

データベースを活用した化学物質の教育・研究情報の
収集・公開システムの開発
**Collection and publishing of educational information and research
information for the chemical material utilizing the database system**

伊藤智博 † ‡, 立花和宏 †, 奥山澄雄 †, 仁科辰夫 † ‡, 尾形健明 †, 鈴木勝人 ‡, 青木和恵 ‡

Tomohiro Ito † ‡, Kazuhiro Tachibana ‡, Sumio Okuyama ‡, Tatsuo Nishina † ‡, Tateaki Ogata ‡, Katsuhito Suzuki †, Kazue Aoki †

tomohiro_ito@ieee.org, c1_lab@yahoogroups.jp, sumio@ieee.org, c1_lab@yahoogroups.jp, ogata@yz.yamagata-u.ac.jp, katsuto@yamagata-u.ac.jp, kazue@yz.yamagata-u.ac.jp

† 山形大学バーチャル研究所・データベースアメニティ研究所
‡ 山形大学 学術情報基盤センター

† Database Amenity Laboratory of Virtual Research Institute, Yamagata University

‡ Networking and Computing Services Center, Yamagata University

〒992-8510 山形県米沢市城南 4-3-16

4-3-16 Jonan, Yonezawa, Yamagata 992-8510, JAPAN

概要

化学物質に関する情報としては、MSDS 制度などによるメーカーが提供する情報、化学大辞典などによる学問体系化によって蓄積された情報、学術論文や特許などによる新規物質に関する情報、研究室などの現在進行中の研究中の物質情報などがある。これらの情報を収集し、解析し、公開するシステムを構築し、教育・研究をサポートするためのツールを開発した。本稿では、データベースを活用し、収集した情報を利用して、教育・研究情報を効率よく公開する Web システムについて述べる。

キーワード

データベース, SEO, 化学情報, Web, 化学物質

1. はじめに

我々は、バックエンドにデータベースを用いた Web

サイトの運用を 10 年近く続けている。当初は教育情報、すなわち、山形大学のシラバスの登録・収集・公開を行うためのシステムであったが、教員の研究情報(研究者要覧)の追加をはじめとした機能拡張を行い、有用な情報を大量に蓄積している。これらの情報はその当初の目的

に利用するだけでなく、他の情報とリンクさせることにより、さらに価値を高めることができる。

近年、PRTR法などの始めとする化学物質の管理機能の追加して欲しいとの要望が寄せられ、2004年度より化学物質の管理機能を追加した。また、2005年度には、山形大学図書館の図書目録を公開している。公開目録数は、公開開始の2004年は、60万冊であり、その後、追加が幾度かなされ、2007年7月には、105万冊をサポートしている。この間に、システムとしても、データベースサーバのバージョンアップや、SSLへの対応、画像データベースへの対応などが行われた。本稿では、インターネットを利用して、化学情報を効果的に発信する試みについて実例をもとに紹介する。¹

2. システム

大学が所有する情報は、専門的な学術情報であり、非常の特色の高い情報である。そのため、非常に特色のあるWebページを作成することは、さほど難しいことではない。しかし、あまりにも専門過ぎることから、多くの人に見てもらおうことができず、結果的にWebサイトを開いた苦勞が報われないことになってしまう。

よいWebサイトを作成しても座して待っていても人は来ないので、多くの人に「このサイトには有用な情報がある」ことを知ってもらわねばならない。このため広告=入り口として、検索エンジン(たとえばGoogle)を効果的に利用することにした。

検索エンジンを効果的に利用するために必要なものは

1. 素材となる良質なデータと体系化
2. データを公開するWebシステム
3. Webを広報するノウハウ (Webページのデザイン)

である。これらは、どれもが重要な要素であるが、1の「素材となる良質なデータと体系化」は一朝一夕で作れるものではないため、最も重要であるかもしれない。

2.1. 素材となる良質なデータと体系化

素材として用意したデータを下記に示す。ポイントはデータを縦横無尽に組み合わせられるようなシステムを構成している点である。

- シラバスなどの講義資料

本システムは当初シラバスデータを保持するために作成されたものであるため1996年から2003年まで山形

¹ <https://a.yamagata-u.ac.jp/>

大学全学のシラバス情報(約3600件/年)をホストしている。2004年からはシラバスのシステムを市販のものに切り替えたため、本システムの業務ではなくなった。²その後、このシラバスシステムは、市販のものでは対応の困難なケースに対応するための機能が追加され、補足用のシラバスシステムとして、有志の先生方が利用している。追加された機能の1つとしては、講義ノートを追加し、毎回の授業のまとめなどを公開し、欠席者への対応などの可能になっている。³シラバスデータには大学の特性をあらわした専門的な言葉が大量に含まれており、この情報を有効に利用しようということが本稿の発端となった。

- 研究者総覧

2000年より山形大学全学の研究者情報(約800名強)をホストしている。研究者情報には各教員の専門とする研究分野のキーワードや論文タイトルを含んでいる。2006年度より、研究者情報のシステムを市販のシステムに移行したため、業務としての役割は終了している。⁴シラバス以上に文系・理系を横断した最先端の専門的な言葉を大量に含んでおり、非常に有用である。

- 図書情報

2005年6月から山形大学附属図書館の蔵書目録を公開している。公開開始時は、105万冊のデータを持っていたが、古い図書に関するデジタル化も進み、2007年7月には、105万冊の図書情報を有している。特に古いデータについてはAmazon等の目録には含まれておらず有用である。⁵

- 化学物質

化学物質データベース化学物質のデータベースとして2900種を超える物質を登録してある。このデータベースが化学関係のデータベースを展開するための基本となっており、後述する体系化とも関連するが、原子、イオン、分子、市販薬品などを体系化して、蓄積されている。⁶

- PRTR法対応データベース

2004年4月から、独立行政法人化に伴い大学もPRTR(Pollutant Release and Transfer Register: 化学物質排出移動量届出制度)法にしたがって特定化学物質の環境への排出量の把握及び管理を行わなければならなくなった[1]。山形大学工学部では、各研究室における該当化学物

² <http://campus2.kj.yamagata-u.ac.jp/syllabus/2007/home.htm>

³ <https://a.yamagata-u.ac.jp/amenity/Syllabus/SyllabusWeb.aspx?nSyllabusID=11111>

⁴ <http://yudb.kj.yamagata-u.ac.jp/index.jsp>

⁵ <https://a.yamagata-u.ac.jp/amenity/Library/YuBooksIndex.aspx>

⁶ <https://a.yamagata-u.ac.jp/amenity/Electrochem/PeriodicTableWeb.aspx>

質の使用量および環境への排出量を把握するために PRTR 法に対応したWEB ベースのデータベースシステムの運用をはじめている。⁷このデータベースは前記の化学物質データベースに加えて、学部内の場所のデータベース、教職員のデータベース、規制物質のデータベースを融合したものである。導入当初は、管理物質が354種類もあることから、システムの利用者は少なく、集計が苦勞したが、現在は、紙媒体による提出とWeb入力のハイブリッド提出によって、集権時間の短縮化が可能になった。

・ 外部 Web ページ

ウィキペディアや試薬メーカーを始めとする外部のWeb ページへのリンクをデータベースで管理して、該当コンテンツごとに、追加することが可能になった。これによって、オフィシャルサイトや詳細な情報の誘導ができることにより、講義ノートなどの幅が広がり、正しい情報を伝えることが容易になった。

・ 研究データベース

研究を支える「研究ノート」と実験で取得したデータをデータベースに蓄積し、一部の情報はWebページで公開されている。また、化学物質データベースである物質情報と関連付けることにより、普遍的な化学物質の情報と新規の実測定を関連付けることによって、普遍的な情報とアクティブな情報の橋渡しを行っている。⁸

・ オブジェクト

画像やソフトウェアなどの情報を公開できるようにファイルを格納するデータベースを開発している。認証情報と連携して、無線LAN用サブクライアントソフトウェアの自動配布と記録を可能にした。⁹

・ 体系化

体系化の基本は、「学問など過去の普遍的な情報＝変わらない情報」と「実験や化学物質管理などの現在の情報＝変化する情報」を相互に組み合わせていることである。それぞれのデータベースは階層構造を有しており、原子性を有している情報から複雑な情報へと展開されている。ただし、全てにおいて、原子性が成立するとは、限らない。例えば、図1に示すように、化学物質の情報は、原子 → 分子・イオン・官能基(単物質) → 材料(混合物) → 試料(実験室で実際に使われるもの)へと展開されている。

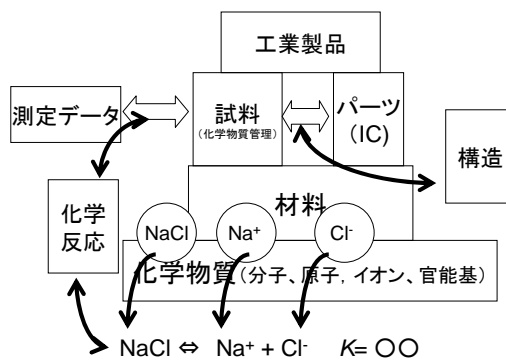


図-1. 化学物質および関連情報の体系化の例

2.2. Web システム

図2にWebシステムの概略を示す[2]。システムはMicrosoft Windows を用いて構成している。Web リクエストを受け付けるWeb サーバにはIIS6.0を用い、機器は物理的に公開用と登録用の2台に分離してある。これは情報登録時には必要に応じてActive Directory を利用した認証を用いている。バックエンドデータベースにはMicrosoft SQL Server 2005(プロセッサライセンス)を用いている。こちらはデータベースサーバの対障害性を向上するために、データベースミラーリング機能を用いて、プリンシパルサーバからミラーサーバへ常に同期が行われている。Web サイトは、開発用システムのサーバ上で開発され、運用Web サイトに発行される。Web サイトの開発言語はasp スクリプトおよびVisual Basic である。Visual Studio .NET が発売されてからはほぼすべてVisual Basic で開発されている。Visual Basic からADO .NET を利用することでデータベースアクセスのCPU パワーをWeb サーバ側で負担することができるようになり、データベースサーバ側の負担を大幅に軽減することができるようになった。また、ASP .NET の導入により、Web のデザインコードとプログラムコード分離が実現したことによって、開発コストの軽減やコードのメンテナンス性が向上した。

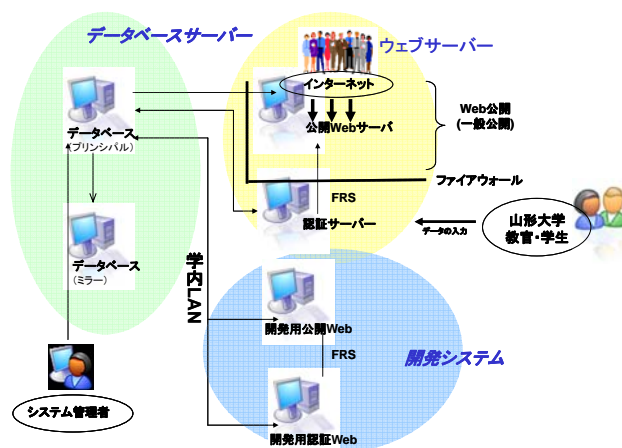


図-2. データベースおよびWebサーバ構成図

⁷ <https://a.yamagata-u.ac.jp/amenity/prtr.aspx>

⁸ <https://a.yamagata-u.ac.jp/amenity/Laboratory/xyGraphIndex.aspx>

⁹ <https://a.yamagata-u.ac.jp/amenity/Knowledge/ObjectIndex.aspx>

3 運用

まず、最初に考えなければならないのは、ユーザーの数である。政治や芸能界の情報とは異なり化学物質の情報は、その情報を必要とする人の数が圧倒的に少ないため、コマーシャルベースの広告には乗りにくい。また、情報が専門的なものであるため、ユーザーが「欲しいと『思う』」ことをトリガーにして情報を提供することが可能である。

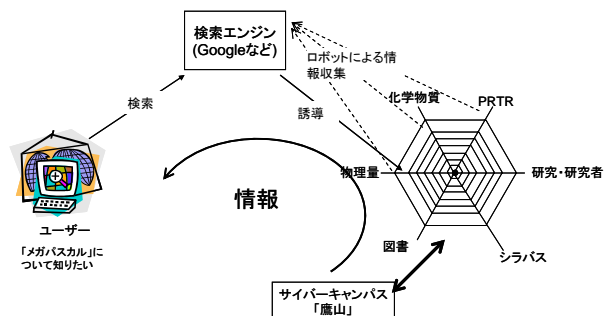


図-3. 検索エンジンを活用した情報提供

図 3 にユーザーに情報を提供するための流れを模式的に示す。ユーザーは、検索エンジン(たとえば Google)で、欲しい情報のキーワードなどを入力して検索を行う。検索エンジンはあらかじめサイト全体をロボットの作業によりスキャンしており、「どこにどのような情報があるか」を知っている。このためユーザーの検索に対して適切な情報への入り口を提供してくれる。このようにすればユーザーに対して情報を適切に提供することができる。しかし、現実的には競合する情報元がある場合にどちらに誘導するかは検索エンジンに寄るところが大きい。このため「検索エンジンを攻略する=SEO」ことが不可欠である。先に述べた化学物質データベースや研究データベースを利用することにより、検索エンジンができるだけ多く我々の Web サイトに誘導するようにした。

表 1 に 2007 年 4 月 1 日から 2007 年 7 月 31 日まで、検索エンジンを經由して我々の Web ページを閲覧したときの検索キーワードを示したものである。

表-1. Web ページを閲覧したときのキーワード

キーワード	セッション数 / 件
1 ニュートン 単位	8,482
2 メガパスカル	6,316
3 元素周期表	3,343
4 周期表	3,066
5 誘電率	2,934
6 パスカル 単位	2,542
7 マイクログラム	2,466
8 mpa	2,367
9 カチオン	1,980
10 ヤング率	1,653

単位や物理量、周期表といった基礎的なデータベースが利用されている。また、専門用語などの難しいキーワードで、検索することは少ない。図 4 に、2007 年 4 月 1 日から 2007 年 7 月 31 日までに Web ページを閲覧した時間帯別の閲覧率を示す。朝 9 時ごろから急激にアクセスが増大し、12 時~13 時までの間に小休止があり、17 時を過ぎると徐々にアクセスが減少している。アクセス元を解析すると国内の教育機関や企業からが主であり、教育や業務での利用が推測される。また、閲覧件数を解析すると、ピークで 11,000 ヒット/時、平均でも 4000 ヒット/時のアクセスがある。トータルでは年間 1000 万ヒット超のアクセスを記録している。この中で検索エンジンを除く、一般閲覧者からのヒット数は、660 万ヒット/年である。また、閲覧件数の約 4.5%が検索エンジンなどを經由しないで、直接閲覧頂いている方である。今後、リピータの方も増やせるような Web ページを作る必要があるであろう。

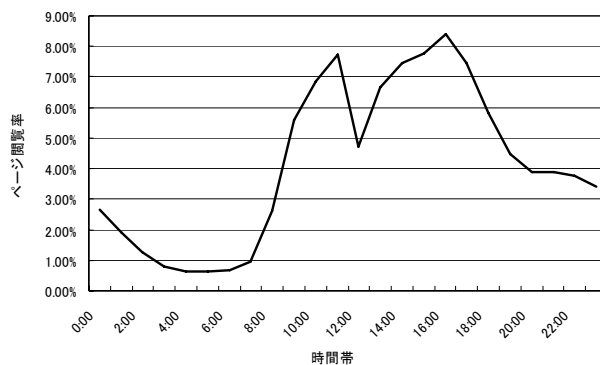


図-4. 時間帯ごとの Web ページの閲覧率

4 まとめ

情報の体系化を図り、それぞれの情報をデータベースに蓄積し、情報発信することによって、検索エンジンのページランキングを上げることができる。ただし、この Web ページの試みは、あくまで、体系化して、質を高めた情報を有効に活用していただくことが目的であり、ページランキングが上がったのは、偶然であろう。我々のサイバーキャンパス「鷹山」の情報によって、少しでも教育・研究活動が楽になることであれば、幸いである。

参考文献

- [1] PRTR インフォメーション広場：
<http://www.env.go.jp/chemi/prtr/risk0.html>
- [2] 立花和宏他: 「サイバーキャンパスの構築とその実用化への問題点の抽出」, 山形大学紀要(工学), 28 (2005) 53: <http://www.lib.yamagata-u.ac.jp/kiyou/kiyoue/kiyoue-28-2.html>