

1. 作成した（取り組んだ）ポータル名若しくは課題
日本古写真メタデータデータベース

2. 概要

長崎大学附属図書館の所蔵する古写真のコレクションは平成 10 年にデータベース化しインターネット上で公開中である。今年度、このデータベースをリニューアルすることとなり、この機会により多くの人にデータベースを利用してもらうため 2 つの新しい試みを行うことにした。

- (1)OAI-PMH 対応と古写真のメタデータの公開

OAI-PMH を実装することで、公開したメタデータを世界中のサイトからハーベスティングしてもらえるようになる。これにより当館サイトの存在を広く知ってもらい、日本研究や写真研究に広く利用してもらうことが期待できる。

- (2)データベース内の検索での連想検索エンジン GETA の使用

GETA を使用して自然文からの検索と選択した写真に類似する写真を表示させる仕組みを作る。これによりデータベースのインターフェイスが使いやすくなるを考える。また、連想検索により写真と写真の間の思いがけないつながりが見付けられる可能性もある。

なお、今年度の計画ではとりあえずデータプロバイダの立ち上げを目標としているが、将来的には古写真に関するメタデータを収集してサービスプロバイダを立ち上げたいと考えている。

3. 演習とその成果 何を計画し、実装して、何ができたか。

- (1) 演習第 1 日

今回の演習時間の中で、ポータルサイトのどの部分を試験実装するかを検討した。上述したように、この企画では OAI-PMH の実装と GETA の実装という 2 つの課題がある。OAI-PMH の実装方法について講師の方々と話しているうちに、扱うデータが静的であるということ指摘された。データ件数は 6,026 件、基本的に内容はフィックスされている。それまで、OAI-PMH の実装は D Space で行うつもりだったが、D Space はコンテンツ管理やユーザー管理のインターフェイスの充実した、機関内で日々増加する情報を登録・発信するのに適したシステムであって、静的なデータならば OAI-PMH の 6 つのリクエストに答えられる CGI を書くという方法もある。今回は演

習時間も限られているということもあり、D Space は時間があったらチャレンジすることにして GETA の実装を中心に取り組むことにした。

(2) 演習第 2 日

GETA の実装。持参した csv ファイルのデータから、古写真の日本語タイトルと日本語解説文を取り出し perl で加工して頻度ファイル freqfile を作成。検索 CGI は、GETA の配布パッケージに含まれる assoc.cgi をカスタマイズした。また CGI の検索結果を、文書識別子ではなく画像のタイトルで表示するために、freqfile を perl で加工してタイトルファイルを作成。日本語データの検索と検索結果でのタイトル表示ができるようになった。(図 1)

(3) 演習第 3 日

引き続き GETA の実装。英語タイトルと英語解説文から英語版の freqfile、タイトルファイルを作成。英語の形態素解析には perl モジュール Text::English を使用した。英語データの検索ができるようになった。(図 2)

さらに、持参した古写真画像ファイルを使って 1 画像 1 ページの html ファイルを作り、CGI を書き換えて GETA の検索結果のタイトルをクリックすると html ファイルを表示できるようにした。(図 3)

また、freqfile 作成時や入力した検索語・文を単語に分割する際に使う形態素解析システム chasen の辞書をカスタマイズし、連結品詞を扱えるようにする方法を教えてください。古写真の日本語解説文中に出てくる年号や英単語もこれで一語として扱えるようになった。

4. 研修で学んだ技術及び知識

- ・ GETA

<http://geta.ex.nii.ac.jp/getaN2001/gdoc/manual.html>

汎用連想検索エンジン GETA(第 2 版)の導入・操作マニュアル。

- ・ 茶筌

<http://chasen.naist.jp/hiki/ChaSen/>

形態素解析システム茶筌に関する情報。パッケージの配布、マニュアル、各種文書やメーリングリストのログもここで見る事ができる。

5. 事前準備として必要と思われるもの (不足していたソフトウェアや予習事項等)

- ・ UTF-8 を扱えるターミナル

OAI-PMH で扱うメタデータの文字コードは UTF-8 のため、ターミナルの文字コードを UTF-8 にする必要がある。Linux はデフォルトで対応できるようだが、私が

持参したマシンの OS が FreeBSD5.3R でフォント等が足りなかったためうまく表示できなかった。

- ・プログラミングの知識

演習ではほとんどの作業を perl で行ったが、私のプログラミングの知識は行いたい処理はわかっているとしてもそれをどう書けばいいのかがわからないというレベルで、結局講師の方に教えていただくことがほとんどだった。

- ・XML と XML プログラミングの知識

今回の演習では、私は結局 OAI-PMH は取り組むことができなかったのだが、もし取り組むならば XML ベースのデータを扱うための知識と必要なソフトウェアは事前に準備しておく必要があると思う。

6. 今後の課題（職場で更に調査する必要のあるもの等）

- ・OAI-PMH の実装についてはこれからである。演習中は結局余裕がなく取り組むことができなかったが、講師の方や参加者の方に教えてもらった以下の方法を試してみたいと考えている。

D Space へのメタデータの一括インポート

<http://www.nii.ac.jp/metadata/irp/dspace-docs-jp/application.html#itemimporter>

XML ファイルベースの簡易データプロバイダ

<http://www.dlib.vt.edu/projects/OAI/software/xmlfile/xmlfile.html>

7. 今後の計画（実際のポータル構築計画等）

現在、4月の公開を目指してデータベースを構築中である。

8. 演習の感想

セミナー参加が決定してから届いた連絡メールで、与えられた課題をこなす演習ではないという文面を見て正直自信をなくしていた。ポータルの要素技術を使ったサイト構築の計画が実際に進んでいるため、自分での課題設定は困難ではなかったのだが、それを達成できるスキルがないという自覚があったからである。

軽井沢での演習が始まってみると、スキル不足で思うように作業が進まないのは確かにあったのだが、わからないことは講師のみなさんや参加者のみなさんにいつでも聞くことができる、また時間を気にすることなく集中して作業できるという恵まれた環境もあって、非常に充実した時間を過ごすことができた。GETA については、丸川先生に 1 から 10 まで教えていただく状態ではあったが、なんとか CGI 検索を動かすことができたので、この経験を手がかりに当館データベースに GETA を組み込み、面白いサイトを作ることができればと思う。

着手発表会、中間報告会と、要所要所で参加者がいまやっていることの進捗状況を

聞く（自分の進捗状況をまとめる）という形式も、頭が整理されてよかった。参加者のみなさんの設定課題がさまざまだったので、要素技術を使ってこんなこともできるのかと参考になることがたくさんあった。

これまで参加したなかで、最もハードでしたが最も得るものの大きかった研修でした。講師のみなさま、NIIのみなさま、そして参加者のみなさまにお礼申し上げます。

図 1 日本語データの検索と検索結果でのタイトル表示

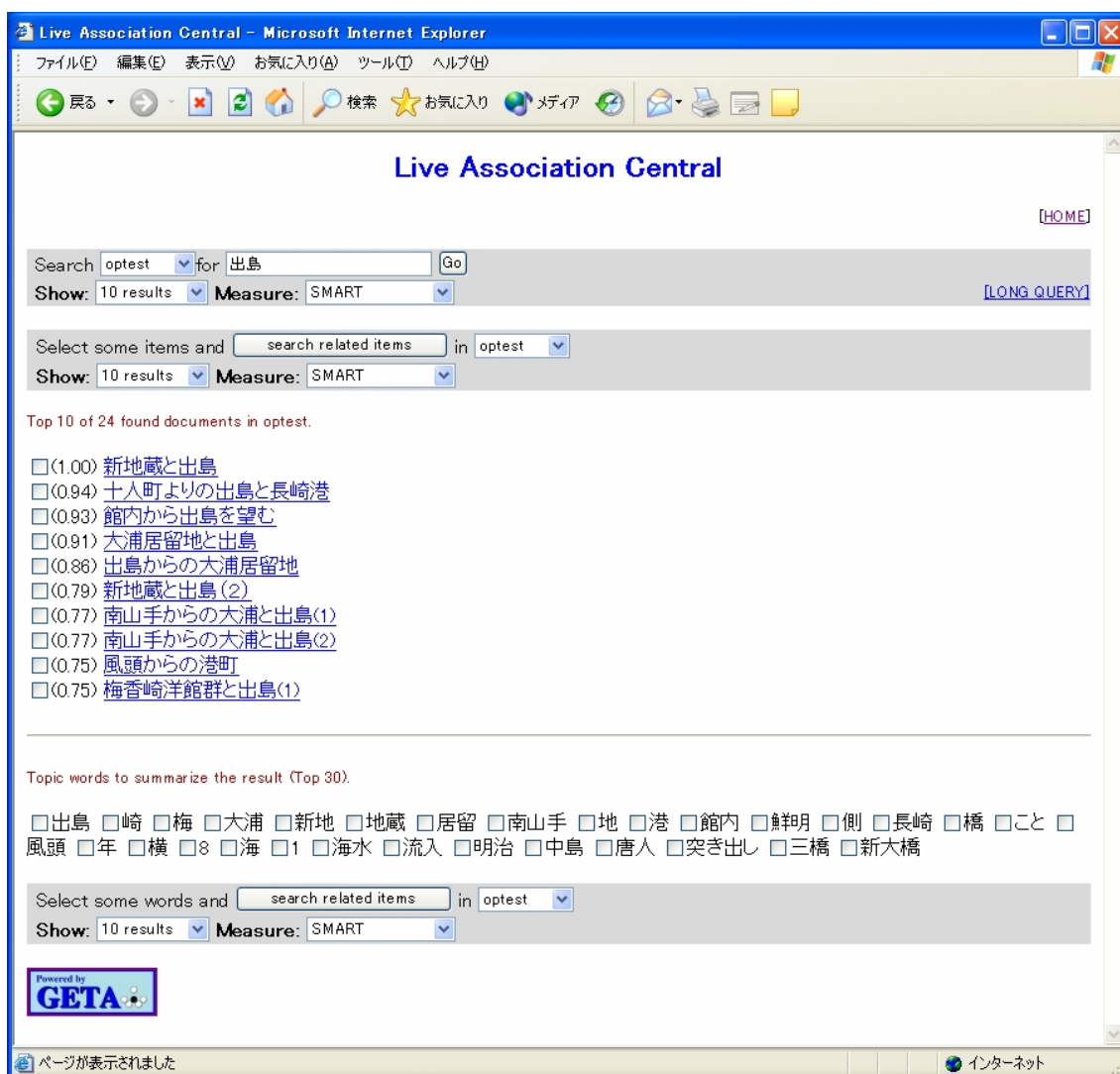


図 2 英語データの検索と検索結果でのタイトル表示

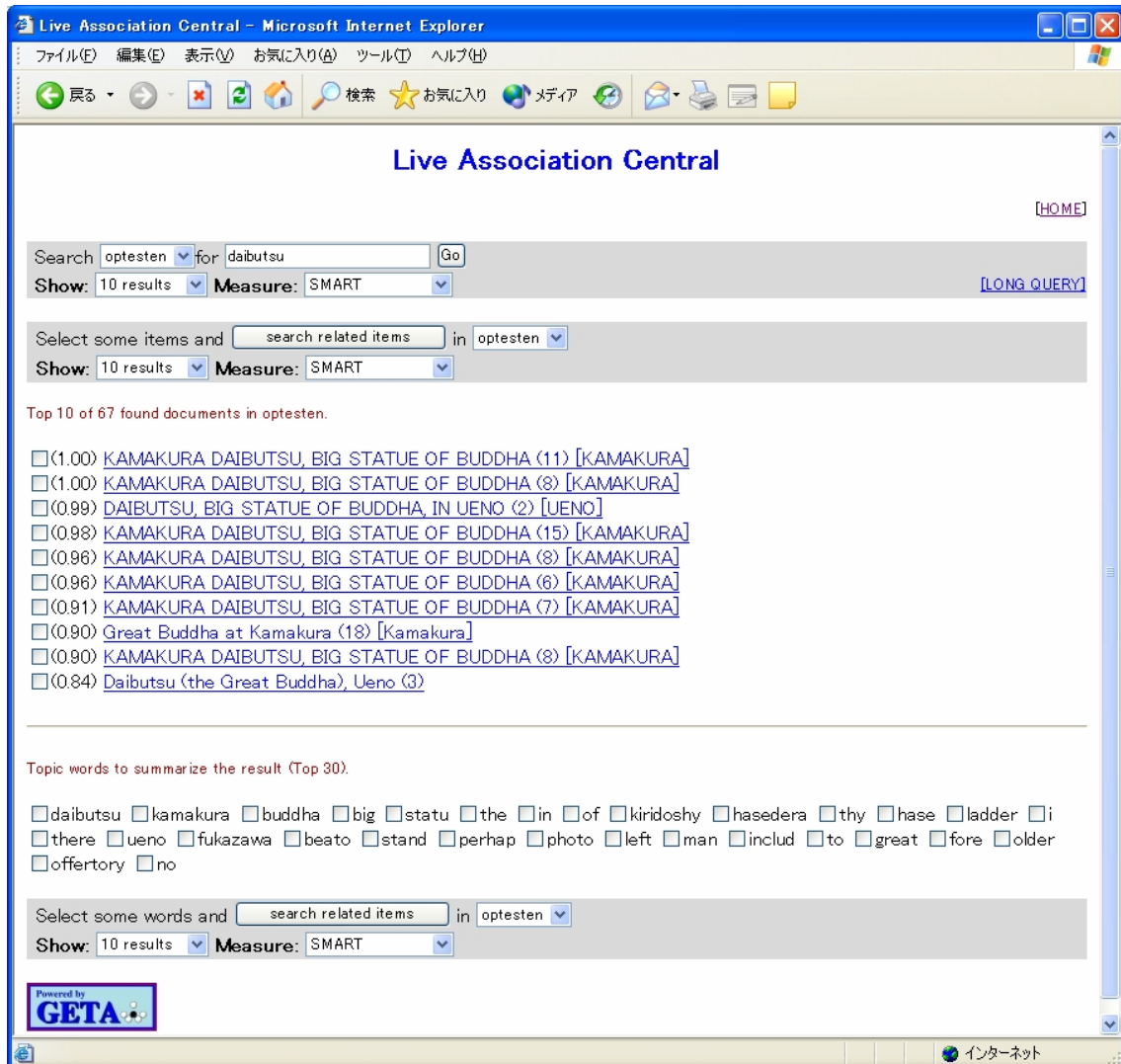


図 3 検索結果のタイトルクリックで実際の画像ファイルを表示



http://133.45.160.33/data/3859.html - Microsoft Internet Explorer

ファイル(F) 編集(E) 表示(V) お気に入り(A) ツール(T) ヘルプ(H)

戻る 進む 印刷 複製 貼り付け 検索 お気に入り メディア



梅香崎洋館群と出島(1)

3859 梅香崎洋館群と出島(1) 長崎 撮影者未詳 年代未詳 カラーなし 解説文無 鶏卵紙着色の1枚もので、DESHIMA NAGASAKIと印字。ファサリ商会のものか。梅香崎の丘から出島を望む。中島川河口は変流工事が終了し陸続きになっている。その横に明治22年(1889)架設の出島橋(明治43年に旧新川口橋がこれに架け替わる)の橋脚が見える。出島後方には明治30年(1897)着工の現長崎駅裏付近の締め切り護岸が見える。左端にピンクの洋館も見えるので撮影は明治30年代中頃。出島2番の旧カピタン部屋の場所に、明治35年(1902)に新築された足場を残すボーデンハウスが見えるので35年の竣工直後の撮影と思われる。イギリス聖公会神学校(中央鐘楼の建物)の左横出島7番に、明治36年(1903)新築される内外クラブはまだ見えない。右側の鐘楼は美以(三位)教会(明治9年建設)。出島後方は新県庁舎。手前梅香崎は左から順に1番長崎郵便局、2番デンマーク大北電信長崎局、3番日本郵船である。風景

ページが表示されました インターネット