

電子図書館用デジタルアーカイブの検索方法の検討

依田平[¶], 小椋正道[†], 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, †京都大学原子炉実験所, ‡神戸大学国際文化学部

本研究では、デジタルアーカイブのコンテンツサービスとして、利用者の欲する部分資料を抽出する手法を検討する。利用者に部分資料を提供するために、資料を分割しそれぞれを独立の資料として取り扱う方法がある。しかし、部分資料を独立の資料として取り扱った場合、AND 検索の際にキーワードの共起が発生しなくなる問題と、表示の際に結果が冗長になる問題が生じる。我々は、資料の分割を位置関係に基づいた包含関係で表し、個々の部分資料を階層的な木構造に関連付け、部分資料に対して検索を行った結果を木構造に基づいて評価することにより、キーワードの共起の問題を解消し、表示に関する冗長性の問題を対処する手法を提案する。さらにツリーの形状を利用することにより、キーワードの関係の違いに対応した細やかな AND 検索を提案する。

A framework for retrieving contents from digital archives.

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

[¶]Graduate School of Cultural Studies and Human Science, Kobe University

[†]Research Reactor Institute, Kyoto University

[‡]Department of Cross Cultural Studies, Kobe University

In this paper, we discuss about the approach to retrieving contents from a digital library archive. To provide the contents of digital archive, the original data can be divided into meaningful sub portions, each of which is treated as an individual data. However dividing portions prevents the co-occurrence of keywords in case of *AND* query which should be matched for the original whole data and often makes the query results redundant. Because, the conventional *AND* query returns contents which include all keywords. We introduce a model that the portions are reorganized as a tree, which is equivalent to the original data. A node of the tree contains one of the portions and the connected nodes are related with a part-of relationship to reconstruct the original data. (So the root node of the tree represents the original data and implicitly includes all the portion.) We arrange the compound query of the tree so as that the result of a query applied to the portions is also evaluated by examining the structure of the tree. Such a compound query avoids the absence of the co-occurrence of keywords and can improve the redundancy of the results. We also propose some extended *AND* queries which depend on the relation of the keywords by examining the structure of the tree.

1. はじめに

電子図書館では、従来の蔵書情報、所在情報、書物の全文情報提供サービスのみならず、様々な形態の資料を提供するデジタルアーカイブサービスが進められている。デジタルアーカイブにおいては、種別やメディアを問わずに網羅的に資料を収容しているものも多い。例えば、神戸大学付属図書館が WWW を利用して公開している「震災文庫」においては、阪神淡路大震災に関する様々な資料を公開することを目的とし、書籍、逐次刊行物、新聞からの抜粋、チラシ、パンフレ

ット、レジュメ、といった種別の異なるものや、一枚もの写真、地図、表、ムービーといったメディアの異なるものまでを含んだアーカイブとなっている[1]。このため、単一の写真や新聞の切り抜きのようにひとつの資料単位が非常に小さいものから、書籍のようにひとつの資料の中に独立した単位として取り扱うことのできる資料を部分資料として多く包含しているまでが混在したアーカイブとなっている。

このようなアーカイブの利用の特徴として、利用者がアーカイブから自分の欲する情報を抽出する際に、
・格納されている資料の種別にこだわらない、あるいは

は、種別を知らない

- ・ひとつの資料をそのものでなく、資料に含まれる部分資料のみを抽出したい

といった要求が多いといった点が挙げられる。

部分資料を抽出するための方法として、元資料を部分資料に分割し、それぞれの部分資料を検索対象として検索する手法がある。しかしながら、個々の部分資料を完全に独立したものとして取り扱い 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。また、単一キーワードによる検索（単純検索）の場合にも、同一資料内の部分資料が多く抽出され、検索結果の数が増大してしまうことも考えられる。

本稿では、このような問題を解決するための検索方法を提案する。提案方法においては、個々の部分情報の包含関係を階層的な木構造によって関連付ける。資料の集合に対して行われる AND 検索に関して、個々の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。結果の掲示に関するユーザ選択肢を準備し、同一の資料に複数の結果がある場合の表示内容を切り替える。

部分資料の包含関係に注目した場合、利用者の欲する AND の意図には大きく 2 種類のものがあると考えられることができる[3]。2つのキーワードが木構造上で直列関係（ルートからの同一パス上にそれぞれのキーワードを含む部分資料が存在する）にあるものを欲している場合と、並列関係にあるものを欲している場合である。利用者のこのような検索の意図を反映できるようにするために、木の形状検索とキーワード検索とが複合されるような、拡張 AND 演算子をいくつか準備する。即ち、本方式においては、利用者は、キーワード、拡張 AND 演算子を用いた条件式、結果表示の選択を行うことによりアーカイブに対する検索を行う。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。

本方式は、資料あるいは部分資料に対して、木構造の接続関係を表す情報を付与する以外は特に規定しない。また、各資料を部分資料に分割する場合にも、特別な制約はない。このため、アーカイブ作成に際して、データの統一性や一貫性を保つためのデータスキーマの制約を設ける必要もなく、大量のデータを人海戦術

でアーカイブにする場合の、作業者の負担を軽減できるなど実用性も高いと考えられる。

2. デジタルアーカイブの構成

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。個々の資料は、文書論理構造などに基づいて分割され、細分化される。資料の細分化の例を図 1 に示す。



図 1 資料の細分化

元々の資料の構成は、部分情報の包含関係を保持する階層的な木構造によって関係付ける。図 2 は図 1 の資料の細分化をツリーで表したものである。ひとつの資料はひとつのツリーを構成する。ツリー中の各ノードは、それぞれ部分資料に対応する。図 1 において、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いツリーに収容する(図 2)。即ち、資料の位置関係による包含関係に基づいたものである。これは part_of の関係である。各ノードには、部分資料のデータならびにノード間の隣接関係を表す情報のみが付与される。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。また、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。例えば、図 1 の、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。このため、ツリーグラフ中のすべてのノードが検索対象となる。



図2 ツリーへの関連付け

3. 利用者の要求

デジタルアーカイブを活用する利用者は、その格納形式や論理構造を熟知している必要はない。そのため複雑な構造を持つデジタルアーカイブに対する利用者の素朴な問い合わせが、どのように評価されるべきか検討しておく必要がある。検討すべきものを大別すると、

- ・ 検索意図の解釈
- ・ 検索結果の表示

の2つが挙げられる。

3-1. 検索意図の解釈

デジタルアーカイブに対する典型的な検索の例の一つとして、「灘地区の火災状況を調べたい」というような検索がある。この場合、利用者の要求は、本などある特定の資料を探すことではなく、灘地区の被災状況が記載されている部分のみを探すことであると考えられる。

1) 単純検索

利用者が「灘」をキーワードとして、アーカイブに対して検索を行う。「灘」に関する部分資料は、書籍中に「灘」に関することが記述されている節、「灘」で撮影された写真、「灘」という項目がある「各地の予想される被害総額」という表等、あらゆるものが考えられる。そのため、「灘」というキーワードを含む部分資料の全てを結果として返すのが望ましい。

2) AND 検索

次に、「灘」AND「火災」で検索を行った場合を考える。このとき、単にノードのみに AND 検索を行っ

ても、一つのノード中にキーワードが共出現するものについてはマッチすることができるものの、複数のノード間にキーワードが分散してしまったものについては、マッチすることができなくなっている。

単にノードに対してのみ検索を行うだけで情報を取りこぼすのは利用者にとって都合が悪く、AND 検索に関しては何らかの形で複数のノードをまとめたものに対して行う必要が生じる。そこで、個々の部分情報は木構造に関連付けられることを利用し、個々の検索結果を木構造に適用することで検索対象となる資料単位の拡大をはかり、検索の正しい評価と部分資料の発見を実現する。

3) 拡張 AND 検索

また、木構造の形状を利用することにより、より利用者の意図に応じた検索を行うことができると考えられる。「灘地区の火災状況」を調べたい利用者にとって、位置関係が包含関係になっているノードが要求するものと一致する可能性が高い。図1の例では利用者の要求するものは、「2章 灘区の様相」に含まれる「2-2 火災・被害」であって、「3-2 火災・被害」ではない。そこで2つのキーワードが包含関係にある場合は、包含関係になっているノード集合のみを検索できる拡張された AND 検索を用意すれば利用者にとってより利便性の図ることができる。

3-2. 検索結果の表示

デジタルアーカイブの検索結果は、的確な部分資料が得られるものでなければならない。しかしながら、利用者がどのような形態の資料を欲しているかによって、検索結果の表示は異なるため、表示にはいくつかの選択肢が必要となる。

ある資料において、キーワード「灘」を持つ部分資料があるノードの子ノードとして、多数見出されたでしょう。「灘」に関することが記述されている資料そのものを探している利用者にとっては、資料そのものを表示すれば良い。次に部分情報を欲している利用者を考える。このとき最も簡単な検索結果の表示は、マッチしたノードをすべて表示することである。しかし、資料を分割することにより独立した資料数が増えてしまったために、検索によって得られたものすべてを表示するのは、検索結果の数が増え表示結果が見にくいものとなる可能性がある。さて、「灘」を持つ部分資料があるノードの子ノードとして多数ある場合、「灘」というキーワードがその親ノードに代表される資料全体にとって大きな意味を持つ可能性が高い。そのため、この親ノードを代表として表示しても、適切な部分資

料を提供するという意味では、子ノードすべてを表示することと同等であると捉えることができる。

これらのことから、同一ツリー内において部分資料がマッチした件数に応じて様々な表示方法が考えられるため、利用者の意図や要求に応じて適切な結果を掲示することができれば利用者にとってより利便性を図ることができる。

4. ツリーに対する AND 検索

ここでは、資料の集合に対して行われる AND 検索に関して、個々のキーワード毎に個々の部分資料に対して検索を行った結果を、ツリーに基づいて再評価する手法を提案する。

ツリーを構成するノードの集合の中に、2つのキーワードが含まれるかどうかを判定することで、ツリーに対する AND 演算を行う。これより、キーワードの共起の問題を解消することができる。

またツリーに対する問合せの際に、キーワードを含んだノードで構成される特定の形状のサブツリーを指定することにより、3.章で述べたキーワード間の包含関係のような AND の意味の違いに対応した検索を行うことができる。そこで、形状の異なるサブツリーに対応した AND 演算子を複数用意して AND 検索を拡張する。

4-1. AND 演算子

AND 演算子は指定した2つのキーワードを含むツリーに対して、キーワードを含む2つのノード(ノード対)から構成される最小サブツリーを抽出する。ここでは、結果のサブツリーを便宜的にサブツリーのルートノードで表し、選出されたノードが重複した場合、重複は数えないこととする。

用意した AND 演算子は次の4つである。ここでは、キーワード「A」、「B」で検索した例に基づいて話を進める。

) 親戚 AND

ツリー内のノードあるいはノード対がキーワードを含むときを親戚 AND とする。これは、一つのツリーに対する一般的な検索方法である。

図3に親戚 AND の例を示す。ここより示す図において、数字はノードの通し番号、楕円はキーワードを含まないノード、四角はいずれかのキーワードを含むノードでノード内に記述している文字が含まれる文字色の付いているノードは抽出されたサブツリーのルートノードとする。

図3の例では、キーワード「A」でノード、キーワード「B」でノードがマッチしている。この2つのノードによって作られる最小サブツリーのルートノードは、ノードである。そのため、結果となるノードはノードとなる。

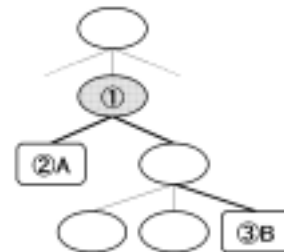


図3 親戚 AND

) 兄弟 AND

共通の親ノードを持つノード対がそれぞれキーワードを含むときを兄弟 AND とする。

図4の例では、キーワード「A」でノード、ノード、キーワード「B」でノードがマッチしている。このとき兄弟 AND の条件を満たしているノード対は、ノード{、}であり、同じキーワードでマッチしているノード{、}ではない。結果となるノードは、ノード{、}の親ノードとなるノードとなる。

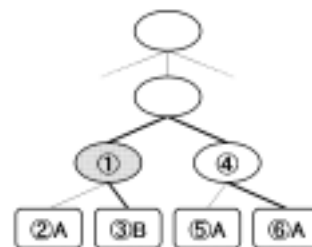


図4 兄弟 AND

) 直列 AND

一つのパス上のノード対がキーワードを含むときを直列 AND 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

のされかたは関係ない。そのため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 では問わないものとする。

結果として返すノードは、ノード間の包含関係から、より絞り込まれた部分を表しているのは下位ノードであるため、含まれるキーワードに関わらず2つのノードの内、下位ノードとする。

図5の例では、キーワード「A」でノード ①、②、③、④がマッチしている。このとき直列 AND の条件を満たしているノード対はノード{ ①, ② }、ノード{ ①, ③ }であり、同じキーワードでマッチしているノード{ ②, ③ }ではない。結果となるノードは、ノード ①、ノード ④となる。

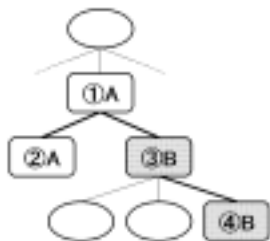


図5 直列 AND

）直列隣接（親子）AND

親子 AND は直列 AND においてノード間の距離が1のときに成り立つ。

図5の例では、ノード{ ①, ② }の対は直列であるものの距離が2であるため、親子 AND の条件を満たさない。そのため、該当するノード対はノード{ ①, ③ }のみであり、結果となるのはノード ① となる。

4-2. AND の多数系

ツリー中にどちらかまたはどちらのキーワードを含んだノードが多数ある場合について考える。ここで、いずれの AND 検索を行うかは、先に選択されているものとする。各ノードについて、自分にとって距離が一番近いものを自分のペアとし、そのペアとの間のできるサブツリーのルートノード（直列・親子 AND では下位ノード）を結果とする。また、距離が一番近いものが複数ある場合、そのいずれをも結果とする。

5. 結果の表示

結果を表示する際に、検索によって得られた結果を利用者にとってより要求を満たした結果として表示するために、もう一度ツリーに基づいて再評価する。評

価方法は複数用意し、利用者が要求に応じて選択できるようにした。

表示方法は検索によって得られた結果が煩雑だった場合、その検索結果をツリーに基づいて集約し、利用者に対して密度の高い情報を提供することを主な目的としている。特に、ツリー内の一部にマッチしたノードが集中していた場合に対し、より効果的に集約が行えることに重点をおいた。そのため用意した表示方法は、検索結果すべてを表示する、ルートノードを表示する、部分資料を集約できるもの5種類の計7種類である。

また、これらの表示方法は、条件を満たしたノード集合に適用され、条件を満たさなかったノード集合については、そのノード集合を検索結果として表示する。

）上方優先方式

検索によって得られたノード集合に、あるノードとその下位にあたるノードがあったとき、上位ノードのみを表示して、下位ノードは表示しない方法である。

図6に上方優先方式の例を示す。ここより示す図において、数字はノードの通し番号、楕円は検索結果として選出されなかったノード、四角は検索によって選出されたノード、色の付いているノードは最終的に結果として表示されるノードである。図6の例では、検索でマッチしたノードは ①, ②, ③, ④ であるが、ノード ② はノード ① の下位ノードにあたるため表示されず、検索結果として利用者に表示されるノードは、ノード ①, ノード ④ となる。



図6 上方優先方式

）単純下方全一致方式

単純下方全一致方式では、ある任意のノードの子ノード（孫以下のノードは含まない）について、子ノード群がすべて検索結果として選出されている場合は、選出されたノードすべてを表示するのではなく親ノード一つを代表として表示する。

図7の例では、検索でマッチしたノードは、ノード ①, ②, ③, ④ である。ここで、ノード ①, ②, ③, ④

はノードの子ノードにあたり、ノードから見て子ノードすべてがマッチしていることになる。そのため、表示されるノードはノードとなる。その結果利用者に表示されるノードは、ノード、となる。

下方全一致方式

これは単純下方全一致方式を可能な限り繰り返し行う方法である。

図7の例で見れば、単純一度下方全一致方式を適用した結果、得られたノードは、ノード、であった。ここで、ノード、はノードの子ノードに当たり、ノードから見ればあたかも自分のすべての子ノードがマッチしているように見える。そこでもう一度、単純下方全一致方式を行い、結果としてノード、を表示する。この作業を単純下方全一致方式が適用できなくなるまで行い、利用者に表示するノードを決定する。

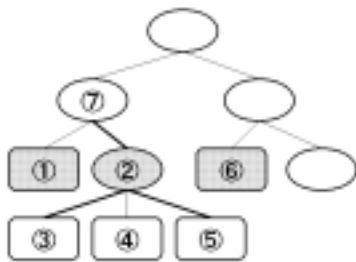


図7 単純下方全一致方式

単純下方部分一致方式、下方部分一致方式

(単純)下方全一致方式では、あるノードの子ノードすべてが検索結果として選出されていたとき、選択されているノードの親ノードを表示する方法であった。(単純)下方部分一致方式は、子ノードがすべて選出されている必要はなく、条件に指定されている割合以上に選出されていた場合、それら子ノードの親ノードを代表として表示する方法である。

6. 絞り込み検索

我々の取り扱っているデジタルアーカイブにおいての絞り込み検索は、一般に想像される絞り込み検索とは異なる。

最初の検索における結果表示はノード単位で表示されるものの、表示されたノードはそのノードをルートノードとするサブツリーを構成することができる。このとき、絞り込む対象は一般に、表示されたノードだ

けではなく、そのノードから構成されるサブツリーであると考えられる。そこで、絞り込み検索を行う対象をサブツリーとした場合、

- ・絞り込み検索で表示される結果が増えることがある
 - ・「A 絞り込み B」と「B 絞り込み A」で結果が異なる
 - ・表示方法の指定の仕方で結果が異なる
 - ・AND 検索 絞り込み検索
- ということが生じる。

我々の行っている検索は、ノードという点を求める検索ではなく、サブツリーの集合という面積を求める検索と捉えることができる。そのため、絞り込み検索は、面積を絞り込む検索と捉えることができる。

ここで図8のツリーに対して、キーワード「A」、「B」で絞り込み検索を行った場合を考える。最初にキーワード「A」で検索を行った結果、表示されるノードはノード、一件である。次に、キーワード「B」で絞り込み検索を行う際、検索対象はノードをルートノードとするサブツリーのため、ノード、となり、検索結果はノード、二件となる。このとき、表示方法を下方部分一致方式にしていなかった場合、表示結果は2件となり絞り込み検索を行ったにもかかわらず表示結果が増えることになる。しかし、図9で見れば面積は絞り込まれている。

また図8のツリーに対して「B 絞り込み A」で検索を行うとする。「B」で検索を行う際に表示する結果に下方部分一致方式以外を指定した場合、絞り込み検索を行えば検索結果は0件となり、キーワードの指定順序によって検索結果が異なることになる。

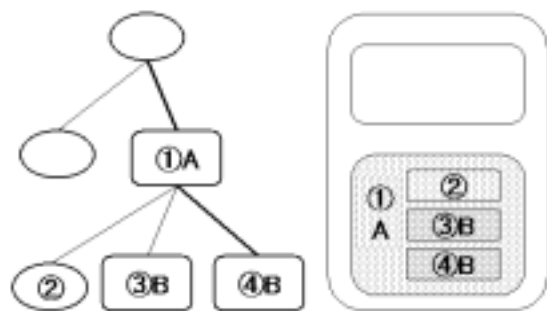


図8 絞り込み1

図9 絞り込み2

7. 評価実験

提案した検索方式について、神戸大学震災アーカイブのデータを用いた実験を行った。

同データの資料に、「兵庫県南部地震による神戸大学

附属図書館被害状況写真集」という写真集がある。この写真集を包含関係によって分割し、木構造に関連付けたものを図10に示す。ただしこの写真集はノード数が多いため、この図では簡略化している。この図の各ノードに記述されている文字は、そのノード中に含まれるキーワードである。また、この写真集において、図中に書かれているキーワードを複数記述されているノードは存在しない。各ノードを詳しく見ると、ルートノードには写真集のタイトルである、「兵庫県南部地震による神戸大学附属図書館被害状況写真集」が記述されている。その子ノードには、神戸大学に所属する各図書館の名称、「人文・社会科学系図書館」、「自然科学系図書館」、「人文科学系図書館」、「国際教養系図書館」、「人間科学系図書室」、「医学部分館」が記述されている。さらに、それぞれの図書館が記述されているノードには、館内見取り図について記述されたノードが一つと、写真に関する情報が記述されたノードが複数ある。このとき、写真に関する情報が記述されたノードには撮影場所が記述されており、その撮影場所は医学部分館が「中央区」である以外は、「灘区」となっている。また、写真に関する情報が記述されたノードの数は、それぞれ、人間・社会：4、自然科学：30、人文科学：24、国際・教養：29、人間科学：15、医学部：56である。

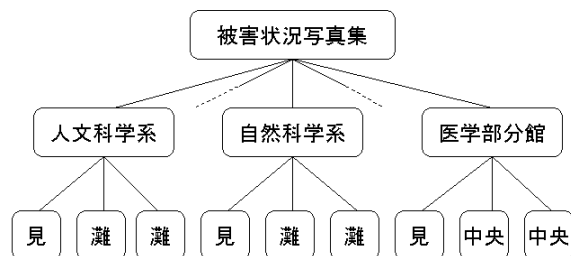


図10 写真集の概念図

この資料に対して「中央区の被害状況」を知りたいという要求のもとに、「中央区」、「被害」という2つのキーワードで検索を行った結果を表1に示す。

この資料において、「中央区」、「被害」を含んでいるノードは包含関係になっており、ノード間の距離は2であるため、マッチすることのできたAND検索は、親戚ANDと直列ANDであった。親戚ANDで抽出されたサブツリーのルートノードは、すべて元資料のルートノードであったため、検索結果は一件となった。

直列ANDの検索結果の56件はすべて写真に関する情報が記述されたノードである。これは、先に示した件数と一致している。この検索結果に上方優先、下

方全一致を適用しても、条件を満たさないため表示するノードは検索結果で得られたノードすべてとなった。下方部分一致（ここでは80%以上マッチで上位ノードを表示することとした）を用いた場合、「医学部分館」と記述されたノードの子ノード57件中56件がマッチしているため、それらの上位ノードの「医学部分館」と記述されたノード一件が結果として表示された。

表示方式がルート表示である場合も、もちろんルートノードが表示されて表示件数は一件となる。ただしこの場合、他の「灘区」に関する写真142件とは区別がつかなくなってしまうため、絞り込まれた部分を提供することはできなくなっている。我々が目標としていた検索は、結果が冗長にならず部分的に絞り込まれた検索結果を利用者に提供可能とすることであった。この写真集に対し、「中央区」、「写真」で直列ANDを指定し、表示方法を下方部分一致とすればその要求を満たした結果を提供することが可能となっている。そのため、一定の効果を得ることができた。

表1：実験結果

AND演算子	親戚	兄弟	親子	直列	直列	直列	直列
検索結果	1	0	0	56	56	56	56
表示方法 (ANY)				ルート表示	上方優先	単純全一致	単純部分
表示件数	1			1	56	56	1

キーワード：灘 被害
表示方法の (ANY) は、いずれの表示方式であってもということであり、単純全一致は単純全一致方式、単純部分一致は単純部分一致方式を示す。

8. 電子図書館用デジタルアーカイブとしての適合性

本検索方式では、資料あるいは部分資料は、単なるテキストであっても、通常のデータベースのように属性を持っていても構わない。ツリー内の各ノードにつけられる属性間に関する特別な制約はない。例えば、ひとつの資料の中に、同じ属性が重複して存在しても構わない。仮に、本の中で他人の本の文章を引用していたとする。この場合、ルートノードには、「著者名」、「題名」などといった属性がありその本の著者名や題名が付与させていると思われる。また、引用の部分をひとつのノードとし、その部分にも引用元の「著者名」や「題名」という属性をつけることができる。この場合、利用者が検索を行う場合に、「著者」="A" 直列AND「著者」="B" などといった条件式を書くことで部分を特定することができる。

ノード間の属性に制約がないことは、個別に作成したツリーを統合する場合にも役立つ。例えば、個別に登録した本がシリーズものであった場合、ツリー構造をあらゆる情報のみを書き換えるだけで、ひとつのツリーとして再構成することができる。この場合、各ノードに付与された属性やその値を変更する必要はない。ルートノードにシリーズの<題名>や<監修者>などを付与するだけでよく、各本の著者名などは子ノードに記述されたままでよい。仮に各本のノード(もともとのルートノード)にもシリーズの<監修者名>が含まれていたとしてもそれを削除しなくともシステムとしては問題がない。

別な言い方をすれば、デジタルアーカイブ作成時において、作成者は、同一資料内の他の部分資料に付与される属性にあまり気をつかわずに各部分資料(ノード)に属性を付与することができ、大量の資料をアーカイブ化する作業が比較的簡単に行えるものと考えられる。

また、通常図書館において付与される分類番号などの様々な属性も検索対象として各ノードに付与することができる。各部分資料がメタデータおよび本文(テキスト)を含むような構成にしても構わない。即ち、メタデータの検索やメタデータ+テキストといった複合データの検索にも本方式は利用できる。

複数人によって共同でアーカイブを作成する場合、データ作成者の主観の違いによって細分化のルールが異なったとしても、part_of 関係でツリーを作成するという条件さえ満たしていれば、親子 AND や兄弟 AND など距離が関係する検索を除けば、多少の精度は落ちるが、システムとしては機能するものと考えられる。即ち、ビューで得られるサブツリーの大きさや数は異なるもの、利用者の欲する部分を示すことは可能となる。

このように、本方式で想定するアーカイブにおいては、データの統一性や一貫性を保つためのデータスキーマの制約を設ける必要もなく、大量のデータを人海戦術でアーカイブする場合の、作業者の負担を軽減できるという特徴をもち、比較的簡便にアーカイブを作成する上で有益な手段となると考えられる。

9. まとめ

本研究では、包含関係に基づいて分割された資料に対して、木の形状を利用した AND 検索の拡張と検索結果に対する複数の表示方法を用意することで、利用者の要求により効果的に応えることのできる、資料の抽出方法の検討を行った。その結果、AND 演算子と表示方法を適切に組み合わせることで、一定の効果を

得ることのできる検索方法を提案することができた。

しかし、利用者が適切な AND 演算子と表示方法を指定することは難しいと想定される。そこで今後は、利用者にとってより使いやすい検索方法となるよう、利用価値の高い AND 演算子と表示方法の組み合わせの検証等を行っていく予定である。さらに、今回あまり述べることができなかった絞込み検索についても考察していく予定である。

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] 渡邊 隆弘:「神戸大学電子図書館システムにおける「電子アーカイブ」の構築」, デジタル図書館 16号, 11 1999.
- [2] 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.
- [3] 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, p.501(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