

研究種目：若手研究 (B)
 研究期間：2007～2009
 課題番号：19700658
 研究課題名 (和文) 世界中の教育コンテンツを検索可能な大規模総合サイトの構築と評価
 研究課題名 (英文) Development and evaluation of a system for retrieving universal learning objects
 研究代表者
 森本 容介 (MORIMOTO YOSUKE)
 放送大学・ICT活用・遠隔教育センター・助教
 研究者番号：00435702

研究成果の概要 (和文)：学習資源の流通・再利用に寄与するため、国内外の学習コンテンツの情報を収集するシステム、および検索システムを構築した。本システムは、国際標準の学習オブジェクト・メタデータをベースに、ライセンスや品質保証を記述できるように拡張したメタデータを利用する。利用可能コンテンツ数の増加に伴い、対象者や利用場面ごとに異なる検索対象レコードや検索インタフェースが必要となる。そこで、目的に応じた複数の検索システムを効率的に開発できる汎用的な検索エンジンを開発した。

研究成果の概要 (英文)：To promote the sharing and reuse of open educational resources, a system for collecting metadata and using it for retrieval has been developed. The system uses metadata extended to describe more information such as rights and quality assurance. With the increase of metadata, search applications adapted to users or usage scenes are needed. To develop a number of such application systems efficiently, a generalized search module for learning object metadata was developed.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	900,000	0	900,000
2008年度	1,400,000	420,000	1,820,000
2009年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,200,000	690,000	3,890,000

研究分野：教育学

科研費の分科・細目：科学教育・教育学 / 教育学

キーワード：学習オブジェクト, メタデータ, LOM, 情報検索

1. 研究開始当初の背景

eラーニングに代表されるように、インターネットを活用した教育・学習活動が一般化してきている。インターネットで学習を行ったり、教育に関する情報を得るためには、まず

教材や情報の所在を探す作業が必要である。これらの情報の検索には、一般的なサーチエンジンを用いることができる。しかし、教育とは関係のない情報も多く含まれる事から適切な検索結果が得られないことがある。また、各コンテンツが体系化されておらず、「C

言語の e ラーニングコースを受講したい」、「制御システムを教えている大学を知りたい」といった場合、適切な検索結果を得ることが困難である。そこで、インターネット上で提供されている学習情報を体系化して、情報提供するシステムが開発されている。代表的なサイトとして、高等教育を主な対象とした NIME-glad(独立行政法人メディア教育開発センター；現在、放送大学 ICT 活用・遠隔教育センター)や、初等・中等教育を対象とした NICER(国立教育政策研究所)があげられる。NIME-glad では、教育コンテンツに LOM を付与し、検索に利用する。LOM とは、Learning Object Metadata の頭字語で、主に検索のために学習コンテンツに付与するメタデータである。LOM は IEEE において標準化されているが、メディア教育開発センターでは、それを改変した NIME LOM を用いている。研究を計画した 2006 年 7 月の時点で、e ラーニングコース 1,994 件、シラバス 103,365 件、公開講座 133 件、学習素材 15,909 件、OCW265 件、FD 及び ID306 件に LOM が付与されている。本研究では、NIME-glad をベースとして、学習コンテンツの共有・再利用をさらに推進することを旨とした。

2. 研究の目的

世界中の教育コンテンツを幅広く扱い、学習コンテンツの共有・再利用を推進するために重要な点は、教育コンテンツの収集システムの確立と検索システムの高度化である。教育コンテンツの収集を 1 つの機関が集中して行うことは、人的リソースや、情報更新への対応の点から、現実的ではない。外部機関やコミュニティとの協力や自動収集手法の確立が必要である。また、検索可能コンテンツ数の増加に伴い、一般的には情報到達性が低下する。検索対象者や利用場面に応じた検索システムが必要となる。そこで、これらの機能を研究開発することを目的とした。

3. 研究の方法

(1) 学習コンテンツの収集方法の確立
質の高いメタデータを効率的に収集する方法を確立する。収集だけではなく、外部へコンテンツ(メタデータ)を発信することにより、学習コンテンツの共有・再利用に資することができる。自動化の研究のほか、外部機関との協力や標準化への対応、コミュニティの構築を行う。

(2) 効率的な検索システムの構築
収集したメタデータを効率的に検索できる

システムを開発する。検索手法の工夫による精度向上のほか、ユーザコンテキストに応じた検索システムの提供が有効であると考えられ、これらを実現するための研究開発を行う。

4. 研究成果

本研究の主要な成果は、メタデータ流通システムの確立、検索システム GLOSS の開発、および汎用的なメタデータ検索ミドルウェアの開発である。それぞれについて詳細を述べる。

(1) メタデータの再設計と収集システムの確立

NIME-glad のメタデータ形式は、IEEE の LOM(IEEE 1484.12.1 - 2002 Standard for Learning Object Metadata)をベースにした独自の LOM(NIME LOM)である。2009 年 1 月 25 日の時点で、日本語の LOM が 153,588 件、英語の LOM が 12,907 件の、計 166,495 件が登録されている。これらのメタデータは、センターの職員が手作業で付与を行ってきた。詳細な作業方法とチェック体制が設けられており、メタデータ品質の統一を試みている。しかし、1 つの機関で全てのメタデータの管理・運用を行うことは、継続性に問題がある。そこで、国内のコンテンツに関しては、コンテンツを所有する機関がメタデータを付与し、それを収集することにより検索サービス・流通サービスを提供するアプローチを採用した。海外のコンテンツに関しては、OAI-PMH によるメタデータハーバスティングを採用し、ハーバスタ(収集システム)とリポジトリ(発信システム)の開発を行った。メタデータの交換は、GLOBE(Global Learning Object Brokered Exchange)の枠組み内で行っている。GLOBE は、高品質の学習オブジェクトを、地球規模で流通・共有再利用させる目的で結成された国際コンソーシアムである。国際的なメタデータ交換のために、標準に準拠したメタデータを採用する必要がある。新しく設計したメタデータは、付与の負担、および重要な項目を考慮し、IEEE LOM で記述可能な項目のうちの一部のみを利用する。また、ライセンスや質保証の項目を記述できるように、拡張を行った。ここで、メタデータの実体は、XML 形式で保存しているのではなく、データベースに格納している。メタデータを流通させる時に、データベースから必要項目のみを抽出し、XML 形式の LOM に整形する。この際、拡張した項目は利用されないで、リポジトリ間での相互運用性が保証される。以下に、再設計を行ったメタデータ項目を示す。

表1 メタデータ項目

項目	IEEE LOM との対応	必須	複数
メタデータのID	3.1 Meta-Metadata Identifier -	○	
タイトル	1.2 General - Title	○	
LO の言語	1.3 General - Language		○
概要	1.4 General - Description		
キーワード	1.5 General - Keyword		○
内容のまとめ	1.8 General - Aggregation Level	○	
LO への寄与者	2.3 Life Cycle - Contribute		○
メタデータの言語	3.4 Meta-Metadata - Language	○	
ファイル形式	4.1 Technical - Format		○
URL	4.3 Technical - Location	○	
動作条件	4.4 Technical - Requirement		
教育分野	5.6 Educational - Context		○
標準的な学習時間	5.9 Educational - Typical Learning Time		
対象レベル	5.10 Educational - Description		
料金	6.1 Rights - Cost	○	
許諾範囲	6.3 Rights - Description		
専門分野 (分類体系 / 値)	9.2.1 Classification - Taxon Path - Source / 9.2.2.2 Classification - Taxon Path - Taxon - Entry		○
ライセンス	(なし)		○
品質	(なし)		○
セット	(なし)		○

表1において、「必須」列が「○」となっている項目は必須項目、「複数」列が「○」となっている項目は複数の値を入力可能な項目である。「ライセンス」には、採用されているライセンス体系とその値のペアを入力する。「品質」には、品質保証体系とその値のペアを入力する。なお、表1には、管理上必要な項目は表示していない。

(2) 検索システム GLOSS の開発

グローバル学習オブジェクト検索システムを新たに設計した。GLOSS は、教員向けの教育用素材と学習者向けのコースウェアを検索できる。論文等の研究寄りのコンテンツやシラバス等は含まない、狭義の教育コンテンツである。想定対象者は、大学教員と生涯学習者を含む大学生以上の学習者である。国内のコンテンツの他、OAI-PMH により海外機関から収集したメタデータも検索できる。また、複数のページから構成される OCV のコースウェアを、まとめて全文検索できる特徴がある。また、GLOSS のユーザ履歴を解析し、コンテンツ推薦機能を開発した。

GLOSS は、検索対象のコンテンツによって、従来の LOM 検索と、コンテンツそのものの全文検索を使い分ける。JOCW(日本オープンコースウェア・コンソーシアム)正会員の大学が運用する OCV サイトから配信されているコンテンツ(以下、「JOCW コンテンツ」)は全文検索を行う。その他のコンテンツは LOM 検索である。ただし、これらのコンテンツの検索結果を区別して提示することはしない。各検索結果のスニペットとして、JOCW コンテンツは検索キーワードを含む前後の文が、その他のコンテンツは LOM の概要が表示される。

JOCW コンテンツに関しては、学習者向けのコースウェアという観点から、コース単位での登録となっている。すなわち、GLOSS で検索できるコンテンツに付与されるメタデータの粒度は、ページ単位ではなく、コース単位である。全 10 回の講義であれば、10 回分をまとめて 1 つのメタデータを付与する。そこで、GLOSS では、1 つのコースに関する全てのページをまとめて、1 つのコンテンツとして検索する。

このため、以下の手順を確立した。

① クローリングによるコンテンツ収集

各大学の OCV サイトをクローリングし、コンテンツを収集する。OCV サイトのトップページ以下全てのページを収集する。

② コースに含まれるファイルの抽出

メタデータには各コースウェアのトップページが登録されているが、収集したコンテンツの中から、コースに属するファイル、すなわち各回の講義資料やシラバスのページを抽出する。抽出対象となるファイルの種類は、HTML・PDF・PowerPoint スライドなど、全文検索に適したものである。

③ 検索インデックスの作成

抽出した複数のファイルからテキストを抽出し、1 つにまとめて検索インデックスを作成する。すなわち、JOCW コンテンツは、トップページ以外に含まれる文字列も検索対象となる。これにより、検索漏れが少なくなり、再現率が上昇することが期待される。反対に

精度は下がる可能性があるが、文書集合が限られたこの種の検索においては、再現率を重視する方がよいと考えた。

以上の手順において、同一コースに含まれるファイルの判定方法は、以下の通りである。

【URLの規則による判定】

各 OCW サイトの調査を行い、URL の規則を求めた。JOCW 加盟大学の 16 大学のサイトを調査し、15 大学について URL の規則を求めることができた。同一のコースが共通の QUERY_STRING をもつサイト、共通のディレクトリに格納されるサイト、またはその組み合わせである。ただし、一部例外ルールを追加した。

【リンク情報による判定】

URL の規則による判定を補完するため、リンク情報を利用する。各コースウェアのトップページからリンクをたどり、それぞれのページについて、URL の規則による判定を実施する。同一コースと判定されれば、そこからさらにリンクをたどる。このように、同一コースと判定されなくなるまで再帰的に判定を行う。

図 1 に GLOSS のスクリーンショットを示す。



図 1 GLOSS の検索結果画面

このほか、GLOSS は、外部機関のリポジトリを検索できる Federated Search クライアント機能、および外部機関へ検索結果を提供する Federated Search サーバ機能を提供している。

(3) メタデータ検索ミドルウェアの開発
メタデータを効率よく検索するためには、対象者に合わせて、検索対象レコードや検索インタフェースを提供することが必要である。このとき、様々な検索システムを用意しなければならない。それぞれのシステムでは、類似した処理を行うため、検索システムごとにシステム開発を行うのは非効率である。そこで、共通の処理を行う部分を検索エンジンとしてまとめ、各検索システムでは、検索エン

ジンを利用することにより、効率的にシステム開発を行えるようにした。本検索エンジンは、Movable Type のプラグインとして実装した。本プラグインは、学習オブジェクトのメタデータを検索する CGI、検索 CGI を起動するフォームをウェブサイトやブログに表示するテンプレートタグ、検索結果を表示するテンプレートタグ、および検索 CGI に検索アプリケーションを追加する機能などを提供する。これらの機能を利用することにより、目的・利用者に応じた様々なアプリケーションを展開することができる。

本プラグインを利用したアプリケーションの一つとして、JOCW コンテンツ横断検索システムを開発した。本システムの検索対象は、JOCW に加盟する大学の OCW サイトで提供されているコースである。各大学から収集したメタデータを、データベースへ登録する際、大学ごとに異なるタグのほか、JOCW を表すタグを付与している。タグは、メタデータ項目の「セット」を利用する。具体的には、「institution:society:jocw」というセットを付与する。すなわち、本検索システムは、「セット」が、「institution:society:jocw」であるメタデータを対象とした検索システムである。この情報は、プラグインの設定ファイルに記載される。



図 2 本システムの詳細検索フォーム

図 2 は、本検索システムの検索オプションである。メタデータ項目のすべてを指定可能にもできるが、多くの検索オプションが表示されると煩雑になる。そこで、このように、利用する検索オプションを検索アプリケーションごとに変えることができる。本検索システムでは、クリエイティブ・コモンズ・ライセンスを指定した検索が可能である。ライセンスに関する情報は、メタデータ項目の「ライセンス」である。「ライセンス」は、ライセンス体系とその値で格納される。ライセンス体系にはクリエイティブ・コモンズ・ライセンスを表す文字列が、その値には、たとえば「表示-非営利-改変禁止」を表す文字列が格納される。本検索システムの検索オプションでは、「改変可」および「商用利用可」のチェックボックスを設けている。たとえば、「改変可」のコースは、「表示」や「表示-継

承」など、複数の値が該当する。このため、検索オプションで指定された値を、具体的な検索式へ展開する必要がある。このような場合にフックメソッドが利用できる。フックメソッドは、パラメータとして渡された検索オプションから検索式を組み立てる前後、およびフォームコントロールへ検索オプションの値を設定する前後に実行される。本システムでは、MIME media type ではなく、「PDF」や「画像」としてメディアの種類を指定するため、および「変更可」と「商用利用可」のオプションからクリエイティブ・コモンズ・ライセンスが付与されたコースを指定するために、フックメソッドを利用し、検索式の書き換えなどを行っている。

図 3 に JOCW コンテンツ横断検索システムのスクリーンショットを示す。



図 3 本システムの検索結果画面

(4) まとめと今後の展望

本研究では、学習コンテンツの共有・再利用を推進することを目的とし、メタデータの流通システムの確立、および検索手法の研究開発を行った。メタデータは、IEEE LOM をベースに新規設計した。持続可能な運用を行うため、国内のコンテンツは、コンテンツを所有する機関がメタデータを付与し、それを収集する方式を確立した。海外のコンテンツは、GLOBE の枠組みの中で OAI-PMH によるメタデータ交換を行うためのシステムを開発した。メタデータを利用した検索システムとして GLOSS の開発を行った。GLOSS は、国内で収集したメタデータのほか、海外から収集したメタデータも検索できる。GLOSS は、複数のページから構成される OCW のコースウェアを、1 つのコースとして全文検索できる特徴を持ち、このための、同一コースの判定手法を構築した。メタデータ増加に伴う情報到達性の低下に対応するため、ユーザコンテキストに応じた検索システムの提供が必要である。こ

のとき、複数の検索システムを効率的に構築するための、メタデータ検索ミドルウェアを開発した。本ミドルウェアは、JOCW コンテンツ横断検索システムなどに活用している。本研究により学習コンテンツの共有・再利用に関するシステムの整備を行うことができた。開発した検索システムは、それぞれの目的に活用されている。コミュニティによるメタデータ付与を継続する必要があることから、コミュニティの拡大が望まれる。このためには、コミュニティに応じた検索システムが重要な役割を果たすと考えられるため、今後はこれらに関する取り組みを進める予定である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 17 件)

- ① 森本容介, 学習コンテンツのメタデータ検索エンジンの開発, 電子情報通信学会ライフインテリジェンスとオフィス情報システム研究会, 2010 年 5 月 20 日, 早稲田大学
- ② Tsuneo YAMADA, OCW Search Services for Lifelong Learners, OCWC Global 2010, 2010 年 5 月 7 日, ハノイ
- ③ 辻靖彦, メタデータの自動生成を目的とした最新シラバスからの情報抽出, 教育システム情報学会研究会, 2010 年 3 月 13 日, 畿央大学
- ④ 森本容介, 「科学技術倫理と知的財産権」の e-Lecture 開発, 電子情報通信学会技術と社会・倫理研究会, 2010 年 3 月 1 日, 宮城県 作並温泉一の坊
- ⑤ 森本容介, 学習コンテンツの持続可能なメタデータ流通と検索システムでの利用, 電子情報通信学会教育工学研究会, 2009 年 12 月 11 日, 琉球大学
- ⑥ 辻靖彦, LOM の自動生成を目的としたシラバス文書の情報抽出, 電子情報通信学会教育工学研究会, 2009 年 12 月 11 日, 琉球大学
- ⑦ Tsuneo YAMADA, Utilization of metadata for quality assurance in the international sharing and reuse of Open Educational Resources, The 23rd Annual Conference of Asian Association of Open Universities, 2009 年 11 月 4 日, テヘラン
- ⑧ 森本容介, Flash Video 形式への変換・録画ツールによるビデオ教材の作成, 日本教育工学会第 25 回全国大会, 2009 年 9 月 19 日, 東京大学
- ⑨ 森本容介, 機能拡張性とコンテンツの

流通再利用性を両立させたeラーニングシステムの開発，教育システム情報学会第34回全国大会，2009年8月19日，名古屋大学

- ⑩ 山田恒夫，OCWにおける連携活動 -グローバル横断検索サービスの開発-，平成21年度工学・工業教育研究講演会，2009年8月8日，名古屋大学
- ⑪ 森本容介，学習コンテンツのメタデータ流通基盤と検索アプリケーションの提案，電子情報通信学会教育工学研究会，2009年3月7日，香川大学
- ⑫ 森本容介，学習コンテンツの流通再利用のための拡張可能なWBTアーキテクチャの検討，日本教育工学会第24回全国大会，2008年10月13日，上越教育大学
- ⑬ 辻靖彦，閲覧履歴とリンク情報を用いた検索システムの情報推薦に関する検討，電子情報通信学会WI2研究会，2008年3月22日，北九州学術研究都市産学連携センター
- ⑭ Aki KOBAYASHI, Issai-Gassai RSS Reader: Distributed Design of Single Action RSS Reader, IADIS International Conference WWW/Internet 2007, 2007年10月7日, Universidade de Trás-os-Montes e Alto Douro
- ⑮ 森本容介，Webベースのビデオオーサリングツールの開発と機関紹介コンテンツへの応用，日本教育工学会第23回全国大会，2007年9月23日，早稲田大学
- ⑯ 小林亜樹，学術・教育コミュニティの形成と洗練された検索のためのソーシャルネットワーク，電子情報通信学会2007ソサイエティ大会，2007年9月13日，鳥取大学
- ⑰ 小林亜樹，アクセス履歴とWebGraphを利用した学習コンテンツの推薦型検索に向けての検討，電子情報通信学会WI2研究会，2007年7月14日，ゆふいん七色の風

[その他]

ホームページ等

<http://gloss.code.ouj.ac.jp/>

<http://keyaki.code.ouj.ac.jp/mt/lorsearch.cgi?app=jocw>

<http://intl.code.ouj.ac.jp/>

6. 研究組織

(1) 研究代表者

森本 容介 (MORIMOTO YOSUKE)

放送大学・ICT活用・遠隔教育センター・
助教

研究者番号：00435702