

SRU/SRWを用いた教育図書館 資料の書誌検索システム

2007年5月25日

情報知識学会2007年度研究報告会

江草由佳(国立教育政策研究所)

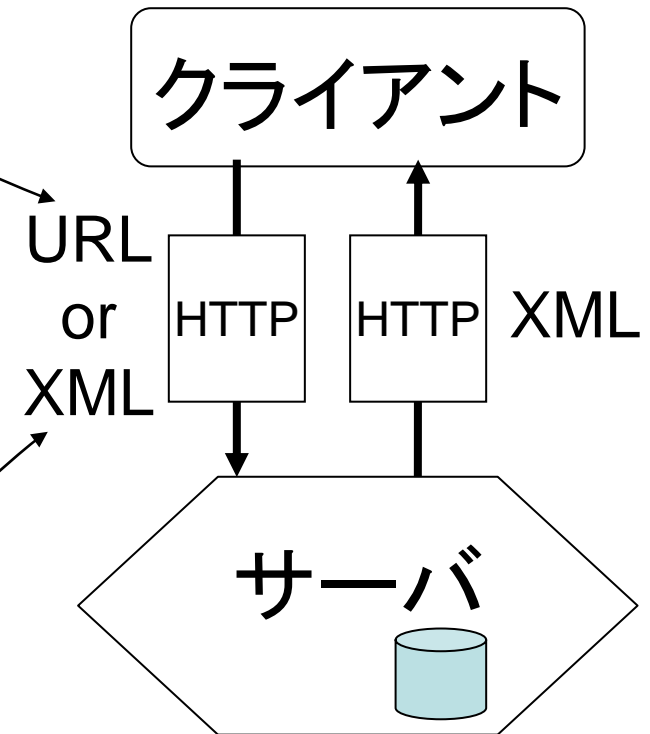
高久雅生(情報・システム研究機構)

背景と目的

- 背景
 - 検索プロトコルSRU/SRWの開発
 - Z39.50の後継
 - WWWとの親和性を重視
 - 日本での構築例
 - 国立国会図書館、筑波大学
- 目的
 - SRU/SRWを用いた教育図書館資料の書誌検索システムの構築
 - 実運用規模におけるSRU/SRWシステムの可能性の検証
 - SRU/SRWプロトコルの特性を考察
 - 特に
 - 日本語検索
 - 文字コード
 - サーバの開発

SRU/SRWプロトコル

- Z39.50の次世代プロトコルとして開発
 - HTTP: 通信プロトコル
 - XML: 要求/応答メッセージの記述に使用
- SRU (Search/Retrieve via URL)
 - REST (Representational State Transfer)フレームワークに基づく
- SRW (Search/Retrieve Web Service)
 - SOAP仕様に基づく検索要求を行う



SRUにおけるリクエストURL例

ベースURL

クエリパート

http://kaede.nier.go.jp/epi?version=1.1&operation=searchRetrieve
&query=%E8%AA%AD%E6%9B%B8

- 「http://kaede.nier.go.jp/epi」サーバに
- version が 1.1のSRUプロトコルを使用した
- searchRetrieve要求において
- 「読書」で検索要求を出している例
 - 「読書」の部分は、UTF-8で表現し、16進数値表現としてURIエンコード
- SRWの場合には、上記のパラメータをSOAP仕様に従ったXML形式でやりとりする

予稿集: 図1

レスポンス例

```
<?xml version="1.0" encoding="UTF-8"?>
<zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>1</zs:numberOfRecords>
  <zs:records>
    <zs:record>
      <zs:recordPacking>xml</zs:recordPacking>
      <zs:recordData>
        <xml>
          <pubdate>1996.3</pubdate>
          <journal_id>370.5-59-29</journal_id>
          <page>p.1~175</page>
          <regdate>1997.9.30</regdate>
          <author>林部一二, 堀井啓幸, 杉本真理子, 佐藤晴雄</author>
          <keywords>教育内容・方法 図書館教育 社会教育・生涯学習 家庭教育</keywords>
          <volnum>29</volnum>
          <book_mapping>教育研究論文索引 1996年版</book_mapping>
          <title>子どもの読書の実態と家庭における指導に関する調査研究</title>
          <journal>日本教材文化研究財団 調査研究シリーズ</journal>
          <paper_id>9603611</paper_id>
        </xml>
      </zs:recordData>
      <zs:recordPosition>1</zs:recordPosition>
    </zs:record>
  </zs:records>
</zs:searchRetrieveResponse>
```

1レコードを返した例

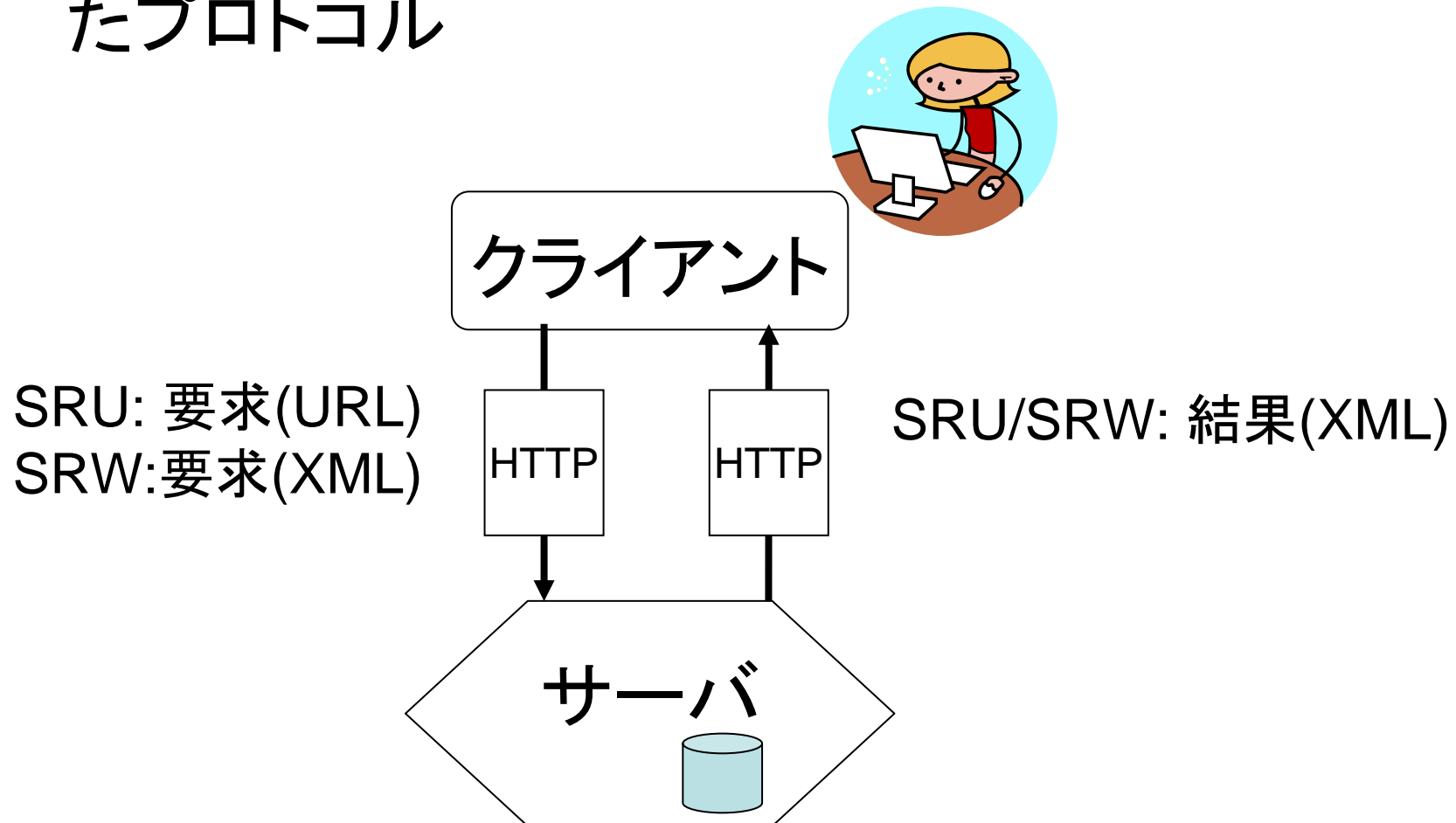
予稿集:図2

検索式:CQL (Common Query Language)

- ・ 人手で書きやすい・読みやすい文法
- ・ 機械処理可能な形式
- ・ 例:
 - 電子図書館
 - 電子図書館 and 調査 and 成立
 - title all “初等 中等“
 - journal exact “現代の図書館“

SRU/SRWのシステム構成

SRU/SRW は WWWとの親和性を考慮したプロトコル



開発システム:EPI-Search*

- 対象データ
 - 約12万件の論文書誌データ
- EPI-Searchサーバ
 - SRU/SRWサーバ
- EPI-Searchクライアント
 - SRUクライアント
 - 開発したEPI-Search*サーバにのみ対応

対象データ：教育研究論文索引

- 教育分野の研究論文書誌データ
- 主に日本語の論文を対象
- EPI-Searchの収録レコード数：122,510件
 - 1968年, 1983～2005年分
- 国立教育政策研究所教育図書館が作成
- Web上で既に検索サービスを提供
 - <http://www.nier.go.jp/homepage/jouhou/tosho-kan/>

教育研究論文索引： すでに提供されている検索サービス

The screenshot shows a Microsoft Internet Explorer browser window displaying the search page for the National Institute for Educational Policy Research of Japan. The page title is "書誌検索値入力(教育研究論文索引) - Microsoft Internet Explorer". The address bar shows the URL: "http://www.nier.go.jp/cgi-bin//retrieve/sr_makehtml.cgi?U-CHARASET=EUC-JP&OGILAN".

The main content area is titled "検索項目入力画面 (教育研究論文索引)" and includes a "多言語表示 | 利用方法" link. Below the title are two buttons: "検索" (Search) and "検索条件のクリア" (Clear search conditions).

The search form contains the following fields:

- 論題= (Topic)
- 著者名= (Author name)
- 掲載誌名= (Journal name)
- 巻号= (Volume number)
- 発行年月: (Publication year/month)
- ISSN= (ISSN)
- 請求記号= (Request number)

Below the search fields, there are options for search conditions and display settings:

- 項目間検索条件: AND (dropdown)
- 表示方法: 書誌 (dropdown), 表示件数: 10 (dropdown)
- 並び替え項目: 昇順 (dropdown), 1 請求記号 (dropdown), 2 請求記号 (dropdown), 3 請求記号 (dropdown)

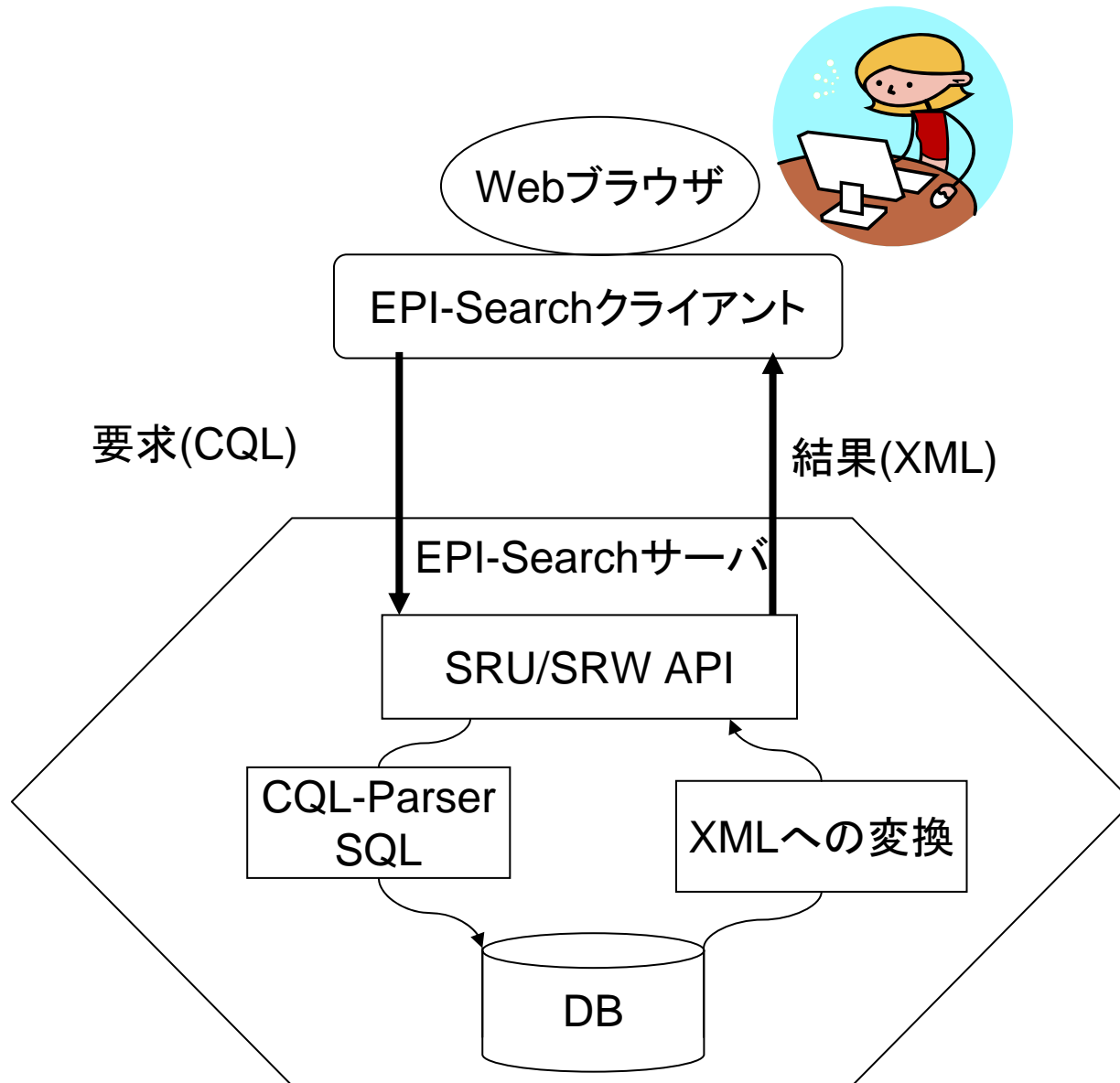
At the bottom of the page, there is a copyright notice: "Copyright (C) 2006 National Institute for Educational Policy Research of Japan All Rights Reserved." and a status bar showing "ページが表示されました" and "インターネット".

教育研究論文索引: 書誌項目

- 論題: title
- 著者名: author
- 著者名よみがな:
author_yomi
- 掲載誌名: journal
- 巻号: volnum
- 掲載ページ数: page
- ISSN: issn
- 請求記号: journal_id
- 発行年月: pubdate
- 文献番号: paper_id
- キーワード: keywords
- 登録日: regdate
- 冊子体: book_mapping

13項目

EPI-Searchシステムの構成



SRU/SRWサーバの機能

- 検索
 - ワード検索、フレーズ検索、完全一致検索
 - ブーリアン検索 (AND, OR, NOT)
 - アクセスポイント指定による検索
 - Dublin Core項目にも対応
- 返戻
 - 独自スキーマのXML

レスポンス例

```
<?xml version="1.0" encoding="UTF-8"?>
<zs:searchRetrieveResponse xmlns:zs="http://www.loc.gov/zing/srw/">
  <zs:version>1.1</zs:version>
  <zs:numberOfRecords>1</zs:numberOfRecords>
  <zs:records>
    <zs:record>
      <zs:recordPacking>xml</zs:recordPacking>
      <zs:recordData>
        <xml>
          <pubdate>1996.3</pubdate>
          <journal_id>370.5-59-29</journal_id>
          <page>p.1~175</page>
          <regdate>1997.9.30</regdate>
          <author>林部一二, 堀井啓幸, 杉本真理子, 佐藤晴雄</author>
          <keywords>教育内容・方法 図書館教育 社会教育・生涯学習 家庭教育</keywords>
          <volnum>29</volnum>
          <book_mapping>教育研究論文索引 1996年版</book_mapping>
          <title>子どもの読書の実態と家庭における指導に関する調査研究</title>
          <journal>日本教材文化研究財団 調査研究シリーズ</journal>
          <paper_id>9603611</paper_id>
        </xml>
      </zs:recordData>
      <zs:recordPosition>1</zs:recordPosition>
    </zs:record>
  </zs:records>
</zs:searchRetrieveResponse>
```

独自スキーマのレコード部分

予稿集:図2

SRUクライアント

教育論文索引・検索: title=言語問題 - Microsoft Internet Explorer

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

教育論文索引・検索

title=言語問題 検索

[title=言語問題]: 5件ヒットしました

1. マレーシアにおける言語政策と言語問題
野津隆志
教育制度研究会 教育制度研究(筑波大学内); 14/15; (1982.11) 【9100435】
キーワード: 教育史・教育事情 アジア諸国 各国の教育
2. 国際教育協力における言語問題の一側面について—英語における「まじめさ」の周辺語彙の構造をケーススタディとして
小原一馬
広島大学教育開発国際協力
キーワード: 教育史・教育事
3. 情報化社会における異文化例として
関口礼子
情報社会試論(図書館情報
キーワード: 教育内容・方法
4. 中国における少数民族文
小川佳万
広島大学大学教育研究セン
キーワード: 教育史・教育事
5. 文化大革命期における延辺朝鮮族自治州の民族教育と言語問題
権寧俊
アジア経済; 43-7; p.23~47; (2002.7) 【0302983】
キーワード: 教育内容・方法 民族教育

検索にかかった時間 0.33秒

ページが表示されました イントラネット

EPI-Searchサーバに特化
Webインタフェース
CQLを利用者が直接入力し、検索できる

予稿集: 図4

実行画面

- 検索例
 - アメリカ
 - title=アメリカ
 - 現代 and 図書館
 - journal exact 現代の図書館

開発環境

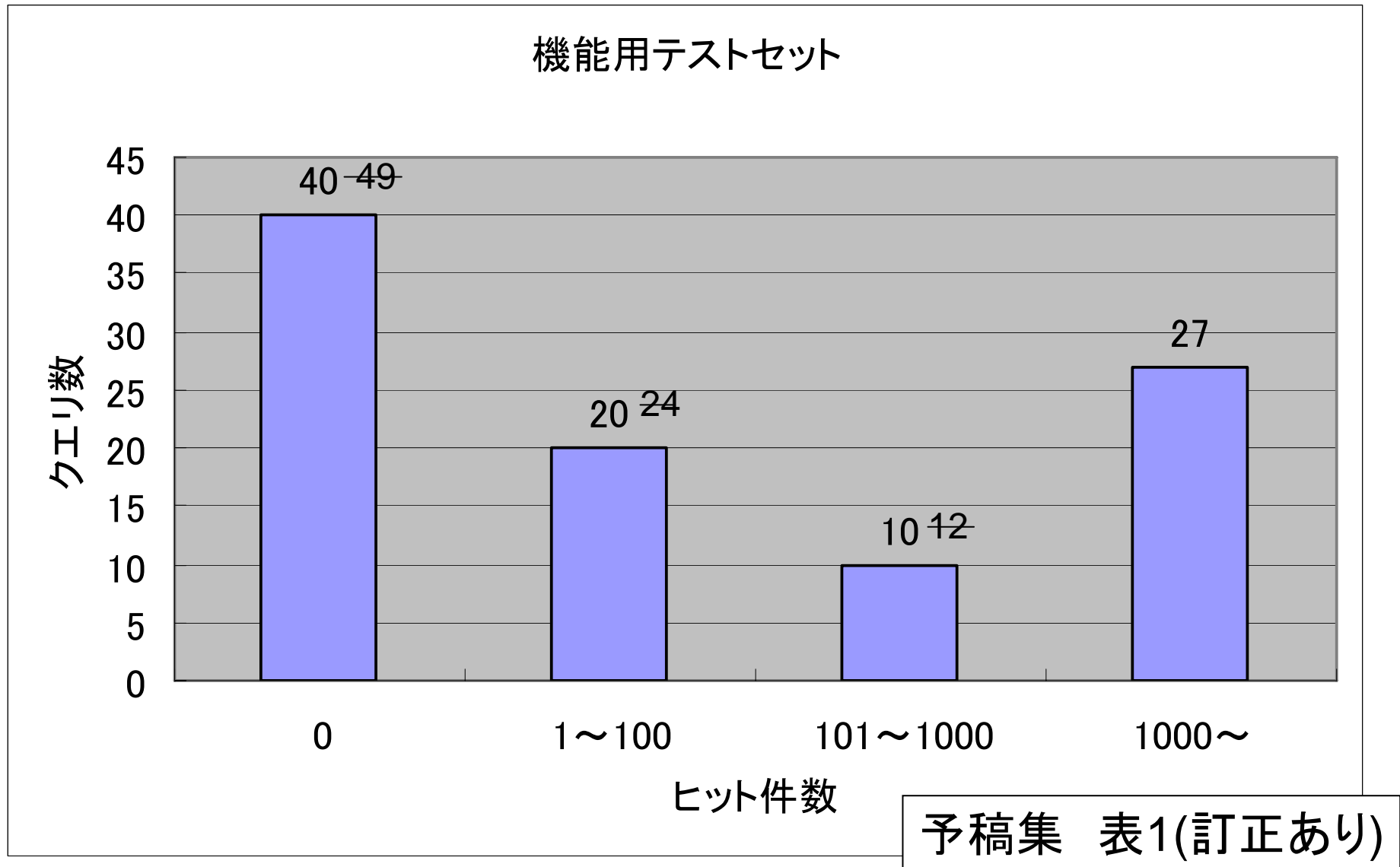
- サーバ側実装
 - YAZ + Net::Z3950::SimpleServer (IndexData社)
 - SRU/SRWプロトコルAPIとしてサーバ側で使用
 - CQL::Parser (Ed Summers氏)
 - CQL検索式の解析 (SQL文への変換は追加実装)
 - SQLite
 - 書誌データ蓄積ツール (RDBエンジン)
- クライアント側実装
 - Apache HTTPD + Ruby CGIスクリプト
 - SRU Rubyライブラリ

クエリ・テストセット

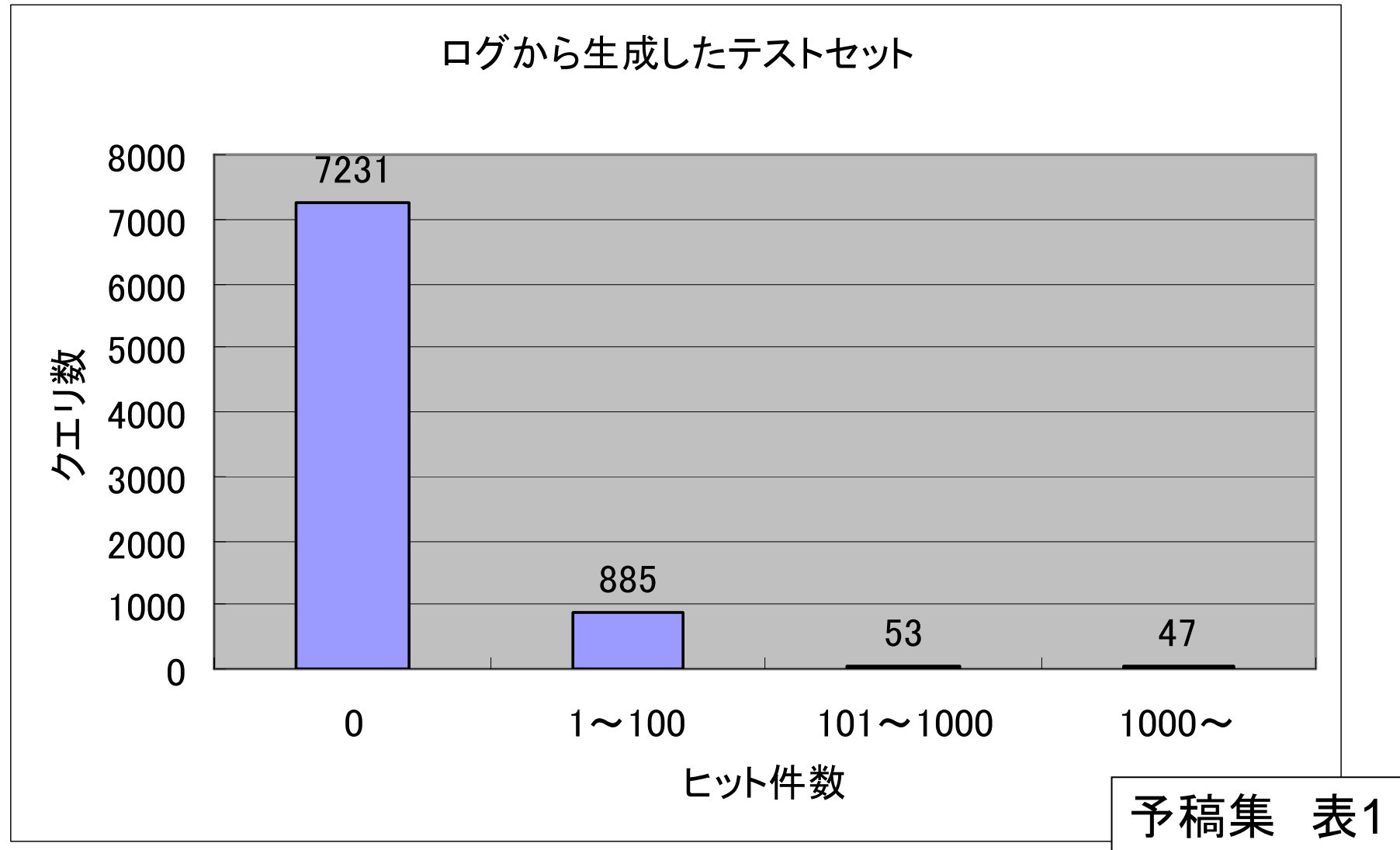
- EPI-Searchサーバがどの程度の検索機能に対応するかを確認
- 機能クエリセット
 - 機能チェック
 - 128クエリ
- 教育図書館OPACログから抽出したクエリセット
 - リアルな検索要求として
 - 8216クエリ

2つのクエリセットの開発

機能用テストセットのヒット件分布



OPACログから抽出したテストセット のヒット件分布



考察(1)

- 日本語検索
 - CQLにおいてword概念が中心に設計されている
 - SRU/SRWが欧米を中心に開発されているため
 - 日本語のような分かち書きされていない言語でのCQLプロファイルの必要性
 - =, any, all などの語に対するリレーション
- 文字コード
 - SRU/SRWはUnicodeのやり取りを指定
 - Unicodeを基本とするXML処理系の普及
 - 文字コードをやり取りする際の煩雑さは減少(Z39.50などと比べて)

考察(2)

- サーバ側における選択 (ServerChoice)
 - サーバ側で選択できる範囲を明示
 - 提供側でもサーバの挙動についての文書を提供し、情報共有を図る必要がある
- 開発が容易
 - RESTベースのWebアプリケーション開発手法をそのまま利用できる
 - 一日で構築できた！
 - ただし、SRU/SRWの基本的な機能の実装だけであれば
 - 開発工数のかかること: 検索システムにおける細かな実装
 - 表示、ソート、エラー処理、ユーザインタフェースカスタマイズなど

まとめ

- 基本機能を持つSRU/SRWサーバとSRUクライアントとの構築
- 実運用規模で開発
 - 約12万件の実データで確認
 - OPACログデータからクエリ生成
- 日本語環境における対応状況の確認