

学術資源共有基盤 WEKO の開発

山地 一禎¹, 青山 俊弘², 武田 英明³

1 国立情報学研究所 学術ネットワーク研究開発センター

2 鈴鹿工業高等専門学校 電子情報工学科

3 国立情報学研究所 学術コンテンツサービス研究開発センター

概要

現在、文献を主とした研究成果へのオープンアクセスが、機関リポジトリを主に進められている。出版による研究成果の公開とは異なり、リポジトリによる公開では即時性と多様性に柔軟に対応することができる。こうした特徴を生かしたリポジトリモデルを検討するために、本研究では、次世代学術資源共有プロトコルに対応した、新しいリポジトリシステム WEKO を開発した。

キーワード

リポジトリ, オープンサイエンス, OAI-ORE, SWORD, コンテンツマネージメントシステム

Development of Scholarly Contents Sharing Platform: WEKO

Kazutsuna YAMAJI¹, Toshihiro AOYAMA², Hideaki TAKEDA³

1 Research and Development Center for Academic Networks, National Institute of Informatics

2 Department of Electronic and Information Engineering, Suzuka National College of Technology

3 Research and Development Center for Scientific Information Resources, National Institute of Informatics

Abstract

Worldwide activities on open access have triggered many Japanese universities to develop institutional repositories. Currently, these repositories contain mostly electronic versions of domestic journal articles. Immediacy and diversity, which are the advantages of the Internet society, should be made available through a repository. This study proposes a repository system, called WEKO, that enables researchers and others to share a variety of scholarly content.

Keywords

Repository, Open Science, OAI-ORE, SWORD, Contents Management System

1. はじめに

近年、機関リポジトリ[1]などの構築が進み、大学などの組織として、積極的に学術情報をインターネット上に流通させる仕組みが急速に普及している[2]. 利用されているシステムの代表的なものとして、マサチューセッツ工科大学とヒューレットパッカード研究所が開発する DSpace がある[3][4]. 最新バージョンである 1.5 系では、Apache Software Foundation が提供する Maven[5]を採用することで、システム開発や管理の面で改善を図っている. また、Manakin と呼ばれるフレームワークの導入により、ユーザインターフェースのカスタマイズ性の向上を実現した[6]. 日本での利用は数少ないが、サウサンプトン大学が開発する EPrints も、欧米では数多くの機関リポジトリで採用されている. 最新バージョンである 3 系では、セルフアーカイビングにおけるメタデータの品質向上を目指した補完機能などの導入により、ユーザインターフェースが飛躍的に向上した[7][8]. こうしたシステムを利用する日本の機関リポジトリは、現在のところ約 100 機関で運用されており、合計すると約 60 万件のコンテンツが登録されている. その殆どが、学術雑誌論文、学位論文、紀要論文、会議発表論文などの文献類である.

一方で、インターネット上における参加型サイトの普及に伴い、従来は文献としてのみ公開されてきた研究成果が、現在では様々な段階、形式で公開されるようになってきた[9][10]. インターネットを利用したオープンサイエンス、オープンイノベーションと呼ばれる潮流は、その黎明期を迎えた段階に過ぎないが、Web2.0 に相当するような新しい科学の枠組みを提供する可能性がある. DSpace や EPrints などのシステムも、最近では研究成果の再利用や共有に主眼をおいた機能拡張がなされており[6][7], リポジトリとオープンサイエンスを融合した新しいモデルの提案や実験も進められている[11]. こうした背景の中で、本研究では、オープンサイエンスが必要とするコミュニティウエアへの発展を可能とする、新しい学術成果共有基盤 WEKO の開発を行った.

2. システム・アプリケーション構成

2.1. 基盤システム

本研究がターゲットとするような、オープンサイエンスを意識したリポジトリ環境を構築する場合、以下に挙げる 2 つのアプローチが考えられる.

- DSpace や EPrints などの既存のリポジトリシステムに、コミュニティウエアあるいはグループウエア機能を充実させる. あるいは、そうした機能をもつ外部システムとのマッシュアップ用 API を設計する.
- 拡張性をもつコミュニティウエアあるいはグループウエアの一機能（アドオン機能）としてリポジトリシステムを構築する.

前者のアプローチでは、既存のリポジトリシステムや外部アプリケーションのコア機能にまで影響を及ぼす可能性が高い. その結果、各アプリケーションのディストリビューションを作ることになり、メンテナンス性に問題が生じる. 一方、後者の場合は、前者

のような問題を回避することができる。しかしながら、基盤とするアプリケーションとしては、オープンソースあるいはオープンアーキテクチャなものを選択する必要がある。これに加え、拡張性の高いフレームワークを提供していることや、安定した開発ロードマップが期待できることが採用の鍵となる。本研究では、後者の選択肢の中でもこうした条件を満足するベースシステムとして、学術機関での導入実績も高い NetCommons[12][13]を採用することとした。

2.2. システム構成

構築するシステムの構成図を図1に示す。ベースシステムの NetCommons には最新バージョンである 2.1 (NC2) を利用した。NC2 では、その基底部として、PHP による web アプリケーションのフレームワークである Maple[14]を採用している。また、必要最低限の管理機能のみをコア部分に装備し、コミュニティウエアとしての諸機能はモジュールという単位でアドオンする形式となっている。モジュールは、管理モジュールと一般モジュールに大別される。用意されている管理モジュールには、「会員管理モジュール」「モジュール管理モジュール」「権限管理モジュール」「ページ管理モジュール」「バックアップモジュール」「セキュリティ管理モジュール」「システム管理モジュール」などがある。本研究では、こうした基本機能やコア部分が提供するライブラリを利用しながら、一般モジュールとしてのリポジトリモジュール: WEKO (スワヒリ語でリポジトリの意)を開発する。

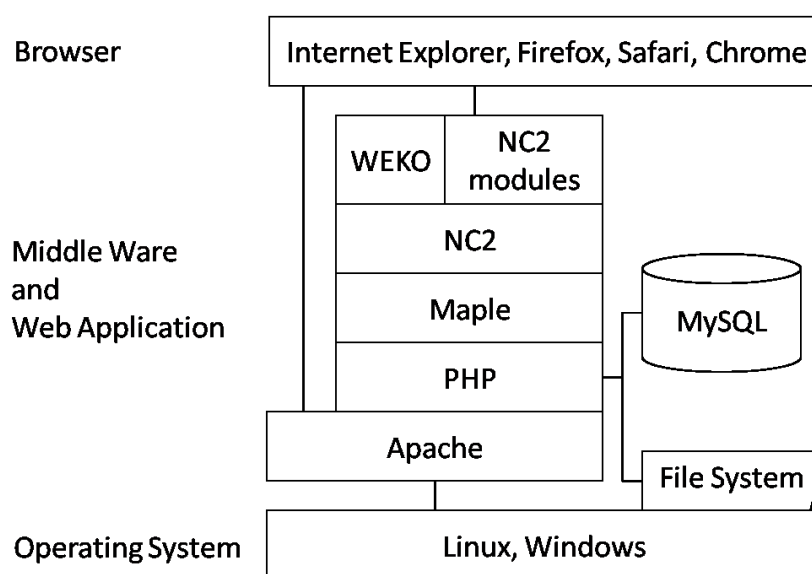


図 1 システム構成図

3. WEKO 基本機能

WEKO では、登録するコンテンツ（アイテム）を、アイテムタイプと呼ばれる大まかなカテゴリで分類したうえで、インデックスと呼ばれるツリー構造をもつキーワードに関連付けて保存する。アイテムタイプは、アイテムを登録する際のメタデータセットを規定する。インデックスツリーは、ディレクトリ検索のためのインターフェースを提供する。こうした動作を基本として、一般ユーザ（Anonymous）、登録ユーザ、管理者の各権限に応じて、以下のような機能を提供する。

3.1. 一般ユーザ

3.1.1. 検索

ディレクトリ検索とキーワード検索の二種類の検索機能を提供する。図 2(A)に示すように、ディレクトリ検索では、WEKO ブロック左側に配置したツリー内のインデックスをクリックすることで、その配下のコンテンツにアクセスできる。図 2(B)に示したキーワード検索では、アイテムを登録する際に入力するメタデータに加え、テキスト、PDF、ワード、エクセル、パワーポイントに限り、アップロードしたファイル内容も対象（全文検索）とする。

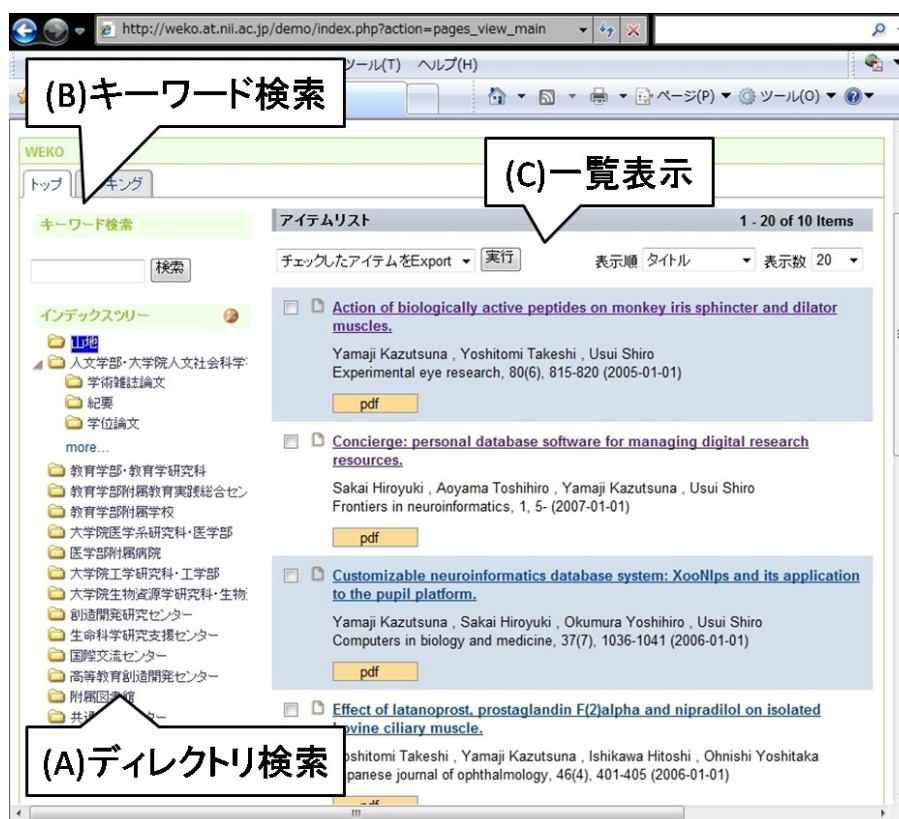


図 2 WEKO トップ画面例

検索した結果は、図 2 (C)に示すように、一覧結果として表示される。この結果に対し、タイトル、登録者、アイテムタイプ、査読日時、更新日時、作成日時でソートすることができ、同一ページ内での表示数も変更することができる。また、リスト左側に提示したチェックボックスをチェックしたアイテムに対し、エクスポートか印刷を実行できる。各アイテムのタイトルは、その詳細ページ（詳細メタデータ）へのリンクとなっている。

3.1.2. ランキング

以下の統計内容を、3.3.7.1 で述べる設定値に基づき表示する。この情報は、ランキングタブをクリックした場合に提示される他、3.3.7.1 で述べる初期表示設定機能により、トップページにも提示することができる。

- 最も閲覧されたアイテム
- 最もダウンロードされたアイテム
- 最もアイテムを作成したユーザ
- 最も検索されたキーワード
- 新着アイテム

3.2. 登録ユーザ

3.2.1. アイテム登録

アイテムの登録は、図 3(A)に示すフローにより行われる。ファイルのアップロードの際には、アクセス制限（エンバゴも設定可能）やライセンスを設定できる。リンク設定は、既に登録されているアイテムと相互関連がある場合に、その関係を記述する機能である。論文とそれに関連する各種コンテンツ（データ、ソフトウェア）などを記述するとき有効である。3.3.7.2 で設定する査読機能を有効にしている場合は、登録されたアイテムは査読待状態になり、登録直後には一般に公開されない。3.3.4 の操作により、管理者がそのアイテムを承認することで公開されることになる。

3.2.2. ワークフロー

自分が既に登録したアイテムの状態（査読状態など）を確認できる。タイトルのリンクをクリックすると、その詳細情報がポップアップウィンドウで表示される。また、編集ボタンをクリックすると、登録した内容を変更することができる。査読機能が有効な場合には、編集されたアイテムは再度査読状態に戻る。

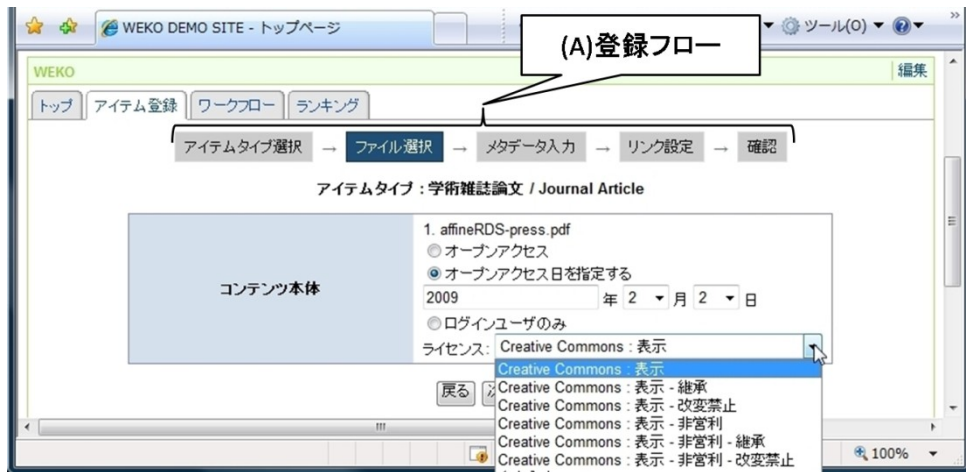


図 3 アイテム登録フロー

3.3. 管理者

3.3.1. アイテムタイプ

アイテムタイプの管理機能では、

- アイテムタイプの作成・変更
- メタデータの OAI-PMH へのマッピング

を設定することができる。インストールした時点で、junii2 フォーマット[2][15]に対応したデフォルトアイテムタイプが用意されており、これをもとにコピー機能などを利用して簡単にメタデータセットをカスタマイズできる。各メタデータのフィールドは、表示順位、必須入力、複数入力、一覧表示の是非などが設定できる。また、アイテムタイプのインポート/エクスポートも可能であり、WEKO リポジトリ間でのアイテムタイプの共有も容易に実現できる。

3.3.2. アイテム管理

あるインデックスツリー配下のアイテムの登録条件を、一括して変更できる。現在変更可能な内容は、以下の通りである。

- オープンアクセスの是非 (含 エンバーゴ期間の指定)
- 会員のみアクセスの是非
- ライセンス

3.3.3. ツリー編集

WEKO ブロック左側に配置されるインデックスツリーの作成・編集を行うことができる。図 4(A)に示したツリーにおけるインデックスの配置は、マウスを使って簡単に編集できる。インデックスの項目名は、同図(B)の機能により日本語と英語を登録でき、NC2 の言語選択

モジュールにより表示内容を切り替えることができる。また、同図(C)の機能によりインデックス自体に公開日を設定することができ、発刊日の制御が必要な雑誌の公開にもかつようできる。同図(D)の表示範囲機能は、ツリーの展開時に設定した上位何件かのインデックスのみを表示する機能である。同図(E)の機能を利用することで、インデックス直下へのアイテムの登録を、NC2 で規定するユーザの権限、あるいは所属するグループで制御することができる。



図 4 インデックストリー管理画面

3.3.4. 査読／承認

登録されたアイテムの状態を変更することができる。3.3.7.2 で設定する査読機能を有効にしている場合は、登録されたアイテムは査読待状態になり、一般には公開されない。査読／承認タブのクリックにより、査読待ち状態のアイテムの一覧が表示される。各アイテムのタイトルの右側のプルダウンメニューから、状態（承認・却下）を変更することができる。変更を確定するために、中央下部にある変更ボタンをクリックする必要がある。

3.3.5. インポート

WEKO 仕様の XML 形式のファイルをインポートできる。インポート形式の仕様は、同 WEKO からエクスポートしたファイルから簡単に確認できる。

3.3.6. ログ解析

リポジトリへのアクセス状況などがグラフ表示で確認できる。解析したい期間を指定した上で、

- アイテム登録数
- ダウンロード数
- 閲覧回数

に対して、日、週、月、および、年単位のスケールでログが集計できる。さらに詳細な解析をしたい場合には、CSV ファイルがダウンロードできる。

3.3.7. 一般管理

一般的なシステムのカスタマイズ項目を、「表示設定」「運用設定」「サーバ設定」に分類して提供する。以下では、各設定内容について説明する。

3.3.7.1. 表示設定

初期表示設定

トップページに表示する内容として、「インデックス配下の内容」か「ランキング」を選択できる。前者の場合には、さらにどのインデックスの内容を表示するかを選択できる。

ランキング管理

ランキングの計算の方法、および、表示順位数と表示項目を設定できる。

3.3.7.2. 運用設定

アイテム管理

Export 機能として添付ファイルまでを対象にするか否かと、査読機能の有効・無効化を選択できる。

ログ解析管理

検索エンジンのクローリングロボットなど、ログ解析から除外する IP アドレスのブラックリストを設定できる。

サイトライセンス認可

登録された IP レンジからのアクセスは、添付ファイルにおいて、

- ファイルアイテム属性タイプのログインユーザのみ公開
- 課金ファイルアイテム属性タイプの各グループへの課金に関する認可処理を介すことなく、ダウンロードが可能になる。

データベース管理

全文検索のためのインデックシングをマニュアルで実行することができる。

3.3.7.3. サーバ設定

OS 設定

WEKO を利用する OS を Linux と Windows から選択できる。

書誌情報簡単登録

ISBN から書誌情報のメタデータを補完する機能で、アマゾンの Access Key ID (AWSAccessKeyId) を設定できる。

OAI-PMH 設定

OAI-PMH の要求 Identify に対する応答のパラメータを設定できる。

サーバ環境

全文検索および PDF ファイルのサムネイル作成のための、外部コマンドへのパスを設定できる。

4. 次世代連携機能

4.1. OAI-ORE

OAI-ORE (Open Archive Initiative – Object Reuse and Exchange) とは、Open Archive Initiative が提案する Web リソースの集合体を記述するための仕様である[16]。WEKO におけるインデックスツリーは、あるキーワードで記述できるアイテムの集合体を表現する。そこで、インデックスのノード毎に OAI-ORE の Atom によりシリアライズされたリソースマップを公開するように実装した。WEKO のリソースマップは、ORE アイコンをクリックすることにより確認できる。リソースマップの HTML 内への埋め込み、および、アイテムの詳細ページ (スプラッシュページ) に対するリソースマップの配置については、今後、検討すべき課題である。

4.2. SWORD

SWORD (Simple Web service Offering Repository Deposit) [17][18]とは、ATOM Publishing Protocol[19][20]のプロファイルとして規定された、リポジトリにアイテムを登録するためのプロトコルである。これにより、ウェブブラウザを介すことなくリポジトリにアイテムを登録できたり、同時に複数のリポジトリへの登録や機械的な登録を簡便に行うことができる。WEKO は、現状の最新バージョンである 1.3 に対応した SOWRD サーバとして機能する。また、図 5 に示した Windows 上で動作するクライアントアプリケーションも用意した。このクライアントを利用することにより、ウェブブラウザから WEKO にアクセスすることなく、アイテムを登録することができる。SWORD では、同図(A)に示したようなサービスドキュメントをサーバから取得し、アイテム登録の条件を知ることができ

る。こうした基本情報に加えて、本クライアントでは、WEKO 上の所望のインデックスにアイテムを登録できるように、インデックスツリーの情報も取得できるような拡張を行った。

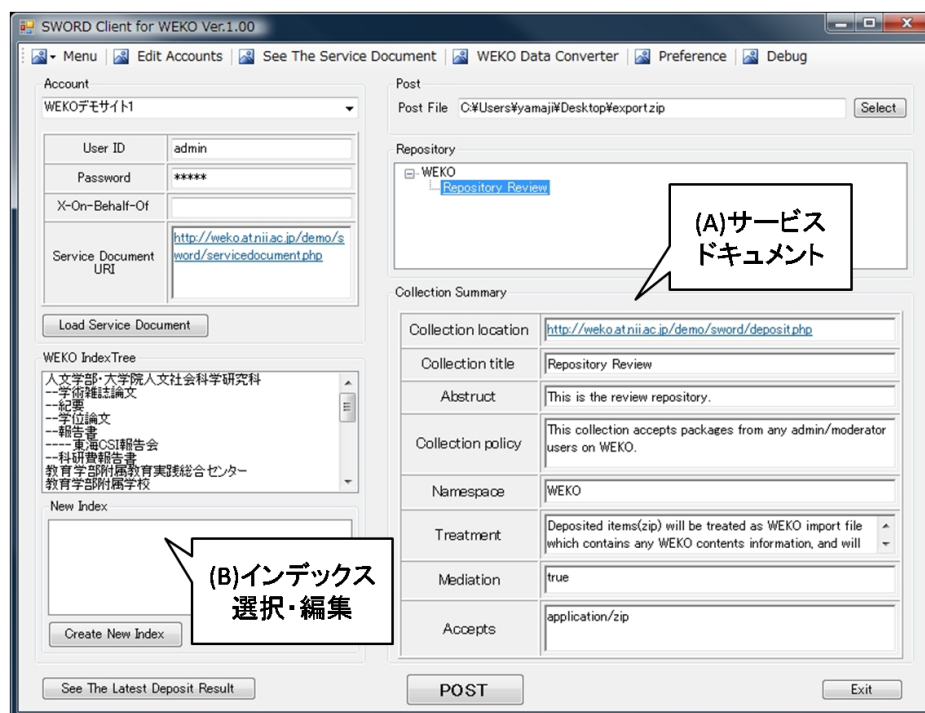


図 5 インデックスツリー対応 SWORD クライアント

4.3. Shibboleth 認証

Shibboleth とは、Internet2 が開発する SAML ベースの認証基盤であり、Web アプリケーションへのシングルサインオン環境を提供する[21]。NC2 への実装としては、権限と所属するグループを属性として取得することで、アイテムへのアクセス制御を可能とした。こうした認可のための情報は、ID プロバイダーからは任意の属性として受け取ることができ、NC2 側のコンフィグファイルで適宜マッピングできるように拡張性をもたせた。

5. おわりに

本研究では、コミュニティウェアとリポジトリを連携させる学術資源共有基盤 WEKO を構築した。今回の開発では、ファイルとメタデータをマネージする従来型のリポジトリ機能に加え、各リポジトリの連携を考慮した最新のプロトコルの実装を試みた。今後、こうした機能を活用し、オープンサイエンス等の新しい研究スタイルを実現するための、実証実験を進めていきたいと考えている。WEKO に関する最新の情報は、<http://weko.at.nii.ac.jp/> から取得することができる。

参考文献

- [1] 時実象一：電子ジャーナルのオープンアクセスと機関リポジトリ-どこから来てどこへ向かうのか：(II)機関リポジトリと研究助成機関の動向, 情報の科学と技術, Vol.57, No.5, 249-255, 2007.
- [2] 杉田いづみ：機関リポジトリについて--NII、そして大学の担当者としての経験から, 日本農学図書館協議会誌, Vol.149, pp.10-18, 2008.
- [3] Smith, M., Barton, M., Bass, M., Branschofsky, M., McClellan, G., Stuve, D., Tansley, R. and Walker, J. H.: DSpace An Open Source Dynamic Digital Repository, *D-Lib Magazine*, Vol.9, No.1, 2003.
- [4] DSpace Foundation : DSpace, <http://www.dspace.org/>
- [5] The Apache Software Foundation : Apache Maven, <http://maven.apache.org/>
- [6] Phillips, S., Green, C., Maslov, A., Mikeal, A. and Leggett, J.: Manakin A New Face for DSpace, *D-Lib Magazine*, Vol.13, No.11/12 2007.
- [7] Millington, P. and Nixon, W. J.: EPrints 3 Pre-Launch Briefing, *Ariadne*, Vol.50, 2007.
- [8] School of Electronics and Computer Science, University of Southampton : EPrints, <http://www.eprints.org/>
- [9] 藤本和則：人工知能研究者から見たオープンイノベーションアプローチ：研究成果のネット公開によるアプローチを中心に, 人工知能学会誌, Vol.20, No.4 pp.463-470, 2005.
- [10] Waldrop, M.: Science 2.0, *Scientific American*, Vol.298, No.5, pp.68-73, 2008.
- [11] Coles, S. and Carr, L. : Experiences with Repositories & Blogs in Laboratories, *Third International Conference on Open Repositories 2008*, 2008.
- [12] 新井紀子: 教育機関向けワンストップサービス構築ソフトウェアNetCommonsについて, 情報管理, Vol.49, No.7, pp.379-386, 2006.
- [13] The NetCommons Project: NetCommons, <http://www.netcommons.org/>
- [14] Maple Project: Maple, <http://kunit.jp/maple/>
- [15] 学術機関リポジトリ構築連携支援事業: メタデータフォーマットJunii2, <http://www.nii.ac.jp/irp/archive/system/junii2.html>
- [16] Open Archives Initiative: OAI-ORE, <http://www.openarchives.org/ore/>
- [17] Allinson, J., Francois, S. and Lewis, S.: SWORD: Simple Web-service Offering Repository Deposit, *ARIADNE*, Vol.54, 2008.
- [18] SWORD project: SWORD, <http://www.swordapp.org/>
- [19] Sayre, R. Atom: The Standard in Syndication, *IEEE Internet Computing*, Vol.9, No.4, pp.71-78, 2005.
- [20] The IETF: The Atom Publishing Protocol, <http://tools.ietf.org/html/rfc5023>
- [21] Internet2: Shibboleth, <http://shibboleth.internet2.edu/>