

2000年1月25日(火)  
修士論文最終発表会

# Z39.50に基づく書誌データ検索 システムの構築

98513 高久雅生  
石塚研究室所属

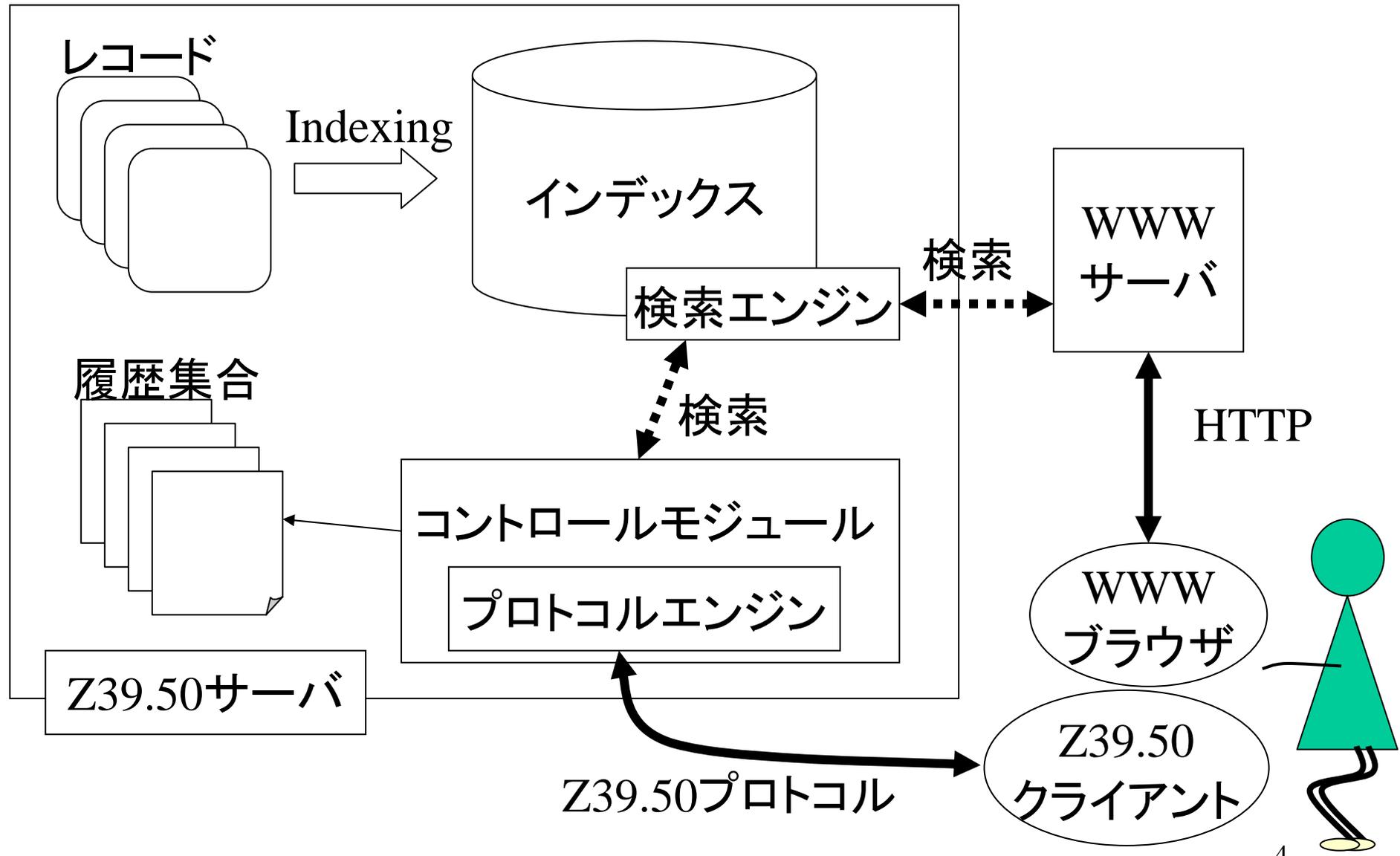
# 研究の背景と目的

- 背景
  - 異種データベース間での利用の問題<sup>[1][2]</sup>
  - Z39.50<sup>[3]</sup>
    - 情報検索のための国際標準プロトコル
    - 欧米を中心に普及：図書館OPAC, 商用データベース
    - 特徴：多種多様なデータベースの利用
- 目的
  - 実験システムの構築
    - 情報資源の共有
    - 相互運用性の向上
    - 日本語検索の問題など

# 実験システム

- 2種類のZ39.50システムの構築
  - ① JAPAN/MARC検索システム<sup>[4]</sup>
  - ② Dublin Core検索システム<sup>[5]</sup>
- 特徴
  - Z39.50とWWWの両者で利用可能
  - 大規模データに対応
  - 相互運用性の向上

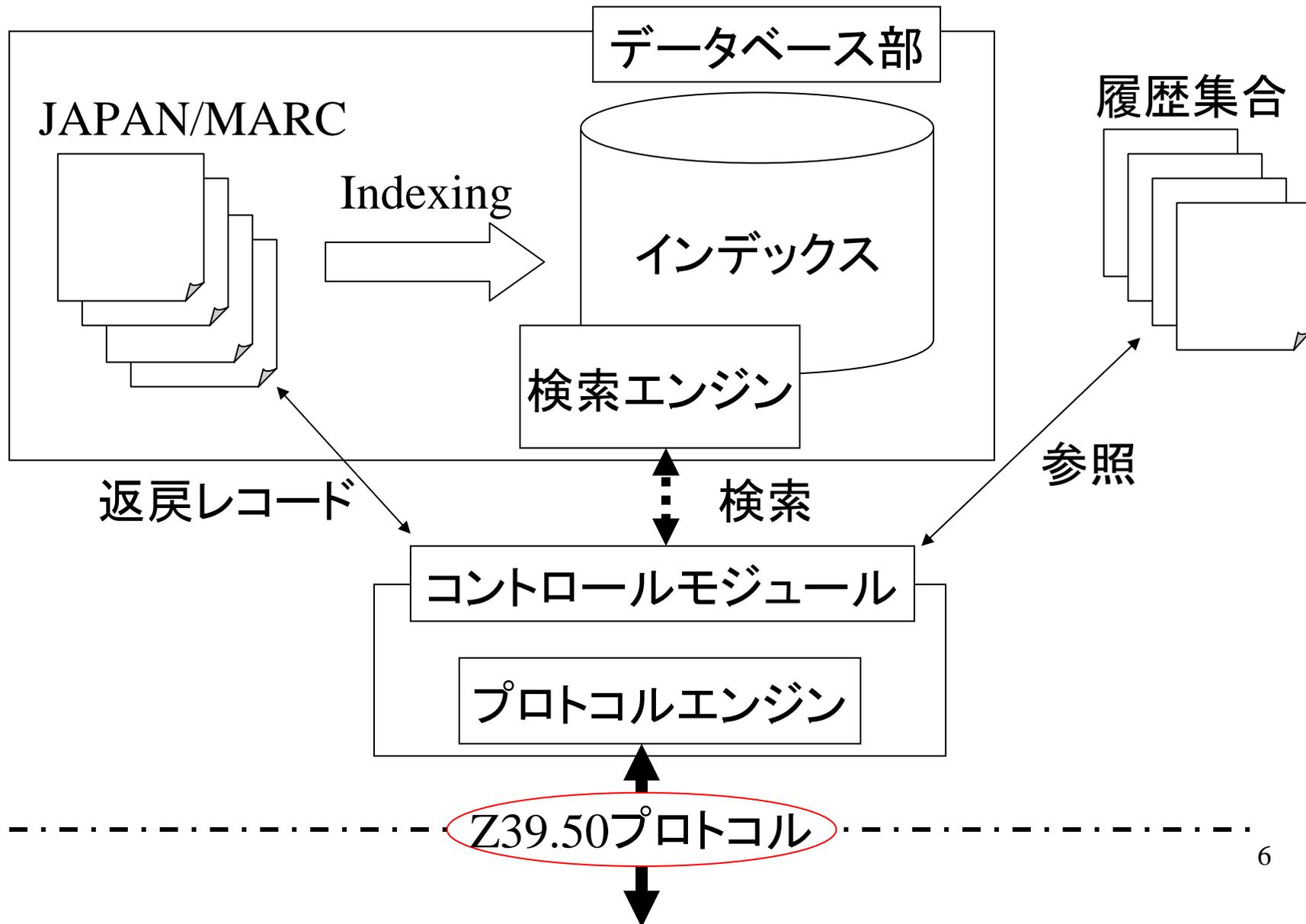
# 実験システムの概念図



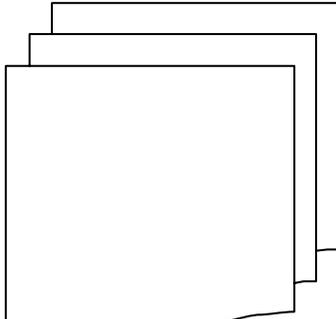
# JAPAN/MARC検索システム

- 日本語書誌データ
  - JAPAN/MARC<sup>[6]</sup>
- 大規模データ
  - 約102万件
- 情報資源の共有
  - 全文検索システムNamazu<sup>[7]</sup>

# 構成

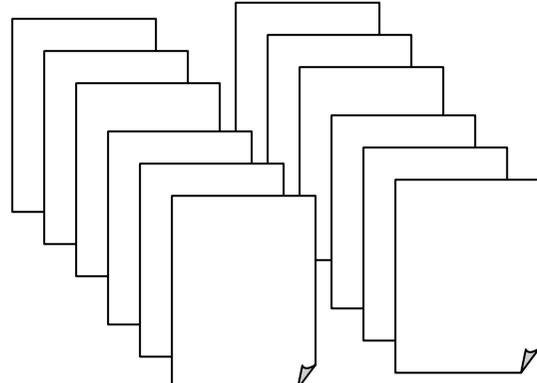
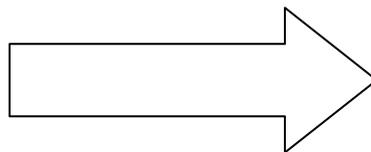


# JAPAN/MARCの前処理



JAPAN/MARC  
レコード  
(配布MT)

jmarcfilter.pl



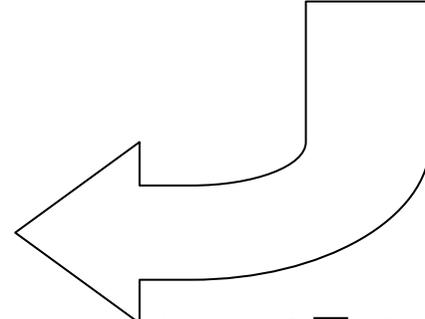
JAPAN/MARC  
テキストレコード  
(1レコード1ファイル形式)

アクセスポイントは、

- 書名(251~259、551~559)
  - 著者名(751~759、791~799)
  - ISBN(010\$A)
  - 出版社(270\$B)
  - 件名(650、658)
  - 注記(350、354、360、377)
  - (上記を含む全て)
- を抽出。



Namazu形式  
インデックス



JAPAN/MARC用フィルター付  
Namazuインデクサ

# Bib-1とJAPAN/MARCの対応

Bib-1	JAPAN/MARC	タグの意味
ISBN	010\$A	ISBN
Title	251-259 (\$A,\$B,\$D) 551-559 (\$A,\$X,\$D)	記述フィールド (書名,副書名,巻次等) 書名アクセスポイント (カタカナ形,ローマ字形,巻次の読み)
Author	751-759 (\$A,\$X,\$B) 791-799 (\$A,\$X,\$B)	著者名アクセスポイント (カタカナ形,ローマ字形,漢字形) 多巻ものの著者名アクセスポイント (カタカナ形,ローマ字形,漢字形)
Subject Headings	650 658	個人名件名標目 一般件名標目
Publisher	270\$B	出版社,頒布者など
Any	任意のフィールド	すべて

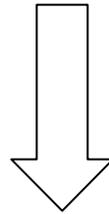
# Dublin Core検索システム

- アトリビュートセットのマッピング問題

[8]

- メタデータによる解決

- Dublin Core, RDF



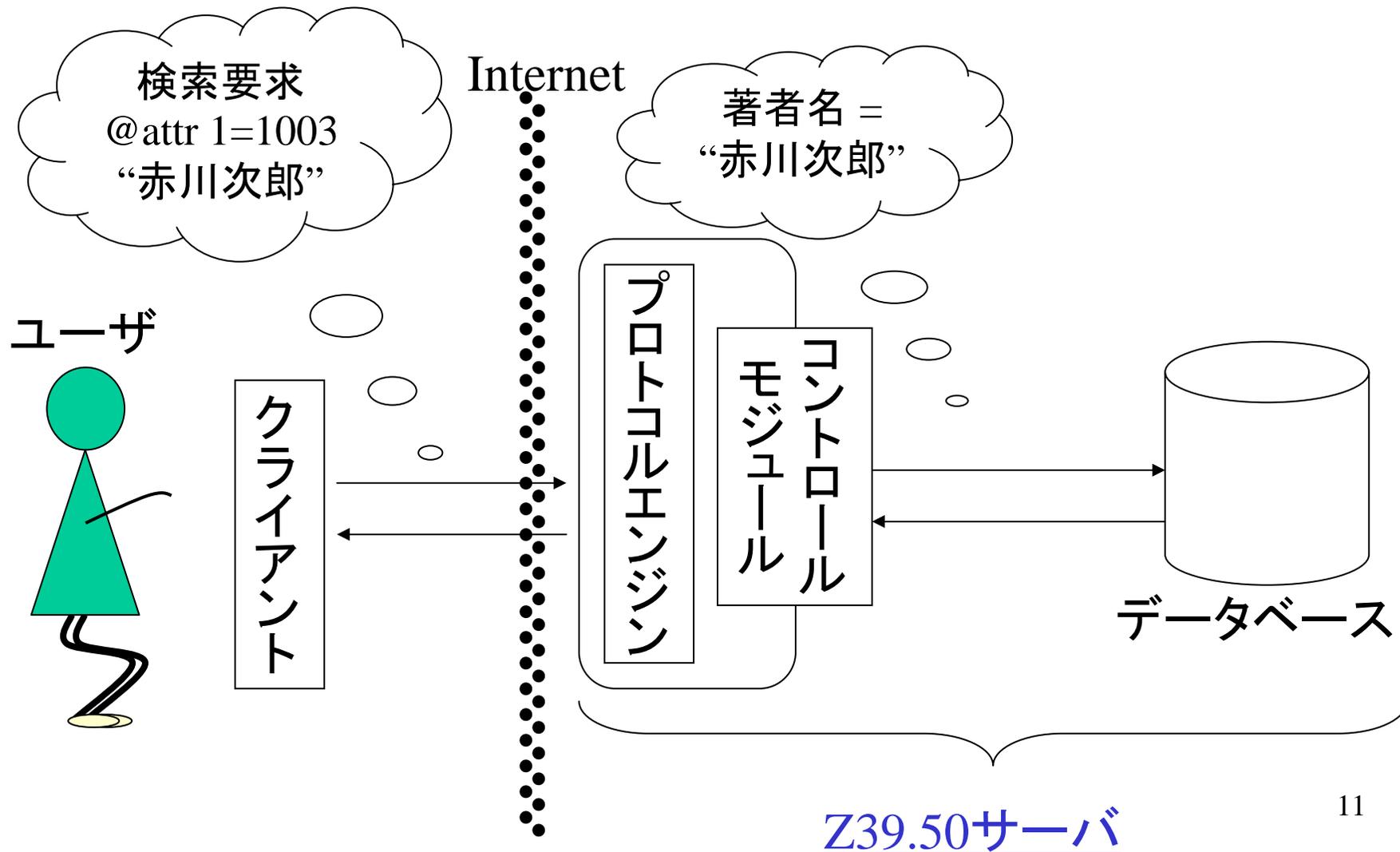
相互利用性の一層の向上

# アトリビュートセット

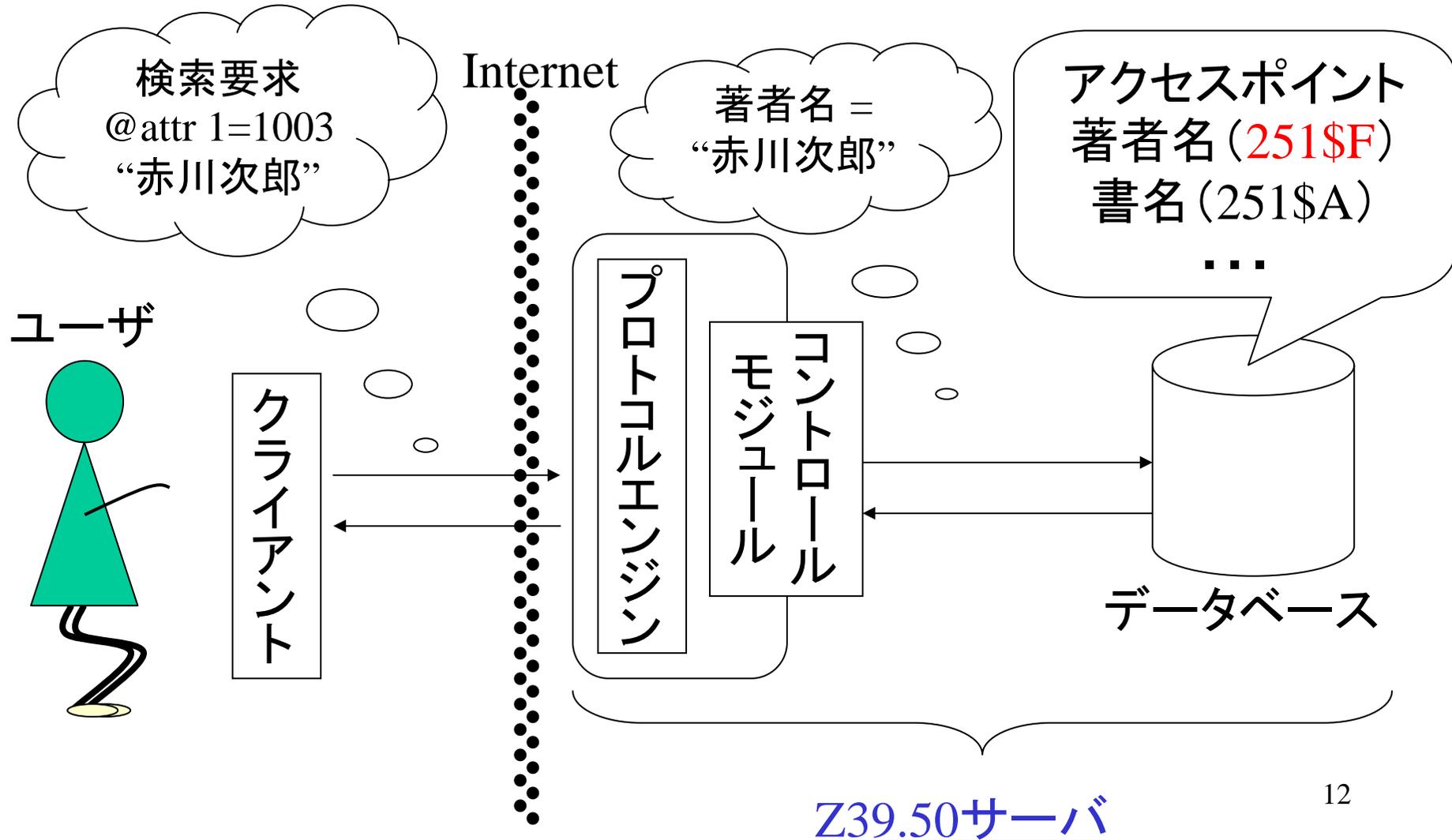
- Z39.50における検索用スキーマ
  - データベースの実装に依存しない論理スキーマ
  - 中心的な役割を担う Bib-1（書誌情報）<sup>[9]</sup>
    - 200以上の網羅的なアクセスポイント

Title	4	Note	63
ISBN	7	Author	1003
Subject heading	21	Any	1016
Date	30	Publisher	1018

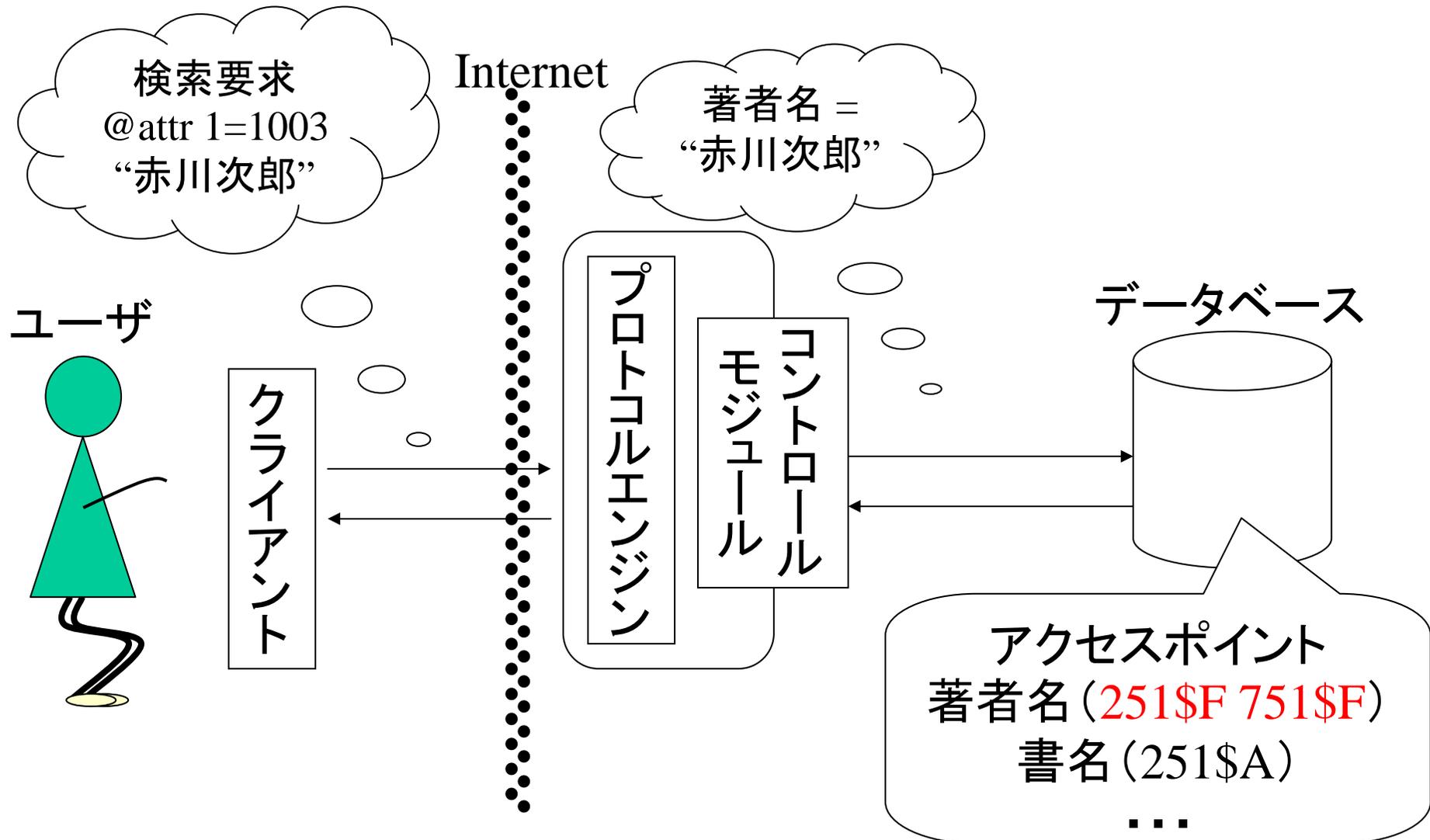
# アトリビュートセットによる検索



# アトリビュートセットの問題点： システムAのマッピング



# アトリビュートセットの問題点： システムBのマッピング



# Dublin Core Metadata Element Set<sub>[10]</sub>

- 15 項目の基本エレメント

Title	Date	Format
Creator	Type	Source
Subject	Identifier	Relation
Description	Language	Coverage
Publisher	Contributor	Rights

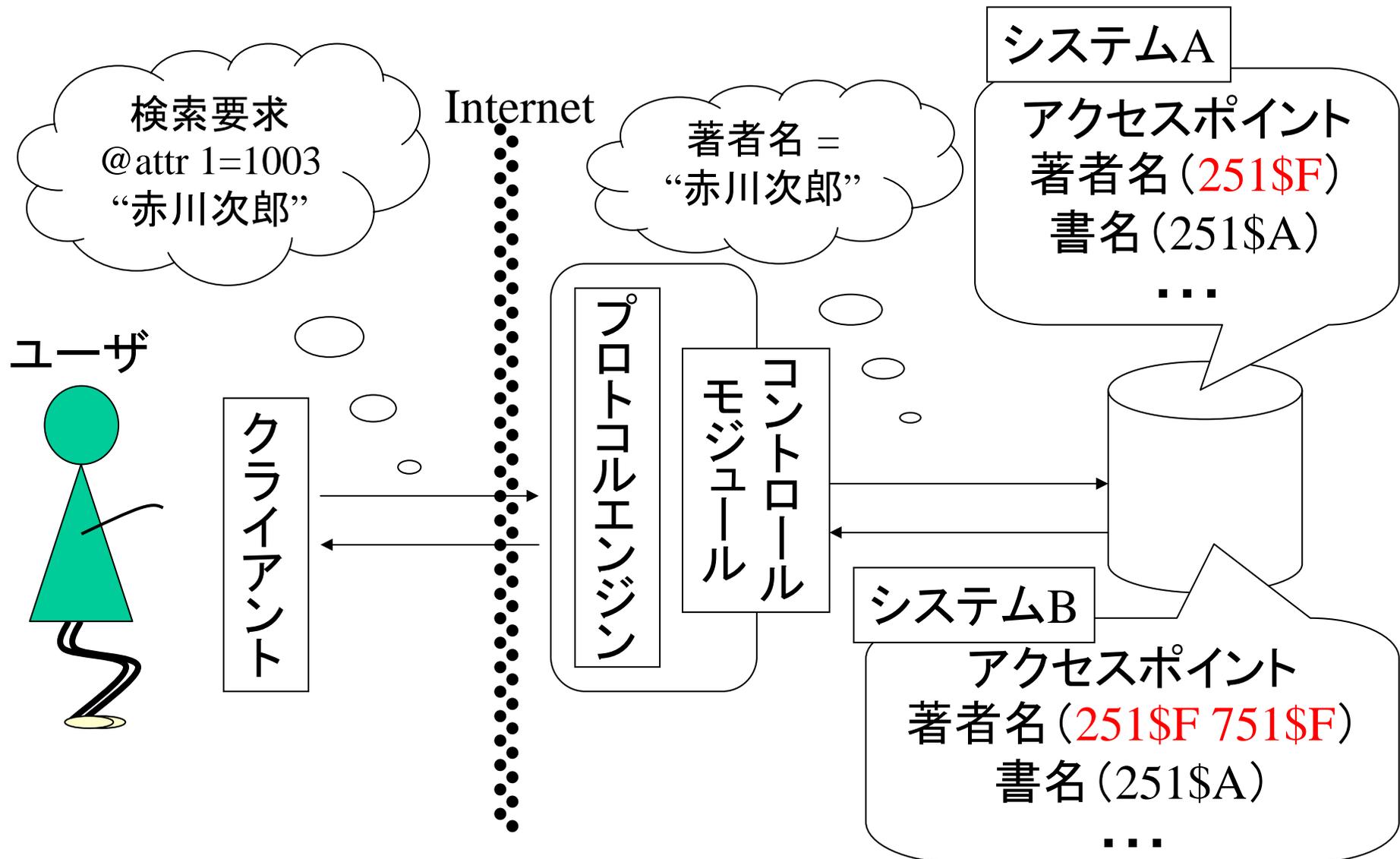
- 情報資源発見のためのメタデータ
- (意味的な)相互利用性

# Bib-1におけるDublin Core

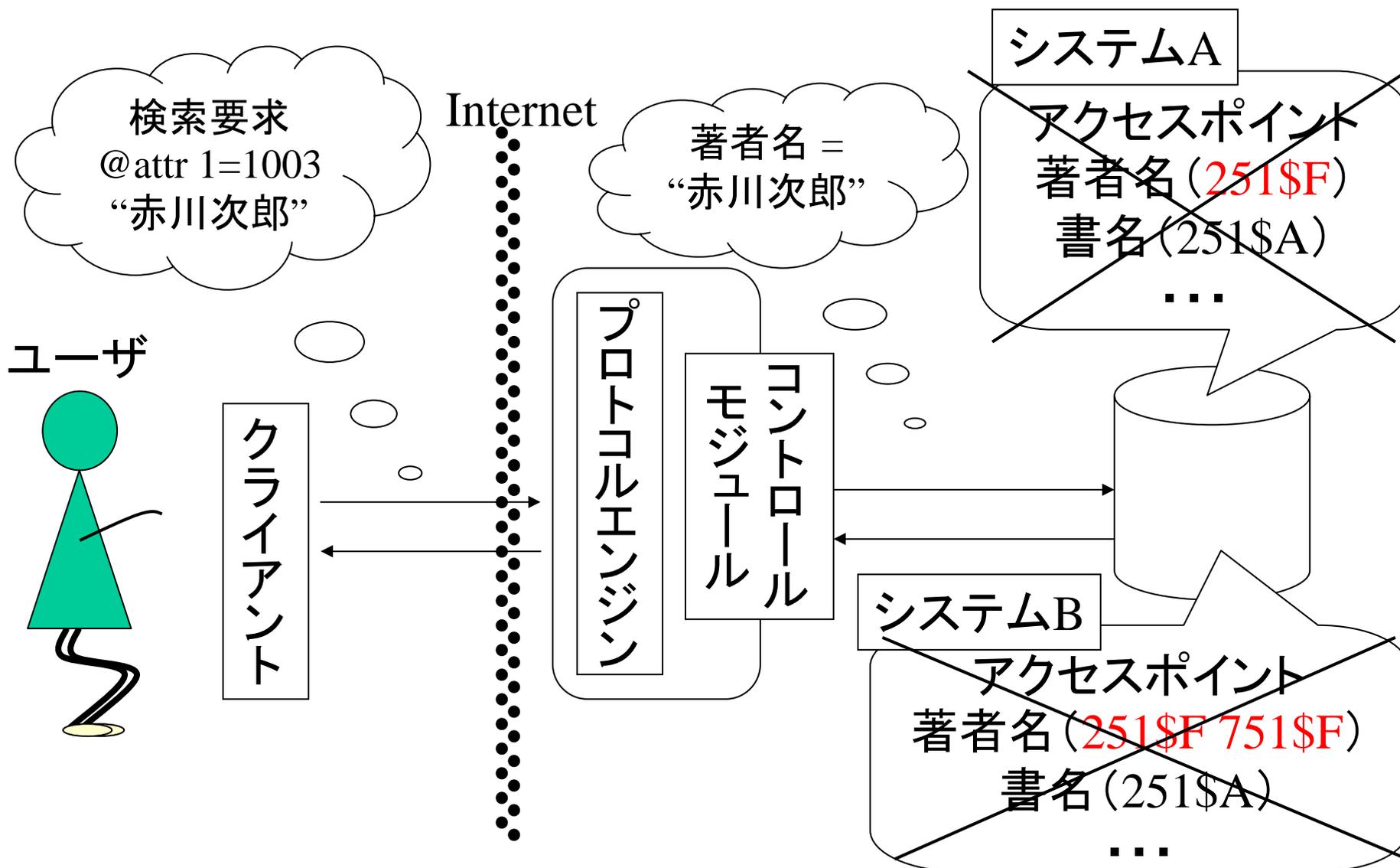
- Bib-1のDublin Coreアクセスポイント<sup>[11]</sup>

Dublin Core	Z39.50 Bib-1 Use Attribute	
	Name	Value
Title	DC-Title	1097
Creator	DC-Creator	1098
Subject	DC-Subject	1099
Description	DC-Description	1100
Publisher	DC-Publisher	1101
Date	DC-Date	1102
Type	DC-ResourceType	1103
Identifier	DC-ResourceIdentifier	1104
Language	DC-Language	1105
Contributor	DC-OtherContributor	1106
Format	DC-Format	1107
Source	DC-Source	1108
Relation	DC-Relation	1109
Coverage	DC-Coverage	1110
Rights	DC-RightsManagement	1111

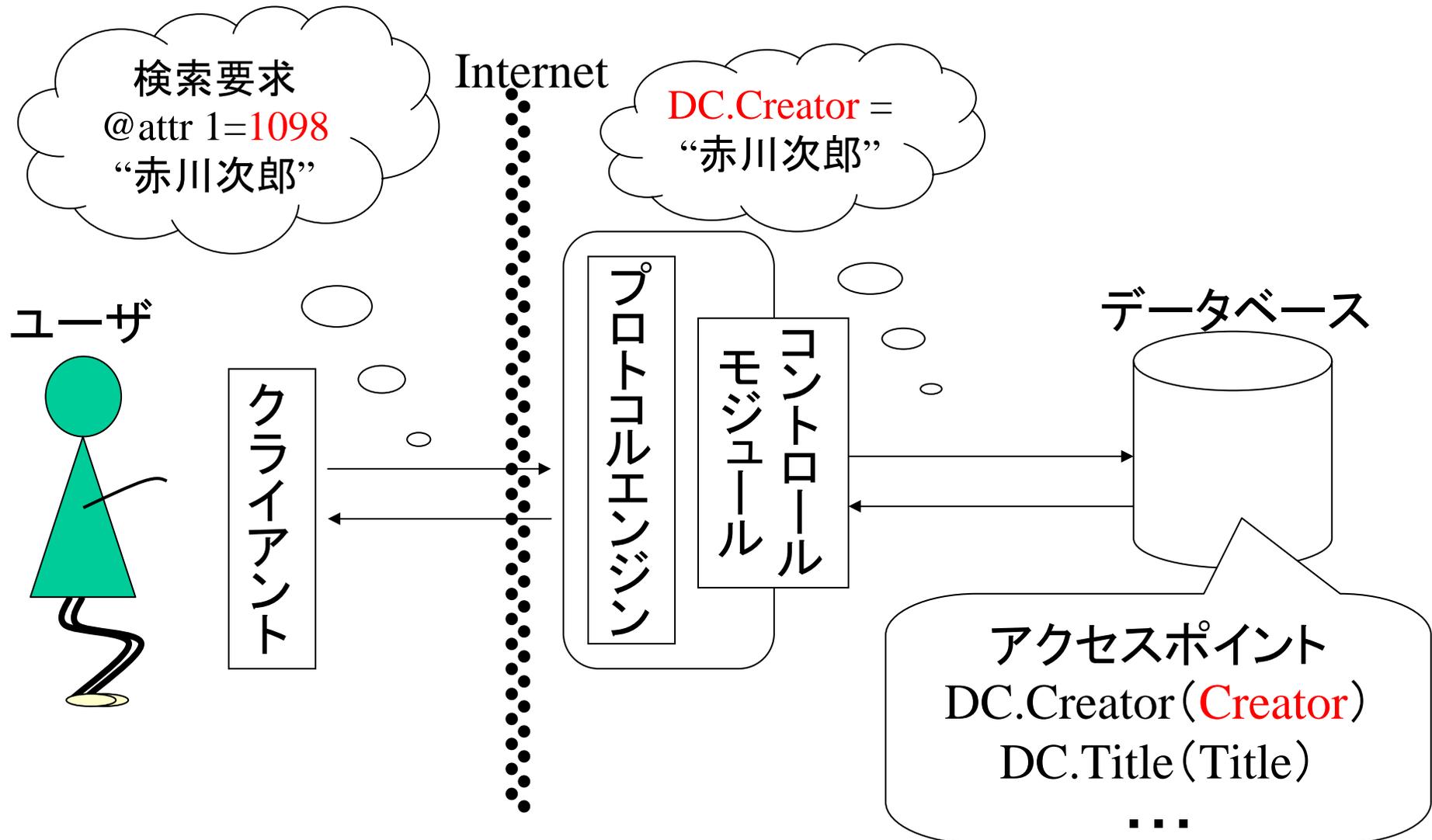
# Dublin Core による解決策(1)



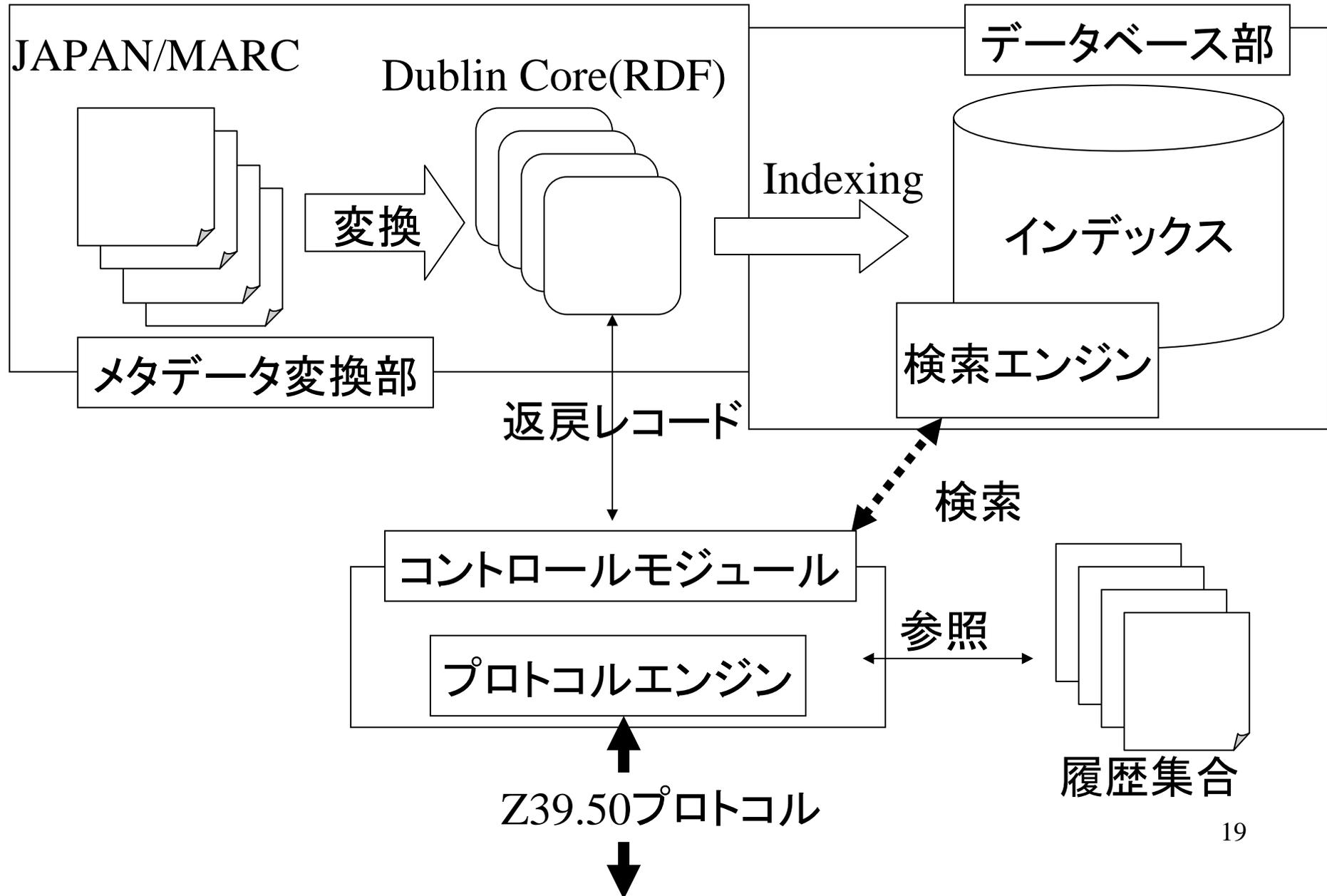
# Dublin Core による解決策 (2)



# Dublin Core による解決策 (3)

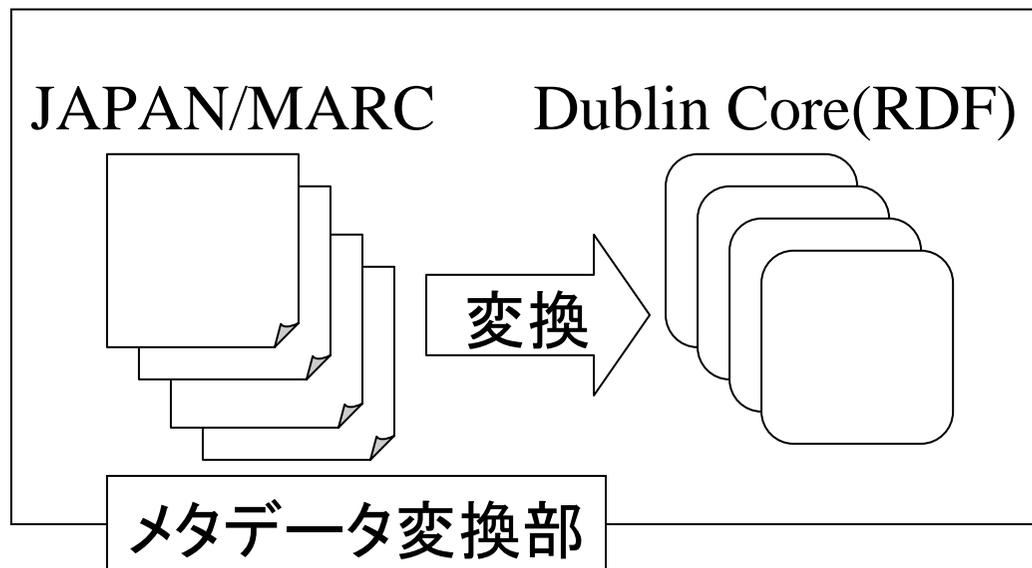


# 構成



# メタデータ変換部

- 既存の書誌データの変換～Dublin Core
  - JAPAN/MARC
- メタデータ記述のための枠組
  - RDF(Resource Description Framework)



# RDF

## (Resource Description Framework)<sup>[12]</sup>

- メタデータ記述のための枠組み
- モデル: ラベルつき有向グラフ
- 記述: XMLでの表現



```
<RDF>  
<Description about="ある本">  
<著者>山田太郎</著者>  
</Description>  
</RDF>
```

# メタデータ変換部: レコード例

00198021725  
020\$AJP  
.....  
270\$A東京  
**270\$D1996. 10**  
275\$A2冊(資料編とも)  
275\$B30cm  
.....  
658\$B十進分類法  
677\$A014. 45  
677\$V9  
685\$AUL655  
.....

```
<rdf:Description about="">  
  <dc:subject>  
    <rdf:Bag>  
      <rdf:li>十進分類法</rdf:li>  
      <rdf:li>014. 45</rdf:li>  
      <rdf:li>UL655</rdf:li>  
    </rdf:Bag>  
  </dc:subject>  
  .....  
  <dc:coverage>東京</dc:coverage>  
  <dc:date>1996. 10</dc:date>  
</rdf:Description>  
</rdf:RDF>
```

# メタデータ変換部: マッピング

Dublin Core	JAPAN/MARC	フィールドの意味
Subject	658\$B	一般件名: 漢字形
	677\$A	NDC: 分類記号
	685\$A	NDLC分類: 分類記号(またはかな付)
Date	270\$D	出版・領布に関する事項: 出版、領布年月
Relation	(該当無し)	
Coverage	270\$A	出版・領布に関する事項: 出版地、領布地等
(該当無し)	001	レコード識別番号: レコードコントロール番号
	020\$A	全国書誌番号: 国名コード

# メタデータ変換部: レコード例

00198021725  
020\$AJP  
.....  
270\$A東京  
270\$D1996. 10  
275\$A2冊(資料編とも)  
275\$B30cm  
.....  
658\$B十進分類法  
677\$A014. 45  
677\$V9  
685\$AUL655  
.....

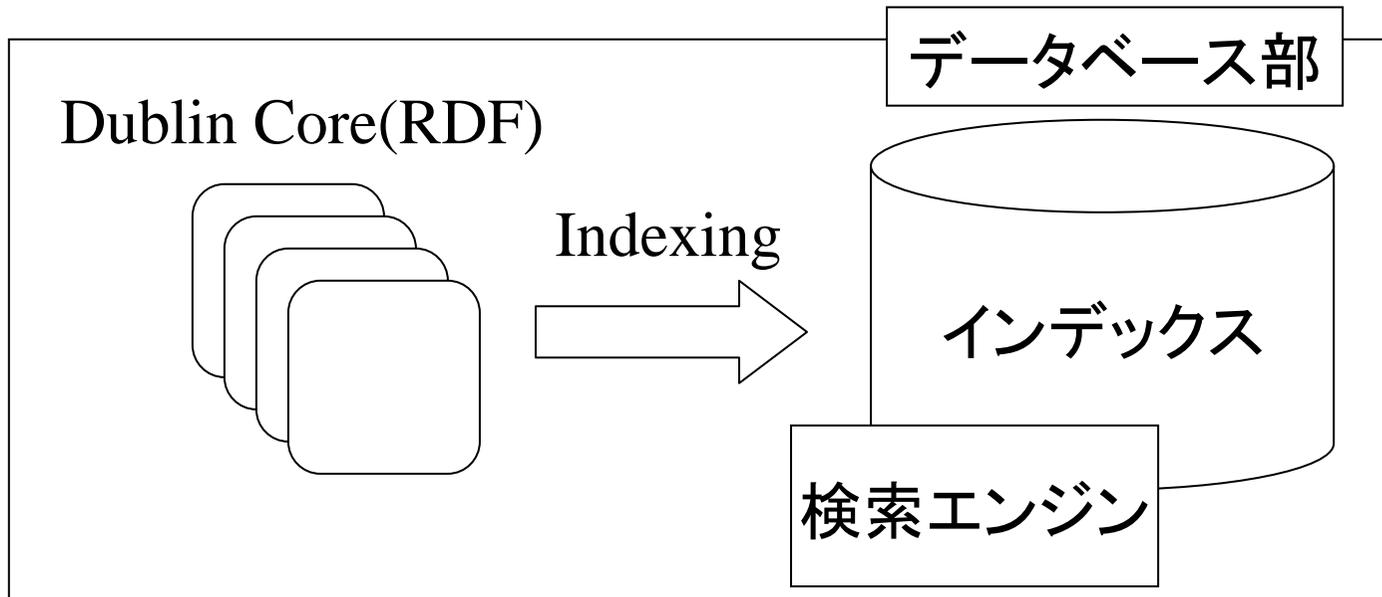
JAPAN/MARC例(資料1)

```
<rdf:Description about="">  
  <dc:subject>  
    <rdf:Bag>  
      <rdf:li>十進分類法</rdf:li>  
      <rdf:li>014. 45</rdf:li>  
      <rdf:li>UL655</rdf:li>  
    </rdf:Bag>  
  </dc:subject>  
  .....  
  <dc:coverage>東京</dc:coverage>  
  <dc:date>1996. 10</dc:date>  
</rdf:Description>  
</rdf:RDF>
```

Dublin Core例(資料2)

# データベース部

- 全文検索エンジンNamazu
  - フィルタ機構
- Dublin Core エLEMENTの抽出
  - RDFの解析



# システムの検索例

## Dublin Core検索

検索画面 <nile.ulis.ac.jp:DC>

File Connect Help

Search Query:  clear

show [ 2 ] No 1 ~ No 2 Full Brief SUTRS

[ 1 ] 検索式: dc.title="河童" ヒットした件数:1  
[ 2 ] 検索式: dc.creator="赤川次郎" ヒットした件数:2

```
<?xml version="1.0" encoding="EUC-JP"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.0/">
  <rdf:Description about="">
    <dc:subject>9 1 3. 6</dc:subject>
    <dc:creator>
      <rdf:Bag>
        <rdf:li>赤川次郎 著</rdf:li>
        <rdf:li>アカガワ, ジロウ</rdf:li>
        <rdf:li>Akagawa, Ziro</rdf:li>
      </rdf:Bag>
    </dc:creator>
    <dc:title>
      <rdf:Bag>
        <rdf:li>小豆色のテーブル</rdf:li>
        <rdf:li>長編青春ミステリー</rdf:li>
      </rdf:Bag>
    </dc:title>
  </rdf:Description>
</rdf:RDF>
```

## JAPAN/MARC検索

検索画面 <nile.ulis.ac.jp:JPMARC>

File Connect Help

Search Query:  clear

show [ 3 ] No 1 ~ No 19 Full Brief SUTRS

[ 1 ] 検索式: ガラス ヒットした件数:667  
[ 2 ] 検索式: 工芸 ヒットした件数:1373  
[ 3 ] 検索式: set=1 and set=2 ヒットした件数:19

19件のレコードが返ってきました

[3] No1 DBname:JPMARC Record type:SUTRS

00183025426  
020\$AJP  
020\$B83025426  
100\$A19830607 1982 H0JPN 1312  
251\$A近代日本のガラス工芸  
251\$B明治初期から現代まで  
251\$F東京国立近代美術館／編  
270\$A (東京)  
270\$B東京国立近代美術館  
270\$Dc 1 9 8 2  
275\$A 1 冊 (頁付なし)  
275\$B 2 4 c m

# まとめ

- Z39.50検索システムの構築
  - 情報資源の共有
  - アトリビュートセット問題の解決
    - Dublin Core メタデータ
    - データスキーマと検索スキーマの一貫性

# 今後の課題

- Z39.50の拡張機能
- Dublin Core – JAPAN/MARC
  - マッピングの共有
  - RDFを利用したより詳細な記述

## 参考文献

- [1] Lynch, Clifford A. The Z39.50 Information Retrieval Standard : Part I: A Strategic View of Its Past, Present and Future. D-Lib Magazine. 1997.  
URL: <http://www.dlib.org/dlib/april97/04lynch.html>.
- [2] 上田修一. Z39.50 とその可能性. 情報の科学と技術. Vol. 48, No. 3, 1998, p. 126-133.
- [3] ANSI/NISO Z39.50-1995. Information Retrieval ( Z39.50 ) : Application Service Definition and Protocol Specification. 1995, 156p.  
URL: <ftp://ftp.loc.gov/pub/z3950/official/>.
- [4] 宇陀則彦, 江草由佳, 高久雅生, 石塚英弘. Z39.50 による日本語書誌データ検索システム. 情報知識学会誌. Vol. 9, No. 2, 1999, p. 1-15.
- [5] 高久雅生, 江草由佳, 宇陀則彦, 石塚英弘. Z39.50 による書誌データ検索システムの構築 : Dublin Core を共通スキーマとして. デジタル図書館. No. 16, 1999, p. 97-106.
- [6] 国立国会図書館. JAPAN/MARC マニュアル : 図書編. 第 2 版. 東京, 国立国会図書館, 1998, 186p. ( ISBN 4-87582-527-7 ).
- [7] 高林哲. 全文検索システム Namazu. last update 1999-09-28.  
URL: <http://openlab.ring.gr.jp/namazu/>.
- [8] Lynch, Clifford A. Building the Infrastructure of Resource Sharing : Union Catalogs, Distributed Search, and Cross-Database Linkage. LIBRARY TRENDS. Vol. 45, No. 3, 1997, p. 448-461.
- [9] Bib-1 Attribute Set. last update 1999-09-24.  
URL: <http://lcweb.loc.gov/z3950/agency/defns/bib1.html>.
- [10] Dublin Core Metadata Initiative. The Dublin Core Metadata Element Set Version 1.1. last update 1999-07-02.  
URL: <http://purl.org/dc/documents/rec-dces-19990702.htm>.
- [11] Ralph LeVan. Dublin Core and Z39.50. Draft Version 1.2. last update 1998-02-02.  
URL: <http://purl.org/DC/documents/notes-levan-19980202.htm>.
- [12] Resource Description Framework (RDF). last update 1999-12-06.  
URL: <http://www.w3.org/RDF/>.