語のペアノードの下位のノードを頂点とするサブツリーを検索することと同等になる．これは，上位ノードにある語の意味が下位ノードに継承していると考え，2 語の意味を持つノード（あるいは部分資料）を AND 検索したことになる．図2において **'灘区'** と **'写真'** で検索を行う場合，直列 AND の結果は **「2 章 灘区」** のサブツリーとなるが，**'灘区'**，**'写真'** の順で絞り込んだ場合は，**「写真 2」** のリーフノードが結果として得られることになる.

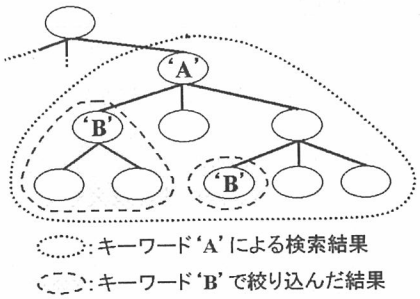

図7　絞り込み検索

6. 表示機構の検討

6-1. 表示機構の改善

集約は，ノード間の相対関係に基づいて行われる．これは，密度の高い部分資料を選出するという点においては，相対関係に着目することが有効であることと，構造が不定なことより利用者が適切な絶対位置の指定を行えないと想定されることによる.

現システムの集約方式では，1つの元資料から複数の部分資料が得られたとしても，それらが包

含関係になっていない場合は，集約されない．このような場合でも検索結果を集約できるように，表示機構を改善する．

集約ルール(ボトムアップ集約)：　ツリーグラフにおいて，あるノード（Aとする）に隣接する下位のノードのうち検索条件を満たしていたノード（部分資料）の数（あるいは，比率）が一定以上であった場合，Aが検索条件を満たすものと見なし，Aのみを検索結果とする．□

上記ルールを，最も下位のノード（葉ノード）から順次上位側に適応し，結果とするノードを決定する．この集約方式をボトムアップ方式と呼ぶ．これに対し現システムで採用している集約方式をトップダウン方式と呼ぶことにする．

6-2. 複合方式

トップダウン方式は，ある部分資料が検索結果として表示されるとき，それに含まれる部分資料が表示されることを防ぐ表示方法である．是に対し，ボトムアップ方式は，独立した部分資料が同一資料内の近隣部分に多数存在する場合に，それらを個別に結果表示されることを防ぐ方式である．この両方のルールを同時に適応すれば，より効果的な集約ができると考える．

この方式を複合方式と呼ぶ．

複合方式は，それぞれの集約ルールを適用する順序によって表示結果が異なるが，ここではボトムアップ －トップダウン手順を複合方式に採用する．

7. 評価実験

提案方式について，プロトタイプのシステムを作成し実験を行った．データは，2000年11月現在の神戸大学電子図書館のメタデータを用いた．データ数は，元資料数が 40429 件，ノード数が 97555 件，部分資料への分割が行われている資料数が 3092 件である．なお，メタデータに記載されている<TREE>項目などの上位ノードの情報のコピーは省いた．

今回行った評価実験は以下の2つである．

1)利用者がキーワードと *AND 演算子*を指定することで意図する部分が抽出できているか？

2)複合方式によって検索結果一覧が，「適当な数」かつ「適当な部分」なものとなっているか？

7-1. AND 検索の評価

いくつかのキーワードのペアに対してAND検索の評価を行った．表1に**‘明石’and‘被害状況’**，**‘芦屋’and‘被害状況’**，**‘明石’and‘芦屋’**の実験結果を示す．なお，AND検索の評価においては，拡張AND検索がどの程度部分資料を選出できるのかを確認するために，集約は行っていない．

表1に示すように，**「明石」**を含むノード数は421件，**「芦屋」**を含むノード数は983件，**「被害状況」**を含むノード数は273件であった．表1の項目5は指定したキーワードを含んでいる資料の数，項目6は指定したキーワードを共に含んでいるノードの数で，両者を比較すればノードのみに対する検索では，本来該当されるべき資料をかなり得損ねていることがわかる．項目7は*標準AND*で得られた部分資料数を表しており，項目5との比較をすれば，検索結果がいくつかの部分資料に分かれて抽出されていることがわかる．

利用者の意図が**「明石の被害状況」**，**「芦屋の被害状況」**，**「明石と芦屋に関する資料」**であった場合に，拡張AND検索がどの程度有効かを検証した．

表1　拡張AND演算の検索結果の違い

1	単語A	明石	芦屋	明石
2	単語B	被害状況	被害状況	芦屋
3	単語A (N)	421	983	421
4	単語B (N)	273	273	983
5	A and B (M)	14	30	53
6	A and B (N)	2	12	23
7	標準AND(N)	23	56	89
8	直列AND(N)	3	14	24
9	並列AND(N)	20	43	65
10	兄弟AND(N)	6	15	34

(N)：ノード数　(M)：資料数

「明石の被害状況」に関しては，表1の**‘明石’and‘被害状況’**で*直列AND*の項目を見れば，検索結果は3件となっている．結果の資料を実際に確認したところ，同一ノードに記述されていたものが2件，直列に並んでいるノードのペアに記述されていたものが1件であった．これらはすべて**「明石の被害状況」**に関する資料，例えば「兵庫県南部地震による震災の記録」の**「被害状況」**に含まれる**「明石市にある附属幼稚園の東塀が倒壊」**，であった．また，*並列AND*および*兄弟AND*の結果を確認したところ，*直列AND*によっても得られるものを除いて，これらは**「明石の被害状況」**とは関係のない資料であった．これらのことは**「芦屋の被害状況」**に関しても同様で，この結果，*直列AND*は関係を持つ2語を調べる時に有効であることが確認できた．

'明石' and '芦屋' で*兄弟 AND* の項目を見れば，検索結果は 34 件であった．これらの資料を確認したところ，「**街の再生へ決意新たに ： 自治体首長にきく**」の資料中に「**明石市長の声**」，「**芦屋市長の声**」といった同じ内容を併記したような資料を得やすいことが分かり，同じ観点からまとめた資料を探すときに有効であると考えられる．これに対して*標準 AND* や*並列 AND* の結果には両者の内容に関係のないものが多かった．しかし，利用者によっては，資料に偶然出現するキーワードの組み合わせを指定する場合，例えば，利用者が資料を過去に見たことがあり，記憶を元に検索を行う場合などにはこれらの演算が必要とされる．

このように 4 種類の AND 演算を上手く使い分けることによって，利用者の意図を反映させた結果を得ることが可能となる．

7-2. 集約方式の評価

次にキーワードを固定し，表示方法を変えることによる検索結果一覧個数の変化から，集約方式の検証を行った．この実験において，ボトムアップ方式の条件は，隣接する下位ノードが 2 個以上マッチしている場合とした．

表 2 に表示方法を変更することによる，結果表示数の差を示す．（なお，キーワード **'六甲台'**，**'被害状況'**，**'予知'** を持つ資料数はそれぞれ 60 件,155 件，127 件であったが，この内キーワードを含んだノードが一つしか検出されない資料は，集約に関係がないため検証対象から省いた．）

キーワードを **'六甲台'** とした場合を検証する．現システムで採用されているトップダウン方式のみを適用した場合，10 件の資料に対する表示が 124 件となり，あまり集約できていない．これは，**'六甲台'** を葉ノードのみ（98 件）に含んでいる写真集があったためである．これに対し，ボトムアップ方式での集約は表示件数が 24 件となり，かなり集約された結果となった．

複合方式では 17 件となり，この 17 件を実際に確認したところ，資料そのものまで集約されたものは 3 件，部分として表示されていたのは 14 件であり，部分が適度に取り出せていることが分かる．

また，資料そのものまでに集約された 3 件の中には，ボトムアップの条件設定を集約の度合が高くなるように設定したことに起因するものがある．表示数を減らすことよりも，密度の高い部分を表示しようと意図するならば，条件設定を調整することによって可能となる．

'被害状況' や **'予知'** をキーワードとした場合も，程度の差こそあれ同様の傾向であり，複合方式の有効性を確認することができた．

表 2 表示方法の違いによる結果表示数の差

1	キーワード	六甲台	被害状況	予知
2	資料数	10	21	25
3	ノード数	161	139	118
4	トップダウン	124	102	59
5	ボトムアップ	24	60	48
6	複合	17	42	41

複数ノードがマッチした資料のみを対象

8. まとめ

本稿では，デジタルアーカイブに対する効率的な検索方法として，部分資料間の相対関係を有効に使う方式の提案を行い，その有効性を検証した．

今後は，3 語以上入力した場合の処理の方法や，表示機構における適切なパラメータの設定等を研究する予定である．

謝辞

本研究にあたり，アーカイブデータを提供していただいた神戸大学附属図書館，ならびに震災文庫システムについて快く解説していただいた同図書館の渡邊隆弘氏に深く感謝いたします．

参考文献

[1] 依田平，小椋正道，大月一弘，森下淳也，清光英成:「電子図書館用デジタルアーカイブの検索方法の検討」，情報処理学会研究報告 70 号，2001，p469-476.

[2] 渡邊 隆弘:「震災アーカイブにおけるメタデータの設計」，情報処理学会シンポジウムシリーズ 17 号 人文科学とコンピュータシンポジウム論文集，2000，p89-96.

[3] 渡邊隆弘:「震災文庫」のこれまでとこれから－電子図書館を中心に，Academic Resource Guide 55 号，2 2000.

[4] K. Tajima,K. Hatano,T. Matukura,R. Sano, K. Tanaka:Discovery and Retrieval of Logical Information Units in Web,(invited) Proc. of WOWS, (in conj. with ACM DL'99), Berkeley, CA, Aug. 1999, pp 13-23.

[5] http://dublincore.org/

[6] 山本昭:「ブール検索における and の使用法と意味論 －共出現の諸ケースと，検索者側での対応」，情報の科学と技術 50 巻 10 号，2000，p.501.