

軽量オントロジーを用いた簡易型 Web コンテンツ管理システムの開発

Development of Web Contents Management System based on Light-Weight Ontology

古崎 晃司
Kouji KOZAKI

飯田 哲也^{*1}
Tetsuya IIDA

來村 徳信
Yoshinobu KITAMURA

溝口 理一郎
Riicro MIZOGUCHI

大阪大学産業科学研究所

The Institute of Scientific and Industrial Research, Osaka University

The research of nanotechnology is extended in various domains, and each domain intertwines with each other closely. The Structuring Nanotechnology Knowledge project, which is a NEDO funded national project, has been carried out. The goal of the project is to build a material-independent platform for supporting development of innovative nano-materials, which is called "Structured Knowledge Platform for Nano-materials and Products". It consists of many contents as follows: conceptual design systems for nano-materials, simulators, electronic textbook, scientific papers/patents database, and so on. These are developed as web-based systems. However, domains and grain sizes of the contents are different from each other. Therefore, we need a consistent framework to interoperate them efficiently. In this background, we aim to develop a web contents management system based on a light-weight ontology, which provides a common vocabulary for contents. This article outlines the system with its functionalities.

1. 研究背景・目的

ナノテクノロジーに関わる知識は多岐にわたっており、それらを領域横断的に俯瞰することは非常に重要である[小宮山 04]。そこで、多彩な研究領域にまたがる知識を構造化することで、複数領域間での研究成果や知識の交流を実現し、新規分野開拓の促進することを目的とした、新エネルギー・産業技術総合開発機構(NEDO)による「材料技術の知識の構造化」プロジェクトが行われている。このプロジェクトでは、様々な領域の知識を構造化して計算機に格納する「知識の構造化プラットフォーム」の開発が進められている[NMC]。知識の構造化プラットフォームには、分野別知識の構造化システム、機能-構造-プロセスの概念設計システムをはじめ、シミュレーター、電子教科書、データベース群など、既に多くのコンテンツが格納されている[古崎 04]。

しかし、増大する多くのコンテンツを効率的にシステムに格納

し、相互に連結させて有効利用するには、これらを統合的に管理する枠組みが必要となる。そこで本研究は、オントロジー工学に基づいてナノテクノロジー知識を構造化し、コンテンツの種類や対象とする研究領域によらず統一的に管理することを実現する計算機システムの開発を目的としている。本稿では、その基盤システムの一つとして開発を進めている、軽量オントロジーを用いたコンテンツ管理システムについて報告する。

2. 知識の構造化プラットフォーム

2.1 知識の構造化プラットフォームの構成

知識の構造化プラットフォームは、実験データや各種文献などを収集・構造化したデータベース群からなる「データ層」、データ層で収集された知識を分析し様々な視点から構造化した「情報層」、それらの知識を一般化しプラットフォームの知識基盤を与える「知識層」から構成される(図1)。各階層に格納されているコンテンツは、様々な形式のデータや文献、シミュレーターや知識システムなど、多岐にわたっており、現在も既存コンテンツの拡充や新しいコンテンツの開発が進められている。これらすべてのコンテンツはWebベースでの公開を想定して開発されており、一部のユーザには試験的に公開されている。

2.2 コンテンツ管理システムの位置づけ

知識の構造化プラットフォームにおいては、増大するコンテンツの効率的な管理と有効利用を促進するため、コンテンツを統合的に管理するシステムの開発が行われている。

ここでシステムに要求されるのは、1)コンテンツをプラットフォームに格納する為の統一的かつ効率的な枠組みを提供し、さらに、2)各コンテンツをその内容の関連性に基づいて有機的に連結させることにある。この実現により、様々なコンテンツ拡充の効率化や、コンテンツの利用しやすさの向上、さらにはその結果として領域を横断した新発想の促進が期待される。この要求仕様の考察に基づき、以下の2つのシステムを中心に開発が進められている。

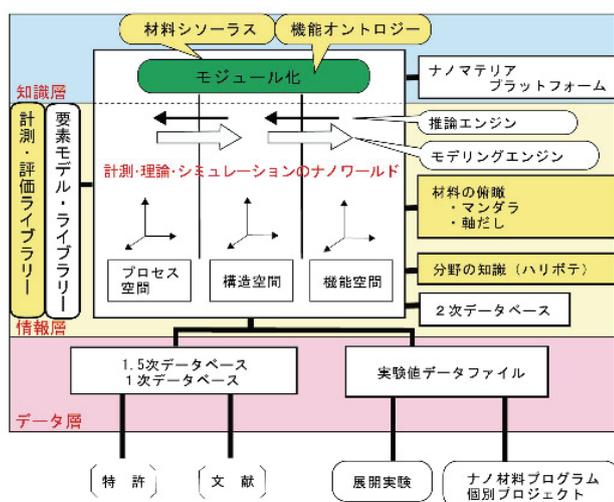


図1 知識の構造化プラットフォーム[NMC]

連絡先: 古崎晃司, 大阪大学産業科学研究所
ナノバイオ知能システム分野, 大阪府茨木市美穂が丘 8-1,
TEL: 06-6879-4304, e-mail: kozaki@ei.sanken.osaka-u.ac.jp

^{*1} 現在, 浜松ホトニクス株式会社

(1) Nanotech Index Ontology(NIO)管理システム

プラットフォームに格納されるコンテンツを、その「内容」に対する一貫した知識語彙体系を提供する軽量オントロジーである Nanotech Index Ontology (NIO)に基づいて管理する。これにより、プラットフォーム内の全コンテンツの情報を、「内容」という知識に直結した観点から、統一的に管理することが可能になる。

(2) 知識ネットワークシステム

コンテンツに含まれる知識を構造化し、計算機処理可能なネットワーク(グラフ)形式で管理する。構造化された知識をネットワーク表現という一貫性のある形式で可視化することで、様々な領域の知識を一貫性を持って扱うことができる。またネットワークを辿ることで、関連する知識をコンテンツの枠を越えて有機的に結合することが可能となる。ここでネットワークのノードで現される要素知識は、NIO 管理システムによって管理される。

これらのシステムに共通している考え方は、コンテンツの作成・管理方法を一元化し、計算機を用いた意味処理が可能な形式でプラットフォームに格納する点にある。次章以降、NIO 管理システムのプロトタイプとして開発した、軽量オントロジーを用いたコンテンツ管理システムについて述べる。

3. 軽量オントロジーを用いたコンテンツ管理

知識の構造化プラットフォームでは、様々な種類のコンテンツを統一的に管理する必要がある。そのため、特定のデータベースなどに格納されたデータのように、一定のスキーマを用いてコンテンツを管理することは難しい。そこで本研究では、コンテンツの内容を表すための一貫した知識語彙体系として、インデックスオントロジーと呼ばれる軽量なオントロジーを用いることで、コンテンツを管理する枠組みを開発した。これにより、ユーザは、「内容」という知識に直結した観点から、コンテンツ管理を行うことが出来る。

3.1 インデックスオントロジー

インデックスオントロジーとは、コンテンツ管理のためのインデックスとしての利用に特化したオントロジーで、コンテンツ内で使用されている概念(語彙)を is-a 関係のみを用いて体系的に記述する。より厳密なオントロジーでは is-a 関係以外に、全体-部分関係、属性、公理など詳細な概念定義の記述も行われるが、本システムではコンテンツの分類情報が利用の中心となるため、is-a 階層の持つ情報が通常にも増して大切であり、その他の詳細な記述に欠けていたとしても十分に利用することができる。

3.2 軽量オントロジーを用いたコンテンツ管理

本研究では、軽量オントロジーを用いたコンテンツの管理機能として、「コンテンツの内容に基づく管理機能」および「関連するコンテンツの連結機能」という2つの機能を実現するための枠組みを開発した。以下、それぞれの枠組みについて述べる。

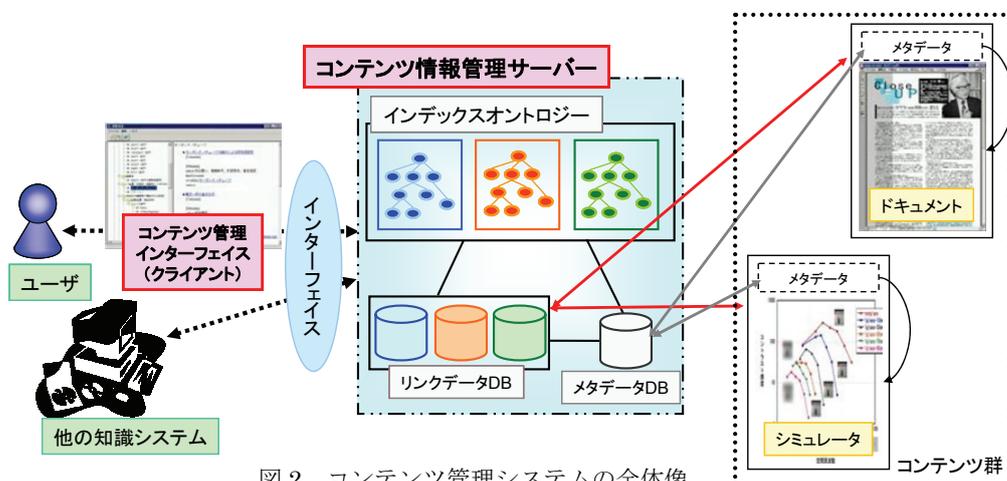


図2 コンテンツ管理システムの全体像

(1) コンテンツの内容に基づく管理の枠組み

コンテンツを管理するには、その種類や格納された場所、特定の ID など様々な情報を利用することができる。しかし、これらの情報はコンテンツの種類に応じて異なることが多いため、様々な種類のコンテンツの統一的な管理には適していない。そこで、本枠組みではインデックスオントロジーで定義された概念と、その概念が表す内容に関するコンテンツを対応付け、その対応関係を管理することで、コンテンツをその内容に基づき統一的に管理する。

この枠組みを実現させるために、「リンクデータ」と呼ばれるデータモデルを導入する。リンクデータとは、インデックスオントロジーの各概念に対して 1 つ存在するデータで、その概念に対応付けられたコンテンツのリストを保持するものである。必要なコンテンツの情報を取得する際には、インデックスオントロジーからコンテンツの「内容」を表す概念を選択し、その概念に対する「リンクデータ」を参照することでコンテンツの一覧が得られる。

なお、リンクデータで保持するコンテンツの情報としては、Web システムにおける URI のような、コンテンツを一意に同定することができる情報が必要となるが、それは本枠組みをシステムとして実装する際に検討すればよい。

(2) 関連するコンテンツを連結する枠組み

ナノテクノロジーの分野では、様々な領域の知識が互いに密接に関係しているため、知識の構造化プラットフォームにおいて、コンテンツ間の関連を管理することが重要となる。そこで本枠組みでは、コンテンツ間の関係を管理するために、各コンテンツと関連する他のコンテンツのリストを用意する。関連するコンテンツのリストは、その内容を表すインデックスオントロジーの定義した概念のリストとして管理する。この概念のリストは、(1)で対応付けられたものとは異なり、そのコンテンツが主に扱っている内容だけではなく、それに関連するより広い範囲の内容を表す概念を含める。

この枠組みを実現させるために、各コンテンツに対してメタデータを用意し、そのコンテンツに関連する内容を表す概念のリストを保持する。関連するコンテンツの情報を取得する際には、このリストにある概念を介して、先に述べた「内容に基づくコンテンツ管理の枠組み」を用いる。

なおメタデータには、コンテンツの種類、タイトル、作成者など、コンテンツに関する様々なといった情報を記述することも可能であるが、その形式や内容は枠組みの実装時に検討すればよい。

4. コンテンツ管理システムの開発

4.1 コンテンツ管理システムの構成

3章で述べた枠組みに基づいて開発したコンテンツ管理システムの全体像を図2に示す。本システムは、「コンテンツ情報管理サーバー」と「コンテンツ管理インターフェイス」から構成されている。「コンテンツ情報管理サーバー」は3章で述べた2つの枠組みを実装したシステムで、インデックスオントロジーに基づき、リンクデータとメタデータを用いてコンテンツの情報を管理している。他のシステムとはインターフェイスを介して接続され、必要に応じてコンテンツの情報を提供する。「コンテンツ管理インターフェイス」はコンテンツ管理サーバーにアクセスし、コンテンツの情報を管理するためのグラフィカルユーザーインターフェイスを提供するクライアントシステムである。

本システムは、JAVAを用いて実装しており、サーバーとクライアント間の通信にはhttpプロトコルを利用している。

以下の節では、「コンテンツ情報管理サーバー」と「コンテンツ管理インターフェイス」について述べる。

4.2 コンテンツ情報管理サーバー

コンテンツ情報管理サーバーは、インデックスオントロジーおよび、リンクデータDB、メタデータDBから構成されている。以下、それぞれの概要を述べる。

(1) インデックスオントロジー

インデックスオントロジーは、筆者らが開発を進めているオントロジー構築・利用環境「法造」[古崎 02a]を用いて構築した。インデックスオントロジーを本システムで処理する際には、法造で構築したオントロジーを操作するためのAPI [古崎 02b]を用いて実装した。

また、法造のAPIの代わりに、Jena[Jena]を用いることで、OWLやRDF(S)で記述したオントロジーをインデックスオントロジーとして利用することができる。

(2) リンクデータDB

リンクデータDBは、インデックスオントロジーの各概念に対して1つのファイルをCSV形式データとして作成しシステムに格納したデータベースであり、コンテンツを識別する情報としては、URLを利用している。リンクデータの各CSVファイルの内容は1行毎にコンテンツのタイトル、コンテンツのURL、コンテンツに対するコメントを記述したものとなっている。



図4 NIO管理システム

(3) メタデータDB

メタデータDBは、各コンテンツのメタデータに対応する1つのファイルをCSV形式データとして作成し、ファイルシステムに格納するデータベースである。メタデータを各コンテンツに直接付与するのではなく、データベースでまとめて管理することにより、直接メタデータを付与することが難しい外部のWebコンテンツに関するメタデータも扱うことができる。メタデータの各CSVファイルには、関連するコンテンツの内容を表すインデックスオントロジーの概念のリストの他に、コンテンツの作成者、コンテンツの種類、コンテンツの出典などを記述している。

4.3 コンテンツ管理インターフェイス

図3にコンテンツ管理インターフェイスの画面例を示す。ユーザはこの画面上でコンテンツの必要な情報の表示や、リンクデータおよびメタデータの追加や削除、編集といった各種操作を行う。

このウィンドウは、インデックスオントロジーの概念をis-a階層に沿って木構造で表示する画面(「Tree Pane」)とインデックスオントロジーの概念に対応づけられているコンテンツの情報を表示する画面(「Link Pane」)から構成される。Tree Paneで概念を選択すると、選択された概念に対応づけられたコンテンツの一覧を、コンテンツ情報管理サーバーから取得しLink Paneに表示する。Link Paneに表示される情報は、コンテンツ情報管理サーバーから取得したメタデータに記述された情報である。さらに、Link Pane上で表示されたHyper Linkをクリックすることで、対応づけられたコンテンツや、関連するコンテンツ情報の表示ができる。

4.4 知識の構造化プラットフォームにおける利用

本節では、本研究で開発したコンテンツ管理システムが、知識の構造化プラットフォームにおいて、どのように利用されているかを述べる。

インデックスオントロジーとしては、知識の構造化プロジェクトにおいて、文献などから抽出したキーワード群や材料に関するシソーラスを元に開発されたNanotech Index Ontology (NIO)を用いた。NIOは、プロセス、構造、機能、材料、応用に関する総概念数約2300の概念階層からなる。このNIOを用いて3章で述べた枠組みを実装したNIO管理システムを開発した(図4)。NIO管理システムは、4.3節で述べたコンテンツ管理システムを拡張した形で実装したシステムとなっている。このシステムは、コンテンツ情報の管理機能に加え、拡張された機能として、

- NIOとして用意した5種類のインデックスオントロジーを切り替えて利用する機能、
- プラットフォーム内のコンテンツを俯瞰的に表示する機能

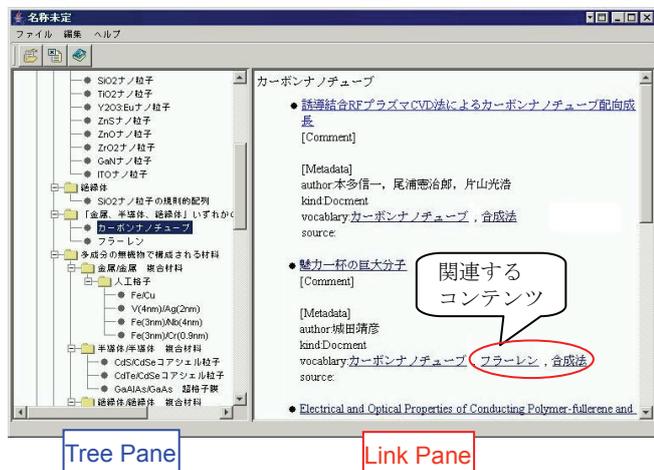


図3 コンテンツ管理インターフェイス

- 他システムからクエリーに応じて、指定された概念に対応するコンテンツ情報の表示する機能を持つ。現在は、知識の構造化プロジェクトと連携していた NEDO のナノカーボンプロジェクト、およびナノ粒子プロジェクトが対象としている領域を、NIO の概念階層を表示する Tree 上で示すことができる。これは、プラットフォームのユーザに、多くの中から必要なコンテンツを探し出す、入口の一つを提供することに相当する。

また NIO は、コンテンツに含まれる知識の関連をネットワーク上に表現する「知識ネットワークシステム」や、機能概念オントロジーに基づく「材料設計支援システム」[垂見 06]など、プラットフォーム内の他のシステムの共通語彙としても用いられる。これらのシステムと連携することで、コンテンツ間のより複雑な連携が可能となる。

5. 今後の展望と課題

5.1 想定される利用例

本システムが扱うことのできるコンテンツの内容は、インデックスオントロジーの対象領域によって決定される。よって、インデックスオントロジーを差し替えることにより、様々な領域を対象としたコンテンツ管理を行うことができる。また本システムの枠組みは、Webコンテンツに限らず多様な種類のコンテンツに適用できる。このように本システムは、汎用性が高いシステムだと言え、様々な利用方法が想定できる。例えば、個人やグループの研究業績、一度読んだ参考文献などの管理への利用が想定される。このように、別の利用方法に適用する際には、その目的に応じたシステムの拡張を行うことで、より効果的な管理が可能となる。

5.2 今後の展望

本稿で述べたコンテンツ管理システムは、小規模なプロトタイプとして開発されているが、今後の展望として、以下のような改良を計画している。

(1) Web サービス化

現在、リンク DB およびメタデータ DB を、SQL によるデータベースとして再開発を進めている。それに伴い、データの形式を再考すると共に、情報管理サーバーの Web サービス化を計画している。これにより、大規模なコンテンツの管理が安定して行えることが期待される。

(2) インデックスオントロジー・メタデータの詳細化

インデックスオントロジーとして、詳細なオントロジーを用いることで、より深く内容に踏み込んだコンテンツ管理が可能になる。筆者らの研究室では、機能概念オントロジー[来村 02]をはじめとして、様々なオントロジーの開発を進めているが、それらのオントロジーをインデックスオントロジーとして利用することで、その研究成果をコンテンツ管理システムに活かすことができる。その際には、「法造」およびその API を利用することで、オントロジー工学の基礎理論[溝口 05]に基づいたシステム開発が容易に行える。

またメタデータに関しては、Dublin Core[DCMES] など標準化された形式も提案されているが、本研究ではコンテンツに関する形式的な情報よりも、その内容に注目した管理を目指している。よって、メタデータを詳細化する際も、知識ネットワークや機能分解木を利用するなど、コンテンツの内容を詳細に表すメタデータの拡張を中心に検討している。

(3) コンテンツ情報の分散・統合

本システムで用いるリンクデータおよびメタデータを、ユーザ毎に別々の DB で管理すると、各ユーザの視点に基づいたコンテンツ管理が行える。さらに、各ユーザの DB を統合することで、複数の視点を統合したコンテンツ管理が実現される。共通のインデックスオントロジーを利用している場合は、このような統合が容易に行える。しかし、異なるインデックスオントロジーを用いている場合は、オントロジーの統合に関する技術が必要とされる。

(4) Web クローリングツールとの連携

本システムは、コンテンツ情報を 1 か所で集中管理する枠組みを採用している。しかし、Web 上のメタデータを自動的に収集するクローリングツールと連携させ、コンテンツの情報を収集することで、セマンティック Web で目指している Web 上に分散した多数の Web コンテンツに対しても、同じ枠組みを適用することができる。

6. まとめ

本稿では、知識の構造化プラットフォームにおいて、多種多様なコンテンツを、軽量オントロジーを用いた統一的な枠組みの元で管理するシステムの概要を述べた。本システムは、インデックスオントロジーをコンテンツの内容に対する共通知識体系として用いて、コンテンツを統一的視点から管理することを可能とする。それにより、領域を横断した知識の構造化を実現し、コンテンツの効果的な利用に貢献する。

今後は、システムの Web サービス化を行うと共に、オントロジーをより厳密なものに拡張することで、さらなる知的な支援を行うことなどを検討している。

参考文献

- [DCMES] DCMI: Dublin Core Metadata Element Set, Version 1.1: Reference Description, 1999, DCMI Recommendation
- [Jena] Jena - A Semantic Web Framework for Java, <http://jena.sourceforge.net/>
- [来村 02] 来村徳信, 溝口理一郎: オントロジー工学に基づく機能的知識体系化の枠組み, 人工知能学会誌, Vol.17, No.1, pp.61-72, 2002
- [古崎 02a] 古崎晃司, 来村徳信, 池田満, 溝口理一郎: 「ロール」および「関係」に関する基礎的考察に基づくオントロジー記述環境の開発, 人工知能学会誌, Vol.17, No.3, pp.196-208, 2002
- [古崎 02b] 古崎晃司, 来村徳信, 佐野年伸, 松本慎一郎, 石川誠一, 溝口理一郎: オントロジー構築・利用環境「法造」の開発と利用 -実規模プラントオントロジーを例として-, 人工知能学会誌, Vol.17, No.4, pp.407-419, 2002
- [古崎 04] 古崎晃司, 来村徳信, 渡邊英一, 山形尚子, 溝口理一郎: ナノテク材料技術の知識の構造化プロジェクトにおけるオントロジーの利用, 人工知能学会研究会資料 SIG-SWO-A303-03, 2004
- [小宮山 04] 小宮山宏: 知識の構造化, オープンナレッジ, 2004
- [垂見 06] 垂見晋也, 古崎晃司, 来村徳信, 溝口理一郎: 方式知識蓄積機能・過去の設計過程の保存機能を用いた統合的設計支援システムの開発, 第 20 回人工知能学会全国大会, 3B4-03, 2006
- [溝口 05] 溝口理一郎: オントロジー工学, オーム社, 2005
- [NMC] 知識の構造化プラットフォーム, <http://mandala.t.u-tokyo.ac.jp/>