

# メタデータを活用した ポータルサイトの構築

NTT コミュニケーション科学基礎研究所

平松 薫

---

# 目次

- ポータルサイトとは？
  - Yahoo! の変遷と wikipedia における定義
  - メタデータとデータベースとセマンティックウェブ
- ポータルサイトの構築事例
  - メリーランド大 Mindswap  
<http://www.mindswap.org/>
- メタ情報を用いた検索サービスの事例
  - 地理関係に基づく連続的情報検索

---

# ポータルサイトといえは

- Yahoo! (<http://yahoo.co.jp>)
- MSN (<http://www.msn.co.jp>)
- インフォシーク (<http://www.infoseek.co.jp>)
- Excite (<http://www.excite.co.jp>)
- goo (<http://goo.ne.jp>)
- Livedoor (<http://www.livedoor.co.jp>)

# http://www.yahoo.com/ の場合

1996.10.20 ~ のページが以下のように蓄積.

The screenshot shows the Internet Archive Wayback Machine interface in Microsoft Internet Explorer. The search results are for the URL <http://www.yahoo.com/>. The search results are displayed in a table format, showing the number of pages archived for each year from 1996 to 2003. The table is titled "Search Results for Jan 01, 1996 - Apr 10, 2005".

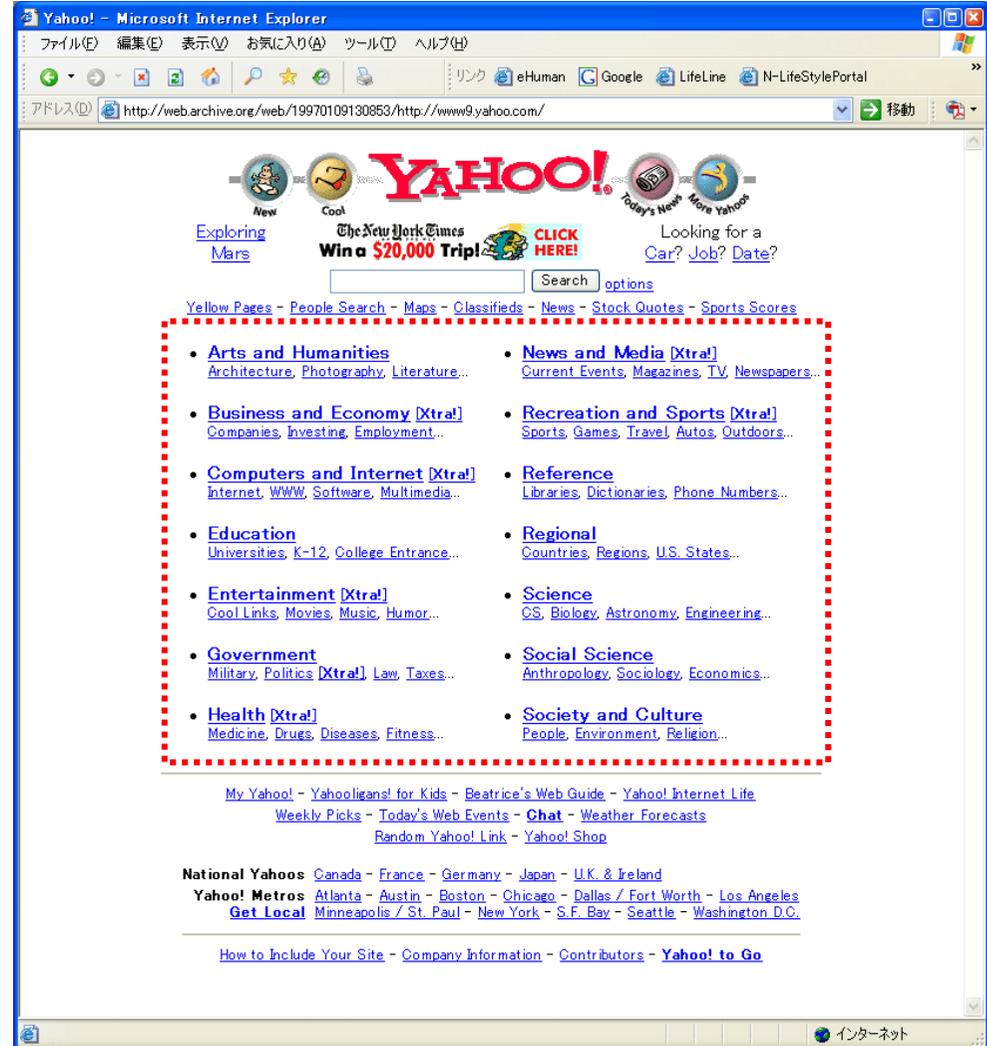
1996	1997	1998	1999	2000	2001	2002	2003	
8 pages	17 pages	20 pages	15 pages	124 pages	1937 pages	140 pages	90 pages	1
<a href="#">Oct 20, 1996</a> *	<a href="#">Jan 09, 1997</a> *	<a href="#">Feb 14, 1998</a> *	<a href="#">Jan 16, 1999</a> *	<a href="#">Feb 29, 2000</a> *	<a href="#">Mar 01, 2001</a> *	<a href="#">Jan 25, 2002</a> *	<a href="#">Jan 22, 2003</a> *	<a href="#">Jan 22, 2003</a> *
<a href="#">Oct 22, 1996</a> *	<a href="#">Jan 24, 1997</a> *	<a href="#">Feb 15, 1998</a> *	<a href="#">Jan 17, 1999</a>	<a href="#">Feb 29, 2000</a> *	<a href="#">Mar 05, 2001</a> *	<a href="#">Feb 02, 2002</a> *	<a href="#">Jan 25, 2003</a>	<a href="#">Feb 02, 2003</a> *
<a href="#">Nov 28, 1996</a> *	<a href="#">Feb 09, 1997</a>	<a href="#">Jun 30, 1998</a> *	<a href="#">Jan 17, 1999</a>	<a href="#">Mar 01, 2000</a> *	<a href="#">Mar 06, 2001</a> *	<a href="#">May 25, 2002</a> *	<a href="#">Jan 26, 2003</a>	<a href="#">Feb 02, 2003</a> *
<a href="#">Dec 20, 1996</a> *	<a href="#">Feb 27, 1997</a> *	<a href="#">Jul 03, 1998</a> *	<a href="#">Apr 17, 1999</a> *	<a href="#">Mar 01, 2000</a> *	<a href="#">Mar 07, 2001</a> *	<a href="#">May 25, 2002</a> *	<a href="#">Jan 27, 2003</a> *	<a href="#">Feb 02, 2003</a> *
<a href="#">Dec 21, 1996</a>	<a href="#">Apr 16, 1997</a>	<a href="#">Jul 04, 1998</a> *	<a href="#">Apr 17, 1999</a> *	<a href="#">Mar 02, 2000</a> *	<a href="#">Mar 31, 2001</a> *	<a href="#">May 26, 2002</a>	<a href="#">Feb 02, 2003</a> *	<a href="#">Feb 02, 2003</a> *
<a href="#">Dec 26, 1996</a> *	<a href="#">Apr 18, 1997</a> *	<a href="#">Jul 05, 1998</a> *	<a href="#">Apr 18, 1999</a> *	<a href="#">Mar 02, 2000</a> *	<a href="#">Mar 31, 2001</a> *	<a href="#">May 27, 2002</a>	<a href="#">Feb 04, 2003</a> *	<a href="#">Feb 04, 2003</a> *
<a href="#">Dec 27, 1996</a>	<a href="#">May 17, 1997</a>	<a href="#">Jul 05, 1998</a> *	<a href="#">Apr 18, 1999</a> *	<a href="#">Mar 02, 2000</a> *	<a href="#">Apr 01, 2001</a> *	<a href="#">May 28, 2002</a> *	<a href="#">Feb 13, 2003</a>	<a href="#">Feb 13, 2003</a>
<a href="#">Dec 28, 1996</a>	<a href="#">May 21, 1997</a> *	<a href="#">Jul 05, 1998</a> *	<a href="#">Apr 21, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 04, 2001</a> *	<a href="#">May 29, 2002</a> *	<a href="#">Feb 16, 2003</a>	<a href="#">Feb 16, 2003</a>
	<a href="#">Jun 05, 1997</a>	<a href="#">Dec 12, 1998</a> *	<a href="#">Apr 27, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 04, 2001</a> *	<a href="#">May 30, 2002</a> *	<a href="#">Feb 17, 2003</a>	<a href="#">Feb 17, 2003</a>
	<a href="#">Jun 06, 1997</a> *	<a href="#">Dec 12, 1998</a>	<a href="#">Apr 28, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 04, 2001</a> *	<a href="#">May 31, 2002</a>	<a href="#">Mar 21, 2003</a>	<a href="#">Mar 21, 2003</a>
	<a href="#">Jun 30, 1997</a> *	<a href="#">Dec 12, 1998</a>	<a href="#">Apr 30, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 05, 2001</a> *	<a href="#">Jun 03, 2002</a> *	<a href="#">Mar 24, 2003</a>	<a href="#">Mar 24, 2003</a>
	<a href="#">Jul 03, 1997</a> *	<a href="#">Dec 12, 1998</a> *	<a href="#">May 08, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 05, 2001</a> *	<a href="#">Jun 06, 2002</a> *	<a href="#">Mar 29, 2003</a>	<a href="#">Mar 29, 2003</a>
	<a href="#">Jul 23, 1997</a> *	<a href="#">Dec 12, 1998</a>	<a href="#">Oct 10, 1999</a> *	<a href="#">Mar 03, 2000</a> *	<a href="#">Apr 10, 2001</a> *	<a href="#">Jul 02, 2002</a> *	<a href="#">Mar 30, 2003</a>	<a href="#">Mar 30, 2003</a>
	<a href="#">Oct 07, 1997</a> *	<a href="#">Dec 12, 1998</a>	<a href="#">Oct 10, 1999</a> *	<a href="#">Mar 04, 2000</a> *	<a href="#">Apr 11, 2001</a> *	<a href="#">Jul 03, 2002</a> *	<a href="#">Mar 31, 2003</a>	<a href="#">Mar 31, 2003</a>
	<a href="#">Oct 08, 1997</a> *	<a href="#">Dec 12, 1998</a>	<a href="#">Nov 16, 1999</a> *	<a href="#">Mar 04, 2000</a> *	<a href="#">Apr 14, 2001</a> *	<a href="#">Jul 03, 2002</a> *	<a href="#">Apr 07, 2003</a>	<a href="#">Apr 07, 2003</a>
	<a href="#">Oct 14, 1997</a> *	<a href="#">Dec 12, 1998</a>		<a href="#">Mar 04, 2000</a> *	<a href="#">Apr 29, 2001</a> *	<a href="#">Jul 06, 2002</a> *	<a href="#">Apr 09, 2003</a>	<a href="#">Apr 09, 2003</a>
	<a href="#">Oct 22, 1997</a> *	<a href="#">Dec 12, 1998</a> *		<a href="#">Mar 04, 2000</a> *	<a href="#">May 01, 2001</a> *	<a href="#">Jul 12, 2002</a> *	<a href="#">Apr 10, 2003</a>	<a href="#">Apr 10, 2003</a>
		<a href="#">Dec 12, 1998</a>		<a href="#">Mar 05, 2000</a> *	<a href="#">May 03, 2001</a> *	<a href="#">Jul 13, 2002</a> *	<a href="#">Apr 19, 2003</a>	<a href="#">Apr 19, 2003</a>
		<a href="#">Dec 12, 1998</a> *		<a href="#">Mar 05, 2000</a> *	<a href="#">May 03, 2001</a> *	<a href="#">Jul 16, 2002</a> *	<a href="#">Apr 21, 2003</a>	<a href="#">Apr 21, 2003</a>
		<a href="#">Dec 12, 1998</a> *		<a href="#">Apr 07, 2000</a> *	<a href="#">May 03, 2001</a> *	<a href="#">Jul 17, 2002</a> *	<a href="#">Apr 22, 2003</a>	<a href="#">Apr 22, 2003</a>
		<a href="#">Dec 12, 1998</a> *		<a href="#">Apr 07, 2000</a> *	<a href="#">May 03, 2001</a> *	<a href="#">Jul 17, 2002</a> *	<a href="#">Apr 22, 2003</a>	<a href="#">Apr 22, 2003</a>

# http://www.yahoo.com/

(1996.10.20)



(1997.1.9)

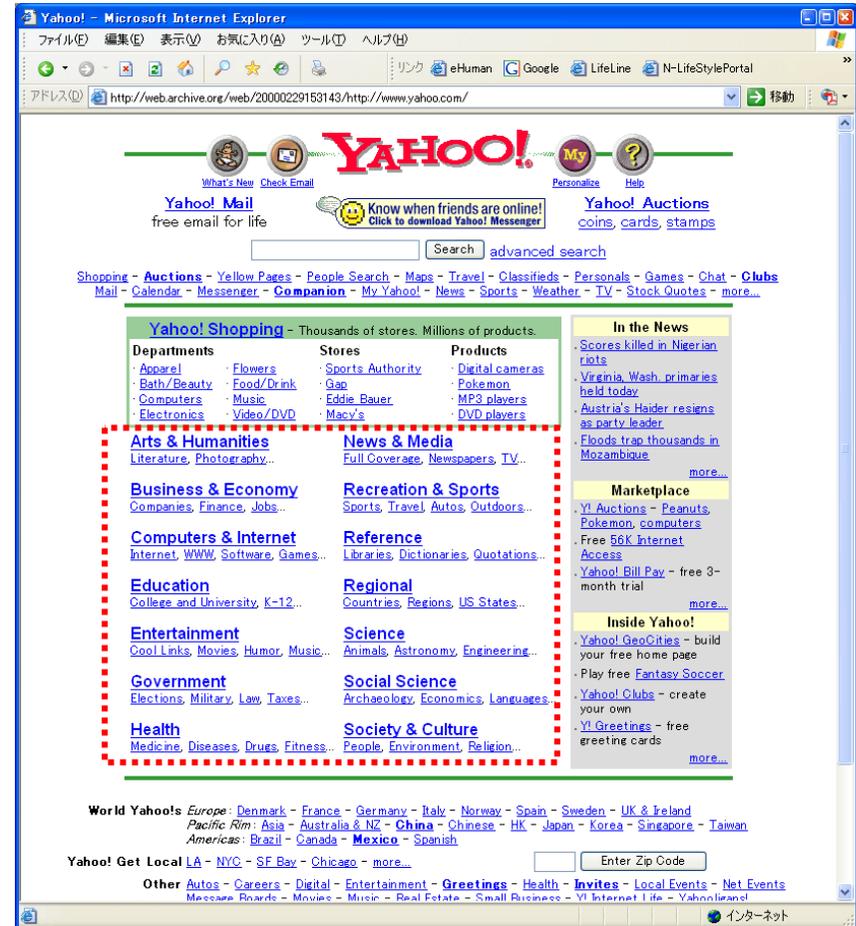


# http://www.yahoo.com/

(1999.4.17)



(2000.2.29)



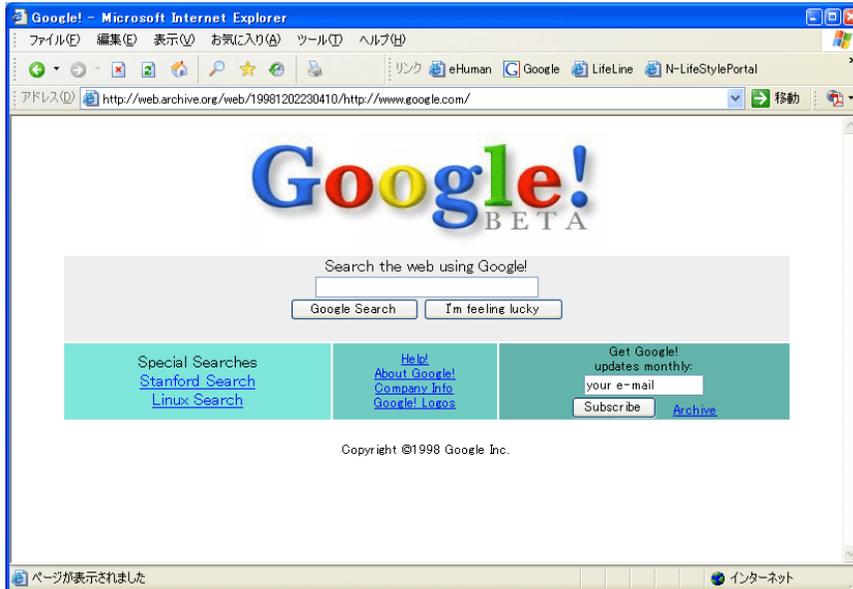
# http://www.yahoo.com/

(2002.8.1)

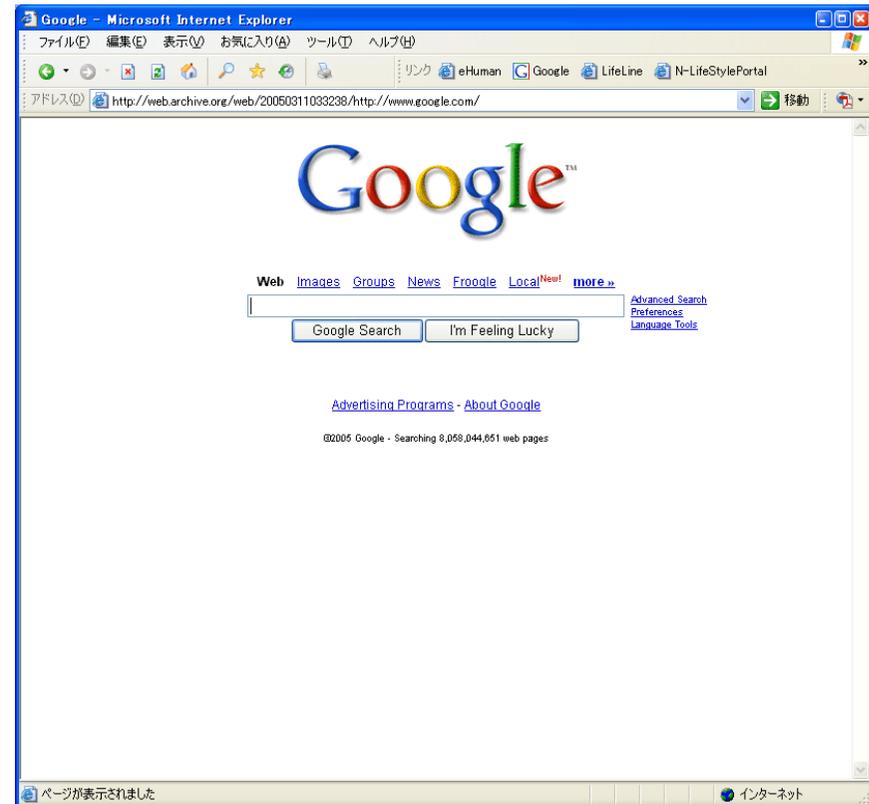
(2004.4.11)

# http://www.google.com/

(1998.12.2)



(2005.3.11)



---

# http://www.yahoo.com/ の変遷

- Web Directory は縮小傾向
  - 「ディレクトリ = ポータル」から「利便性」の追求へ
- 自社サービスの広告を挿入
  - ショッピング, オークション,  
メッセンジャー, メール, GeoCities
- ニュース, 天気予報を挿入

# ポータルサイトとは？

出典: フリー百科事典『ウィキペディア (Wikipedia)』

- ポータルサイト (portal site)とは、WWWにアクセスするときの玄関口となるウェブサイトのこと。
- 検索エンジン、ウェブディレクトリ、ニュース、オンライン辞書、オークションなどのサービスを提供し、利用者の便宜を図っている。
- ポータルサイトのビジネスモデルは、サイトの集客力を生かして広告や有料コンテンツで収入を得ることである。1996年以降のインターネットブームに乗じて、多くのポータルサイトが乱立したが、徐々に統廃合が進んでいる。  
(後半へと続く)

# ポータルサイトとは？

出典: フリー百科事典『ウィキペディア (Wikipedia)』

(前半からの続き)

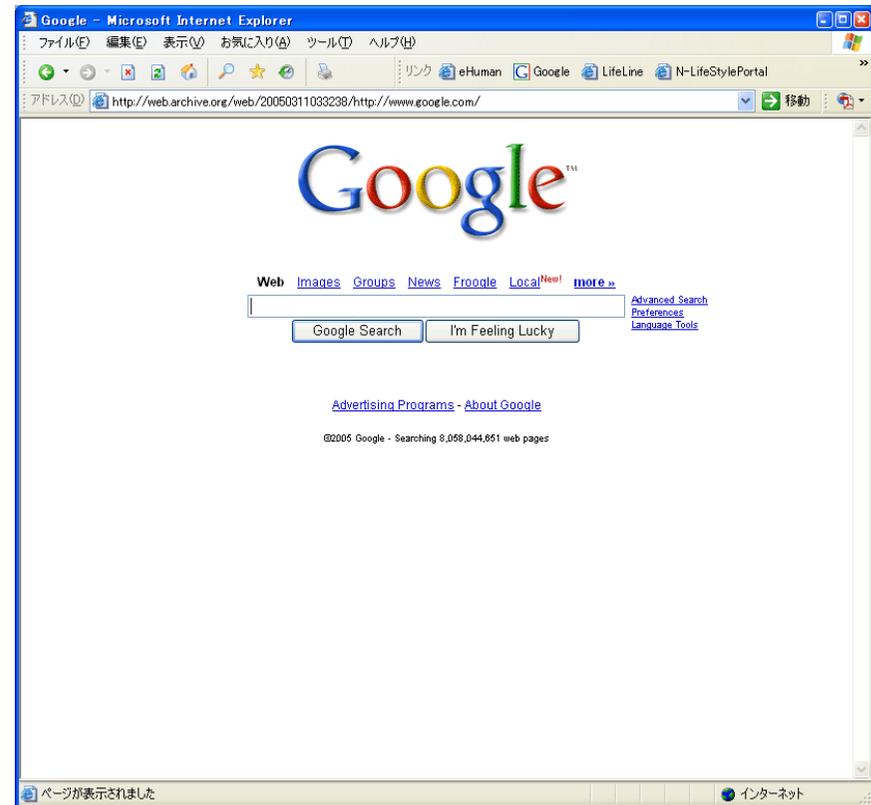
- 初期のポータルサイトは自前で検索エンジンやウェブディレクトリを運用していたが、情報の肥大化に対応できず[アウトソーシング](#)が多くなった。
- 生き残りをかけて、特定の地域サービスに特化した地域ポータルサイトや、プロバイダーのサービス情報サイト、育児、環境、オルターナティブカルチャー、音楽、女性の生き方などにテーマを絞ったポータルサイトも。
- 不特定多数のアクセスがあるだけに、こうしたポータルサイトにアダルト情報を持ち込むことの是非も問題になっている。

# ポータルサイト = 玄関口 ?

- お気に入りポータルサイト = 見慣れた玄関口
  - 習慣的に繰り返しアクセス
  - ユーザの定着, コミュニティの形成
- 見知らぬポータルサイト = 見知らぬ玄関口 ?
  - 知人の紹介でもない限り利用しない
  - セールスマンの飛び込み営業??

# キーワード検索

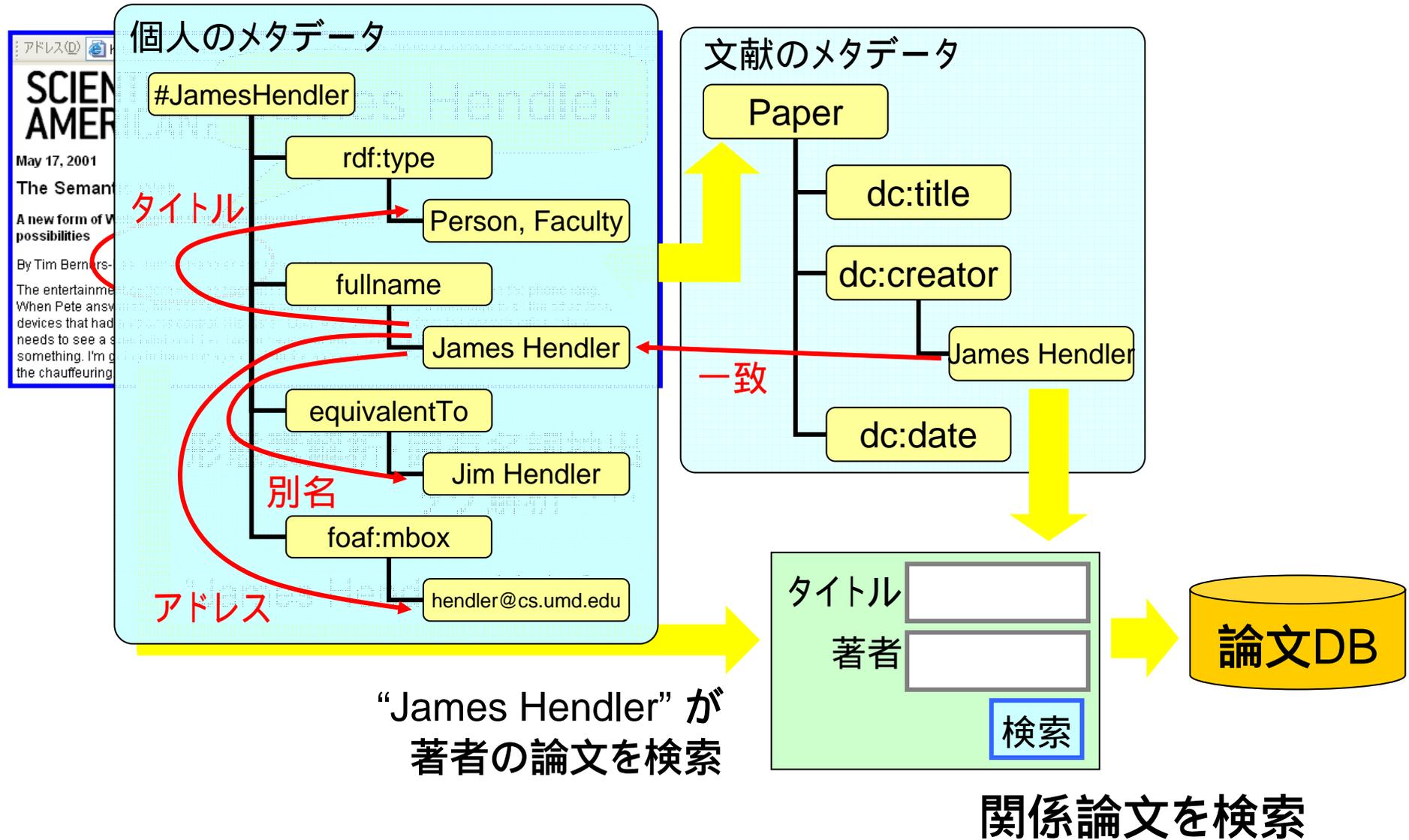
- 必要な情報を含んだ Web ページを検索可能
  - ポータルサイトが提供する重要な機能の一つ
  - ポータルサイトをスキップできてしまう



# ただし限界も存在

- 「Jim Hendler が執筆した論文は？」
  - CiteSeer.IST を利用すべき？
- 「金閣寺の近くの停留所は？」
  - 京都の地理に明るければ名前もわかるけど...
  - 地理的関係を考慮した問い合わせ
- 「4/25に慶応大学で開催されているイベントは？」
  - AIシンポジウム 4/25 開催, × 健康診断 4/18 ~ 4/28
  - 日時・場所を指定した問い合わせ
- 「山口先生の所属学科は？」
  - どの「山口先生」か？ いつの情報か？
  - 状況を考慮した質問応答

# ポイントは情報間の関連付け



# 限界への挑戦

- Semantic Web のアプローチ
  - データに対してメタデータを付与
  - 機械処理可能なメタデータと共通オントロジー
  - 第三者所有のデータベース, サービスの利用
  - 書式検証, 関連概念の推論, など
- 統計・学習・自然言語処理によるアプローチ
  - 特定属性の抽出, カテゴリ分け, など

URI/RDF/RDFS

OWL

WSDL/OWL-S

メタデータ生成を支援

---

# W3C による標準化動向

- RDF, RDFS, OWL
  - W3C Recommendation (10 Feb 2004)
- WSDL 2.0
  - W3C Working Draft (21 December 2004)
- OWL-S
  - Semantic Markup for Web Services Note Published (2004-11-22)

---

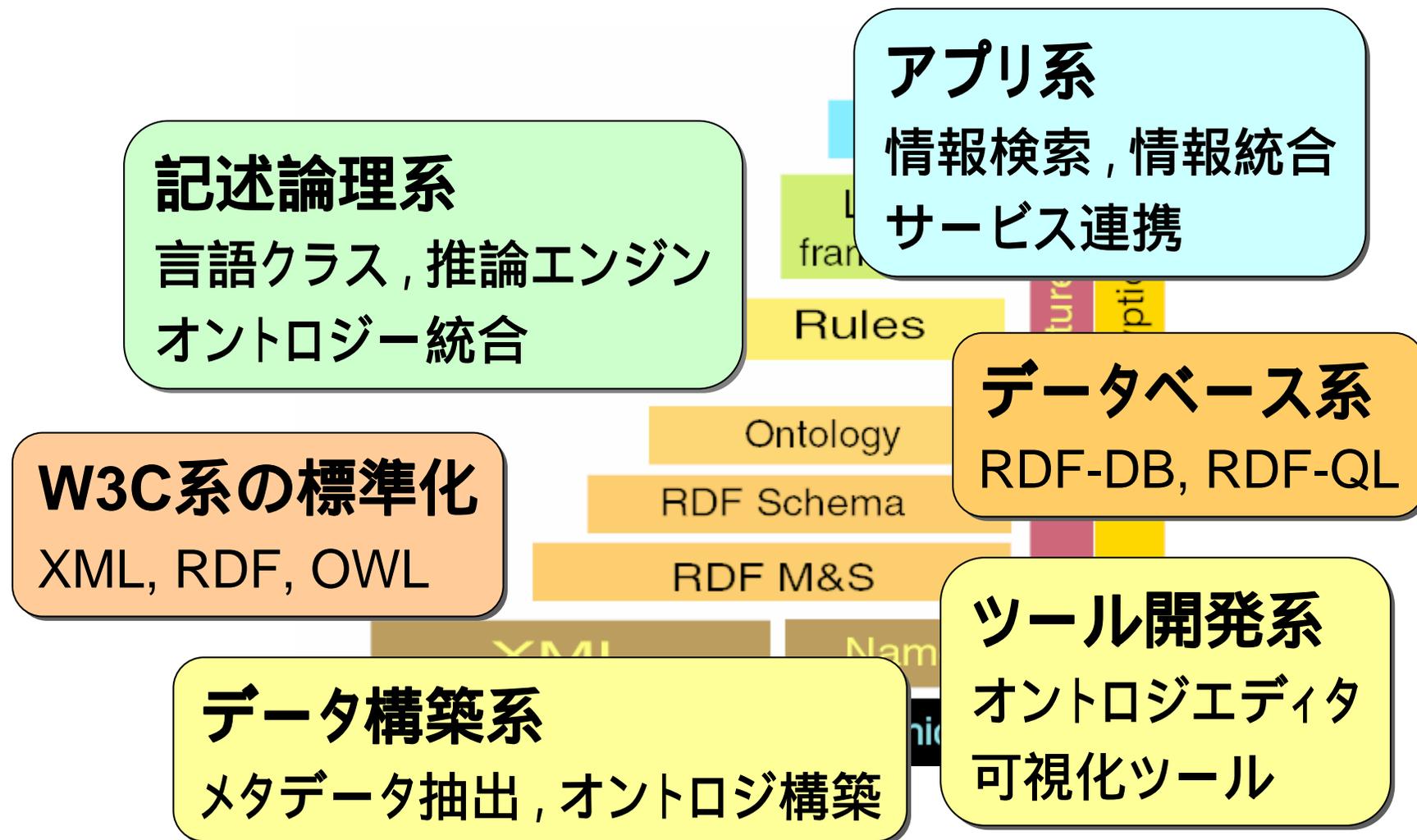
構築事例：

<http://www.mindswap.org/>

---

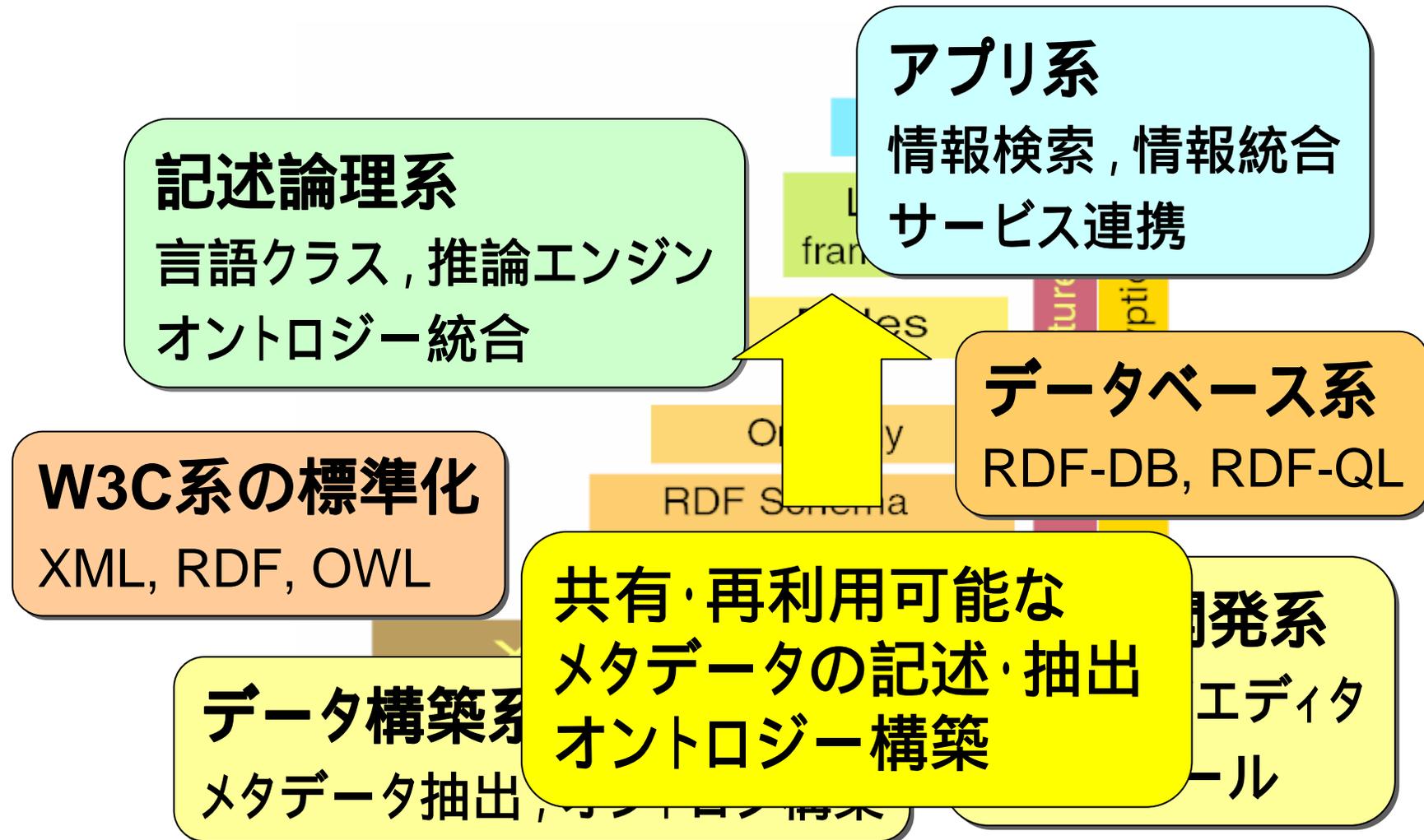
参考文献：J. Golbeck, et al., *Handbook of Human Factors in Web Design*, chapter Organization and Structure of Information using Semantic Web Technologies, 2003.

# Enabling Technologies - Layer Cake



by Eric Miller, <http://www.w3.org/2002/Talks/www2002-w3ct-swintro-em/slide9-0.html>

# メタデータ・オントロジの果たす役割



by Eric Miller, <http://www.w3.org/2002/Talks/www2002-w3ct-swintro-em/slide9-0.html>

# 表示用ページを機械処理可能に

- Human Readable な現在の Web ページを
  - HTML はレイアウトなどを指定する表示用タグセット
  - (表示レイアウトや字面から意味抽出も可能)

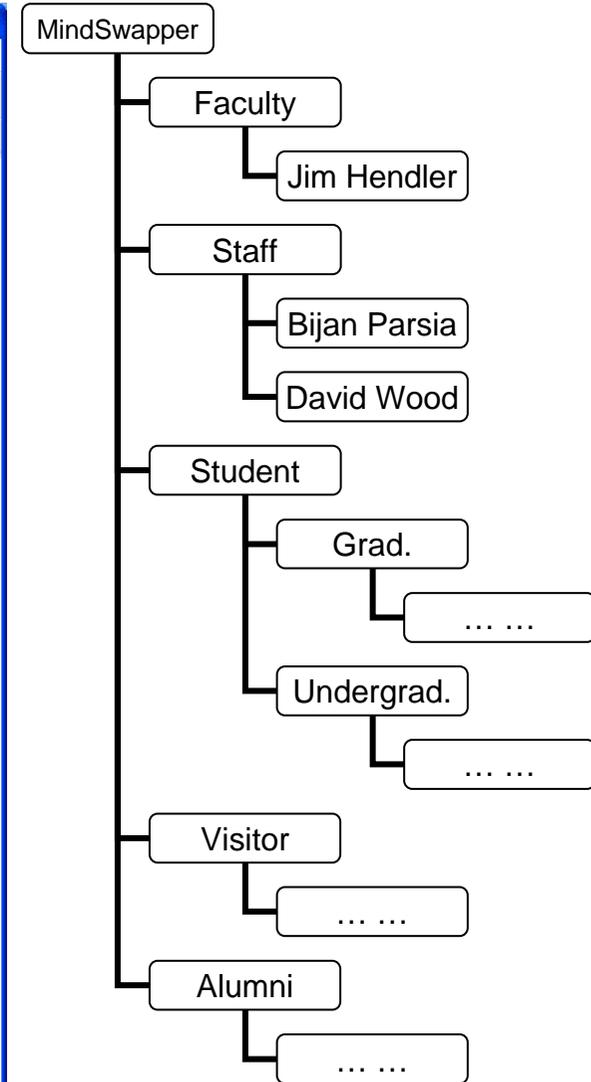
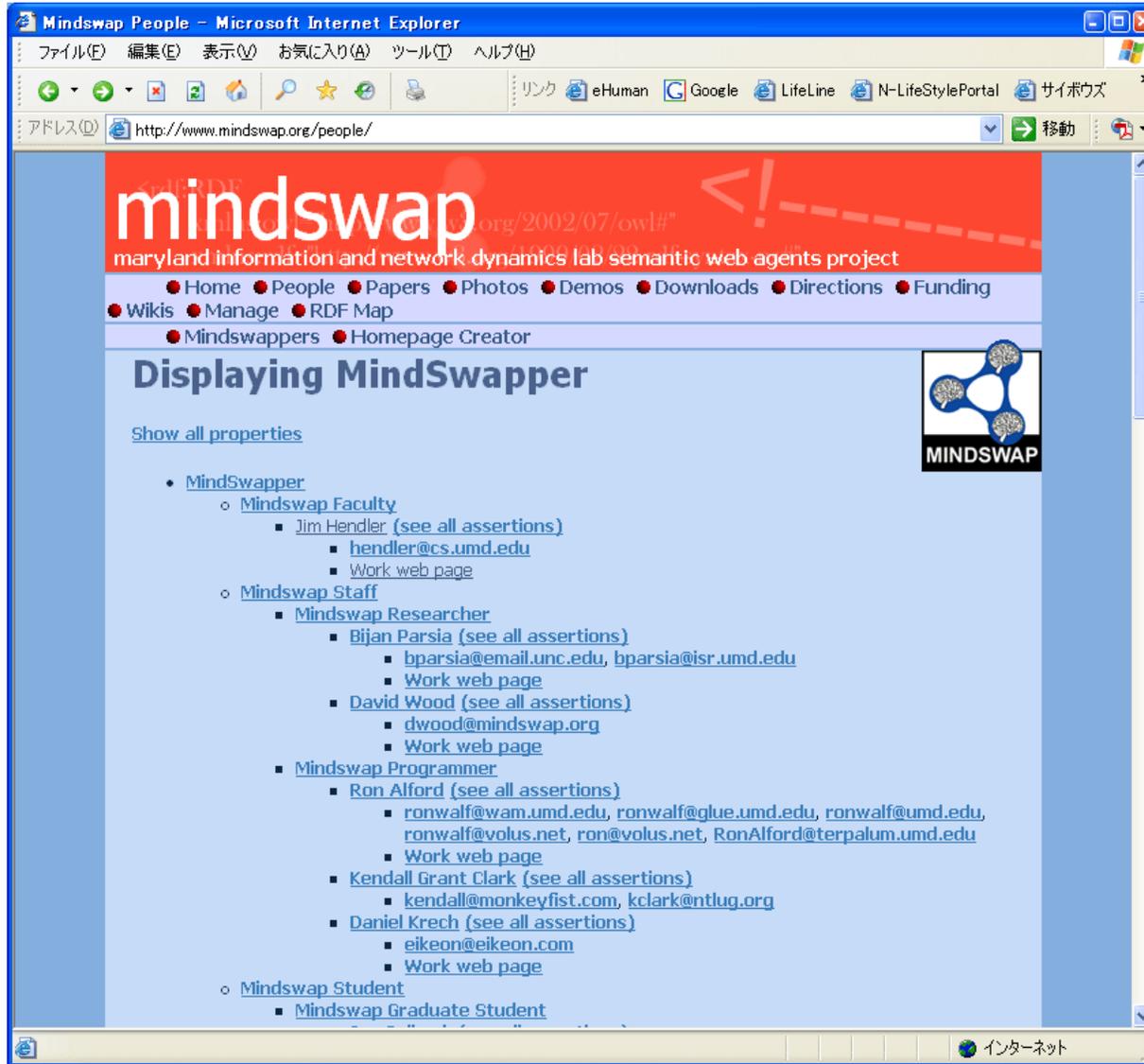
(情報発信者の努力によりメタデータを付与して...)

- Machine Readable な Web ページに
  - URI の付与, クラスを指定, Property を記述
  - 第三者のためにスキーマ, オントロジーを公開

# http://www.mindswap.org/



# mindswap - person



# mindswap - kaoru

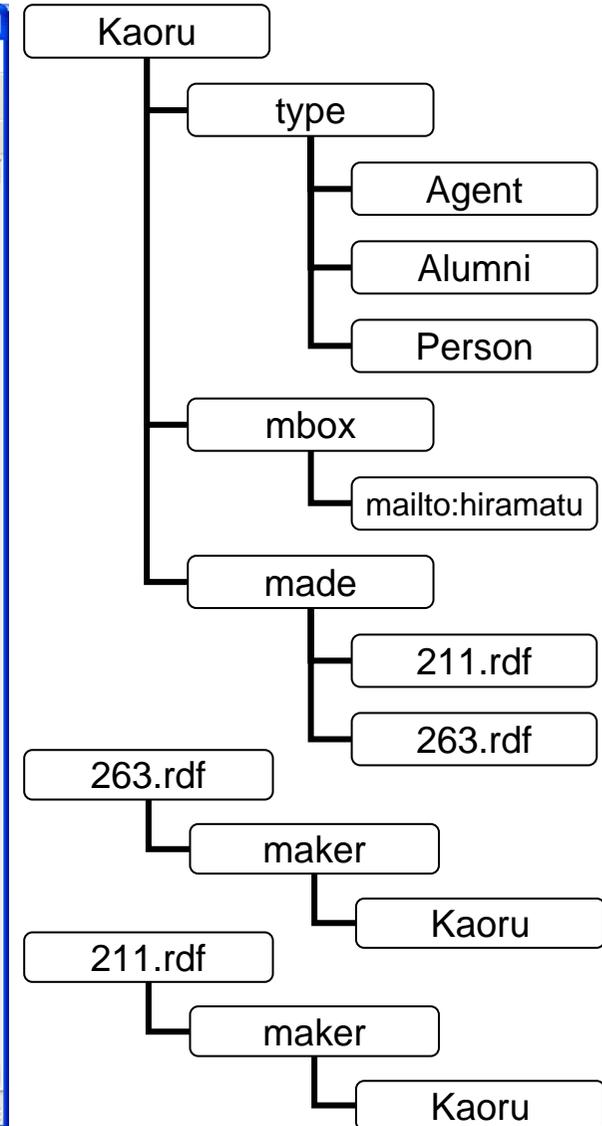
mindswap  
maryland information and network dynamics lab semantic web agents project

- Home
- People
- Papers
- Photos
- Demos
- Downloads
- Directions
- Funding
- Wikis
- Manage
- RDF Map
- Search Instances

**Kaoru Hiramatsu**

Depictions:

Kaoru Hiramatsu	type	Agent, Mindswap Alumni, Person
	personal mailbox	<a href="mailto:hiramatu@umiacs.umd.edu">mailto:hiramatu@umiacs.umd.edu</a>
	made	<a href="http://owl.mindswap.org/2003/submit-rdf/211.rdf">http://owl.mindswap.org/2003/submit-rdf/211.rdf</a> , <a href="http://owl.mindswap.org/2003/submit-rdf/263.rdf">http://owl.mindswap.org/2003/submit-rdf/263.rdf</a>
	work info homepage	<a href="http://www.kecl.ntt.co.jp/csl/sirg/people/hiramatu/index-e.html">http://www.kecl.ntt.co.jp/csl/sirg/people/hiramatu/index-e.html</a>
	depiction	<a href="#">Jim and Kaoru</a>
		<a href="http://owl.mindswap.org/2003/submit-rdf/263.rdf">http://owl.mindswap.org/2003/submit-rdf/263.rdf</a> , <a href="http://owl.mindswap.org/2003/submit-rdf/211.rdf">http://owl.mindswap.org/2003/submit-rdf/211.rdf</a>
	maker	Kaoru Hiramatsu
	depicts	<a href="#">Jim and Kaoru</a>



# 表示用HTMLファイルの生成

263.rdf

```
<rdf:RDF .... >
  <rdf:Snippet rdf:about="">
    <creator rdf:parseType="Resource">
      <foaf:mbox rdf:resource="mailto:hiramatu@umiacs.umd.edu"/>
    </creator>
    <dc:date>08/21/2003</dc:date>
  </rdf:Snippet>
  <Paper rdf:about="http://www.isi.edu/info-agents/workshops/ijcai03/papers/hiramatu.pdf">
    <rdf:type rdf:resource="http://www.isi.edu/webscripiter/bibtex.o.daml#inproceedings"/>
    <authors>Hiramatsu, K, Akahani, J, Satoh, T</authors>
    <creator rdf:parseType="Resource">
      <foaf:mbox rdf:resource="mailto:hiramatu@cslab.kecl.ntt.co.jp"/>
    </creator>
    <dc:identifier>http://www.isi.edu/info-agents/workshops/ijcai03/papers/hiramatu.pdf</dc:identifier>
    <dc:title>Two-phase Query Modification using Semantic Relations based on Ontologies.</dc:title>
    <bib:author>Hiramatsu, Kaoru and Akahani, Jun-ichi and Satoh, Tetsuji</bib:author>
    <bib:title>Two-phase Query Modification using Semantic Relations based on Ontologies.</bib:title>
    <bib:url rdf:resource="http://www.isi.edu/info-agents/workshops/ijcai03/papers/hiramatu.pdf"/>
  </Paper>
</rdf:RDF>
```



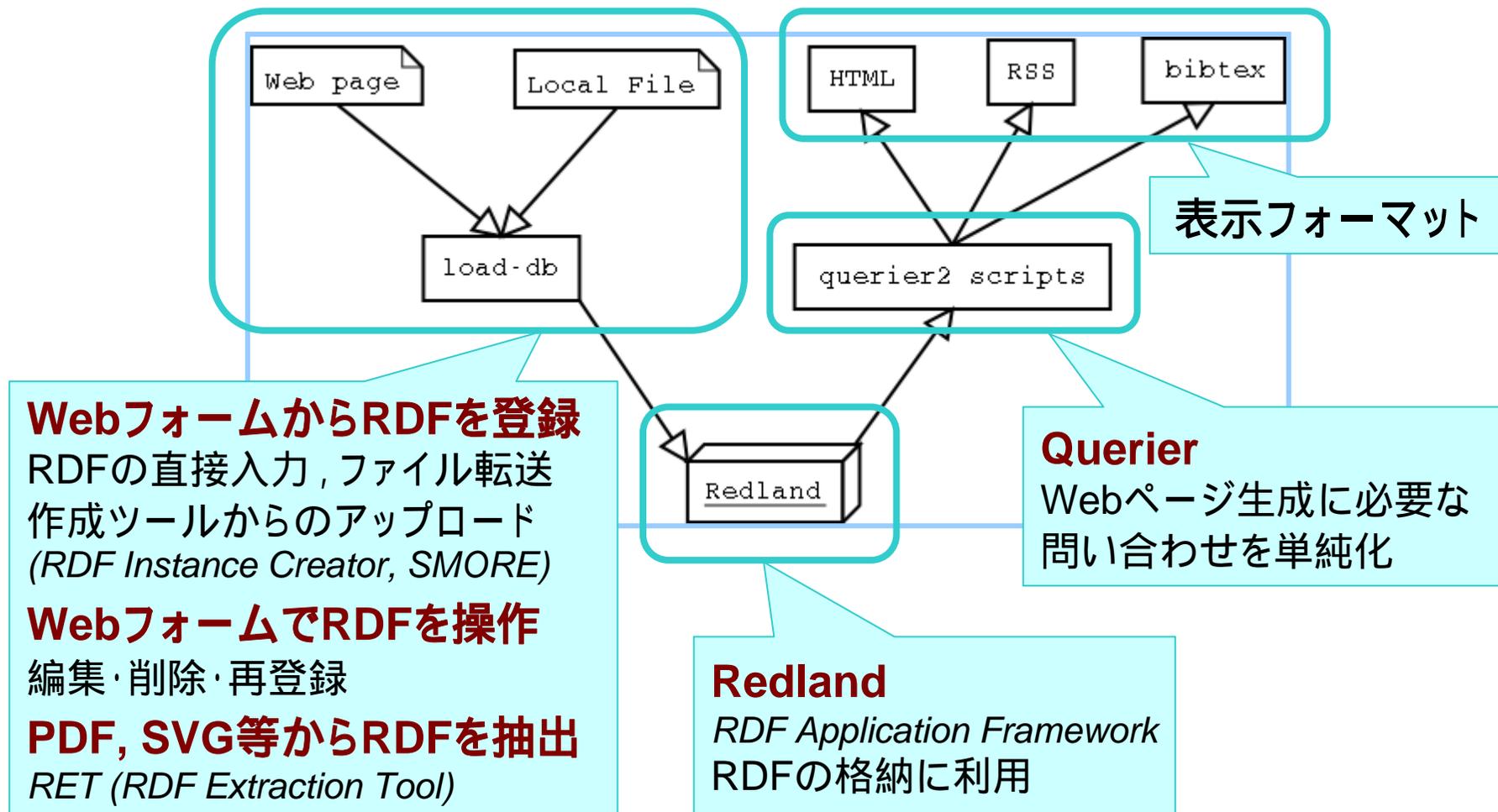
Mindswap  
- papers

Mindswap  
- people  
- Kaoru

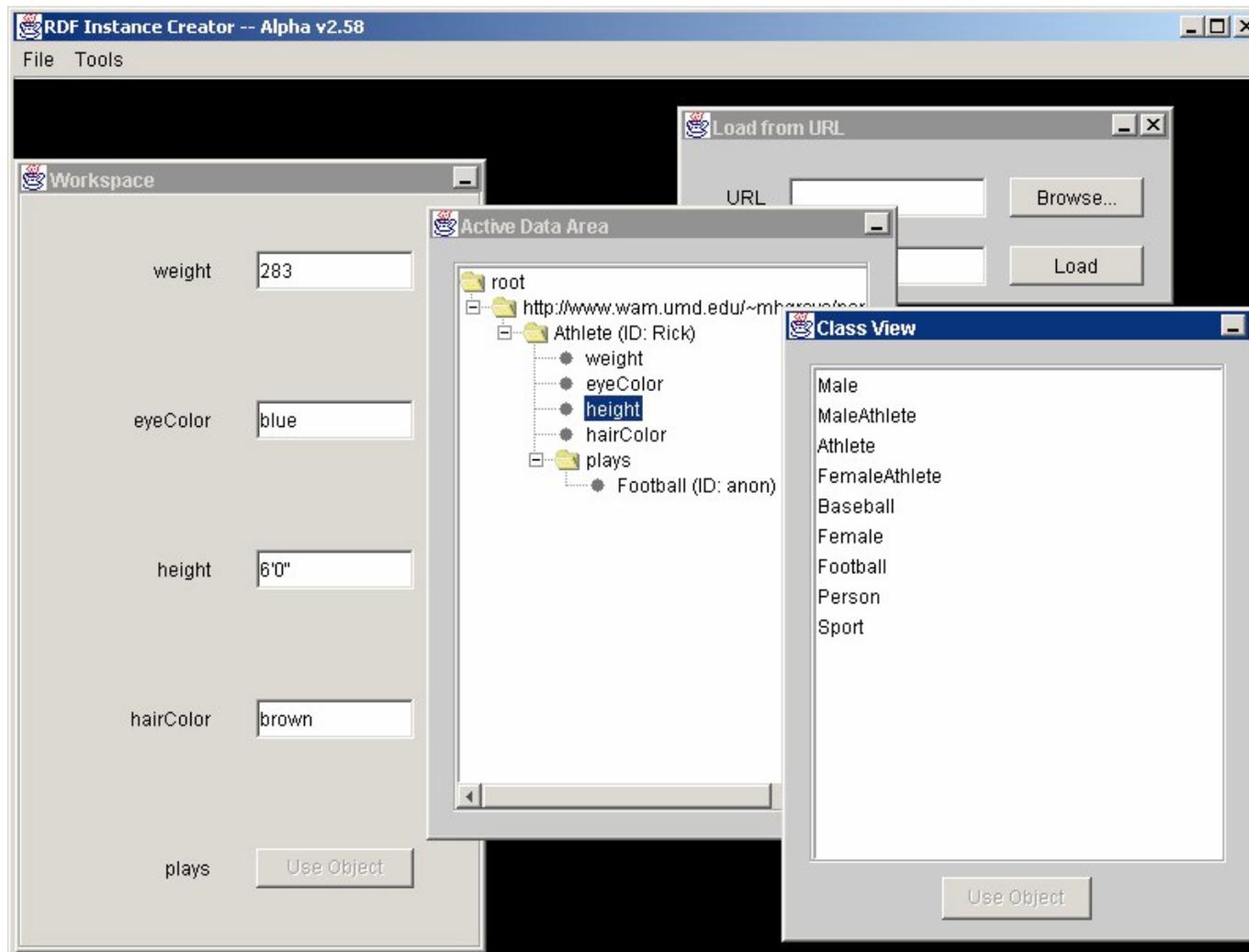
Mindswap  
- people  
- (co-author)

# Architecture Description

by Ron Alford, Amy Alford (Mindswap, UMD)  
<http://www.mindswap.org/2003/OwlSite/>



# RDFデータ生成用ツール



RDF Instance Creator (RIC) by J. Goldbeck (UMD)

# RDFデータ生成用ツール

The screenshot displays the SMORE v3.0b interface, which is used for generating RDF data from images. It consists of several windows:

- PhotoSMORE:** A window showing an image of an orangutan with several yellow circles and polygons overlaid on it, indicating detected shapes. Below the image, it shows the last selected shape (Circle) and its coordinates, along with a button to copy coordinates to the clipboard and a button to add triples.
- Options:** A window for configuring the tool. It includes fields for "Choose Ontology" (No Ontology Selected), "Choose Class" (No Class Selected), and "Selected Text" (bulbous\_forehead). Below these are fields for "Subject" (orangutan), "Predicate" (<animal:hasImagePart>), and "Object" (bulbous\_forehead). Buttons for "Add Triple to Tree/Table" and "Hide Advanced" are also present.
- RDF:** A window showing the generated RDF code. The code defines shapes like "bulbous\_forehead" and "belly" and their relationships to the main image and other parts of the orangutan.
- Triples:** A window showing a table of generated triples. The table has columns for Subject, Predicate, and Object.

Subject	Predicate	Object
orangutan.jpg	<image-ont:hasSubImage>	belly
belly	<rdf:type>	<image-ont:SubImage>
belly	<image-ont:hasShape>	Circle2
Circle2	<rdf:type>	<image-ont:Shape>
Circle2	<image-ont:hasCoordinates>	(188.0,184.0) (188.0,148.0) (5...
orangutan.jpg	<image-ont:hasSubImage>	orangutan
orangutan	<rdf:type>	<image-ont:SubImage>
orangutan	<image-ont:hasShape>	Polygon1
Polygon1	<rdf:type>	<image-ont:Shape>
Polygon1	<image-ont:hasCoordinates>	(8.0,181.0) (8.0,180.0) (9.0,17...
bulbous_forehead	<image-ont:description>	a bulbous forehead
belly	<image-ont:description>	a large belly
orangutan	<image-ont:description>	a type of monkey - orangutan
orangutan	<rdf:type>	<animal:Monkey>
orangutan	<animal:weight>	150 pounds
orangutan	<animal:hasCoat>	orangutan_coat
orangutan_coat	<animal:color>	burgundy
orangutan	<animal:hasImagePart>	belly
orangutan	<animal:hasImagePart>	bulbous_forehead

Semantic Markup, Ontology, and RDF Editor (SMORE)  
by A. Kalyanpur (UMD)

# Software Projects in mindswap

<http://www.mindswap.org/downloads/>

- RIC (RDF Instance Creator)
  - RDF生成 , AgendaRIC など
- ConvertToRDF, RDF Web Scraper
  - Excel(csv) , 表データをRDFに変換
- SMORE (Semantic Markup, Ontology, and RDF Editor)
  - メタデータ作成環境 (MailSMORE, PhotoStuff など)
- SWOOP (Web Ontology Browser and Editor)
  - Webブラウザをメタファとしたオントロジーエディタ
- Pellet (タブローアルゴリズムに基づくOWL DL reasoner)
  - 検証 , 分類 , 継承の確認 , など

---

# BBQ – July 20, 2003

To: semweb@mindswap.org  
From: Jim Hendler <hendler@cs.umd.edu>  
Subject: MIND SWAP barBque - July 20

If you can read the below, than you are invited:  
(ns headers omitted)

```
:MindSwapPicnic a :SocialEvent;  
  :invitedPersons (You YourSignificantOther YourSortOfSigOther YourOffSpring);  
  sw:date 2003-07-20^^xsd:date;  
  sw:dayofweek "Sunday";  
  sw:time 1500^^xsd:integer;  
  sw:timeinhumanterms "3:00 PM";  
  geo:Locationname "My house";  
  geo:Location <@ @directions to follow>;  
  jh:Whattobring "Nothing required, beverages if you insist on something";  
  jh:WhatNotToBring "food, unless you really know the kosher rules"  
  jh:WhatIfIdoKnowThoseRules "Assume fleischig (Basar)"  
  :RSVP "please do".
```

Look forward to seeing you there  
-JH

---

# 構築事例： 地理情報に基づく連続的情報検索

---

参考文献： K. Hiramatsu, et al., Querying Real World Services through the Semantic Web, 3rd International Semantic Web Conference (ISWC2004), LNCS 3298, pp.741-751, 2004.

# 情報検索における利用方法

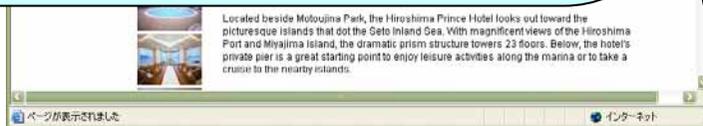
## ■ キーワード検索



入力キーワードとの合致



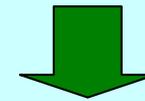
- ・ 質問解析性能の向上
- ・ 分類木を利用した  
上位・下位・類似概念の  
検索



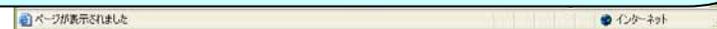
## ■ 地図検索



地図へのプロット



- ・ メタデータに基づく可視化  
(概念関係, 履歴など)
- ・ 第三者のツールを  
利用した可視化
- ・ ユーザのナビゲーション



# セマンティックウェブ技術を利用した 実世界情報検索

## ■ メタデータ・オントロジーを利用して

□ 自然文の質問を解析

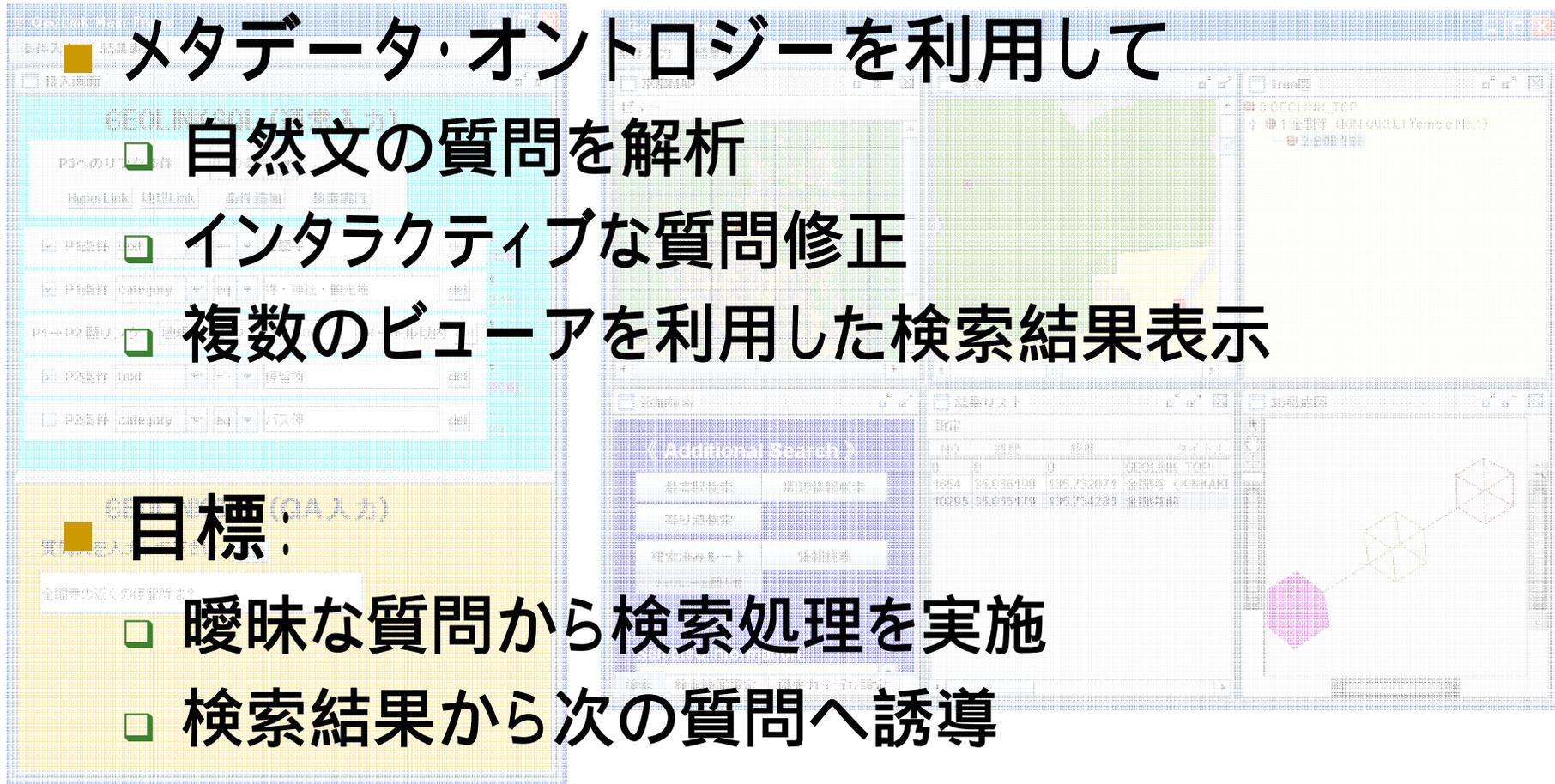
□ インタラクティブな質問修正

□ 複数のビューアを利用した検索結果表示

## ■ 目標:

□ 曖昧な質問から検索処理を実施

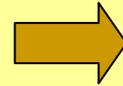
□ 検索結果から次の質問へ誘導



# Webページからのメタデータの抽出

## Webページの収集

京都市関連の  
Webページを収集



掲載確認  
(著作権, プライバシ)

フォームによる  
掲載申込み



Webページ

形態素解析  
パターンマッチング

オントロジ  
スキーマ

メタデータ  
データベース

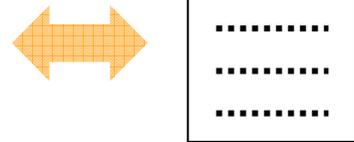
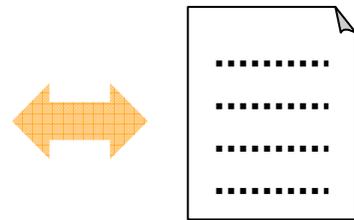
メタデータ抽出

# Webページのメタデータ

実世界の事物



Webページ



メタデータ

URL  
名称  
位置情報  
拝観時間  
拝観料  
...

URL  
名称  
位置情報  
時刻表  
料金  
...

オントロジー

辞書  
シソーラス  
分類体系  
地理索引  
電子地図

# メタデータを利用した検索エンジン

## ■ 構成要素

- キーワードベースの検索エンジン  
「金閣寺」 → 「金閣寺」を含むWebページの集合
- RDFデータベース(メタデータを管理)  
「寺社」カテゴリー → カテゴリーに属する集合
- 地理関係, リンク関係  
distance(pageA, pageB) < 250m  
link(pageA, pageB)

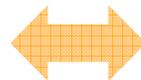
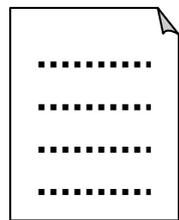
## ■ 検索結果

- “pageA  $\text{--}(\textit{link})\text{--}$  pageB” の条件を満たす集合

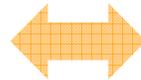
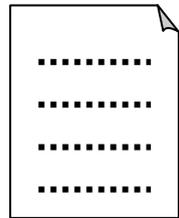
# Webページ間の地理的距離の計算

Webページ

メタデータ



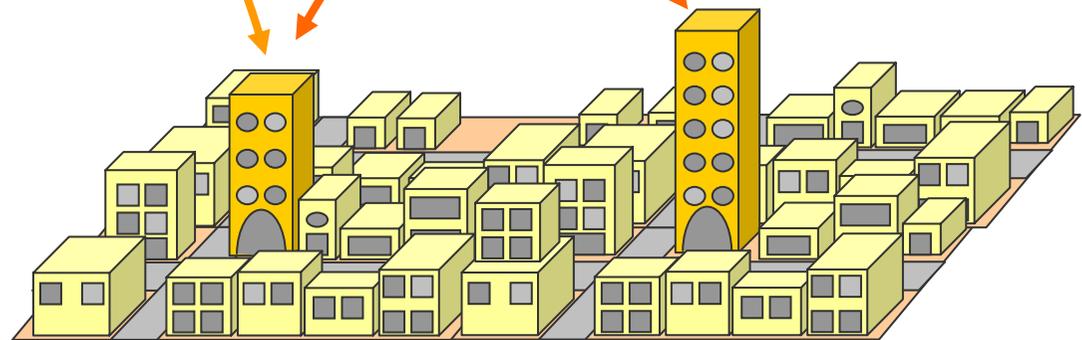
URL  
名称  
位置情報  
...



URL  
名称  
位置情報  
時刻表  
...

地図上に位置を射影

直線距離を計算



# 検索例

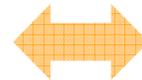
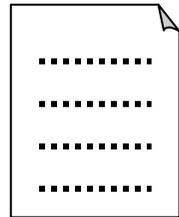
電子地図上に  
情報の位置表示

自然言語処理で  
質問文を解析

「金閣寺の停留所は？」

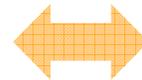
■ 期待される結果:

- 「金閣寺」の情報
- 「停留所」の
- 距離が250m以内



選択した検索結果から  
引き続き検索を発展させ  
ていくことが可能

< 250m

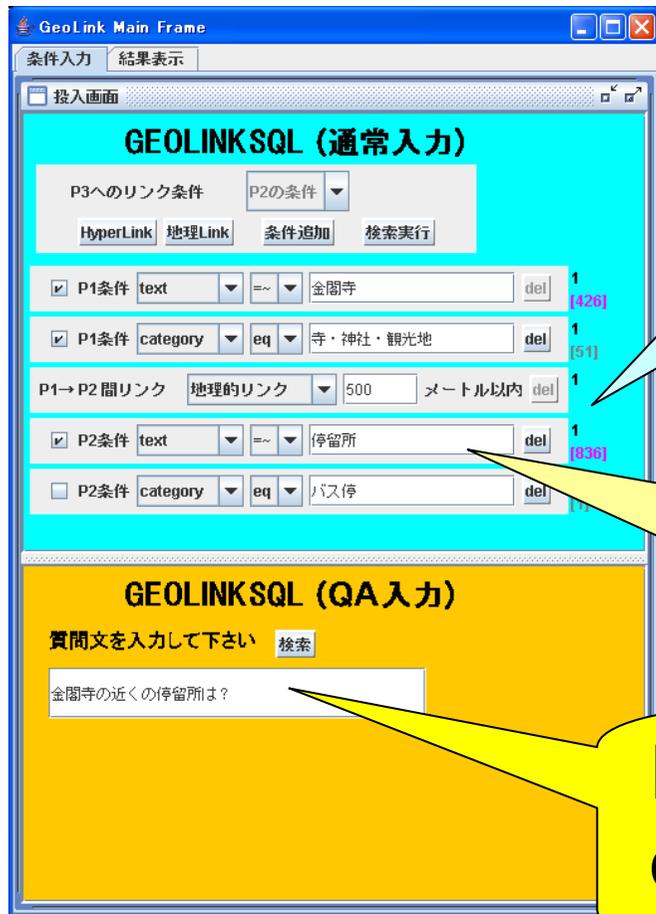


# 自然言語による質問の解析

- 形態素解析とパターンマッチングによって質問のタイプを解析
  - 例) 駅のそばのスターバックスは？
    - Type: Find-cafe, Target: Starbucks, station  
Constraint: distance is less than 250m
- 解析結果を拡張SQL形式に変換

```
1 SELECT cafe.url
2 FROM station,distance_link,cafe
3 WHERE
4   station.text =~ /station/i
5   distance_link AS Distance(station,cafe) < 250
6   cafe.text =~ /Starbucks/i
```

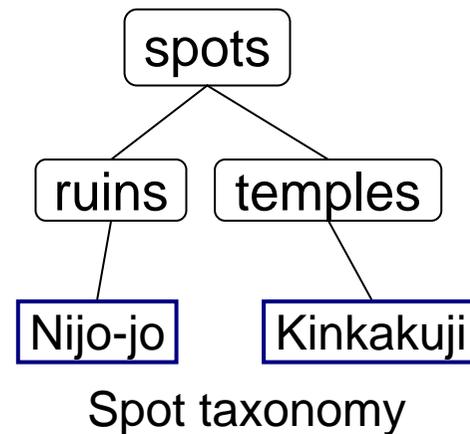
# Overview of our user interface



Queries are revised along ontologies according to number of results

Translate it style query

Extracts keyword query type us



*Enter queries in “(controlled) Japanese”*

# Overview of our user interface

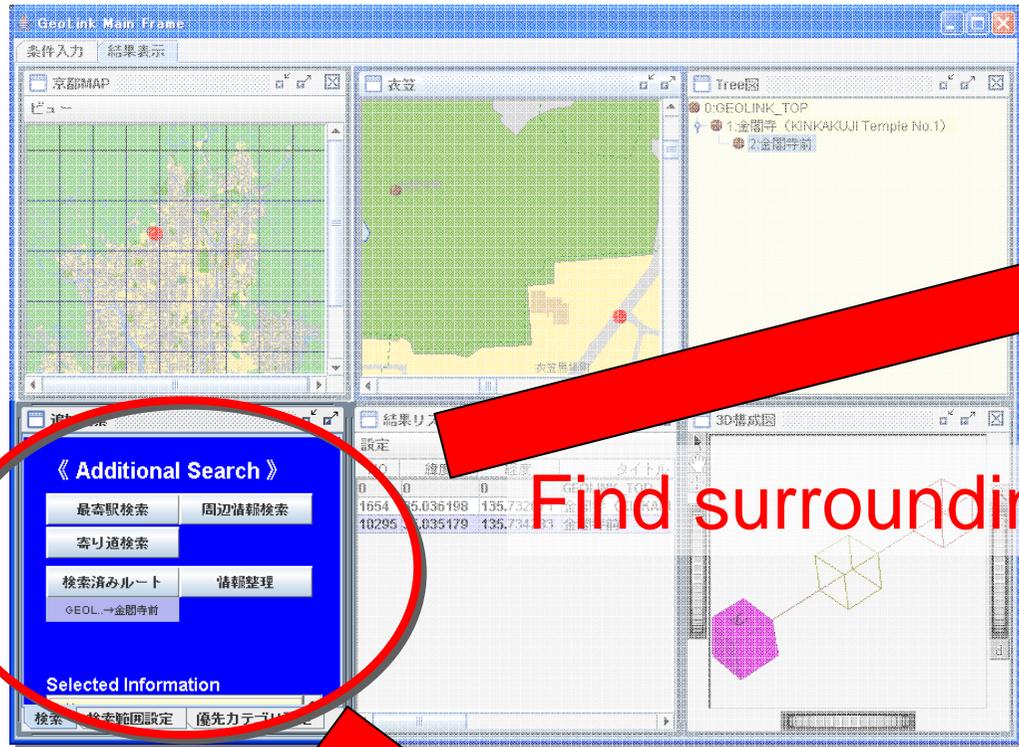
The screenshot displays the GeoLink Main Frame interface with several panels:

- Map Views:** Includes a '京都MAP' (Kyoto Map) view showing a city map with a red marker, and a '3D構成図' (3D Structure Diagram) view showing a 3D wireframe model of a building.
- Table View:** A '結果リスト' (Result List) table with columns for NO, 緯度 (Latitude), 経度 (Longitude), and タイトル (Title). The table contains three rows of data.
- Tree Views:** A 'Tree図' (Tree Diagram) view showing a hierarchical tree structure with nodes for '0: GEOLINK\_TOP', '1: 金閣寺 (KINKAKUJI Temple No.1)', and '2: 金閣寺前'.
- Input Panels:** Two panels for entering search conditions: 'GEOLINKSQL (通常入力)' (Normal Input) and 'GEOLINKSQL (QA入力)' (QA Input).

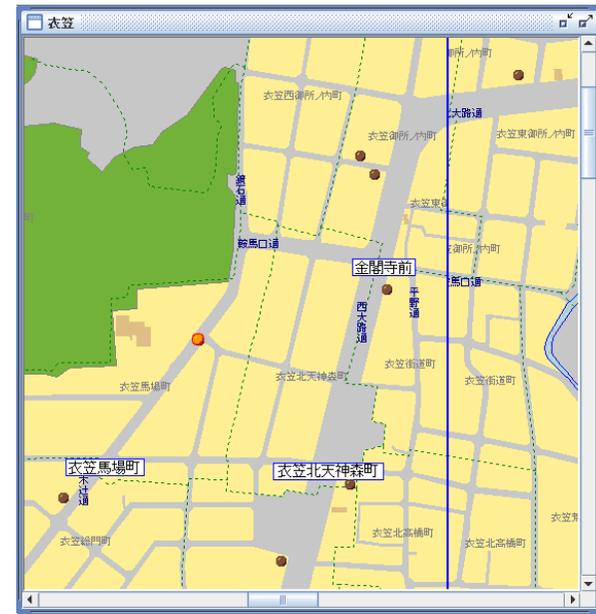
Yellow callout boxes highlight the 'Map Views', 'Table View', and 'Tree Views' components.

*Displaying results in multi-view*

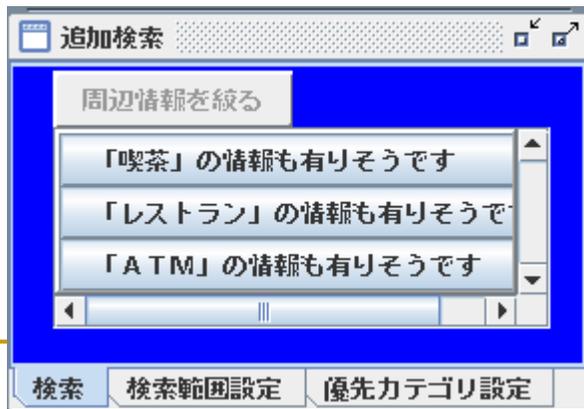
# Overview of our user interface



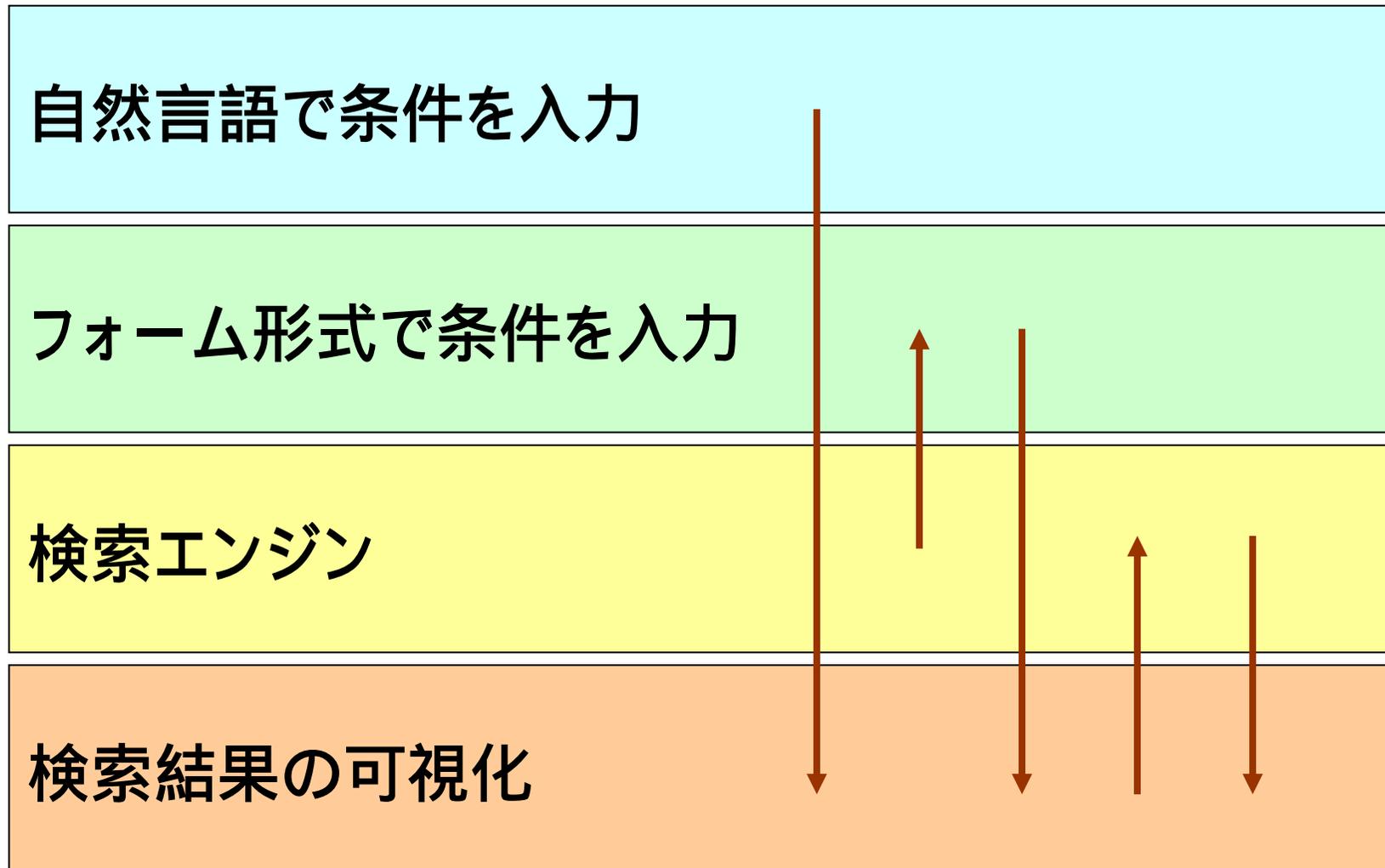
Find surroundings



Show alerts based on profile



# 検索時のユーザの状態遷移



---

# まとめ

---

# ポータルサイトの連携に向けて

- メタデータ抽出**作成**, (大規模)オントロジーの構築
  - 自動処理方法, 既存知識の形式変換  
(質問処理で利用, ユーザナビゲーションのガイド)
- Webサービスの連携
  - 韓国語映画評論 + 韓日翻訳 → 韓国映画評 (日本語)  
(サービス記述による自動連携, 連携方法の共有)
- 家電機器, オフィス機器の連携
  - Webサービスの応用, Task Computing

---

## 今後の課題

- 機械的コンテンツ統合は受け入れがたい面も  
例) Google News に対する訴訟や苦情
- ビジネスモデルはあるのか？  
課金, ユーザ管理, プライバシー, セキュリティ

# (おまけ) 今後の予測

- ページ閲覧数を増加させるための対策
  - (これまで) 代表的なディレクトリに登録
    - Yahoo!, MSN, goo に登録すると効果抜群
  - (現在) 検索エンジン最適化(SEO)の実施
    - 「Title, H1 タグにキーワード」「相互リンク」など
  - (今後) メタデータに基づく評価指標を重視
    - 情報の確かさ, 個人・組織間の信頼関係
    - 情報源への近さ(一次引用, 二次引用, …)
    - 編集技術・コンセプトに対する加点