

グループ研究発表（整理技術研究グループ）

セマンティック Web と資料組織法

- 概念体系管理の今後

渡邊隆弘（神戸大学図書館）

0 . はじめに¹

資料組織法の動向

- ・記述目録法の活発な変化
FRBR、目録諸規則の改訂（1990年代後半～）
AACR 全面改訂など、さらに抜本的な動き
- ・主題アクセス法の停滞
分類・件名・シソーラスには目立った動きがない

「セマンティック Web」の登場

- ・セマンティクス（意味）情報を付加した Web
Web の「情報組織化」
図書館の視点から見た Web
従来の Web : 「操作対象資料」（目録の対象）
セマンティック Web : 「組織化技術」として相通じる
- ・主題アクセス、概念管理の今後につながる可能性？

本発表

セマンティック Web と資料組織法の接点、とりわけ主題アクセスへの応用可能性について基礎的考察

1 . セマンティック Web の提唱と開発²

セマンティック Web の基本構想

- ・ T. Berners-Lee（Webの発明者）による³

¹ グループ活動記録等は <http://www.tezuka-gu.ac.jp/public/seiken/> に蓄積している。

² 2004 年後半から、以下のものなど、いくつかの邦文図書が出版されている。

神崎正英『セマンティック・ウェブのための RDF/OWL 入門』森北出版, 2005.1. 224p

情報処理相互運用技術協会編『セマンティック Web 入門』オーム社, 2004.11, 242p

また、わが国では情報処理相互運用技術協会（INTAP）の「次世代 Web 委員会」が調査・啓蒙活動を活発に行っており、多くの情報が Web 上で公開されている。

<http://www.net.intap.or.jp/INTAP/s-web/>

³ Web は 1989 年にスイスの研究機関 CERN で Berners-Lee によって開発された。

T. Berners-Lee 著 高橋徹訳『Web の創成 : World Wide Web はいかにして生まれどこに

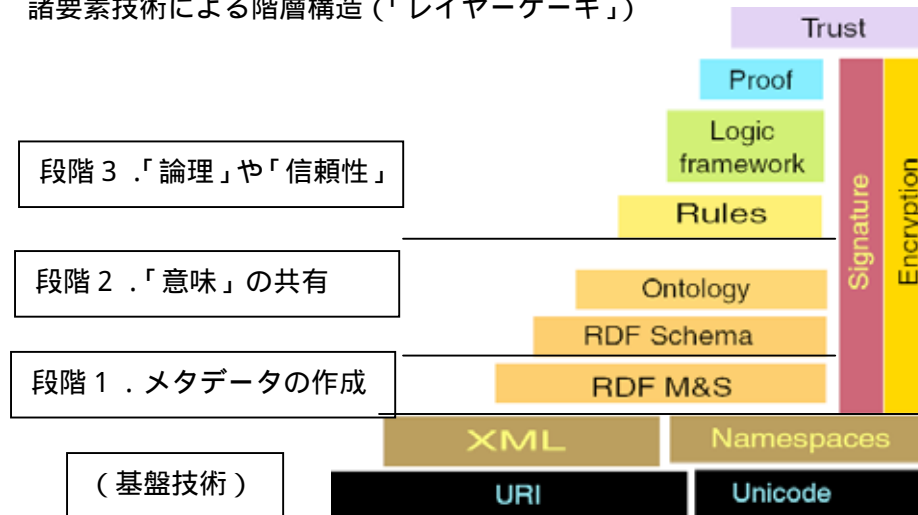
Web創始当時から暖めていた考え方

2001.5 発表論文 「自分で推論する未来型Web」(邦訳題)⁴

- ・ エージェント・ソフトウェアが人に代わって問題解決のためのデータ収集・判断・評価を行う Web 世界
データ収集・判断・評価の自動化⁵
- ・ 文書構造やレイアウト情報だけでなく、「意味 (Semantic)」情報も交換
- ・ 人だけではなく機械に理解できる Web データ
リソースの内容を明示的かつ標準化された形式で共有
具体的には、「メタデータ」の付与と活用
- ・ Web の分散・非集権環境は堅持
- ・ 標準化活動は、W3C (The World Wide Web Consortium)で⁶

セマンティック Web の諸要素技術

- ・ 諸要素技術による階層構造 (「レイヤーケーキ」)



段階 1 .メタデータの作成 - 共通的な構文枠組み

- ・ RDF モデル&シンタックス (M&S) 層
Resource Description Framework
- ・ メタデータを明瞭・論理的に表現するデータモデル⁷

向かうのか』毎日コミュニケーションズ, 2001

⁴ Berners-Lee, T. et al. "The semantic Web" *Scientific American*. 284(5), 2001. pp. 34-44. <http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>

[邦訳] 村井純ほか訳「自分で推論する未来型ウェブ」『日経サイエンス』. 31(8), 2001, pp.54-65

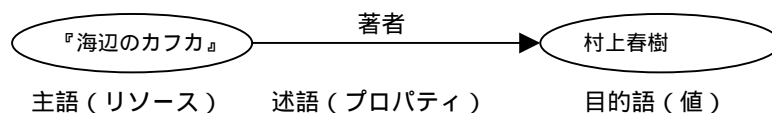
⁵ Berners-Lee論文では「家族の病気を治療してくれる病院探し」を例にとり、患者や家族の諸条件をもとに、ソフトウェア(エージェント)が病気や病院に関するWeb上の情報を勘案して最適解を提示してくれるという未来図を描いている。

⁶ "Semantic Web" <http://www.w3.org/2001/sw/>

RDF, OWL, Rules の独立した WWW ページも W3C トップページよりたどれる。

⁷ RDFはセマンティックWebではじめて提唱されたものではなく、ダブリン・コアの開発と同時期にメタデータの構文枠組みとして登場した。データ要素(タイトル、著者など)の

- どのようなメタデータ規則にも適用できる汎用性
モデルと、それに対応した構文記述（XML による）を規定
- 原理はきわめて単純
「主語、述語、目的語」のセット・・・「トリプル」



- 単純だが、汎用性
複雑な構造も「トリプル」の組み合わせで表現可能（樹状のグラフとなる）
- 既に一定の実用段階にある

段階2 . 「意味」の共有 - プロパティや値の相互関係を伝達

- RDF スキーマ層（RDF Schema）とオントロジ層（Ontology） 具体的には後述
- プロパティ情報を誤解なく共有
異なるメタデータ規則のプロパティ「責任表示」と「作成者」の対応は？
- 値情報を誤解なく共有
「自動車」と「乗用車」の対応関係は？
あるページ中の単語「鍋」は料理なのか調理器具なのか？
- 標準化作業はほぼ終わった段階

段階3 . 「論理」や「信頼性」 - 推論・判断の共有

- 具体化はまだこれから

ルール層（Rules）	共通基盤となる論理式の定義
ロジックフレーム層（Logic Framework）	個々の枠組みに応じた論理式
プルーフ層（Proof）	処理の履歴や理由を共有
トラスト層（Trust）	情報の信頼度評価法を共有

2 . セマンティックWebと資料組織法⁸

資料組織法との親和性

- 「相互利用性」 = セマンティック Web の目的
「分散環境での情報生産を前提に、どう確実に共有していくか」
- セマンティック Web の出発点はメタデータの付与
素朴に図書館目録になぞらえて説明されることも

標準化ではなく、その構文表現（記述法）を統一することのみを目的としている。

⁸ いずれも短いものだが、セマンティックWebと図書館・図書館学との接点について、以下の文章で文献紹介を中心にまとめた。

渡邊隆弘「セマンティックウェブと図書館」『カレントアウェアネス』281, 2004.9

<http://www.ndl.go.jp/jp/library/current/no281/doc0006.htm>

渡邊隆弘「セマンティック Web」『図書館・情報学研究入門』（三田図書館情報学会編 勁草書房発行），2005.10. pp.79-81

米欧における論調

- ・ 図書館情報学から取り上げた文献は 10 本以上 (2001 ころ ~)
- ・ 主に論じられる接点
 - 典拠コントロール
 - 主題アクセス
 - ポータル、リポジトリ、電子図書館
- ・ どちらかという「図書館がセマンティック Web に貢献できる可能性」がよく語られる
 - 規則・ツール (目録規則、MARC、分類表) 構築・維持の伝統
 - 典拠コントロールや主題管理の伝統
- ・ 例 : ミラー他「図書館とセマンティック Web の将来」⁹
 - 図書館コミュニティの経験
 - メタデータの構文・構造や意味における伝統 (MARC、目録規則)
 - 典拠コントロールや異種データ取扱いの経験
 - 利用者行動や「ビッグデータ」を扱ってきた経験
 - 図書館こそが貢献できる場
- ・ 全体的な感想
 - 「楽観的」 図書館情報学の障壁拡大 ?
 - 「発散的」 各論者が思い思いの論点に引き付け

セマンティック Web と資料組織法の伝統との差異

集権と分散

図書館 : 目録規則、統制語彙を統一化してデータを標準化
分散させるならなお、ルールの精緻化

セマンティック Web : あくまで分散、非集権を前提

情報の粒度

図書館 : ドキュメント単位のアクセスを前提とする主題索引

セマンティック Web : 知的エージェントの推論に資するオントロジ
もっと細かなレベルでなくては

差異を意識しながら、資料組織法への応用可能性を探るという方向性も考えられる

3 . オントロジによる概念管理

オントロジとは¹⁰

- ・ Ontology もともとは哲学用語「存在論」
- ・ 1980 年代以降、人工知能等の分野で注目
- ・ 定義 : 「概念化の明示的な規約」

⁹ Miller, Eric et al. "Libraries and the future of the semantic Web : RDF, XML, and alphabet soup". *Cataloging the Web : metadata, AACR, and MARC21*. Jones, Wayne et al., ed. Lanham, Scarecrow Press, 2002, 57-64

¹⁰ セマンティック Web に限定せず、広くオントロジを概説した図書に、次のものがある。
溝口理一郎『オントロジー工学 : 知の科学』オーム社, 2005.1. 275p

対象世界の基本概念や概念間の関係を体系づけたもの

セマンティックWebにおけるオントロジ¹¹

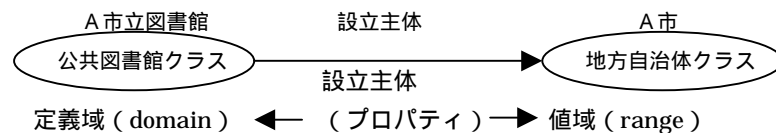
- ・ 「意味の共有」を扱う2層の標準化
 - RDFスキーマ層 = RDF語彙記述言語 (RDFスキーマ)¹²
 - 比較的基本的なレベルの意味情報 (クラス表現や階層構造)
 - オントロジ層 = オントロジ記述言語 (OWL: Web Ontology Language)¹³
 - RDFスキーマの基礎の上に立って、より精密な意味情報管理
- ・ 標準化作業はほぼ終わった状態 (実装の際の構文記述はXML)

セマンティックWebにおけるオントロジの目的

- ・ メタデータの「プロパティ情報」の共有
 - 異なるメタデータ規則のプロパティ「責任表示」と「作成者」の対応?
 - 目録規則やメタデータ規則の相互運用性に通じる
- ・ メタデータの「値情報」の共有
 - 「自動車」と「乗用車」の対応関係は?
 - あるページ中の単語「鍋」は料理なのか調理器具なのか?
 - 主題アクセスツールに通じる
- ・ オントロジ標準化の目的 - 構築のルール・文法を標準化
 - 普遍的、集権的なオントロジ構築ではない
 - 異なるオントロジ間の相互運用性 (共有や併合)

OWL (Web Ontology Language) の基本構造

- ・ 基本要素
 - クラス (Class)
 - リソースをグループ化したもの ex. 「図書館」「哺乳類」
 - 個体 (Individual)
 - クラスのメンバーとなるインスタンス
 - ex. 「公立図書館」クラスのメンバー=A 市立図書館、B 県立図書館...
 - プロパティ (Property)
 - リソース間を関係づけるもの。クラスとは独立して定義
- ・ クラスとプロパティの設定により、対象世界を表現
 - RDFの発展型



¹¹ 2000年に国際標準化 (ISO 13250) された「トピックマップ (TopicMap)」という技術も、セマンティックWebとの関連で脚光を浴びている。様々な「トピック」間を「関連 (association)」で結ぶ方式で、OWLの考え方と相通じる面がある。

¹² W3C. *RDF Vocabulary Description Language 1.0: RDF Schema* <<http://www.w3.org/TR/rdf-schema/>>

¹³ W3C. *Web Ontology Language (OWL)* <<http://www.w3.org/2004/OWL/>>

OWL (RDF スキーマを含む) における概念定義の基本

- ・ 関係性記述を基本とする意味の定義
コトバによる定義には頼らない
コンピュータに理解可能な定義表現
- ・ クラス、個体の基本的関係
クラスと個体の所属関係
「公立図書館」クラスのメンバー = A市立図書館、B県立図書館...
クラス間のサブクラス関係
「公共図書館」クラスは「図書館」クラスのサブクラス
Class (ex:公共図書館 partial subClassOf (ex:図書館))¹⁴
プロパティの設定により、クラス間・個体間の諸関係を表現
逆に、プロパティは対象クラスの設定によって定義

OWL における関係表現の諸相 (例)

- ・ クラスの論理組み合わせ
Class (ex:電子鍵盤楽器 complete intersectionOf (ex:電子楽器 ex:鍵盤楽器))
クラスの論理積 (論理和、論理差の表現もある)
「互いに素」の定義も DisjointClasses (ex:哺乳類 ex:鳥類 ex:爬虫類)
- ・ プロパティにも諸関係
階層関係 「提携」 と 「資本提携」
反対関係 「捕食」 と 「非捕食」
- ・ プロパティの性質を通じて関係表現
推移型 「下部組織」 A B、B C ならば A C
対称型 「姉妹都市」 A B ならば B A
関数関係 「設立主体」
定義域 (公立図書館) が決まれば地域 (自治体) は一意

4 . オントロジと主題アクセスツールの可能性

図書館の世界の主題アクセスツール

- ・ 近年、それほど大きな動きがない
- ・ 分類法
列挙型分類、ファセット分類
- ・ シソーラス
ISO2788
- ・ 件名標目表
シソーラス化が進む
(LCSH は 1988、BSH は 1999、NDLSH は 2004)

¹⁴ OWLには、実装時に用いられる「交換構文」(RDF/XMLによる) と、モデル的な「抽象構文」(セマンティクスを関数風のスタイルで記述) がある。本稿では後者を例示的に示すが、記述法には触れない。詳しくは注 2) 文献等を参照されたい。

シソーラス (ISO2788) の概念管理

- ・ スコープノート (SN) による限定
- ・ 基本関係として、等価関係、階層関係、連想関係
- ・ 等価関係 (USE, UF)
 - 優先語と非優先語の関係
- ・ 階層関係 (BT, NT)
 - 種類関係 鳥 / オウム 胃腸薬 / 整腸剤
 - 一部の全体部分関係 耳 / 内耳 軍団 / 師団
 - 身体組織、地理的位置、学問分野、社会構造の4種
 - 事例関係 山脈 / アルプス山脈、ヒマラヤ山脈...
 - 階層関係の定義を厳密に
 - 一般的な全体部分関係 (自動車 / タイヤ) は階層としない
- ・ 連想関係 (相互に RT)
 - 等価・階層以外の定義・・・明確な定義はない
 - 学問分野とその対象 地震学 / 地震
 - 操作と道具 速度測定 / 速度計
 - 動作と生産物、行為と受動者 染織 / 染物 栽培 / 作物
 - 性質に関する概念 毒物 / 毒性
 - 因果関係を持つ概念 病気 / 病原体
 - 事物と対抗物 植物 / 除草剤
 - ...

オントロジの手法による改善の可能性

- ・ より機械処理になじむ概念管理
 - スコープノートに頼らず、他概念との関係性による定義
 - = 機械操作可能な定義表現
- ・ 階層関係以外の関係構造表現
 - RT はシソーラス手法の大きな弱点
 - (相互関係を示すだけなので、付けすぎると混乱)
 - プロパティの自由な設定による克服可能性
 - より豊富な関係情報による概念のネットワーク
 - 階層関係も含めて、より精緻な関係構造

5 . おわりに

今後の課題

- ・ さらに考察すべき事項
 - 階層関係の区分原理の表現
 - 統語論的關係の表現
 - など
- ・ より具体的な検討
 - 必要なプロパティは何か、などディシプリンごとに検討の必要
- ・ より明快で機械操作可能な概念体系管理を目指して