

L-026

文書管理システム ONTDOC でのオントロジー・メタデータの利用

Content Management System ONTDOC Using Ontology Metadata

年岡 晃一 鈴木 裕利†
Kohichi Toshioka Yuri Suzuki

1. はじめに

インターネットやイントラネットがますます普及して行く中で WEB 上に公開されている情報は爆発的に増加している。Google などの全文検索エンジンはキーワード検索として漏れがない検索機能を提供しているが、反面膨大なヒット文書を返して来るのでユーザはさらにキーワードを付加して絞込みを行わなければならない。WEB 上での商品販売購入システムの e-shopping などの用途では役に立ちそうなページが1つ見つければ事足りることが多いのでこれでも十分である。しかし、上位ランク文書でも1つ1つ文書を読んで目指す情報があるかチェックしなければならない。また網羅的な検索が必要な特許検索や技術論文の検索、また絞込みで特定の探すべき情報が決まっていなくてキーワード指定が出来ない場合の問題が指摘されている。このような要求に応えるべくメタデータを概念体系として表現しようとするセマンティック Web/オントロジーの試みが盛んに行われている[1]。

我々は学習支援システム構築の中でセマンティック Web オントロジーのような情報構造を使って効率的な学習、情報交換・共有を行うことを目指している。その Web アプリケーション ONTDOC で、形式言語としての RDF/OWL で表現されたメタデータの知識情報の構築・管理・交換抽出を行う Web アプリケーションの構築を行った。本稿では、そこで試みている知識関係を俯瞰するインターフェイス、クラスやプロパティ値を辿っていくことによる絞込み検索の方法、及びそれらを実現するためのメタデータをオントロジーとして付与するインターフェイスを提案する。

2. 構築システム

ユーザシナリオとして、オントロジー表現からのタクソノミー構造を使った絞込み検索、内外の Web 文書に対するメタデータ付与機能、各 Web 文書内の部分に対するメタデータ付与、個人利用のメタデータを考える。文書メタデータの管理、構築、参照は、図1に示されるような形態となる。

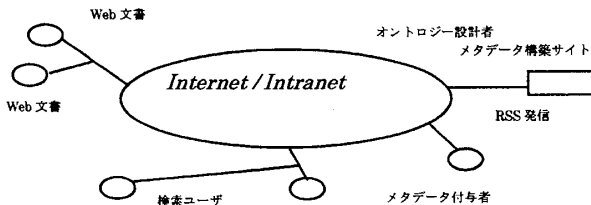


図1. Webにおけるメタデータサイトと参加者

関与するユーザとして、次のグループが考えられる。

†中部大学 工学部 情報工学科

- オントロジー設計者
- メタデータ付与者
- 検索ユーザ (一般利用者)

オントロジー設計者はタクソノミー、クラス、プロパティ及びそれらの相互運用性・オントロジー情報の流通の設計を行う。メタデータ付与者はオントロジー・デザイナーほどオントロジー構成に詳しくないがドメイン情報に精通しメタデータ付与や情報の編集を行う。一般利用者はオントロジーの概念詳細を意識する必要なくメタデータを操作できることが望ましい。

(1) RDFとシステム構成

構築中の ONTDOC のセマンティック Web 関連のデータ及び機能についてシステム構成を図2に示す。

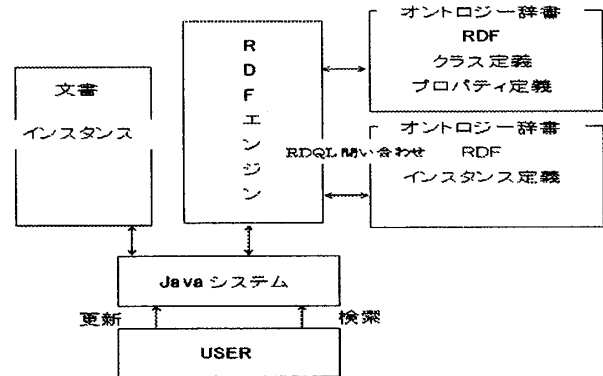


図2. ONTDOCのシステム構成

RDF エンジンには「RDF パーサー」と「RDF クエリ」から成るビジネスロジック部である。RDF パーサーは、Protégé[2]から作られた RDF/OWL 文書からなるタクソノミー定義文書を解析し N-triples[3]形式へ変換の上関係データベース(以後 RDB)へ格納する。RDF クエリは、ユーザからの検索要求に対し RDQL[4]形式で問い合わせを実行する。ユーザへの良好なるインターフェイスを提供すべく Ajax による非同期通信を行う。文書インスタンスでは、イントラネット文書はテキストとして RDB に、またインターネットの外部文書は URI 表現をやはり RDB に格納する。外部文書はプロキシとしての動作が要求される場合もしくは ID 属性で部分文書としてオントロジーを付与したい場合は Java の URLconnection 経由で内容が取得される。

(2) オントロジー

使用するオントロジーを表1、2に示す。

(3) 文書管理

ONTDOC は文書管理機能として、講義・演習の資料公開、オンラインテストなどの機能を持つ。図3にユーザインターフェイス画面を示す。

表1. クラス

クラス	説明	親クラス
機能	文書のサブクラス。或る機能に関する記述。機能の目的、適用領域、サンプル、操作対象などのプロパティを持つ	文書
文書	ONTDOCに登録された文書。一般Webページも可能	rdfs:Class
領域	機能が適用される領域。親領域のプロパティによりツリー構造を持たせることができる	文書
用語	機能の操作対象、入力されるもの、出力されるものを含め機能が扱う対象となる用語	文書

表2. プロパティ

プロパティ	説明	定義域 (domain)	値域 (range)
目的	機能が実現しようとしている目的	機能	rdfs:Literal
適用領域	機能が動作する適用範囲	機能	領域
サンプル	サンプルコードまたは手順の記述	機能	文書
操作対象	機能が操作する対象を用語で指定する	機能	用語
タイトル	文書タイトル	文書	rdfs:Literal
作者	文書作成者	文書	rdfs:Literal
作成日	文書作成日	文書	rdfs:Literal
更新日	文書更新日	文書	rdfs:Literal
URI	文書のURI	文書	rdfs:Resource

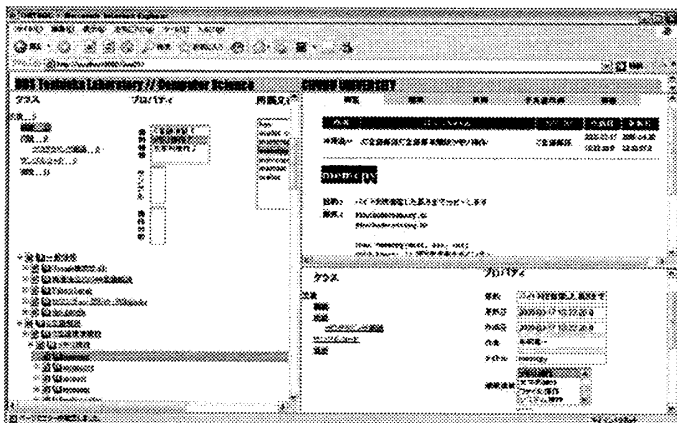


図3. ONTDOCのユーザインターフェイス (メタデータ付与と絞り込みの操作画面)

左ペイン下部の文書ツリーでは全文書が表示されている。現在 memcopy と題する文書が選択されているのを受けて右ペイン上段にその文書自身の内容が表示されている。ONTDOCには外部の一般Web文書もURI記入でONTDOC文書として登録可能で同様にここに表示される。編集タブをクリックすることで選択文書の編集が可能となる。Wiki風にタグによる指定を記入することで文字飾りやテーブル表示、ファイル転送やアンケート調査の動的処理機能を各文書に持たせることができる。

(4) メタデータ付与の操作

選択文書に対応するメタデータは右ペイン下段に示されている。各文書のメタデータはオントロジーのクラスのイ

ンスタンスとして生成・登録される。メタデータの登録や変更は、所属するクラスの変更も含めここで行なえる。クラスを変更するとそれに応じて所属するプロパティも増減して表示される。タクソノミー定義も変更されると動的に画面も変更される。非同期通信の Ajax で実装されているので画面遷移も押さえられ良好なインターフェイス使用感を得ている。クラス、プロパティは推移律が成り立てばツリー構造として表示される。

この画面では、選択文書 memcopy のクラスとして機能クラスが選択されており、それが持つプロパティ群が右側に表示されている。目的、更新日、作成日、作者、タイトルは値域 rdfs:Literal として定義され文字列入力となる。インスタンス登録ボタンが押されると N-triples の関係データベース内の値が変更される。適用領域、サンプル、操作対象は ObjectProperty として定義されているので値域は或るクラスのインスタンスを入れることとなりここでは ONTDOC 内に登録されたインスタンス群がセレクト内の選択候補として提示される。

このインターフェイスによりメタデータ付与者もオントロジーやクラス概念をそれ程知らなくても或いは意識することなくメタデータの付与を各文書に行うことが可能となる。ONTDOCに登録された外部のWeb一般文書もメタデータの付与・参照を同様に行うことができる。蓄積されたメタデータはRSSによりWeb発信されることになる。

(5) 絞りこみ検索の操作

左ペイン中央部ではクラス・プロパティ・所属文書と題する3つのセレクトがあり絞り込み検索を行う。領域クラスの11インスタンス、機能クラスの15インスタンスがメタデータを持つ文書として存在していることを示している。絞り込みで機能クラスが選択されているので関連するプロパティ値及び所属している memmove, memcopy と題する文書群が閲覧文書候補として掲げられている。

3. まとめ

開発したセマンティックWeb機能付き文書管理システム ONTDOC を紹介した。各文書にメタデータとしてオントロジーのクラスやインスタンスを付与などの知識構築ができること、それを利用した絞り込み検索の操作手順が示された。構成がゆるく知識構築の拡張性が高いこと、プロパティ値の欠損からの知識補充の必要箇所の発見、網羅的検索、継承関係・プロパティ関係の提示による情報マップは従来の検索を大きく補完すると期待される。

今後は RDF/OWL の汎用性、複数のオントロジー定義の取り込み、フレームワークとしての公開に対応していく考えである。

参考文献

[1] S. Staab et al. Semantic Community Web Portals Proceedings of the 9th. International WWW Conference). 2000
<http://www9.org/w9cdrom/134/134.html>
 Annotea <http://www.w3.org/2001/Annotea/>
<http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707>
 GRDDL <http://www.w3.org/TR/2007/PR-grddl-20070716/>
 [2] Protégé <http://protege.stanford.edu/>
 [3] N-triples <http://www.w3.org/2001/sw/RDFCore/ntriples/>
 [4] RDQL -A Query Language for RDF,
<http://www.w3.org/Submission/2004/SUBM-RDQL-20040109/>