

すべての図書館コンテンツを一つのプラットフォームに

工藤 絵理子

九州大学附属図書館 e リソースサービス室
kudoe@lib.kyushu-u.ac.jp

概要：次世代 OPAC は、表紙・内容・目次情報の提供、適合度によるソート、絞込み検索、スペルチェック／サジェスト機能など、利用者目線のユーザインターフェースにより注目を集めてきた。また最近では、従来からの冊子資料に加え、e リソースや機関リポジトリ・デジタルコンテンツ等の、図書館が提供する多様なコンテンツへの一元的アクセスを実現したディスカバリ・インターフェースも登場してきている。九州大学附属図書館では、海外のオープンソース・ソフトウェアである eXtensible Catalog (XC) によってディスカバリ・インターフェース Cute.Catalog を導入し、2010 年 4 月に試験公開した。本稿では、XC の概要と選定過程について説明すると共に、本学での導入プロセスや課題の解決、今後の展望について述べる。

キーワード：ディスカバリ・インターフェース、次世代 OPAC、eXtensible Catalog、XC、Cute.Catalog、Summon、Web プラットフォーム

1. 開発の背景

図書館の提供するコンテンツは従来の OPAC で検索可能な冊子体の資料にとどまらず、電子ジャーナル、電子ブック、文献データベース、デジタルコレクションなど多様化しており、これらの膨大なコンテンツの中から必要な情報へ到達することは、利用者にとって困難になってきている。そこで九州大学附属図書館では、2つのコンセプトとともに、ユーザに必要な情報へとナビゲートする、OPAC に変わる新しい検索プラットフォーム「Cute.Catalog」を開発することとした。

コンセプトの一つは「アクセス統合」である。資料によって異なるプラットフォームで提供されていたメタデータを一つのインターフェースに統合し、ただ一つの検索ボックスから、図1のように、図書館の持つあらゆるデータソースの統合検索が可能になることをめざした。

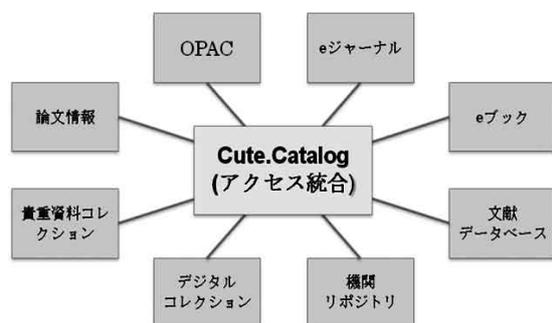


図1 Cute.Catalog のアクセス統合機能

もう一つが「次世代 OPAC」機能^[1]である。表紙画像や文字情報の視覚化、内容・目

次情報の表示、適合度ソート、ファセット検索、スペルチェック・サジェスト機能、利用者参加型機能など、Webの世界では一般的となっている Web2.0 の機能を取り入れた利用者目線のユーザインターフェースの構築を行った。

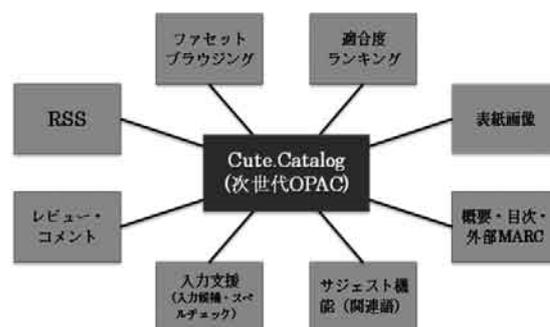


図2 Cute.Catalog の次世代 OPAC 機能

こうした Google の一元的検索に代表される「アクセス統合」機能と、Amazon に代表されるユーザナビゲート志向を取り入れた「次世代 OPAC」機能を併せ持ったディスカバリ・インターフェースの開発を進めることを決定した。^[2]

2. 選定までの経緯

Cute.Catalog は、eXtensible Catalog (以下 XC)^[3] というオープンソース・ソフトウェアにより作られている。XC は、米国のロチェスター大学がアンドリュウメロン財団やパートナー機関からの助成を受けて、綿密なユーザ動向調査を元に、図書館資源の発見・メタ

データ管理のためのツールとして開発された。現在は eXtensible Catalog Organization (以下 XCO) により運営されている。XC は、本学が設定した以下の選定条件を満たしていた。

- アジア言語の検索に対応し、日英インターフェースの切替え表示が可能なこと。
- 多様なデータソースからのメタデータが取り扱えること。
- 所蔵詳細表示やパーソナルサービスが、ディスカバリ・インターフェース単独での運用にて可能なこと。
- デザインや機能拡張に柔軟性があり、かつ持続的なアップデートが容易な設計であること。そのために、国際的な技術標準が積極的に採用されていること。

XCO により、本学が目指す実装、特にアジア言語や日本独特な状況への対応について連携して取り組む合意が得られたため、本学での採用を決定した。現在は、NTT データ九州と連携して XCO の開発運営コミュニティへ参加している。

3. eXtensible Catalog ソフトウェア

XC は、ユーザインターフェース (Drupal Toolkit)、メタデータ管理ソフトウェア (Metadata Services Toolkit)、接続ソフトウェア (OAI Toolkit および NCIP Toolkit) の 3 つの要素から構成されている。それぞれが独立して開発されているため、ディスカバリ・インターフェースを提供するだけでなく、特定のニーズを満たすために必要な要素のみ使用することも可能である。各ツールキットの設定はインターフェース上の管理画面から行い、特別なカスタマイズを行わない限り、プログラミングスキルは不要である。

3.1 ユーザインターフェース

Drupal Toolkit は、図書館システムの目録データや機関リポジトリ、Web サイト上のコンテンツなど、図書館の提供するコンテンツを統合して、単一のインターフェースからのアクセスを提供するためのツールキットである。コンテンツマネジメントシステム (Content Management System, CMS) である Drupal^[4] の拡張モジュールの一つで、拡張性や自由度の高さから、世界的に図書館を

含め多くの機関でウェブサイトの構築に採用されている。他のコミュニティが開発したさまざまな拡張モジュールを組み込むことも可能であることから、ディスカバリ・インターフェースとしてだけでなく、Web サイトや認証システムとの連携によるパーソナルサービスを提供するための図書館総合プラットフォームにもなりえるだけの高い機能を備えている。

インターフェースは全文検索エンジン Lucene/Solr を用いた検索機能やファセットブラウジング^{注1}をはじめ、NCIP Toolkit との連携による貸出状況のリアルタイム表示や認証システムとの連携によるパーソナルサービスなど、豊富な機能を備えている。また Drupal Toolkit へのデータ登録は、OAI-PMH^{注2[5]} によるデータハーベスティングによって行う。

3.2 メタデータ管理ソフトウェア

Metadata Services Toolkit (以下 MST) は、さまざまなデータソースから OAI-PMH によりハーベストしたメタデータを統合・整形し、Web アプリケーション上で利用可能にするためのツールキットである。メタデータスキーマは、現在のところ MARCXML^{注3} と Dublin Core^[6] (以下 DC) の 2 形式に対応している。集約されたメタデータは、Normalization Service と呼ばれる正規化を経たのち、FRBR 化された構造を持つ XC スキーマに標準化される。XC スキーマは、DC、RDA^[7]、および XC 独自の要素から構成される。標準化されたデータは、Drupal Toolkit などの外部アプリケーションから、OAI-PMH によるハーベスティングを行うことで利用できる。

注1 目録情報から標準化して生成されたキーワードを著者、フォーマット、ジャンル、言語、出版年、配架場所などのカテゴリーごとに提示させ、検索結果から必要なものを抽出する絞込検索に使用される。

注2 特定のアプリケーションに依存することなく、XML 形式を用いたメタデータを自動的に交換できる国際標準プロトコル。

注3 国際的な標準書誌データ形式である MARC21 を XML に対応させたもので、Library of Congress Network Development and MARC Standard Office によって開発された。XML 化された MARC レコードは、さまざまな XML メタデータとの相互変換が可能。

3.3 システム間接続ソフトウェア

接続ソフトウェアには、OAI ToolkitとNCIP Toolkitの2つのツールキットがある。

OAI ToolkitはOAI-PMHを使って、図書館システム内のメタデータをMSTに取り込むことができる。図書館システムとMST双方のレコードは、このツールキットにより自動的に同期をとり、最新の状態に保たれる。

NCIP Toolkitは、NCIP^{注4[8]}を動かすためのインターフェースを図書館システムに追加するためのツールで、ユーザインターフェース(Drupal Toolkit)など、外部アプリケーションがリアルタイムに図書館システムとデータ交換を行うことを可能にしている。今のところ国内の図書館システムではNALIS^[9]にのみ対応しているが、非対応の図書館システムについても、Javaプログラミングによってドライバを追加可能である。

4. Cute.Catalog 全体像

XCソフトウェアを実装したCute.Catalogの全体像は図3のとおりである。本学が所蔵する図書・雑誌のメタデータはCATP^{注5}形式に準拠しているため、まずはこれらデータソース①(電子ジャーナル、電子ブック、図書館Webサイトなど)上のメタデータをMARCXML形式に変換する必要がある。本学にて作成したマッピング表に基づき、MARCXMLファイルに合成したものを図書館システムにて生成し、OAI Toolkitを通じてMSTへ受け渡している。

データソース②に分類される機関リポジトリやデジタルコレクション上のデータは、DC形式で直接MSTに取り込んでいる。ただし、CiNiiやScopusなど論文データベースのメタデータは、九州大学所属研究者による文献のみあらかじめAPIで取り出してお

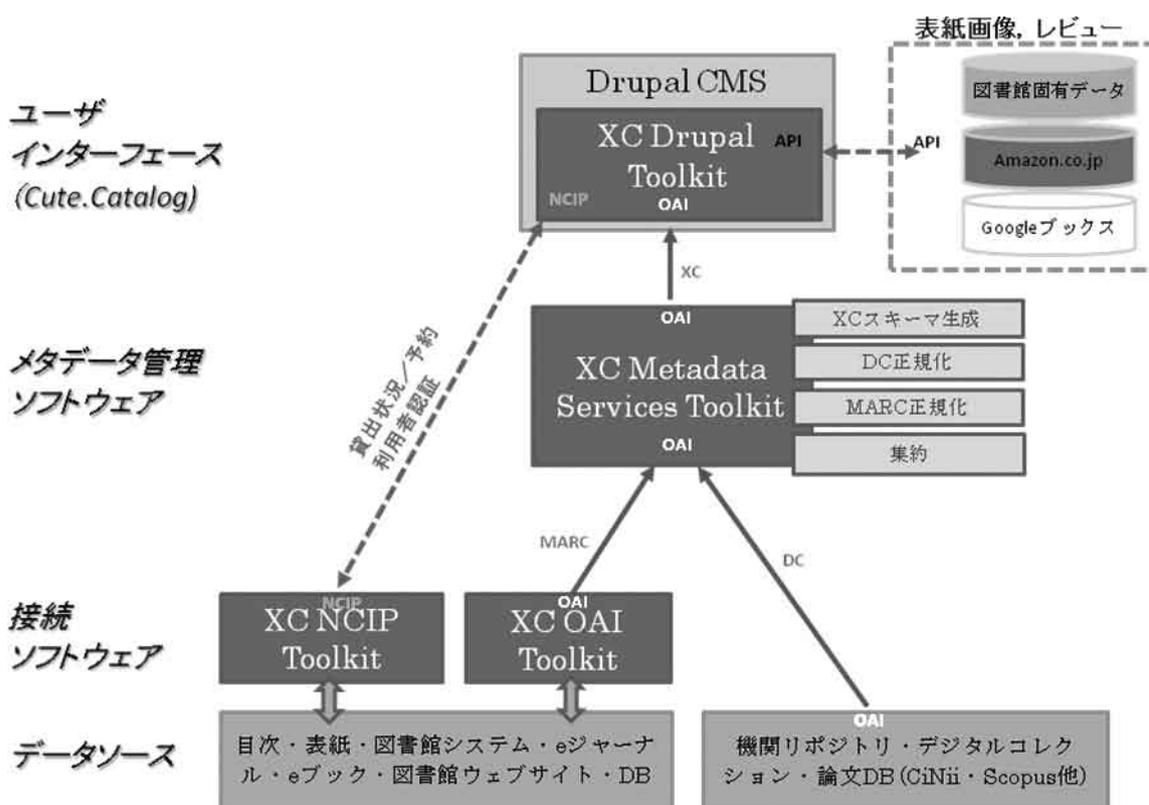


図3 Cute.Catalog 全体像

注4 NISO Circulation Interchange Protocolの略。図書館システムなどの間で、資料の貸出状況や利用者認証を伴うサービスに必要な情報をリアルタイムに交換する国際標準プロトコル「NISO/ANSI Z39.83」として規定されている。

注5 Cataloging information Access & Transfer Protocolの略。国立情報学研究所が運営する日本最大の総合目録所在情報データベースである「NACSIS-CAT」において、クライアントとサーバ間のメッセージ交換方式を規定するプロトコルである。

いたものを MST への受け渡し対象としている。この DC 取り込み部分については、NTT データと本学も開発に参加している。こうしてさまざまな形式に準拠していたメタデータが XC スキーマに統一され、再び OAI Toolkit によって、ユーザインターフェースである Drupal Toolkit へデータが受け渡され、ユーザによる検索が可能となるデータとなる。

ユーザインターフェースの実装例は、図 4 以下のとおりである。



図 4 検索画面

トップページは検索ボックスのみのシンプルな作りとし、目録情報だけでなく、目次・概要も検索対象としている。インターフェースの言語は Drupal の翻訳機能を使い、元の英語表記に日本語の翻訳をつけることで、日英の切替えを行っている。

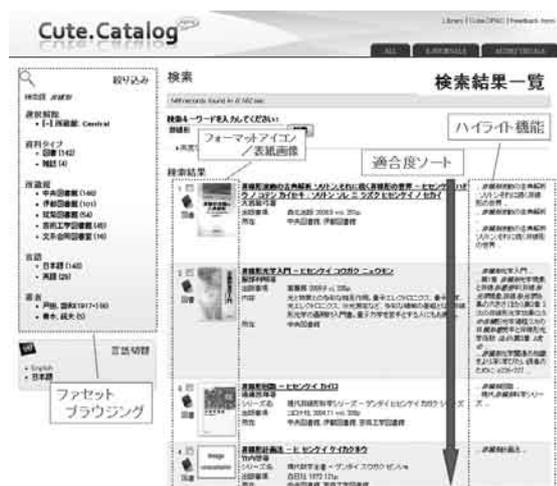


図 5 検索結果一覧

検索結果は適合度順に並べられ、館ごとの所在場所や内容が表示される。ビジュアル化

を重視し、資料フォーマットや Amazon の表紙画像も表示させた。右側の表示はハイライト機能とよばれるもので、入力されたキーワードにマッチした個所の前後を表示する。

図 5 画面左のファセットブラウジングが最大の特徴である。ファセットとは切り口という意味で、利用者をナビゲートする部分である。検索結果が膨大であっても、関連するファセットエレメントを選んでいくことで、検索結果を絞り込むことができる。現在、資料タイプ、所蔵館、本文の言語、著者を表示しているが、出版年、主題などのキーワードファセットも追加する予定である。

詳細表示画面 (図 6) では、ユーザにとって重要な情報 (資料タイプ、出版者、内容・目次情報) を上部に配置した。Amazon Web サービス API を利用し、表紙画像、当該資料の Amazon 詳細ページへのリンクおよび Amazon レビューを取得・表示している。なお、e リソース (電子ジャーナル、電子ブック) の詳細表示には、本文へのリンクが表示される。

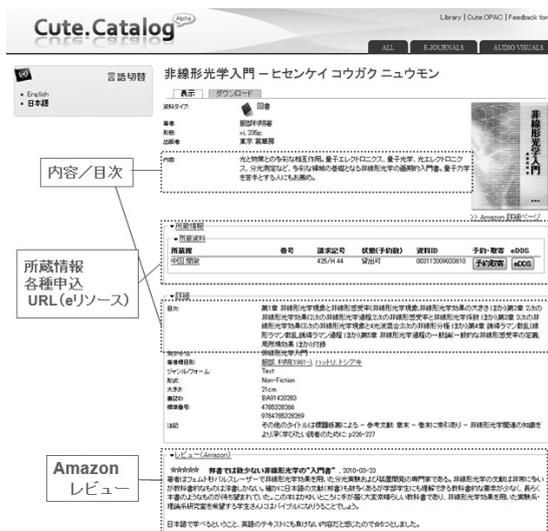


図 6 詳細画面 (図書)

5. 日本語環境への対応

XC は海外製のため、日本独自の環境への対応を行った。主なものは以下の 3 点である。

5.1 日本語インデクシング

日本語での検索を行うためには、日本語イ

ンデクシングに対応した検索エンジンのチューニングを行う必要がある。そこで、Drupal Toolkit が採用する Lucene/Solr に対して、以下の調整を行った。

- ① 異体字変換
- ② Unicode 正規化 (NFKC)
- ③ 文字列の置換 (土耳其からトルコなど)
- ④ 日本語固有の正規化 (平仮名からカタカナへの変換等)
- ⑤ KAKASI によるローマ字文字列の作成
- ⑥ CJK Tokenizer による 1 文字毎のトークンの切り出し (未知語対応)
- ⑦ Sen によるトークンの切り出し (形態素解析)

5.2 日本語ヨミの扱い

XC は海外製のソースであるため、日本語検索には欠かせない「ヨミでの検索」という概念が存在しない。また、XC も準拠している、アメリカ議会図書館が開発した MARC21 形式がもはやメタデータの世界標準フォーマットともいわれる中で、日本では CATP など国内仕様のフォーマットを使用している。このような背景により、日本で XC を機能させていくためには、MARCXML・XC スキーマを CATP で機能している「ヨミ」に対応させねばならない。そこで、MARCXML 上の複数言語による記述の役目を持つフィールド (880 番フィールド) へマッピングし、CATP 上のオリジナルフィールドとリンク関係を持たせた。

5.3 主題ファセット

資料の内容に係わる主題ファセットについても、アメリカ議会図書館作成の LCSH^{[10][11]}注6 が世界的に広く採用されている。しかし、北米に関わるキーワードは豊富でも、アジアに関わるものは少ないといった地域性の偏りもあるため、日本の資料のナビゲートにはそぐわない。そこで、NDC (日

注6 米国議会図書館件名標目表。資料の主題や形式についてのキーワードを、一定の規則に基づく統制語で作成したものを「件名」と呼ぶ。件名として採用した統制語が「件名標目」となり、これを一定の規則や順序によって配列したものが「件名標目表」となる。この件名標目を書誌データに付与することにより、統制されたキーワードによる、高精度な検索が可能となる。

本十進分類法) 記号から抽出したキーワードや、国会図書館作成の NDLSH (国立国会図書館件名標目表)^[12] の導入などを検討している。

6. 今後の課題と展望

6.1 グローバルスタンダード

今回の Cute.Catalog 構築によって、改めて気づいた点がある。すなわち、国際規格の積極的な採用の重要性である。技術進歩の激しい昨今の状況では、システム構築においてもグローバルスタンダード (国際標準) を採用してゆくことがますます必要になってくる。

XC においては、表 1 のように、通信プロトコルやメタデータフォーマットにはグローバルスタンダードといわれる規格を、検索エンジンやインターフェースなどには、オープンソースの中でも世界規模で幅広く受け入れられているものを採用している。そのため、機能改善や追加の必要があっても、世界中のユーザからなるコミュニティ全体で対応することができる。もちろん、オープンソースだからといってコストが削減できるわけではないし、維持・発展させるにはコストや人材がかかる。しかしながら、知識やノウハウの蓄積、(図書館主体で行える形での) 実現性の高さ、国際化など、得られる知識や経験は非常に大きい。

表1 XCにおけるグローバルスタンダード

機能	国際標準
メタデータ	MARCXML, Dublin Core (DC terms)
メタデータ受渡し	OAI-PMH
図書館システムのリアルタイム参照/更新	NISO Circulation Interchange Protocol - Z39.83 (NCIP)
検索エンジン	Apache Solr / Lucene
インターフェース	Drupal

6.2 Summon との連携

Cute.Catalog の開発と並行して、世界的に注目を集めている Web スケール・ディスカバリの一つである Summon^[13] の導入も進めている。これは、電子ジャーナル、電子ブック、文献データベースなど、ライセンス契約

している学術コンテンツをインデクシングサーバに集め、Googleのような一括検索を提供するツールである。Webスケール・ディスカバリは図書館の書誌・所蔵レコードや機関リポジトリを同時に扱うことで、学術情報全体への迅速なアクセスを実現している。連携の具体的な方法は、XCのMSTを使ってローカルコンテンツのメタデータを集約・正規化し、単純にこれをSummon側に渡すやり方で行う。

Webスケール・ディスカバリを検索の入口として導入し、XCをローカルコンテンツへのアクセスやパーソナルサービスを提供する受け皿にできれば、ユーザの快適性をさらに高めることもできる。

6.3 今後の展望

Cute.CatalogおよびXCは進行中のプロジェクトであり、著者名典拠対応、適合度のカスタマイズ、スペルチェック・サジェスト機能、RefWorks^[4]へのダイレクトエクス

ポートなど、現在実装すべき機能がまだ残されている。これらは2011年4月の正式リリースに向け、現在実装中である。

こうした機能を発展させることにより、最終的に私たちが目指しているものは、コンテンツの統合だけでなく、すべての図書館サービスを図書館インターフェース(Web)上で一元化することにある。図7で示しているように、これらすべてを取り込んだ「図書館Webプラットフォーム」の開発も世界規模で始まっている。Cute.Catalogも同様に、将来的にはその検索機能をDrupalインターフェースによって本学Webサイトやパーソナルサービスと統合させ、「九州大学附属図書館Web」としてシンプルな形に進化させることを狙っている。これにより、世界中の膨大な学術情報に効果的にアクセスできること、機関リポジトリ上などにある大学の学術成果物への容易なアクセスを提供してその価値を高めること、の2点を実現させていきたいと考えている。

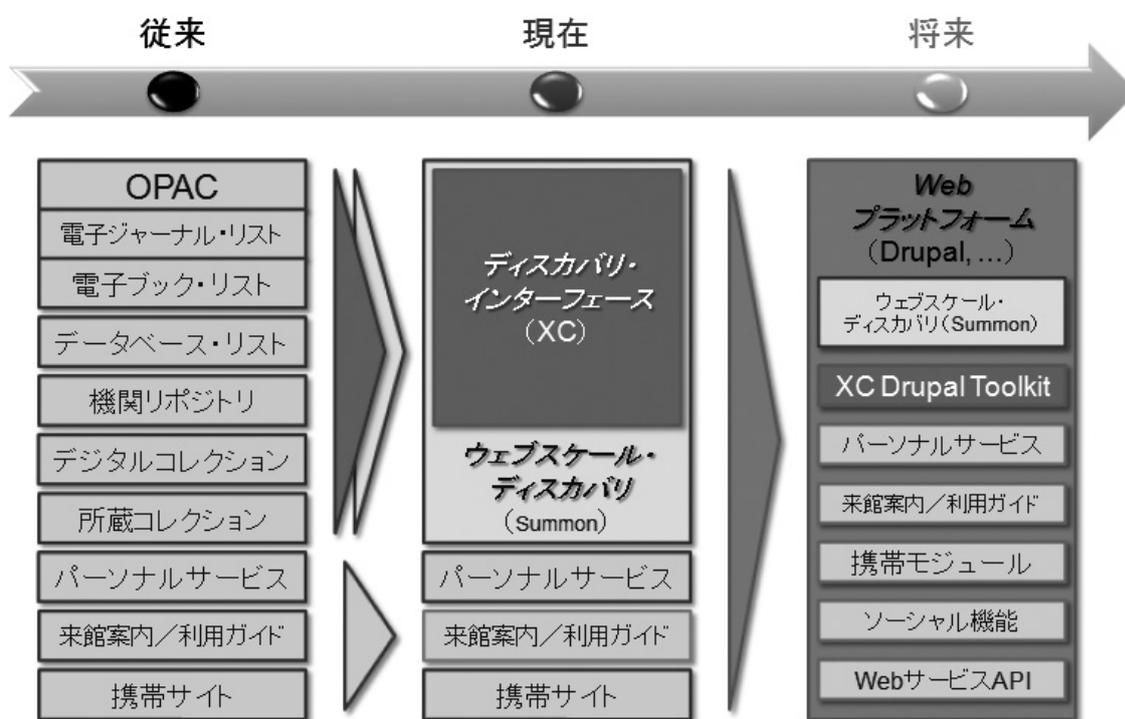


図7 九州大学附属図書館の目指す方向

参考文献

- [1] 工藤絵理子, 片岡真, 次世代 OPAC の可能性—その特徴と導入への課題—, 情報管理, 51 (7), pp.480-498, (2008) .
- [2] 兵藤健志 et al, 九州大学附属図書館における Cute.Catalog のデザインと開発, 情報管理, 53 (6), pp. 311-326, 2010.
- [3] eXtensible Catalog,
<http://www.extensiblecatalog.org>, (accessed 2011-02-09) .
- [4] Drupal, <http://drupal.org/>, (accessed 2011-02-09) .
- [5] The Open Archives Initiative Protocol for Metadata Harvesting, <http://www.openarchives.org/OAI/openarchivesprotocol.html>, (accessed 2011-02-09) .
- [6] Dublin Core Metadata Initiative, <http://dublincore.org>, (accessed 2011-02-09) .
- [7] RDA : Resource Description & Access Toolkit, <http://www.rdatoolkit.org>, (accessed 2011-02-09) .
- [8] NISO Circulation Interchange Protocol (NCIP)
<http://www.niso.org/workrooms/ncip/continuous>, (accessed 2011-02-09) .
- [9] NALIS, <http://www.livesolutions.info/nalis/index.html>, (accessed 2011-02-09)
- [10] Vocabulary Descriptions, Library of Congress,
<http://id.loc.gov/descriptions/>, (accessed 2011-02-09) .
- [11] 上綱秀治, 図書館員のコンピュータ基礎講座,
<http://www.asahi-net.or.jp/~ax2s-kmt/n/ref/subject.html>, (accessed 2011-02-09) .
- [12] 国立国会図書館, 書誌データ作成ツール : 分類・件名 (NDLC, NDLSH など)
http://www.ndl.go.jp/jp/library/data/ndl_ndlsh.html, (accessed 2011-02-09) .
- [13] Summon, <http://www.serialssolutions.com/summon>, (accessed 2011-02-09)
- [14] RefWorks. <http://www.refworks.com/>, (accessed 2011-02-09)