

Dublin Core Metadata Element Set における多言語への対応

永森光晴, Thomas Baker[†], 阪口哲男, 杉本重雄, 田畑孝一

図書館情報大学,

GMD German National Research Center for Information Technology[†]

Email: {nagamori, saka, sugimoto, tabata}@ulis.ac.jp, thomas.baker@gmd.de[†]

概要

Dublin Core は、インターネット上の多種多様な情報資源のメタデータを記述するためのメタデータ基準で、ほとんどの情報資源に共通な 15 の基本要素を定義している。Dublin Core の基本要素の役割を定義している参照記述は、英語での記述をもとにして、20 以上の言語に翻訳されている。現在、DCMI (Dublin Core Metadata Initiative) では Dublin Core における多言語への対応に関して議論がおこなわれている。

我々は、様々な言語で記述された Dublin Core の参照記述を相互に関連づけて管理するためのレジストリを提案し、そのプロトタイプを作成した。Dublin Core の参照記述は RDF Schema を用いて記述し、表示には MHTML (Multilingual-HTML) アプレットを用いることでどの端末でも文字化けすることなく表示することを可能にした。

Dublin Core Metadata Element Set in Multiple Languages

Mitsuharu Nagamori, Thomas Baker[†], Tetsuo Sakaguchi,
Shigeo Sugimoto, Koichi Tabata

University of Library and Information Science,

GMD German National Research Center for Information Technology[†]

Email: {nagamori, saka, sugimoto, tabata}@ulis.ac.jp, thomas.baker@gmd.de[†]

Abstract

Dublin Core, an emerging metadata standard for describing resources on the Web, has already been translated into more than twenty languages. This paper proposes an architecture for linking these multiple versions of the standard into a registry. The registry uses the Resource Description Framework (RDF), an emerging standard for expressing metadata schemas, to allow these versions to share a single namespace and a single set of machine-readable tokens. The prototype, implemented as a Java servlet, holds versions of Unqualified Dublin Core in fourteen languages other than English.

1 はじめに

現在、インターネット上には非常に多数の情報資源がある。図書館や博物館、企業といった情報資源を扱うコミュニティは、これまで個別に情報を蓄積し利用してきた。しかし、こうしたコミュニティがインターネット上の情報資源を共有するようになるに従い、様々な情報資源に関する記述能力を持つ共通のメタデータの記述形式が必要となってきた。Dublin Core Metadata Element Set (一般的に Dublin Core と呼ぶ) はインターネット上の多種多様な情報資源のメタデータを記述するためのモデルである。Dublin Core では、ほとんどの情報資源に共通な、分野に関わらない基本的な 15 の基本要素を定義している [4]。

Dublin Core は英語版を元にして、これまで 20 以上の言語に翻訳され、さらに現在 6 つの言語で翻訳の準備が進められている。このように、さまざまな言語で記述した場合、Dublin Core の Semantic Interoperability を損なわないために、Dublin Core の各要素が異なる言語で記述されていても意味的に等価であると索引付け管理する必要があると考えられる。

本稿では、このような管理機能を実現するために、RDF Schema を用いて Dublin Core Element Set を記述し、様々な言語に翻訳された Dublin Core Element Set を提供するレジストリ設計とプロトタイプ作成をおこなった。以下では、Dublin Core における多言語への対応、レジストリの概念、そしてレジストリの実現について述べる。

2 Dublin Core における多言語への対応

Dublin Core はインターネットのような巨大な情報空間から分野や言語によらずに情報資源を見つけ出すという要求にこたえるために開発されてきたメタデータである。そのため、情報資源のタイトルや主題といった要素の意味は言語とは無関係に、世界共通のものとして与えら

れている。また Dublin Core は素人にもメタデータを記述することができるという観点から、簡潔で少数の基本要素 (Core Element) のみによるメタデータ記述を目指してきた。1997 年 10 月にヘルシンキで開催された第 5 回 Dublin Core Metadata ワークショップにおいて、タイトル、出版者、主題といった 15 項目の基本要素が定義された (DC-Simple と呼ばれている)。

Dublin Core はもともと英語で定義され開発が進められてきたが、タイトル、出版者、主題といった 15 項目の基本要素は、現代の言語であればどのような言語でも表現することは可能である。もし技術的用語に適切な訳が存在しなければ句として説明することも、あるいは新しい語を作ることも可能である。また、実際に Dublin Core を利用して情報資源のメタデータを記述することを考えた場合、メタデータを記述する人が用いている言語で Dublin Core を理解できることが望ましい。例えば日本人であれば日本語で、タイ人であればタイ語で Dublin Core の要素の意味を理解できたほうが、効率よくメタデータを作成できるであろう。

現在、Dublin Core は様々な地域の人々に受け入れられ、数多くの言語に翻訳されている。1999 年 10 月現在でアラビア語、アルメニア語、中国語、チェコ語、デンマーク語、オランダ語、フィンランド語、フランス語、ドイツ語、ギリシャ語、インドネシア語、イタリア語、日本語、韓国語、ノルウェー語、ポルトガル語、パンジャブ語、ロシア語、スペイン語、スウェーデン語、タイ語、トルコ語に訳されている。さらにビルマ語、ハンガリー語、クメール語、ネパール語、ルーマニア語、ウクライナ語で翻訳準備が進められている。

Dublin Core のコミュニティでは、Dublin Core における多言語への対応 (Dublin Core in Multiple Languages, DCML と呼ばれる) について、1997 年 3 月にオーストラリアのキャンベラで開催された第 4 回 Dublin Core Metadata ワークショップ [15] において、英語以外の言語で記述された Dublin Core の管理に関する議論がおこなわれた。現在、多言語への対応は Dublin Core Metadata Initiative (DCMI) [5] のワー

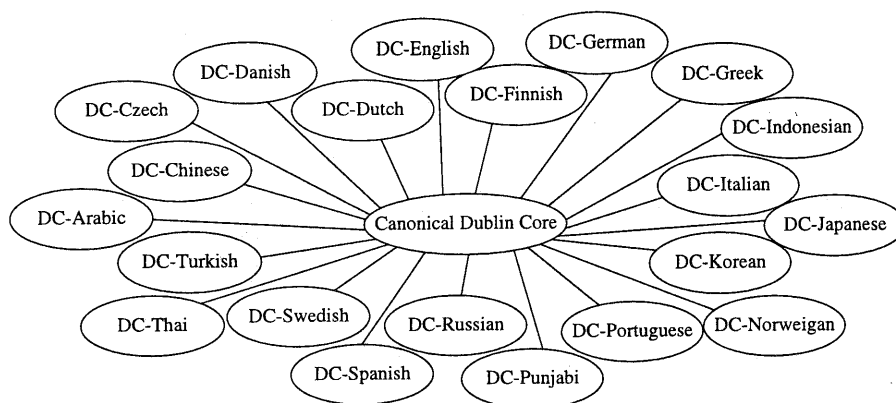


図 1 DCML の概念図

キンググループ (DCML Working Group) で本稿の著者のひとりである Thomas Baker を中心に議論されている [6] [7].

2.1 DCML (Dublin Core in Multiple Languages)

Dublin Core はもともと英語で議論され開発されてきたが、DCML では、Dublin Core が持つ分野、地域、言語によらず情報資源のメタデータを記述するという観点から、全ての言語で記述された Dublin Core の参照記述が同等であると考えている。つまり、言語から中立なユニバーサルな Dublin Core (Canonical DC) を中心に、全ての Dublin Core を相互に結びつけて考えている (図 1)。

現在 DCML ワーキンググループでは、多くの言語で記述された Dublin Core の参照記述を管理するためのレジストリ (DCML レジストリ) に関する議論と開発がおこなわれている。本研究では、その DCML レジストリのプロトタイプとして、Dublin Core の参照記述を管理し、利用者に提供するレジストリ の設計と作成をおこなった。

3 DCML レジストリ

DCML レジストリでは、英語も含め、日本語やタイ語といった様々な言語で記述された Dublin Core の参照記述を管理する。DCML レジストリが持つべき機能には以下のものが考えられる。

- (1) Dublin Core の参照記述を閲覧することができる、
- (2) Dublin Core の参照記述を登録することができる、
- (3) それぞれの言語に独自のサブエレメント (Qualifier) を追加することができる

本稿では、まず図 1 にしたがって各言語の Dublin Core の参照記述を DCML レジストリに登録した。そして機能 (1) を実現し、利用者が Web ブラウザを介して Dublin Core の参照記述を閲覧ができるレジストリを作成した。利用者による参照記述の登録や、それぞれの言語へのサブエレメントの追加に関しては、現在 DCMI で議論が行われているところである。そのため機能 (2) と (3) の実現は今後の課題である。

DCML レジストリでは、全ての言語において、タイトル、著者、主題といった 15 の基本要素を共有するために、それぞれの要素にユニークな token を与える。token は、言語に依存せず、

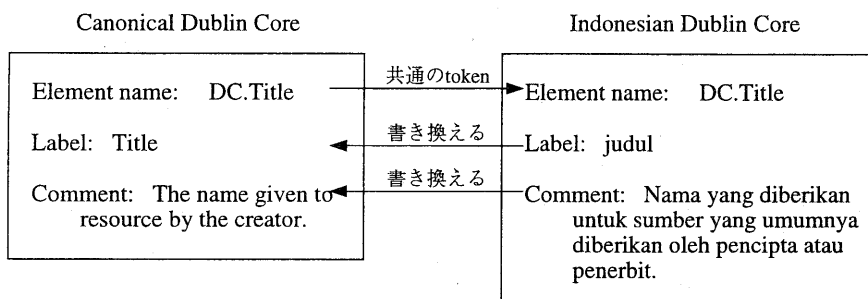


図2 Canonical Dublin Core と翻訳した Dublin Core との関係

機械処理を目的とした名前であればどのようなものでも構わない。各言語の Dublin Core 参照記述では、token を用いて Canonical DC の要素との関連を示す。各言語では、基本要素を表すラベルとその要素の解説をそれぞれの言語で記述する。例えば、インドネシア語の Dublin Core (DC-Indonesian) と Canonical DC との関係は図 2 のようになる。Canonical DC と DC-Indonesian 双方でタイトルという要素は DC.Title という token を共通に用いている。一方、DC-Indonesian では、インドネシア語を用いてメタデータを作成する人のために、ラベルとその解説をインドネシア語で記述する。

DCML レジストリは、利用者からの要求に従って各言語の Dublin Core の参照記述を利用者に送る (図 3)。例えば、タイ語が要求されれば、DCML レジストリはタイ語で記述された Dublin Core の参照記述を利用者に送る。この場合、英語やドイツ語といった、いわゆる Latin-1 フォントであれば一般的に問題なく利用者の Web ブラウザで表示することができるだろう。しかし、日本語、タイ語、韓国語などの場合、利用者の Web ブラウザにフォントがインストールされていなければ、それらの言語で記述された Dublin Core の参照記述を表示することはできない。フォントがなく表示できない場合、DCML レジストリは、Multilingual HTML (MHTML) アプレット [12] を用いて各言語の表示をおこなう。MHTML アプレットは、フォントのインストールを必要とせずに

様々な言語で書かれた HTML 文書を表示することを可能にした Java アプレットである。現在は日本語、中国語、韓国語、タイ語など、10 以上の符号化方法に対応している。DCML レジストリは、MHTML アプレットを用いることで、利用者の Web ブラウザのフォントの有無に関係なく Dublin Core の参照記述を提供することができる。

さらにフォントがなく Java アプレットも使えない Web ブラウザの場合、DCML レジストリは各言語の参照記述が入った画像ファイルを提供する。

4 DCML レジストリの実現

DCML レジストリは、Web ブラウザを介して利用者の要求を受け取り、その要求に従った言語で記述された Dublin Core の参照記述を返す。DCML レジストリは、当該言語で書かれた基本要素の要素名の一覧と各要素の解説を利用者に送る。現在 DCML レジストリで提供しているのは、英語、フィンランド語、フランス語、ドイツ語、ギリシャ語、ノルウェー語、ポルトガル語、スペイン語、中国語 (China)、中国語 (Taiwan)、インドネシア語、韓国語、タイ語、日本語の 14 言語である。プロトタイプでは、RDF Schema [14] 定義に基づいて Dublin Core の参照記述を作成した。また、Dublin Core の参照記述は、Dublin Core 1.0 [4] に従った。図 4 は、Canonical DC と日本語で記述された Dublin

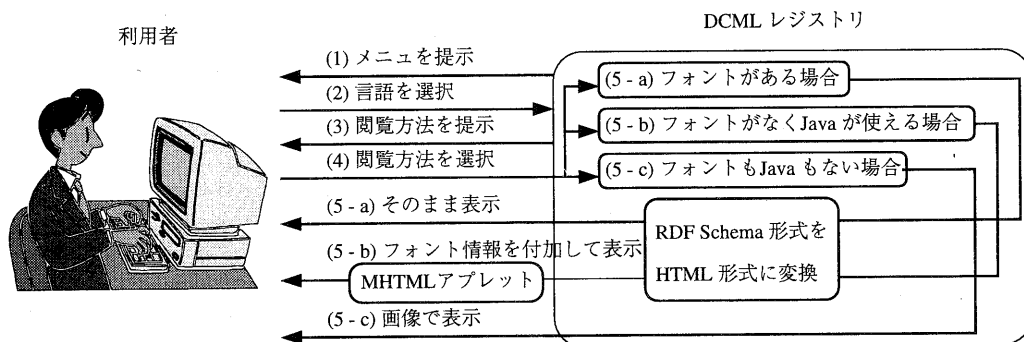


図3 DCML レジストリの処理の流れ

Core の参照記述の一部分である。図4の (a), (b) それぞれは、タイトル要素の記述例である。

各 Dublin Core の参照記述は RDF Schema を用いて記述されているので、DCML レジストリプログラムは、その記述を HTML 形式に変換して利用者に送る。利用者は、Web ブラウザの条件にあわせて適切な閲覧方法を選択する。その閲覧には、次の三通りの方法がある。(1) Web ブラウザが当該言語のフォントを持っている場合。RDF Schema 形式で記述された Dublin Core の参照記述から HTML 形式に変換したものを通常の通常の HTML 文書と同様に表示する。(2) フォントが無く、Web ブラウザが Java をサポートしている場合。MHTML アプレットを利用して HTML 形式に変換した Dublin Core の参照記述を表示する。(3) フォントが無く、Java もサポートしていない場合。Dublin Core の要素名一覧とその解説が入った画像を表示する。

本稿では、DCML レジストリプログラムの実装にサーブレット [10] を用いた。http サーバにはサーブレットを扱うために jserv [11] を付加した apache [1] を用い、RDF Schema を操作するために XML のパーサである IBM XML Parser for Java [9] を利用した。

4.1 利用例

利用者は、まず DCML レジストリの中から参照する言語を選択する (図5)。次に Web ブラウザの機能にしたがって閲覧方法を選択する (図6)。図7はブラウザにフォントが無く、Java をサポートしている場合で、図6で No.2 を選んでタイ語の Dublin Core の参照記述を表示している。この場合、MHTML アプレットを用いているので、利用者はタイ語のフォントを持っていない Web ブラウザでも文字化けすることなくタイ語で記述された Dublin Core の参照記述を見ることができる。

5 おわりに

本稿では、Dublin Core における多言語への対応について述べ、DCML レジストリの設計とそのプロトタイプ作成をおこなった。本稿で作成した DCML レジストリは、図書館情報大学で公開している [8]。詳しくはそちらを参照されたい。

本稿の DCML レジストリは要素名一覧とその解説を表示するだけであるが、今後は検索機能や各言語から任意の同じ要素を取り出して表示する機能などを付加することが考えられる。また、DCML レジストリの分散化も今後の課題である。

今後、DCMI においてサブエレメントを表す

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntaxns#"
  xmlns:rdfs="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#"
  xmlns:dc="http://purl.org/dc/elements/1.0/">
  <rdf:Property rdf:ID="title">←
    <rdfs:label>Title</rdfs:label>
    <rdfs:comment>The name given to the resource, usually by the
    Creator or Publisher.</rdfs:comment>

    </rdf:Property>
    .. (そのほかの要素の記述) ..
  </rdf:RDF>

```

共通のtokenを用いる

(a) Canonical Dublin Core の記述例

```

<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntaxns#"
  xmlns:rdfs="http://www.w3.org/TR/1999/PR-rdf-schema-19990303#"
  xmlns:dc="http://purl.org/dc/elements/1.0/">
  <rdf:Property rdf:about="http://purl.org/dc/elements/1.1/title">←
    <rdfs:label>タイトル</rdfs:label>
    <rdfs:comment>当該情報資源に与えられた名前。一般には作者もしくは公開者によ
    って与えられる。</rdfs:comment>

    </rdf:Property>
    .. (そのほかの要素の記述) ..
  </rdf:RDF>

```

(b) 日本語のDublin Core 参照記述の記述例

図 4 RDF による Dublin Core の記述

Qualifier 等, より細かなメタデータ記述のための記述形式の扱いが決まれば, DCML レジストリに各言語独自のサブエレメントを追加する機能 (例えば日本語であれば Creator に読みのサブエレメントを追加するなど) を付加することなども考えられる。

参考文献

- [1] the Apache Software Foundation, <http://www.apache.org/>
- [2] Baker, Thomas. Dublin Core in Multiple Languages: Esperanto, Interlingua, or Pidgin?, International Symposium on Digital Libraries 1997.
- [3] Baker, Thomas. Languages for Dublin

Core Metadata, D-lib Magazine, <http://www.dlib.org/dlib/december98/12baker.html>

- [4] Dublin Core Element Set, 1.0: Reference Description, <http://purl.org/dc/elements/1.0/>
- [5] Dublin Core Metadata Initiative, <http://purl.org/dc/>
- [6] Dublin Core in Multiple Languages Working Group, <http://purl.org/DC/groups/languages.htm>
- [7] DCML Working Group のワーキングドラフト, <http://purl.org/DC/documents/woking-drafts/wd-multilingual-current.htm>

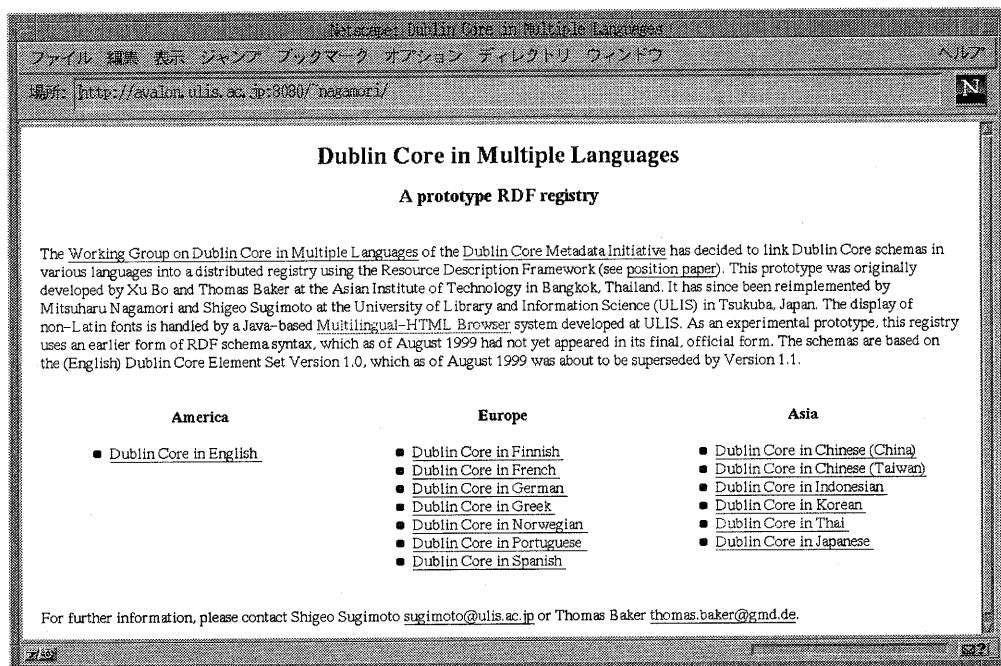


図 5 DCML レジストリのトップページ

- [8] DCML レジストリのプロトタイプ, www.dlib.org/dlib/june97/metadata/06weibel.html
<http://www.dl.ulis.ac.jp/DC/>
- [9] IBM XML for Java, <http://www.alphaworks.ibm.com/>
- [10] Java Servlet API, <http://www.javasoft.com/products/servlet/>
- [11] Java Apache Project, <http://java.apache.org/>
- [12] Multilingual-HTML Browser Project, <http://mhtml.ulis.ac.jp/>
- [13] Resource Description Framework (RDF), <http://www.w3.org/RDF/>
- [14] RDF Schema, <http://www.w3.org/TR/PR-rdf-schema/>
- [15] Weibel, Stuart. The 4th Dublin Core Metadata Workshop Report, <http://www.w3.org/2000/05/21-dc4/>
- [16] eXtensible Markup Language(XML), <http://www.w3.org/XML/>

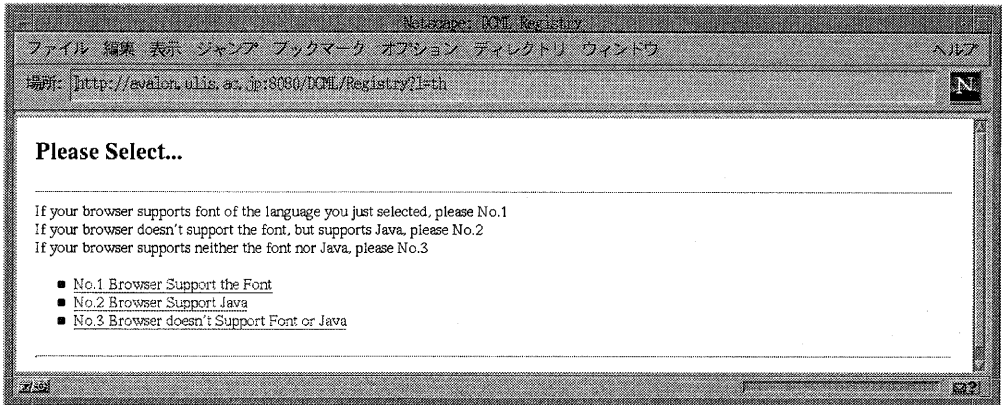


図 6 言語選択画面

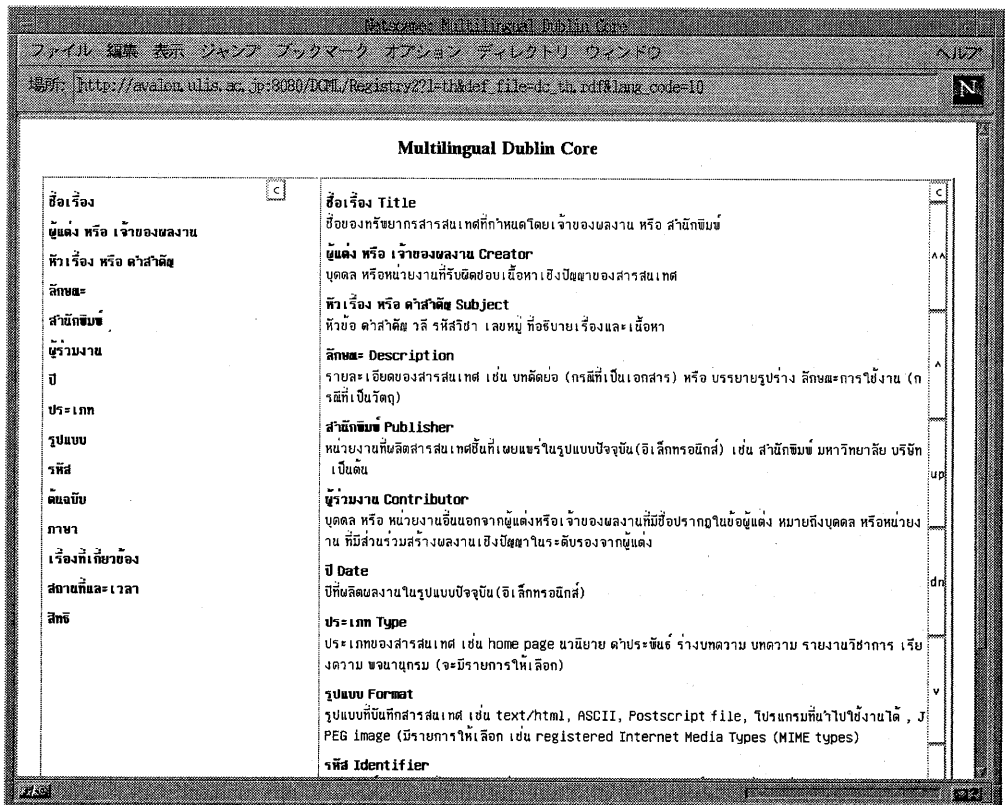


図 7 タイ語 Dublin Core 参照記述の表示例