

TDL から携帯端末へのコンテンツ配信システム開発

宮崎 明美[†] 今井さやか^{††} 金森 吉成^{††} 首藤 伸夫^{†††}

[†] 群馬大学 大学院 工学研究科 情報工学専攻 〒 376-8515 群馬県桐生市天神町 1-5-1

^{††} 群馬大学 工学部 情報工学科 〒 376-8515 群馬県桐生市天神町 1-5-1

^{†††} 日本大学 大学院 総合科学研究科 〒 102-0073 東京都千代田区九段北 4-2-1 市ヶ谷東急ビル 6 階

E-mail: [†]{akemi,sayaka,kanamori}@dbms.cs.gunma-u.ac.jp, ^{††}shuto-nobuo@arish.nihon-u.ac.jp

あらまし 我々が開発している津波デジタルライブラリ(Tsunami Digital Library: TDL)では、過去の津波に関する書籍や新聞記事の資料を XML 文書の形で公開している。これを利用した津波災害教育を考える。すなわち、海岸沿いや過去の被害地域の地名をキーワードとして検索を行い、その地域の被害状況を提示することで、住民や観光客に災害に対する注意を喚起する。住民や観光客の所在地から検索地名を特定する必要があるため、携帯端末で資料の閲覧、検索を行う。ここで、携帯端末の画面では表示可能文字数や処理能力に制約があるため、携帯端末に適した重要なコンテンツのみを提供することが必要となる。本研究では、このためのシステムの設計と実装を行った。

キーワード XMLDB, コンテンツ配信, 情報検索, 携帯端末, 津波デジタルライブラリ

Developing contents delivery system from TDL to mobile equipments

Akemi MIYAZAKI[†], Sayaka IMAI^{††}, Yoshinari KANAMORI^{††}, and Nobuo SHUTO^{†††}

[†] Department of Computer Science, Graduate School of Engineering, Gunma University

1-5-1, Tenjin-cho, Kiryu, Gunma 376-8515, Japan

^{††} Department of Computer Science, Faculty of Engineering, Gunma University

1-5-1, Tenjin-cho, Kiryu, Gunma 376-8515, Japan

^{†††} Advanced Research Institute for the Sciences and Humanities, Nihon University

6F, Ichigaya Tokyu Building, 2-1, Kudan-Kita 4-chome, Chiyoda-ku, Tokyo 102-0073 Japan

E-mail: [†]{akemi,sayaka,kanamori}@dbms.cs.gunma-u.ac.jp, ^{††}shuto-nobuo@arish.nihon-u.ac.jp

Abstract We are developing “Tsunami Digital Library(TDL)” which consists of reports, books, newspapers and etc. on the tsunami occurred in the past. These documents in TDL are represented in XML structures and are opened through Web. Also, we have proposed the system for a public education on the tsunami disaster mitigation using TDL. That is, when we input a coast name or the damage area name of the tsunami occurred in the past as keywords for TDL, the results of damage in the region are retrieved, and are displayed to catch the attention of citizens or tourists in the area. Therefore, it is necessary to retrieve the region name from the place where these people stand on. Then, we use mobile equipments to retrieve and show the results. There are restrictions of the number of display characters and the computing powers in mobile equipments. We have to give some important contents to fit mobile equipments. This paper describes the design and the implementation on our system.

Key words XMLDB, contents delivery, information retrieval, mobile equipment, Tsunami Digital Library

1. はじめに

我々は津波に関する多くの資料を総合的に管理し、一般に公開する津波デジタルライブラリ(Tsunami Digital Library: TDL)[1][2][3]を提案・構築している。

津波研究者ではない一般の人々はどの資料にどのような事実が記載されているかを知ることは難しい。膨大な量の資料がた

だ提供されるだけでは、目的の資料を検索することが困難である。そこで TDL では、津波のメカニズムの学習効果を高めるため、過去の津波資料を利用した津波教育教材 [4] を提供している。津波教育教材はインターネットを通して閲覧することができる。このシステムは全く津波のことを知らない初心者や津波に関する基礎知識を得ることを目的としている。そのため、対象者は津波の危険性がある海岸沿いに住む人も危険性がない

山辺に住む人も含まれ、全員に等しく津波情報を提供している。

一方、海岸沿いの住民や観光客にとって最も高い関心事は、自分の住む地域や自分が居る地域(所在地)で起こった過去の事例であると考えられる。そこで、TDL に格納されている資料の中から、所在地の地名が書かれた資料のみを検索し、より所在地に密着した情報を提示することで、効率的に住民や観光客の防災意識向上を促すことにする。観光客などその地域に住んでいない者が外出先で資料検索を行うことを可能にするためには、容易に持ち運びできる端末で本システムが使用できなければならない。よって、本システムは携帯端末に向けた配信を前提として考える。しかし、携帯端末では、入力装置の制限や、一度に出力できるデータ量の制限、処理能力の制限がある。そこで、携帯電話にインターネットに接続したサービスが多数あるということに注目した。例えば、携帯電話用ブラウザを用いた Web コンテンツ閲覧機能や、基地局情報を用いた位置情報取得機能などである。本システムではこれらの機能を有効に利用し、利用者が地名入力のわずらわしさを感じずに、最小限の操作だけで資料を閲覧できるよう工夫した。また、データ量の制限に対して、地域に関連する重要なデータのみを出力できるように、資料検索システムを構築した。

2. システム概要

TDL はインターネットに接続された端末で Web ブラウザを用いて閲覧することを前提として開発している。サーバはスクリプト言語 (PHP) プログラムを用いてデータベースにアクセスし、XML 文書を Web ブラウザに適した形式に整形し、クライアントに配信している。クライアントが携帯端末になっても、基本的な考えは同じである。

図 1 は携帯端末用コンテンツ配信システム構成の概要図である。検索処理の流れを説明する。

- (1) 利用者は GUI として携帯端末用 Web ブラウザを用いて検索要求を送信
- (2) インターネットを経由し、サーバが検索要求を受信
- (3) サーバは検索要求を満たすデータが存在するかデータベースに問い合わせる
- (4) 存在すればサーバはデータベースからデータを取得し、HTML テキストデータへ加工
- (5) 加工したデータを検索結果としてクライアントに返信
- (6) クライアントは検索結果を受信

3. TDL のコンテンツ

3.1 コンテンツの種類

本システムでは、TDL に格納されている資料の中から、過去の津波被害を記した重要なデータとして、書籍と新聞記事に注目した。書籍と新聞記事として TDL に格納されているデータを、携帯端末画面表示に適したデータに加工し、加工したデータを検索する方法を第 4 章に示す。

TDL に格納されている書籍、新聞記事については次の通りである。書籍は、論文、報告書、体験談集などからなり、テキストデータとして入力した。このテキストデータを構造化し、

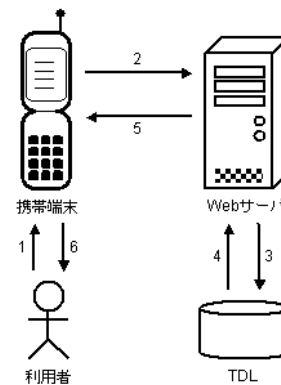


図 1 システム構成概要図
Fig. 1 System outline

XML 文書を作成した。1つの書籍に対して1つのXML文書がある。テキストデータとして入力できない図、地図、写真は、書籍のページをスキャンし画像データとして保持している。また、表については公開方法の違いから画像データとして保持する他に、書籍のXML文書とは別に表ごとのXML文書を作成した。1つの書籍に対して複数の表があるため、表のXML文書は複数存在する。新聞記事は、各新聞から津波に関する記事のみからテキストデータを入力し、XML文書化を行った。書籍と同様に図、地図、写真、表は、画像データとして保持している。

3.2 書籍XMLデータ構造

図2は書籍のXML文書のデータ構造である。ルート要素の下に書籍情報<metadata>と、本文情報<substance>を子要素にもツリー型の構造と捉えることができる。

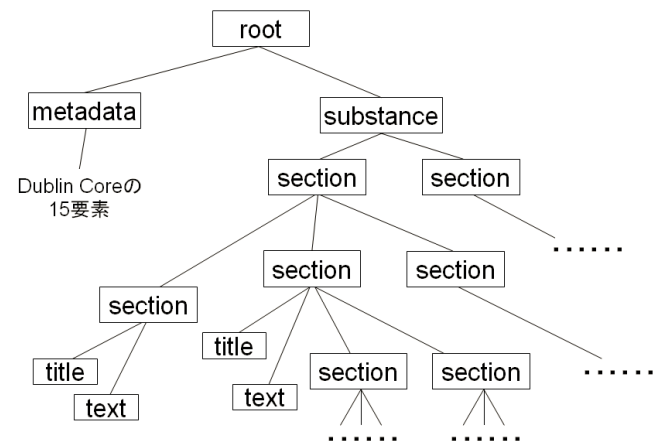


図 2 書籍 XML データ構造
Fig. 2 XML data schema of documents

書籍情報<metadata>には、Dublin Core [5] で定義されている 15 要素を利用した。Dublin Core はインターネット上で提供される様々な文書のメタデータ規則として提案・開発された。各要素名と書籍情報の定義を表 1 に示す。

本文情報<substance>は、子要素に書籍の章や節にあたる<section>を複数持ち、<section>は章や節のタイトル名<title>と章や節の内容<text>と入れ子の<section>を

表 1 Dublin Core 各要素名と定義

Table 1 Dublin Core Metadata Element Set (DCMES)

要素名	定義	要素名	定義
Title	書籍名	Type	型, ジャンル
Creator	作成者, 著者	Format	データ形式
Subject	主題, キーワード	Identifier	文献識別子
Description	概要	Source	関連文献への参照
Publisher	出版社, 発行社	Language	言語
Contributor	編集者	Coverage	対象範囲
Date	日付	Relation	関連文献への参照
		Rights	権利管理

持つ形になっている。

3.3 被害記録表 XML データ構造

被害記録表は地域名とその被害記録が書かれているものを前提とする。本システムでは、特定地域の特定被害種別の被害値を、複数の表から検索し抽出する。表を XML 構造化する場合、データの意味から見たタグ付け方法とデータ利用、データ処理から見たタグ付け方法が考えられる。タグ付けをデータの意味から考えると、被害記録表の被害種別が書かれた行(タイトル行)を要素名としたツリー構造で表すことになる。複数の書籍から収集した被害記録表は、タイトル行の入れ子構造や各被害種別の名称、地名の記述方法が異なるため、それぞれ異なったスキーマを用いなければならない。しかし、本システムでの利用を考えると同一のスキーマで表すことが望ましい。従って、データ利用、データ処理から見たタグ付け方法を採用した。

図 3 は被害記録表 XML 文書のデータ構造である。ルート要素は表の情報<metadata>と<rowdata>, 表の備考情報<note>を子要素に持つ。表の一行の情報を<rowdata>を頂点としたサブツリーで表す。表の一行とは、地名とその地域で調査された被害種別、被害値をセットにした情報である。地名は<area>の子要素として以下の要素に記述される。複数の被害記録表の異なった地名記述方法から、地名が表す意味を正しく得られるように、都道府県名<ken>, 郡名または市名<gunshi>, 町村名<choson>, 字<aza>の要素に地名を分類して記述する。要素<damage>はタイトル行の入れ子構造に従って階層構造を作る。<damage>は属性値として被害種別名<dtype>を持ち、リーフ要素には親の<damage>の属性値に対応した被害の状況を数値で記述する。各<rowdata>に被害種別名を記述することで、特定地域が含まれる<rowdata>を頂点としたサブツリーを抽出するだけで、被害種別名と被害値両方を容易に取得できる。

3.4 新聞記事 XML データ構造

図 4 は新聞記事の XML 文書のデータ構造である。書籍の場合と同様に、ルート要素の下に記事の情報<metadata>と、本文情報<page>を子要素にもつツリー型の構造と捉えることができる。本文情報<page>は、子要素に新聞の面番号<num>と、新聞の一つの記事にあたる<article>を持つ。<article>は記事のタイトル名<title>と記事の内容<text>と入れ子の<article>を持つ形になっている。

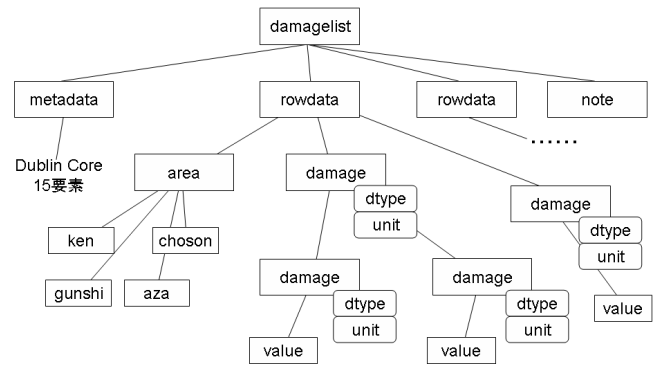


図 3 被害記録表 XML データ構造

Fig. 3 XML data schema of table

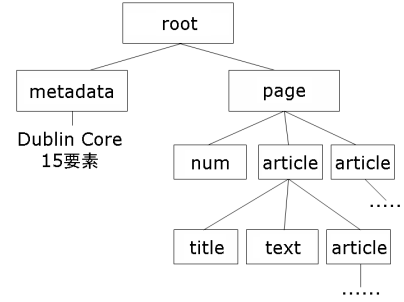


図 4 新聞記事 XML データ構造

Fig. 4 XML data schema of newspaper

4. 検索システムの構成

4.1 書籍検索

書籍の内容と情報をインターネットを通して携帯端末のブラウザに配信しようとした場合、書籍の要約文を提示する方法と、書籍の全文を提示する方法がある。書籍の要約文とは、書籍の目次情報や本文情報から、書籍全体のあらすじとして1つまたは少数の文章を作成することを意味する。主にこの要約文は、書籍のオンラインショッピングでの書籍紹介や、図書館等のオンライン総合目録情報サービスに使用されているものと同様のものである。また、書籍の全文を配信する場合は、書籍の要約文のみではなく書籍の目次、本文など、実在の本を手にとって見るのと同様な情報を、オンラインで得ることができる。しかしながら、文章量が少ない要約文のみを配信する場合、書籍全体の内容を詳しく提供することは不可能である。一方、文章量が多い全文を配信する場合、スクロールとページ遷移のわずらわしさや通信速度に問題がある。

4.1.1 要約文の必要性

携帯端末のディスプレイで一度に表示できる文字数は、数百文字程度である。書籍によっては一冊で数百~千ページ程あり、読みたい箇所に辿り着くまでに多大なスクロールとページ遷移が必要となる。一方、要約文を表示する場合は、少ない操作で書籍の内容を閲覧できる。しかし、従来の要約文は、書籍の紹介やあおり文句といった意味合いが強い。書籍の概要を知りたいという利用者にとって、満足のできる要約文ではない。さらに、従来の書籍要約方法は、携帯端末で閲覧するという前提は

特に設けず、PC用ブラウザ等比較的大きな画面での閲覧を対象にしているため、あらかじめ決めた要約率(要約後の文章量 ÷ 要約前の文章量)で要約を行っている。そのため、要約された文章は元の文章が長ければ長く、短ければ短いものとなる場合が多い。しかし、携帯端末で閲覧する場合、一度に表示可能な文字数の制限から、元の文章の長さに関わらず均一な量の要約文を得たい。

4.1.2 要約文作成

要約文を書籍の目次情報に従って章ごと、節ごとの区切りによって作成する。利用者は携帯端末から書籍の目次情報を閲覧し、読みたい章・節のタイトルをサーバにリクエストすることで、その章・節の要約文のみを閲覧できるようにする。

要約文を書籍の目次情報に従って作成するため、要約前の書籍の階層構造と要約後の書籍の階層構造は同じとした。要約方法は、まず一番深い階層での文章の要約を行う。この作成した要約文から、一つ上の階層の要約を行う。繰り返していくことで、最終的に本全体の要約文が一つ作成できる。このすべての要約文を用いて、利用者に階層ごとの表示を提供する。文書の書籍情報と階層ごとの要約文を一つの文書として表現し、どの階層について作成されたものか、という情報とともに管理しておく。ここでは、書籍情報の更新が無いこと、検索速度の点から、あらかじめ階層ごとのXML文書を作成しデータベースで管理する、という方法をとる。

図5は書籍をデータベースに登録するまでの流れである。

- (1) 実在の書籍から、書籍内の文章をデジタル化しテキストファイルを入力
- (2) テキストファイルから、図3のXML文書のデータ構造に従って、XML文書に整形
- (3) 全文XML文書をデータベースに格納
- (4) XML文書から要約文XML(複数)を作成
- (5) 要約文XMLをデータベースに格納

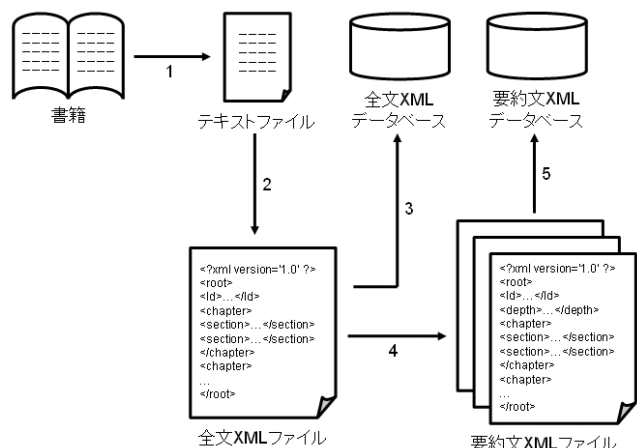


図5 書籍をデータベースに登録するまでの流れ
Fig.5 Flow from documents into database creations

図6は要約文XMLデータベースのスキーマである。要約文検索処理では、IDとDepthを主キーとして検索し、検索結果のXML文書を返す。次に、書籍のXML文書から、書籍の階層情

報に従って要約文を作成する手順を、図7を用いて説明する。

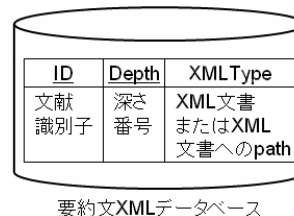


図6 要約文XMLデータベースのスキーマ
Fig.6 Database schema of XML outline documents

本文情報は、章・節ごとに章・節タイトルと章・節本文をXMLスキーマに基づいて入れ子の形式で表現する。図7(a)のような階層構造の場合、図7(b)のように入れ子構造とする。<section>は必ず1つの章・節タイトル<title>と0つまたは1つの章・節本文<text>と0つまたは複数の子要素<section>を持つ。このとき、親要素<substance>から見て子孫要素<section>の関係を、深さ番号(Depth)として表す。Depthのナンバリング例を図7(c)に示す。

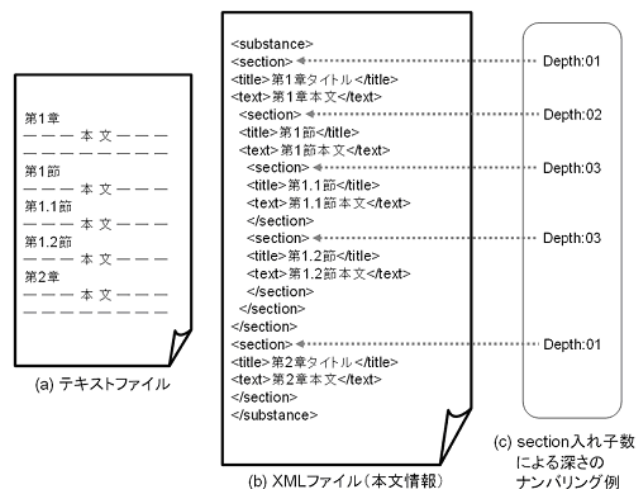


図7 section入れ子数による深さのナンバリング例
Fig.7 An example of depth number

要約は、要約前XML文書の<text>タグ内に記述された各章・節本文に対して行う。まず、一番深い階層、つまり子要素の<section>を持たない<section>の<text>タグ内の文章をそれぞれ特定文字数に要約する。特定文字数は、多くの携帯端末で1画面に表示できる量として、130字とする。一番深い階層以外の<text>タグ内の文章を削除する例を図8(a)に示す。次に、要約した階層の一つ上の深さの階層で、要約を行う。この時、<section>タグはその章・節のタイトル、章・節の本文、子要素の<section>を持っている。子要素の<section>内には、すでに要約された<text>がある。要約は、章・節の本文、子要素の要約文を繋ぎ合わせ、その中で特定文字数に要約する。この作業を繰り返し行うことで、一番深い階層から一番浅い階層までの要約文を作成できる。例を図8(b)と図8(c)に示す。一番浅い階層の要約文をすべて合わせ、その中で一つの要約文を生成すると、書籍全体の要約となる。

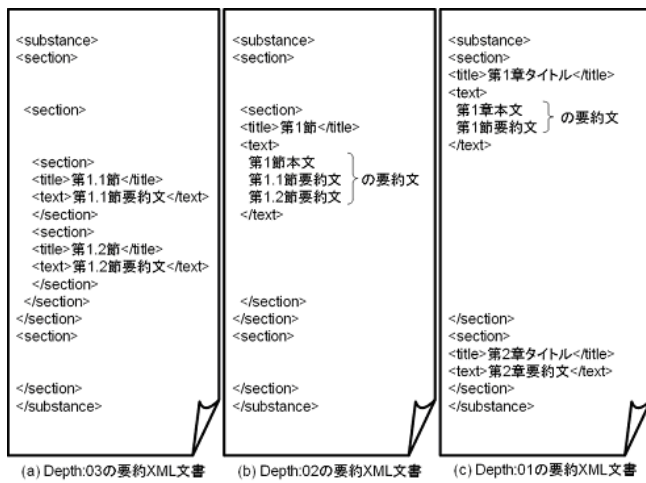


図 8 深さごとの要約文 XML 例

Fig. 8 XML outline documents depend on each depth

4.1.3 要約文作成アルゴリズム

要約文の作成は、以下のアルゴリズムを用いた。このアルゴリズムは、`<text>`タグに囲まれた要約前の文章を引数として受け取り、重要度の高い文字列を最大 3 つ出力する手続きである。

```

proceduer summary($text, $title);
// $text: 要約対象の文章
// $title: 要約対象の文章の題名
var sentence : String[5400];
var weight : Integer[5400];
var num : Integer[3] := {-1,-1,-1};
var s, i sum: Integer;
begin
  @titleword := $title内に含まれる重要単語のリスト;
  @sentence := $textを句点で分割した各文字列;
  for $n := 0 to count(@sentence) do
    begin
      // $sentence[$s]の重要度を調べる
      $weight[$n] :=
        $sentence[$n]の出現位置 / 定数; // ※重み1
      $weight[$n] := $weight[$n] +
        $sentence[$n]に含まれる重要単語の出現回数 / 文字列長; // ※重み2
      $weight[$n] := $weight[$n] +
        $sentence[$n]に含まれる@titlewordの出現回数 / 文字列長; // ※重み3
      // 上位3位の重要度と比較し、高ければ$numを更新
      if $weight[$n] >= $weight[$num[0]]
        $num[2] := $num[1]
        $num[1] := $num[0]
        $num[0] := $n;
      elseif $weight[$n] >= $weight[$num[1]]
        $num[2] := $num[1]
        $num[1] := $n;
      elseif $weight[$n] >= $weight[$num[2]]
        $num[2] := $n;
      end if
    end
  end
  for $i := 0 to count(@num) do
    begin
      $sentence[$num[$i]]を出力;
      $sum = $sum + StrLen($sentence[$num[$i]]);
      // 出力した文字列の文字数が合計130字を超えた場合
      if $sum >= 130 then exit; // 終了
    end;
  end.
end.

```

本研究では「重要文抽出」と呼ばれる方法 [7] [8] を採用した。重要文抽出は、与えられた文書の中から、

- (1) 文書中の頻度に基づいてその文書の重要語 (キーワード) を特定する

- (2) 文書中の各文の重要度を、重要語を含む割合で算出する

(3) 最後にある一定の重要度以上の文を出力するという方法である。文章の重要語は、我々が作成した津波に関する重要単語表を用いる。

\$substance の重要度を、3 種類の重みから求める。

- 出現位置による重み (重み 1)

文章中のどの位置に出現したのかを重視する。最初、最後の文章ほど重みを高くする。

- 重要単語の出現回数による重み (重み 2)

我々が作成した津波に関する重要単語表を用い、重要単語の出現が多いほど重みを高くする。文字列の長さに影響されないように、文字列長で割り、1 文字における単語の出現割合を求める。

- @titleword の出現回数による重み (重み 3)

タイトルは本文を最大に要約した文章と言える。そのため、タイトルに含まれる単語は、本文の要約にとっても重要と考えることができる。我々が作成した津波に関する重要単語表を用い、文書のタイトルに含まれる重要単語 (@titleword) を調べる。重要度を調べたい文章に、タイトルに含まれる重要単語が出現した場合に、重みを付加する。文字列の長さに影響されないように、文字列長で割り、1 文字における単語の出現割合を求める。

要約には関らの研究 [9] を参考とし、上記の 3 種類の重みを考慮する我々が考えた式 1 を用いた。i は文章番号、A は \$text における文章数、は定数、w_i は 1 文章における重要単語の出現数、tw_i は 1 文章におけるタイトルに含まれる重要単語の出現数、l は平均文字列長、l_i は 1 文章の文字列長、X_i は文章重みの合計値を示す。

$$\begin{aligned}
 X_i &= (\text{重み 1}) + (\text{重み 2}) + (\text{重み 3}) \\
 &= \frac{i - \frac{A}{2}}{A} + \frac{w_i}{|l - l_i|} + \frac{tw_i}{|l - l_i|} \quad (1)
 \end{aligned}$$

以上のアルゴリズムを用いて XML 文書を深さごとに要約する。一番深い階層では、XML 文書の `<text>` タグ内の文章をそれぞれ引数 \$text として手続きを実行させる。図 8(a) に対応する要約文 XML を得ることができる。次に深い階層では、子要素を持つ `<section>` の `<text>` 要約文と、子要素の `<text>` 要約文を結合することで、複数の要約文を一つの文章とする。その文章を引数 \$text として手続きを実行する。図 8(b) に対応する要約文 XML を得ることができる。この要約文の結合と手続きの実行を繰り返し、一番浅い階層の要約文 XML を得る。

4.2 被害記録検索

4.2.1 被害記録表 XML

報告書には、死傷者数や被害家屋数、被害船舶数などの津波被害の統計表がしばしば登場する。統計表には、地名と被害の種別 (死者数、重傷者数、軽傷者数、倒壊家屋数、流失家屋数、...) が記載されている。統計表をみることで津波の規模や地形の変化が分かる。ただ「多くの被害が出た」といったようなあいまいなイメージではなく、被害の数値を具体的にみるこ

が、防災意識の向上に貢献すると考える。表2は、明治三陸地震津波における岩手県気仙郡（現岩手県陸前高田市）の各地被害概数表 [6] を第3.3節で述べた被害記録表のXMLスキーマに従ってXML構造化し、XSLで整形した表である。

表2 気仙郡各地被害概数表 (XML文書)
Table 2 Kesen-gun damage list (XML data)

地名	人口				戸数			
	人口	死亡	負傷	健在者	戸数	流失家屋	半潰家屋	存在家屋
岩手県気仙郡気仙村	3,651	23	10	3,618	569	35	16	518
岩手県気仙郡高田村	3,489	3	未詳	3,486	616	未詳	未詳	616
岩手県気仙郡米崎村	3,460	12	2	3,446	350	11	50	289
岩手県気仙郡川友村	2,519	260	14	2,245	381	70	5	306
岩手県気仙郡広田村	3,102	500	11	2,591	469	163	未詳	306
岩手県気仙郡末崎村	2,965	606	30	2,329	400	191	同	209
岩手県気仙郡大船渡村	2,304	780	35	1,489	306	105	30	171
岩手県気仙郡赤崎村	2,985	448	68	2,469	389	172	未詳	217
岩手県気仙郡陸里村	2,803	1,458	58	1,268	451	285	100	66
岩手県気仙郡越喜来村	2,449	411	60	1,978	322	113	124	85
岩手県気仙郡吉浜村	1,075	215	9	851	133	32	33	68
岩手県気仙郡唐丹村	2,807	2,100	20	687	474	341	3	10
岩手県気仙郡合計	33,609	6,816	318	26,475	4,860	1,518	361	2,981

PC用ブラウザには、表データをすべて公開した被害記録表検索システムを提供しているが、携帯端末では表全体の情報ではなく特定の地域に関する被害のみが対象となる。表には別の地域の情報や要求されない被害種別も記載されている。そこで携帯端末用の配信システムでは、地名、対象津波名、被害種別を入力すると、必要な情報のみを取捨選択したコンテンツを作成する。検索の流れを図9に示す。

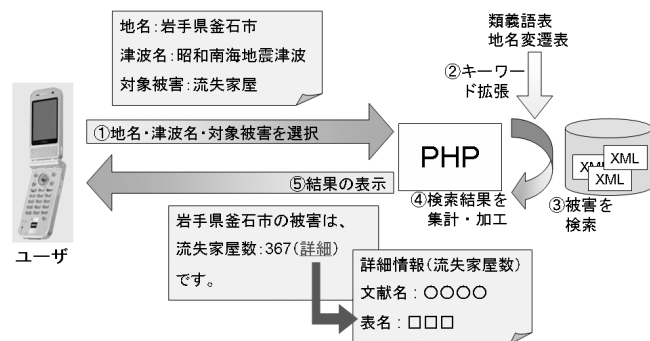


図9 被害記録表検索
Fig. 9 Retrieval of damage list

4.2.2 信頼性の高い表データ選択

被害記録表は、複数の書籍に複数存在するため、地名、対象津波名、被害種別で情報を絞っても、複数のデータが選択されてしまう。被害は、調査した日付によって刻々と変わり、また、国や都道府県などが正式に最終発表したデータと、調査機関が独自に調査したデータは異なる場合がある。そのようなデータをすべて携帯端末で表示すると混乱を招く。ユーザの欲しいものは一番情報に信頼がおけるデータである。そこで、選択された複数の被害記録表データの中から、表の信頼性が一番高いと思われるデータを、検索結果として表示する。

被害記録表の情報は、表名、調査年月日、調査機関名、対象津波名、典拠書籍IDを持つ。これらの中で、調査年月日と調査機関名を比較することで、信頼性を求める。まず、調査年月

日は、日付が新しいものがより正確であるとする。被害情報は日を追うにつれ次々と報告され、そのたびに被害数が増え変わるからである。次に調査機関としては、公共団体（公官庁、地方自治体、研究機関等）や民間がある。公的機関は津波対策策定や、復旧事業の基礎データとして利用するため信頼できると考えられる。

4.2.3 NULL値を補完した信頼性の比較

被害記録表XML文書の中には調査年月日や調査機関名がNULL値となっているものがある。これは書籍の内容に、被害記録表の表データが明記されていないためである。NULL値のものとの値のある被害記録表は単純に比較することができない。そこで被害記録表の情報源である書籍情報から、被害記録表の足りない部分を補完し、被害記録表同士の比較を行う。書籍情報は、書籍のXMLデータベースにアクセスをして、被害記録表の書籍IDを持つデータを特定する。被害記録表の調査年月日がNULL値の場合は、書籍の発行年月日で代用し、調査機関名がNULL値の場合は、書籍の著者・発行者で代用する。表データの補完方法の概念図を、図10、図11に表した。検索結果が図10のように調査年月日と調査機関名がNULL値である場合は、書籍のXMLデータベースから情報を取得し、図11のように表データの補完を行う。

被害記録表のデータ

被害ID	被害名	被害調査者	調査年月日
32_sannikudaishinsaish0058_00	岩手県下町村別被害実数 気仙郡 九戸郡 上閉伊郡 下閉伊郡	岩手県警本部	19330307
33_Sannikuikiyoshin005_0055_00	岩手県盛岡県津波被害 被害表	盛岡県津波被害	19330603
33_Sannikuikiyoshin005_0098_00	(集約)一人及家屋の被害	盛岡県津波被害	19330603
34_SannikuChihotsunami02_0218_01	震災被害状況報告(第39報)内務省警保部(岩手県)	内務省警保部	19340301
34_SannikuChihotsunami02_0228_00	岩手県一般被害調査 下閉伊郡	東北地震研究所	19340301
36_IwateShowaShinsai_00073_00	人の被害表	岩手県	NULL

図10 NULL値がある検索結果
Fig. 10 Lack of table data in retrieval results

被害記録表のデータ

被害ID	被害名	被害調査者	調査年月日
32_sannikudaishinsaish0058_00	岩手県下町村別被害実数 気仙郡 九戸郡 上閉伊郡 下閉伊郡	岩手県警本部	19330307
33_Sannikuikiyoshin005_0055_00	岩手県盛岡県津波被害 被害表	盛岡県津波被害	19330603
33_Sannikuikiyoshin005_0098_00	(集約)一人及家屋の被害	盛岡県津波被害	19330603
34_SannikuChihotsunami02_0218_01	震災被害状況報告(第39報)内務省警保部(岩手県)	内務省警保部	19340301
34_SannikuChihotsunami02_0228_00	岩手県一般被害調査 下閉伊郡	東北地震研究所	19340301
36_IwateShowaShinsai_00073_00	人の被害表	岩手県	NULL

書籍情報

ID	著者/編纂者	タイトル	発行日	巻/号	発行元
32	三陸大震災史料行会	三陸大震災史	19330415	NULL	仙台 東文堂
33	中央気象台	昭和八年三月三日 三陸津波被害と津波報告	19330603	特設特報 第七卷第二号別冊	気象庁
34	東北地震研究所	昭和八年三月三日 三陸地方津波に関する調査報告	19340301	東北大学 地震研究所集報 別冊第1号	東北地震研究所
36	岩手県	岩手県昭和震災誌	NULL	NULL	NULL

図11 NULL値の補完
Fig. 11 Complement of table data

4.3 新聞記事検索

書籍と似たXMLスキーマ構造を持つ新聞記事を検索するためには、書籍と同様に、長文である記事の内容<text>を一定文字数まで要約し、データベースで管理すればよい。しかし、新聞記事には書籍のような深さ階層がない。そのため、Depthという概念は考えず1つのXMLにつき1つの要約文XMLを作成する。検索方法も目次を提示して検索する方法ではなく、地名をキーワードとした全文検索を行う。要約文の作成には利用者が指定した観点に基づき要約する手法 [9] を採用した。

4.4 位置情報取得機能

本研究では、NTT DOCOMO が FOMA 方式 i モードに提供するオープン i エリア [10] を用いて、携帯電話の位置情報を得る方法を採用した。このオープン i エリアを用いて得た地域名をキーワードとして、上記 3 つの検索システムにアクセスすることにより、スムーズな情報検索を行うことができる。

オープン i エリアは、オープン i エリア対応サイトにアクセスして取得される基地局情報をもとに、その基地局がどのエリアに属するかをコンテンツサーバに対して通知する。コンテンツサーバとは、インターネット上にある情報提供者のコンテンツを格納している Web サーバのことである。ここでは、本システムの Web サーバを指す。コンテンツサーバでは通知情報をもとに、該当するエリアに関連する情報を返信する。この際、通知される情報は NTT DOCOMO が規定したエリアコードであり、緯度経度のような詳細なポイントを示す情報ではない。i モード機能を搭載した PDC 方式携帯電話及び FOMA 方式携帯電話を用いて、i モードが利用可能な場所であればこのサービスを利用できる。

(1) エリア情報検索画面表示より i モードサーバの CGI を呼び出すと、i モードサーバは (2) エリア情報送信確認画面を返す。(2) でユーザが認証すれば、コンテンツサーバに対し、エリアコードなどの情報を送信する。コンテンツサーバはエリアコードをもとに、そのエリアに該当する情報を (3) エリア情報検索結果画面として返す。(1) および (3) はコンテンツサーバが、(2) は i モードサーバが携帯電話画面に配信する。(1) でオープン i エリア機能を使用する際、HTML 記述で使用するパラメータを表 3 に示す。

表 3 コンテンツサーバから送るパラメータ

Table 3 Parameters from contents server

パラメータ	内容
URL	"http://www.docomo.ne.jp/cp/iarea?"(固定)
ecode	"OPENAREACODE"(固定)
msn	"OPENAREAKEYS"(固定)
nl	コンテンツサーバへの遷移先 URL
arg1	遷移先 URL に引数を渡したい時に使用 任意のパラメータと引数を指定可能

また、(2) でユーザが認証後、i モードサーバがコンテンツサーバに渡すパラメータを表 4 に示す

表 4 コンテンツサーバが受け取るパラメータ

Table 4 Parameters to contents server

パラメータ	内容
method	"POST"(固定)
action	コンテンツサーバへの遷移先 URL 表 3 nl パラメータの値
ACTN	OK(承諾) ボタンのパラメータ名
AREACODE	基地局エリアコード (5 桁)
表 3 arg1 のパラメータ名	表 3 arg1 の引数名

オープン i エリアでは全国を 505 のエリアに定義している。

それぞれにエリアコードと呼ぶ 5 桁の数字コードとエリア名称を設定している。この定義は NTT DOCOMO が独自に定めたものである。地域によって複数の市町村が 1 エリアに含まれる。そのため、本システムではエリアコードを取得し、含まれる市町村名を検索キーワードとして、上記 3 つの検索を行う。エリアコードから市町村名を取得するため、エリアコードとそのエリアに含まれる市町村名を記述した csv ファイルを作成した。ユーザは目的に応じてエリア内すべての地名に対して検索を行うか、いずれかの市町村に対してのみ検索を行うかを選択をする。

5. システムの実装

本システムを実現するにあたり、Apache、PHP 及び ORACLE10g を用いて実装を行った。図 12 に構成を示す。

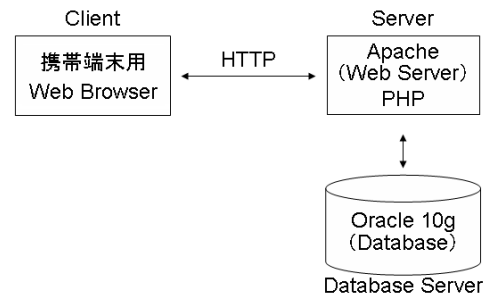


図 12 システム構成

Fig. 12 System structure

5.1 各プログラムの関係

図 13 にサーバ上のプログラムの関係を示す。実線の矢印はユーザの入力により検索が進むことを示し、点線の矢印は自動的にプログラムを呼び出すことを表す。

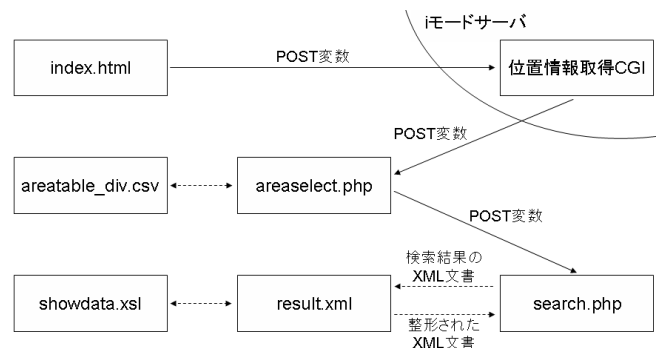


図 13 サーバ上にあるファイルの関係

Fig. 13 Relation among files on server

以下に各ファイルの説明を記す。

- index.html

選択する検索項目をクライアントに表示させる html ファイル。書籍検索、新聞記事検索、被害記録検索の 3 つから選択する。i モードサーバの CGI プログラムに選択した項目を Post 変数として送信する。表 3 のパラメータ arg1 に引き渡す TYPE は、書籍検索、新聞記事検索、被害記録検索のいずれかを示す。

- 位置情報取得 CGI(i モードサーバ)

基地局情報をもとに、どのエリアに属するかをコンテンツサーバの areaselect.php に対して通知する．コンテンツサーバへ Post 変数として送る情報は、5 桁のエリアコードなど表 4 に示したものである．

- areaselect.php

検索キーワードになる市町村名を取得し、search.php に市町村名を Post 変数として送信する PHP．プログラムの処理は、

- (1) エリアコードを i モードサーバから受信
- (2) areatable_div.csv を読み込み、該当する市町村名を抽出
- (3) 抽出した地名を、クライアントに表示
- (4) クライアントが選択した地名を search.php に Post 変数として送信

- areatable_div.csv

分割したエリアコード．

- search.php

クライアントが選択した検索項目の値から結果を取得する PHP．プログラムの処理は、

- (1) areaselect.php でクライアントが選択した地名を検索項目としてデータベースから検索
- (2) データベースから該当する XML 文書を取得
- (3) 取得した複数の XML 文書を 1 つの XML 文書として編集し、result.xml へ書き込む．
- (4) result.xml を呼び出す．

- result.xml

search.php により検索結果が書き込まれるテキストファイル．

- showdata.xsl

XSLT(XML Stylesheet Language Transformations)[11] を用いて、result.xml に書き込まれた XML 文書を整形して表示するためのスタイルシート．

5.2 実行例

図 14 は、岩手県宮古市にて動作確認と撮影を行った写真である．携帯端末から書籍の要約文を閲覧している．

6. まとめと今後の課題

本研究では、携帯端末での閲覧を念頭に置いた書籍、新聞記事、被害記録表の XML 文書化と検索システムの実装、携帯電話の位置情報取得機能を利用した検索キーワード自動取得モジュールの実装を行った．

本研究で使用した資料は、テキストで表される書籍、新聞記事、被害記録表のみであったが、TDL にはこれら以外の資料も多く保持している．とくに災害直後に撮られた被災地の写真や、地震津波災害対策の講演で用いられたプレゼンテーション資料、被災地に溯上した津波を CG で再現した動画など、画像データは防災意識向上に役立つと思われる．今後は、携帯端末での静止画・動画の閲覧システムの実装を考える必要がある．

また、総務省の情報通信審議会[12]によると、2007年4月以降、GPS 機能(緯度経度情報取得機能)を持つ携帯電話が普及すると考えられる．そのため、オープン i エリアのような広域な位置情報ではなく、詳細な位置情報を容易に取得できるよ



図 14 実行結果

Fig. 14 Execution result

うになる．より詳しい地名検索を行うために検索条件として、今後、緯度経度情報を用いる方法も考慮する必要がある．

7. 謝 辞

本研究は文部科学省科学研究費補助金基盤研究(B)(課題番号15300029)、文部科学省大都市大震災軽減化特別プロジェクト III-3 の助成を受けて行った．また、本研究を行うにあたり、新聞記事要約文作成プログラムと書籍要約文作成プログラムについて、群馬大学工学部情報工学科金森研究室の小菅広太さんと中島秀明さんにご協力頂いた．ここに記して謝意を表する．

文 献

- [1] Sayaka Imai, Yoshinari Kanamori, and Nobuo Shuto: "Tsunami Digital Library", ECDL2006, LNCS 4172, pp.555-558, 2006.
- [2] Tsunami Digital Library: <http://tsunami.dbms.cs.gunma-u.ac.jp/>
- [3] 川本健造, 今井さやか, 金森吉成, 首藤伸夫: "統合スキームを利用したデジタルライブラリ検索システムの実装", DEWS2007, 2007.
- [4] 津波デジタルライブラリ 津波教育支援システム: <http://tsunami.dbms.cs.gunma-u.ac.jp/education/>
- [5] Dublin Core Metadata Initiative: <http://dublincore.org/>
- [6] 震災予防調査会: "震災予防調査会報告", 第 11 号, 1897.
- [7] H.Luhn, "The Automatic Creation of Literature Abstracts,," IBM Journal of Research and Development, Vol.2, No.2, pp.159-165, 1958.
- [8] 難波英嗣: "情報抽出を利用した複数文書要約", 知能と情報, 18 巻 5 号, 16-22, 2006 .
- [9] 関洋平, 江口浩二, 神門典子: "利用者の情報要求を考慮した観点に基づく複数文書要約とその評価", 情報処理学会論文誌, Vol.46 No.SIG8(TOD 26) pp.106-119, 2005.
- [10] 株式会社 NTT ドコモ: "オープン i エリア 説明書 第 5.1 版", 株式会社 NTT ドコモ, 2006.
- [11] W3C, XSL Transformations (XSLT): <http://www.w3.org/TR/xslt>
- [12] 総務省: "携帯電話からの緊急通報における発信者位置情報通知機能に係わる技術的条件", 情報通信審議会, 2006.