

リポジトリシステム WEKO のコンテンツ再利用環境 (SarabiWEKO) の開発

青山 俊弘¹, 山地 一禎², 池田 大輔³, 行木 孝夫⁴

- 1 鈴鹿工業高等専門学校 電子情報工学科
- 2 国立情報学研究所 学術ネットワーク研究開発センター
- 3 九州大学 システム情報科学研究院
- 4 北海道大学 理学研究院 数学部門 数学分野

概要

近年, 文献を主とした研究成果のオープンアクセスが機関リポジトリで進められている. 出版社での公開とは異なり, リポジトリではコンテンツの公開時期を柔軟に選ぶことができる. 一方, 機関リポジトリは機関のコンテンツを横断的に網羅する必要があり, コンテンツのタイトルや著者名など機関で定められた属性名に対応する属性値 (メタデータ) を追加する必要がある. したがって, コンテンツに対する属性名を自由に拡張することは難しい. そのため, 学部や研究者個人のページ, 研究者情報データベースなどの業績データベースなどにリポジトリのメタデータを流用することも難しい. そこで本研究では, リポジトリシステム WEKO に対して, リポジトリ上のコンテンツメタデータに付与する形でユーザが任意のメタデータを追加, 表示, 公開するためのソフトウェア SarabiWEKO を開発した.

キーワード

リポジトリ, ショーケース, メタデータ, 再利用

SarabiWEKO: A reuse system of WEKO contents and metadata

Toshihiro AOYAMA¹, Kazutsuna YAMAJI², Daisuke IKEDA³, Takao NAMIKI⁴

- 1 Suzuka National College of Technology
- 2 National Institute of Informatics
- 3 Kyushu University
- 4 Hokkaido University

Abstract

Institutional repositories (IR) have been collected many research articles, theses and university journals. Metadata of these contents are restricted by institutional archiving policy. Thus it is difficult to reuse the metadata for other purpose, e.g. department web pages, researcher databases or field repositories. This study develops new functionality for IR, extending its showcase ability to different sectors in the university. This system, called SarabiWEKO, enables to append specific metadata by the sectors to the contents in the IR.

Keywords

Repository, Show Case, Reuse of contents, metadata

1 はじめに

近年、多くの学術機関等では、セルフアーカイビングによる機関リポジトリを構築し、学術雑誌論文、学位論文、紀要論文、会議発表論文などを公開している。日本の学術機関リポジトリでは、2011年8月31日現在で183機関、1,166,071件のコンテンツが公開されている [1]。第4期科学技術基本計画 [2] においても、大学や公的研究機関における機関リポジトリの構築を推進することが求められており、今後も機関リポジトリの数、コンテンツの数ともに増加するものと考えられる。こうした機関リポジトリを構成するシステムとしては、DSpace[3] や EPrints[4] が多く用いられており、Fedora Commons Repository Software[5] をバックエンドのリポジトリシステムとして利用し、フロントエンドとしてユーザフレンドリな UI を持つ CMS などを組み合わせている例 [6] もある。

一方、多くの学術機関では学部、学科、研究室などでもそれらの Web ページ上に論文の書誌情報や論文自体を業績リストとして公開することは珍しくない。また、機関によっては、機関リポジトリとは独立に研究者情報のページがあり、そこに業績リストを載せる必要がある場合もある。すなわち、機関レポジトリと学部や研究室の Web ページ、あるいは研究者情報ページの両方でコンテンツあるいはそのコンテンツのメタデータを維持する必要がある。また、Web ページを作成するのではなく、OpenSearch などの機能を使い、機関リポジトリ上のコンテンツの検索条件を指定することで、所属する部署のコンテンツを選択して表示することも可能であるが、この場合にはコンテンツのメタデータは機関レポジトリが収集しているものに限られることになる [7]。しかしながら、学部や研究室などでは、単なるコンテンツの一覧ではなく、他の情報も含め、一種のショーケースとしてリポジトリのコンテンツを見せることが重要になる場合がある [8]。この場合、機関リポジトリにあるメタデータだけではなく、他のメタデータを付与して見せるということが必要になる。

例えば、通常機関リポジトリでは、人がどの所属単位に所属しているかといった情報を持っていない。そのため、学部ごとあるいは研究室ごとなどでコンテンツ一覧を表示することは難しい。また、機関リポジトリの枠内では、著者へのフィードバックフォームなどを集めるといったことは、機関レポジトリ内のメタデータを操作することができないため難しい。

そこで本研究では、リポジトリシステムである学術成果共有基盤 WEKO[9] に対して、リポジトリ上のコンテンツメタデータに付与する形でユーザが任意のメタデータをつけ、表示、公開するためのソフトウェア (SarabiWEKO) を開発した。

2 WEKO[9]

WEKO は国立情報学研究所で開発が進められているリポジトリシステムである。WEKO は、DSpace や EPrints などのリポジトリシステムとは異なり、ベースシステムとして NetCommons(NC2) を採用した一つのモジュールして実装されている。NC2 は PHP アプリケーションフレームワーク Maple の上に作られるコンテンツマネジメントシステム (CMS) であり、さまざまなモジュールを組み込むことにより、ラーニングマネジメントシステム (LMS) やソーシャルネットワークサービス (SNS) のような役割を果たすことができるシステムとなっている。WEKO は NC2 にリポジトリシステムを提供するためのモジュールである。NC2 はコミュニティソフトウェアとしての機能が充実しているため、コミュニティベースのリポジトリの運用やリポジトリを中心としたユーザ同士の情報交換を支援することができる。

3 SarabiWEKO

3.1 機関レポジトリコンテンツの再利用

日本における機関レポジトリの多くは、大学等に所属する学部や研究所などの業績をセルフアーカイビングするために利用されている。そのレポジトリシステムのほとんどは、バックエンドのデータベースとして関係データベース (RDB) を用いている。そのため、図書館が中心となって、各学部などからコンテンツを収集し、それらをアーカイブするにあたり属性名の集合 (メタデータセット)、すなわちデータベース、の設計を行う必要がある。コンテンツに追加する属性名は、各学部を横断的に見たとき、最大公約数的な必須項目の属性名の集合、あるいは、任意の入力を許した最小公倍数的な属性名の集合となる。そのため、レポジトリの属性値を再利用する時に、必要な属性名がレポジトリのデータベースに存在しないということも発生し得る。

レポジトリコンテンツの再利用の形として、オープンアクセスとしての機関レポジトリとは別に、学科のページ、研究者総覧、分野レポジトリなどの機関コンテンツコンテンツの部分集合からなるショーケースを作成することが考えられる (図 1)。所属する研究者の経歴や、サプリメントデータ、分野固有の分類などのメタデータをコンテンツに与えることができれば、機関レポジトリのコンテンツ、メタデータの単なる再利用のみならず、付加価値を生み出し提供することになる。SarabiWEKO は WEKO を機関レポジトリに持ち、そのコンテンツ (あるいはコンテンツのメタデータ) に付加価値となるメタデータを付与し、再利用するための機能を提供する。

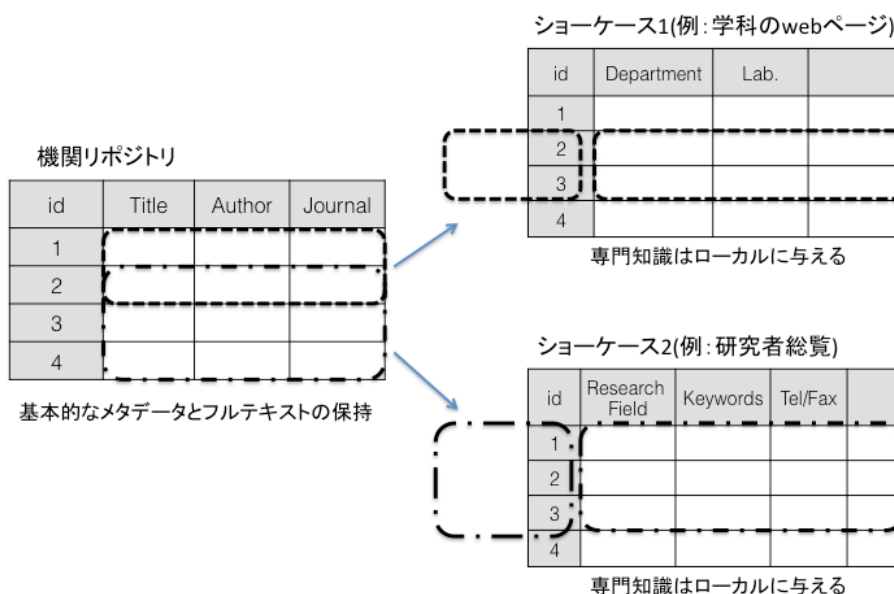


図 1 Examples of showcase using SarabiWEKO

3.2 システム構成

SarabiWEKO はレポジトリシステムである WEKO と密接に関連する NC2 のモジュールである。Sarabi とはスワヒリ語で蜃気楼 (mirage) を意味する。スワヒリ語でレポジトリを意味する WEKO にあるコンテンツの実体を蜃気楼のように自在にユーザに提供することが可能であるシステムであることからこの名がつけられた。

SarabiWEKO は図 2 に示すシステム構成である。SarabiWEKO が管理するメタデータは、WEKO リポジトリが保持するコンテンツのメタデータと、SarabiWEKO が管理するユーザが任意に追加することができるメタデータの二種類となる。SarabiWEKO はこれらのメタデータをテンプレートエンジンによって任意のスタイルで可視化する。また、SarabiWEKO は HTML だけではなく、JSON や XML(OAI-PMH) の形式でも出力することができる。これにより、ユーザ Web サイトでマッシュアップすることが容易となる。

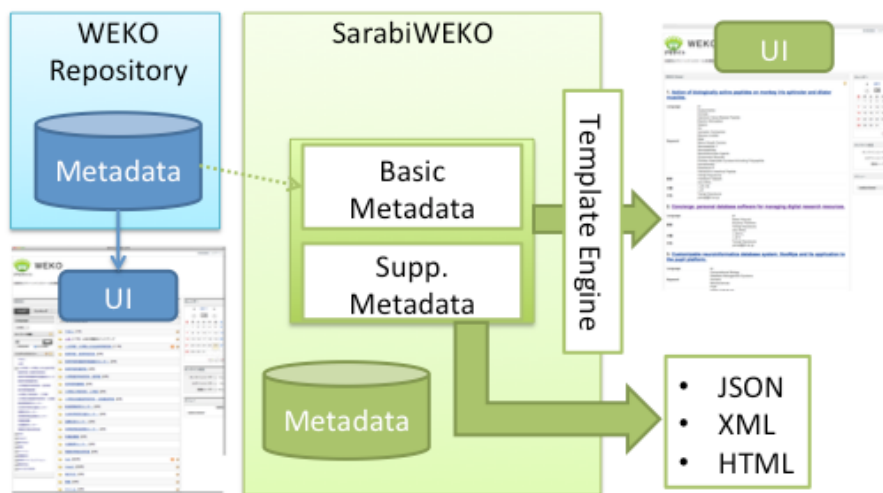


図 2 System architecture of WEKO and WEKO viewer module

3.3 基本機能

SarabiWEKO では、1. SarabiWEKO で表示したいコンテンツを WEKO から選択する (メタデータソース)、2. 取得したコンテンツに対するメタデータセット (属性名の集合, 追加アイテムタイプ) を作成し、メタデータを入力する、3. テンプレートエンジンを用いて、表示するメタデータの選択、およびスタイルを決める、という流れで、リポジトリコンテンツ (メタデータ) を再利用することができる。

3.3.1 メタデータソース

SarabiWEKO で表示する基本となるメタデータのコンテンツを選択する。これは WEKO(基盤リポジトリ)にあるコンテンツから再利用するコンテンツを選択することを意味する (図 3)。コンテンツの取得方法としては、リポジトリ内のコンテンツ ID の指定、コンテンツの所属インデックスからの選択、あるいは WEKO 内のコンテンツメタデータに対する検索がある。これらの条件を複数使い、それぞれの条件で検索されたコンテンツの和集合を得ることもできる。

3.3.2 追加アイテムタイプ

リポジトリから取得したコンテンツに、ユーザ側で追加するアイテムタイプを追加する。アイテムタイプとはメタデータの集合であり、コンテンツに対して複数のアイテムタイプ (メタデータセット) を付与することができる (図 4)。メタデータは Dublin Core, jumi2 フォーマットに対応させることも可能である。また、複数のコンテンツに対して、追加したアイテムタイプのメタデータを一括して登録、あるいは選択したコンテンツの詳細ページでメタデータを編集することも可能である。

図3 Metadata source. Select contents from WEKO repository using WEKO ID, index or searching by metadata.

項目名	属性	Dublin Core	JuNii2	言語	複数可	入替	削除
発表スライド	ファイル	未設定	未設定	指定なし	<input type="checkbox"/>	↑ ↓	削除
発表予稿集	ファイル	未設定	未設定	指定なし	<input type="checkbox"/>	↑ ↓	削除
発表日時	日付	date	date	指定なし	<input type="checkbox"/>	↑ ↓	削除

図4 Editor of additional item types

3.3.3 テンプレート

リポジトリから取得したコンテンツのメタデータと SarabiWEKO にて追加したメタデータは、テンプレートエンジンを用いてコンテンツごとに表示するスタイルを定義することにより表示形式を決定する。NC2 がテンプレートエンジンとして利用している Smarty[10] の文法に基づき、一覧表示にて表示したいメタデータを選び、それらの表示順を指定し、CSS(Cascade Style Sheet) よりスタイルを指定する。

4 ケーススタディ

本節では、開発した SarabiWEKO を使ったケーススタディの一例を示す。リポジトリをコンテンツの一次情報として使うことはどれも変わらないが、追加するメタデータと合わせ、どのようにコンテンツのメタデータを使うかによってさまざまなショーケースを構成することができる。

4.1 サプリメントデータの公開

機関リポジトリで用いているリポジトリシステムでも、論文に付随するサプリメントデータを公開することは可能であるが、機関の運用ポリシーによってはそれらのデータを公開できないこともある。このような場合に、研究室や学部などの Web ページで関連するデータを公開することが可能となる。

4.2 分野リポジトリ

機関リポジトリには機関に所属するさまざまな分野の研究者の業績が蓄積される。学部ごとなどの単位でコンテンツを閲覧、検索することも可能であるが、例えば、数学の論文リスト [11] といったある分野の研究のコンテンツを収集したりリポジトリを構成することは、カテゴリズという観点の付加価値をつけるという意味で重要である。さらに、SarabiWEKO では、こうしてカテゴリズしたコンテンツに独自のメタデータを付与することができる。細かな分類やコンテンツに対するアノテーションなどを付加することでさらなる付加価値をつけることが可能となる。

4.3 研究者業績リスト

研究者データベースは研究者の経歴などとともに研究業績リストからなることが多い。SarabiWEKO を用いることで機関リポジトリにあるコンテンツメタデータを再利用することにより、改めて業績リストを作成せずともよい環境を構築することができる。科研費番号などの著者 ID や機関独自の ID があり、それらが機関リポジトリのメタデータとして登録されている場合には、各研究者ごとの業績リストを作成することは SarabiWEKO を使用せずとも、OpenSearch のクエリを一つ作成するだけで得られるため容易い。しかしながら、機関リポジトリのすべてのコンテンツに著者 ID が割り振られている訳ではない場合には、コンテンツの抜けが生じたり、同姓同名問題が発生する可能性がある。SarabiWEKO では、リポジトリ内にあるコンテンツを検索、あるいは一つずつ指定することができるため、厳密な著者 ID が存在しないリポジトリであっても業績リストを漏れなく作成することができる。また、SarabiWEKO はリポジトリからメタデータを得るため、一度コンテンツ取得条件を作れば、その後はリポジトリにコンテンツを入れるだけで、自動的に業績リストが更新されることになる。

4.4 学部ページ

学部などでは、所属する研究者、教員、学生の代表的な研究成果をまとめ、学部の研究分野の紹介などを行うことがある。このとき、リポジトリのコンテンツのメタデータを再利用し、さらに学部独自の付加価値を添えることで、学部のショーケースを作成することができる。

4.5 認証評価用業績リスト

大学などの認証評価のために研究者の業績リストが求められる場合がある。しかしながら、機関リポジトリで公開しているメタデータでは不足していることが多い。そこで SarabiWEKO を用いて、不足しているメタデータを補い、JSON や XML による電子処理しやすい形でデータを得て電子的に処理する、あるいはそのまま資料として用いるということが可能となる。

5 おわりに

本研究では、リポジトリシステム WEKO のコンテンツメタデータを再利用するためのフレームワークとして SarabiWEKO を開発した。学术论文などの文献を中心としたコンテンツを収集する機関リポジトリは、今後も普及が進むものと考えられる。これらのコンテンツは、各社の検索サイトや、JAIRO のような機関レポジトリのハーフスタや Google 等の検索システムを経由し利用者に提供される。本システムは、機関リポジトリのコンテンツを、学部や研究室といった機関より小さな所属単位の情報源の一つとして用いるため再利用する仕組みを提供する。

機関リポジトリを構築する際に、内部メタデータをどのように設計するかという点は大きな問題である。取りうる一つの方針は、構築機関で要求される機能、所属研究者の要請する機能に応じたメタデータを全て保持することであろう。しかし、機関が要求する可能性のある機能は多岐に渡り、研究者の要請する機能も細分化されている。したがって、必要なメタデータを全て取り入れてしまえば膨大なメタデータを設計しなければならない。

一部の機関では、研究者の要請に沿ったメタデータの設計を実現している。例えば、東京大学および京都大学の機関リポジトリにおける数学分野への対応^{*1}(2006-) であり、一橋大学、アジア経済研究所、北海道大学等における RePEC への対応である。物質・材料研究機構における eSciDoc[12] も同様に位置づけられる。大学情報データベースと連携している機関も一つの例として位置づけられる。これらの機関においては、機関リポジトリを運用する主体としての図書館が機関もしくは研究者の要請を積極的に吸収し、リポジトリの設計段階からリポジトリソフトウェアの内部設計へ手を入れることで実現したと考えられる。これらの機関のように、内部メタデータを豊富にすることは一つの理想だが、このような実装の可能な機関は多くない。運用主体の体力と技術に依存する点が大きく、必ずしも一般化できるモデルにはなり得ない。新たな要請に対しては大きなコストがかかることも問題となりうる。

一方で、最小限のメタデータを機関リポジトリに保持しつつ、機関および研究者の要請するメタデータを外部連携によって利用することが可能であれば、その応用範囲は格段に広がるのが想定できる。一つのありかたとしては OAI-PMH を使い機関および研究者側でハーベストした後に機関および研究者側で用意したメタ

^{*1} <http://repository.dl.itc.u-tokyo.ac.jp/> における dc.subject.msc

データとマージすることであり，例えば数学ポータル^{*2}(2008-) はそのような実装を行っている．要請する側のメタデータおよびを SWORD 等によって投入することも考えられる(北大, 2010-)．これも，機関リポジトリのメタデータを利用する側に大きな負荷がかかることは違いがない．

このように，機関リポジトリのメタデータ設計における二通りの方針を概観すれば，相反する利点と欠点から逃れられない．SarabiWEKO は，機関リポジトリ運用側には運用主体である図書館の持つ目録に相当する最低限のメタデータを保持し，機関および研究者の付与する外部メタデータと連携することで機関リポジトリの活用範囲を拡大するモデルを提案し，実装したものである．

参考文献

- [1] 国立情報学研究所: JAIRO, <http://jairo.nii.ac.jp/>
- [2] 文部科学省: 第4期科学技術基本計画,
http://www.mext.go.jp/component/a_menu/science/detail/_icsFiles/afiedfile/2011/08/19/1293746_02.pdf
- [3] DSpace Foundation : DSpace, <http://www.dspace.org/>
- [4] University of Southampton, UK : EPrints, <http://www.eprints.org/>
- [5] DURA SPACE: Fedora Commons Repository Software, <http://fedora-commons.org/>
- [6] DURA SPACE: Examples Uses, <http://www.fedora-commons.org/about/examples>
- [7] 小野 真由美, 井上 創造, 星子 奈美: 研究者業績情報システムと学術情報リポジトリの連携, 情報処理学会研究報告, Vol. 109, pp. 43-51, 2007
- [8] 国立情報学研究所: 学術コミュニケーションの新たな地平,
http://www.nii.ac.jp/irp/archive/report/pdf/csi_ir_h17-19_report.pdf, 2009
- [9] 山地 一禎, 青山 俊弘, 武田 英明: 学術資源共有基盤 WEKO の開発, デジタル図書館, Vol.36, pp.51-61, 2009
- [10] Smarty: Smarty, <http://www.smarty.net/>
- [11] Takao Namiki and Daisuke Ikeda, Digital Mathematics Library in Japan as an Overlay of Digital Repository Systems, The 6th Annual Conference on Open Repositories, <http://indico.cern.ch/contributionDisplay.py?contribId=10&sessionId=3&confId=132638>
- [12] 谷藤 幹子, 高久 雅生, 大塚 真吾, 轟 眞市: 材料系研究所におけるリポジトリシステムの実践と将来, 情報管理, Vol. 51, No. 12, pp. 888-901, 2009

^{*2} <http://eprints3.math.sci.hokudai.ac.jp/>