

平成 17 年度情報処理軽井沢セミナーレポート

宮崎大学 学術研究協力部 情報図書課

寸田五郎

1. 作成した（取り組んだ）ポータル名若しくは課題

課題① IMAP サーバを用いた図書館ポータル

課題② GETA による連想検索 OPAC

2. 概要

課題① IMAP サーバを用いた図書館ポータル

ポータルの各レコード（例：リンク、貸出中図書、文献複写依頼等）をメールとして保存する。利用者は一般的なメールクライアントを使用してポータルを利用する。IMAP サーバを用いることで次のようなメリットがある。

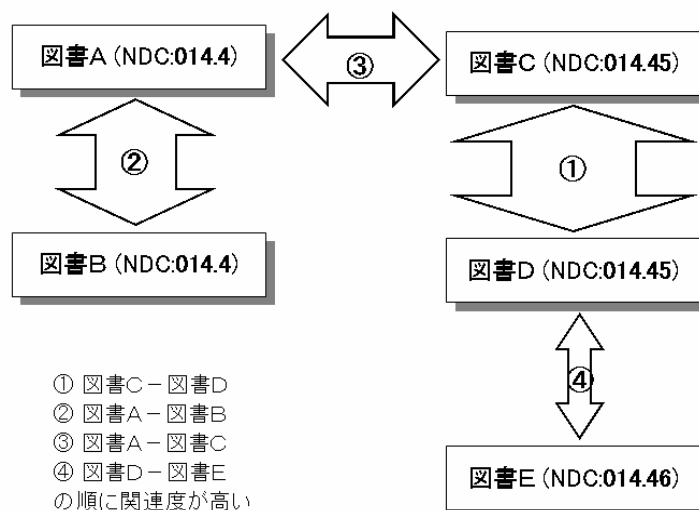
- ・ ユーザインターフェースを作成する手間が省ける
- ・ 利用者がユーザインターフェース（メールクライアント）を選択することができる
- ・ 複数のポータルを横断的に利用することができる。また、ポータル間でデータを移動することができる

研修ではリンク集のための電子ジャーナルリストを OAI-PMH によるメタデータハーベスティングをおこない、DOAJ から取得した XML データをメールに変換して IMAP サーバに格納した。

課題② 連想検索エンジン GETA をインストールし、Webcat Plus のような連想検索 OPAC を作成する。

Webcat Plus では、書名のほか、図書の目次情報や紹介文に含まれる単語を用いて連想検索を行っている。しかし、通常はこのようなデータを収集するのは困難である。そのため、分類(NDC)をもちいた連想検索を試みた。

NDC は階層構造をもち、桁数が大きいほど細かい分類である。2冊の図書で分類が一致する場合、その図書間に関連性があり、なおかつ分類記号の桁数が大きいほど関連性が高いと考えられる。また分類が一致しなくても、分類記号の前方で一致する桁数が大きいほど関連性が高いと考えられる。このため、「分類が一致するかどうか」「分類記号で前方一致する部分の大きさ」によって図書間に関連性を反映するように GETA の頻度ファイルを作成することにした。



3. 演習とその成果

演習第1日

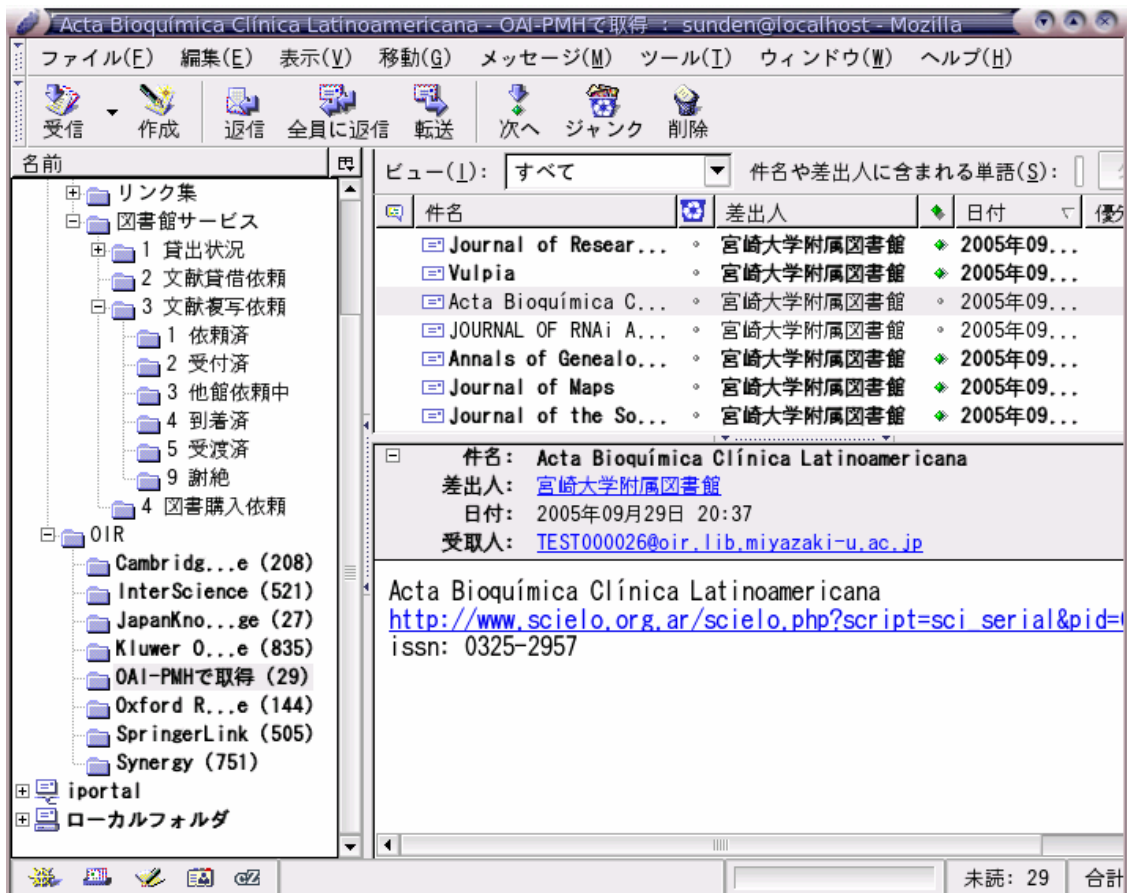
OAI-PMH の使用方法を確認。

XML 解析のためのツールの下調べなど。

演習第2日

Perl モジュール Net::OAI::Harvester でハーベストを試みるも、失敗。PHP で試みる。

phpdom をインストール後、domxml_open_file 関数を用いてメタデータ取得し、XML をテキストデータに変換することができた。



(DOAJ から取得したメタデータをメールとして保存)

演習第3日

GETA をインストールした。make するとエラーがでて失敗するので、「testsuit/tst.c」の「varargs.h」を「stdarg.h」に書き換えてインストールした。図書を文書、分類記号を単語とみなした頻度ファイルを作成した。

```
@BA72668882
5 NDC:673.8
2 NDC:673
@BA6936674X
6 NDC:673.86
2 NDC:673.8
1 NDC:673
@BA71131818
4 NDC:675
```

(頻度ファイルの一部：分類記号の桁数が大きいほど頻度が高い)

4. 研修で学んだ技術及び知識

- OAI-PMH によるメタデータハーベスティング
- PHP による XML 解析
- 連想検索エンジン GETA の使用方法

【参考 URL】

- DOAJ のサイト : <http://www.doaj.org/>
- GETA のサイト : <http://geta.ex.nii.ac.jp/>

5. 事前準備として必要と思われるもの

課題 2 については研修開始後に設定したため、当然ながら準備不足でした。次の点が必要だと思われる。

- 課題を決める
- 環境構築
- 事前にインストール等を試み、問題点を明らかにする
- サンプル用のデータを用意する

6. 今後の課題（職場で更に調査する必要のあるもの等）

課題①②ともに図書館システムとの連携が必要である。そのほか、課題①ではメールクライアントによっては日本語がうまく表示されない場合があり、IMAP メールサーバの仕様等の調査が必要である。課題②については NDC 以外の分類についても適用が考えられる。また、タイトル等による連想検索とどのように結合するか検討が必要である。

7. 今後の計画（実際のポータル構築計画等）

GETA を利用した連想検索 OPAC は、目録データだけでなく電子ジャーナルや web 上の情報源も収録し、学術情報ポータルの一機能として構築できるのでは・・・

The screenshot shows a Microsoft Internet Explorer browser window displaying the OPAC search results page. The page has a light green header with search input fields. The main content area is light blue and contains a list of 10 search results, each with a checkbox and a link. The results include titles like 'オンライン・サーチャーのためのChemical Abstractsの情報検索' and '情報検索の基礎知識'. On the right side, there is a pink sidebar with sections for '関連ワード' (Related Words) and '関連分類' (Related Categories). The '関連ワード' section has checkboxes for '検索', '探索', 'ABSTRACTS', 'オンライン', '道案内', '各種', 'ウェイ', '東京学芸大学', '配信', and 'アブストラクト'. The '関連分類' section has checkboxes for 'NDC:007.5', 'NDC:015.2', 'NDC:015', 'NDC:016.5', and 'NDC:014.9'. At the bottom of the page, there are buttons for '関連資料検索' and '関連ワード・分類検索'.

(作成中の連想検索 OPAC)

8. 演習の感想

3日間俗世間から隔離され、研修に集中することができました。講師および受講者の方々から貴重な意見をいただくことができ、大変有意義でした。

9. 備考、その他

講師およびスタッフの方々には毎晩遅くまで面倒みていただきありがとうございました。軽井沢ではなかったのが残念でしたが、武蔵嵐山もそれなりによかったです。NIIの図書室の図書を貸出していただくと、受講生にとってはありがたいと思います。