

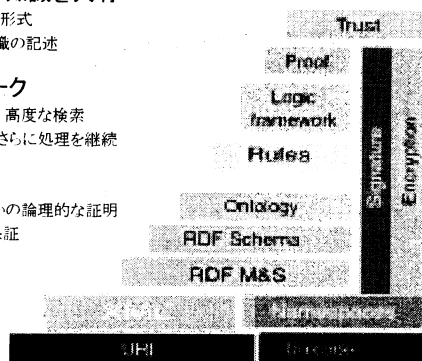
RDFとセマンティック・ウェブの現在

- RDFとオントロジーの役割
 - セマンティック・ウェブの情報とRDF
 - RDFS/OWLのクラスとプロパティ
- RDFの利用とメタデータ
 - RSSによるメタデータのパッケージ
 - FOAFを介したデータの連携
 - メタデータの作成、抽出とfolksonomy

2005-05-19
神崎正英

セマンティック・ウェブのレイヤーケーキ

- URI, Unicode, XMLを基盤に
 - ウェブの蓄積をベースにグローバルなデータ(OpenWorld)を扱う
- マシンが読める形でデータや知識を共有
 - RDFによるシンプルで柔軟なデータ形式
 - RDFS, オントロジーによる語彙と知識の記述
- 推論規則と論理フレームワーク
 - 述語論理, 記述論理を用いた推論, 高度な検索
 - 検索結果をエージェントが判断してさらに処理を継続
- 署名と証明と信頼
 - エージェントがなぜこの処理をしたかの論理的な証明
 - 電子署名と暗号化によるデータの保証



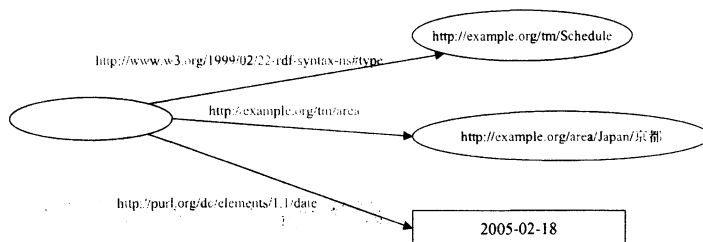
パーナードリーによる「レイヤーケーキ」の図。厳密な技術階層というよりは、セマンティック・ウェブを構成する技術の開発ステップを階層的にプレゼンテーションしたものだ。

セマンティック・ウェブの段階的応用

- 高度なリソース探索
 - 全文検索 v.s. メタデータによる検索
 - 最も分かりやすいセマンティックウェブのアプリケーション(出発点)
 - 鶏と卵: 誰がメタデータを与えるのか
- 語彙の共有とデータの多角的利用
 - URIを用いた語彙表現によるデータの連携
 - ウェブログだけ、SNSだけの閉じたサービスではなく、互いのデータが連動する
- サービスの自動発見と連動
 - 単なる検索ではなく、推論によるサービスの発見
 - 単一のサービスの自動化だけではなく、複数サービスを組み合わせ、連携する
- 信頼のウェブ
 - マシンに探索、サービス処理を任せ、人間が確認、判断する
 - データの信頼性、SPAM、プライバシー、セキュリティ上の問題

RDFのデータモデル

- グラフによるトリプルの表現
 - 主語(S)-プロパティ(述語:P)-値(目的語:O)の3者関係=トリプルを基本にしたモデル
 - ノード(円)とリンク(矢印)による有向ラベル付きグラフとして表現
 - グラフのつながりによる「意味ネットワーク」の表現
- トリプルの各要素をURI参照で識別する
 - 同音異義語などの曖昧さを避け、グローバルにリソースを識別する
 - ウェブ上で分散してリソースを記述することが可能(グラフをグローバルに集約・併合できる)



RDFスキーマでのクラス、プロパティ表現

- シンプルな階層関係の表現が基本
 - rdfs:subClassOfとrdfs:subPropertyOf
 - ツリーをたどった簡単な推論を可能にする
- プロパティの適用範囲の表現
 - プロパティはクラスに從属せず、グローバル
 - rdfs:domain, rdfs:rangeで定義域(主語のクラス)、値域(目的語のクラス)を示す
 - 値を制約するのではなく、推論を可能にする

```
<rdfs:Class rdf:ID="MusicalWork"/>

<rdfs:Class rdf:ID="Symphony">
  <rdfs:subClassOf rdf:resource="#MusicalWork"/>
</rdfs:Class>

<rdf:Property rdf:ID="composer">
  <rdfs:subPropertyOf
    rdf:resource="http://purl.org/dc/elements/1.1/creator"/>
  <rdfs:domain rdf:resource="#MusicalWork"/>
  <rdfs:range rdf:resource="#Person"/>
</rdf:Property>
```

オントロジー言語OWLの基本機能

- OWLによる知識記述と推論
 - RDFSよりもきめ細かなクラス定義と推論のできるプロパティ
 - 記述論理による推論を可能にする知識ベースの記述
- ウェブ上でのスキーマ/オントロジーの相互運用
 - 別々のコミュニティで作られるオントロジーの共有性、相互運用性
 - オントロジーの改訂や統合が可能な発展性の確保
 - 複数のオントロジーやデータの間の不整合の検出
- OWLの3つのサブ言語
 - OWL DL: 記述論理に基づく推論が可能なオントロジーの記述
 - OWL Lite: 実装を容易にするために使用できる語彙に制限を設けたDLのサブセット
 - OWL Full: OWLの語彙と機能をフルに利用できる。最も表現力が高いが、計算の決定可能性が保証されず、厳密な推論には向かない
- OWL DL/Liteと型分離
 - 記述論理に従うために、クラス、プロパティ、インスタンスの型を明確に分離する
 - 同じ名前(URI)がクラスとインスタンスの両方を表すことはできない(OWL Fullでは可能)
 - プロパティを個体値型とデータ値型に分離する

クラス公理の記述

- さまざまなクラス表現の組み合わせ
 - 名前付け表現 (存在公理)
 - 全てのインスタンスを列挙する
 - クラスのインスタンス集合の論理組合せ (既存クラスの組合せで新しいクラスをつくる)
 - プロパティの制約 (値、出現回数)
- 同等なクラスと分離クラス
 - オントロジーのマッピングと推論
 - owl:equivalentClass: 同じインスタンスを持つ
 - owl:disjointWith: 共通のインスタンスがない

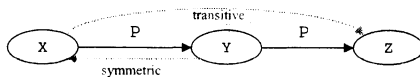
```
<owl:Class rdf:ID="StringEnsemble">
  <owl:disjointWith rdf:resource="#WindEnsemble"/>
  <rdfs:subClassOf rdf:resource="#MusicalGroup"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:onProperty rdf:resource="#player"/>
      <owl:allValuesFrom rdf:resource="#StringPlayer"/>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
```

OWLプロパティの表現と推論

- 同等関係、反対関係
 - owl:equivalentProperty: 同じ主語・述語の組み合わせの集合に対応する
 - owl:inverseOf: 反対の関係を表すプロパティ

```
<owl:ObjectProperty rdf:ID="composes">
  <rdfs:domain rdf:resource="#Person"/>
  <rdfs:range rdf:resource="#MusicalWork"/>
  <owl:inverseOf rdf:resource="#composedBy"/>
</owl:ObjectProperty>
```

- 論理的な関係
 - プロパティPが、X P Y. という関係を記述するとき、Pが次のタイプの場合:
 - owl:FunctionalProperty Xに対してYが高々1つに定まる
 - owl:InverseFunctionalProperty Yに対してXが高々1つに定まる (IDとして働く)
 - owl:TransitiveProperty X P Y. かつ Y P Z. ならば X P Z. が成り立つ
 - owl:SymmetricProperty X P Y. ならば Y P X. も成り立つ



オントロジーのマネジメントと共有

- **オントロジーの管理情報**
 - owl:Ontology = オントロジーヘッダを記述(オントロジー自身を表すクラス)
 - owl:versionInfo = テキストによる注釈情報
 - owl:priorVersion = 古いバージョンと結び付けておく
- **オントロジーの共有、再利用**
 - owl:imports = 外部オントロジーのグラフを取り込んで利用可能にする
 - 名前空間宣言と異なり、実際にそのオントロジーを取り込む
 - 推移的性質を持つ(インポート先にowl:importsがあれば、さらに取り込む)

```
<owl:Ontology rdf:about="">
  <owl:versionInfo>
    演奏会オントロジー v 1.44 2005-02-18 etc...
  </owl:versionInfo>
  <owl:imports rdf:resource="&music;" />
  <owl:imports rdf:resource="&foaf;" />
  ...
</owl:Ontology>
```

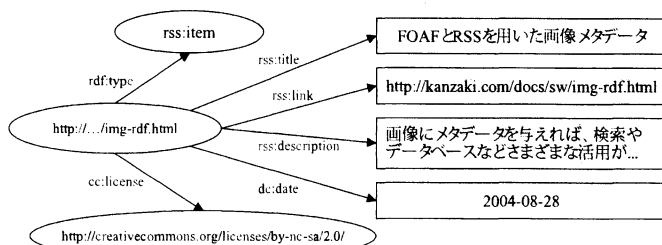
RDFの利用例1: RSS

- **RDF Site Summary**
 - サイトの更新情報、ニュースなどをRDFのフォーマットで配信(RSS 1.0)
 - かつての「フレッシュ型」情報配信の発展形
- **基本語彙以外に様々なデータを加えて配信できる**
 - Dublin Coreを使って日付、作者などの基本書誌情報
 - Creative Commonsを使ってライセンス情報
- **シンプルなので自動処理が簡単**
 - ウェブツールなどが自動生成
 - サイトの更新情報との連動も簡単

```
<rss:item rdf:about="http://kanzaki.com/docs/sw/image-rdf.html">
  <rss:title>FOAFとRSSを用いた画像メタデータ</rss:title>
  <rss:link>http://www.kanzaki.com/docs/sw/image-rdf.html</rss:link>
  <rss:description>画像にメタデータを与えれば、検索やデータベースなど
    さまざまな活用が可能になります。...</rss:description>
  <dc:date>2004-08-28</dc:date>
  <cc:license
    rdf:resource="http://creativecommons.org/licenses/by-nc-sa/2.0/" />
</rss:item>
```

RSSデータのRDFグラフ

- 主語は“item”という型を持つリソース
 - item要素の子要素がプロパティを示す
 - rdf:about属性の値が主語のURI
- プロパティ要素の内容がプロパティ値
 - プロパティの値がオブジェクトリソースの場合はrdf:resource属性でURIを示す

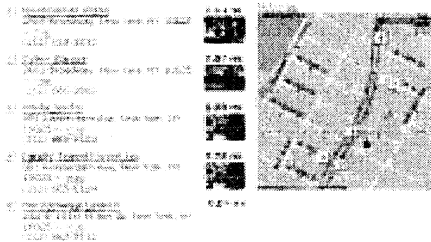


ウェブ文書の基本メタデータとしてのRSS

- Dublin Coreとの組み合わせで基本書誌データはカバー
 - rss:title, rss:descriptionはDCのサブプロパティ
 - 大多数のRSSがdc:creator, dc:date相当のメタデータを持つ
 - 多くのRSSがdc:subjectとして主題標目を提供
- RSSの集積とウェブ文書データベース
 - 個別のRSSフィードを購読するだけでなく、多数のRSSを収集・蓄積するサービス
 - Technorati, bulkfeeds, feedbackなど
 - フィード単位ではなく、item単位でデータを利用できる
 - メタデータ項目を利用した、主題検索など
- ウェブ文書検索の標準メタデータとしてのRSS
 - ウェブログ以外のウェブ文書もRSSを採用しつつある
 - 検索結果などをRSSとして提供するサービス

RSS:メタデータのコンテナとしての可能性

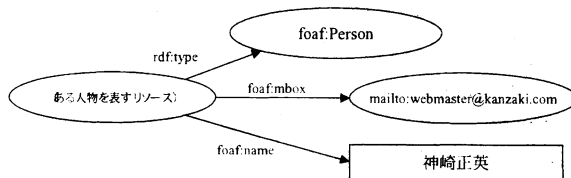
- 専用リーダーやブラウザ組み込みなどの閲覧環境
 - 多くの利用者がRSSを取得して閲覧する環境を備える
 - 本来は「サイト要約」だが、それ以外にもさまざまなメタデータを配信可能
- 統一された基本フォーマットと拡張性
 - RDFに基づく基本形が確立しているため、多くのツールで簡単に利用できる
 - RDFの仕組みを利用して、さまざまな語彙を組み合わせて配信可能
- RSSを利用した新しいデータ配信
 - 時間軸を持つ情報としてのRSS: イベント・新刊情報などの未来形の情報/カレンダーの配信
 - さらに位置情報や写真を組み合わせた、レストラン情報など



a9.comで検索結果をYellowPageボタンを使って地図表示した画面。RSSにはなっていないが、位置情報を加えると、同様の表現が可能になる。

RDFの利用例2: FOAF

- 人を中心としたメタデータをRDFで記述
 - 人についての情報と、知人関係を記述
 - 人間にURIを割り当てる一般的な方法がないので、通常は空白ノード
 - 個人をメールアドレスで特定 (IFPとして機能する)



```
<foaf:Person>
<foaf:name>神崎正英</foaf:name>
<foaf:mbox rdf:resource="mailto:webmaster@kanzaki.com"/>
</foaf:Person>
```

FOAFによるデータの記述

- 人物情報や交友関係などのポキャブラリ
 - 関心事項、写真、ホームページなどの多彩な語彙
 - 人物を記述する機会が多いので、他の語彙との組み合わせで利用範囲が広い
 - ソーシャルネットワーキングの流行にもマッチする
- 交友関係の連鎖で、さまざまなRDFデータがつながる
 - 収集したデータから人間関係マップができる
 - `rdfs:seeAlso`はRDFのエージェントのためのリンク(htmlのa要素のような役割)

```

<foaf:Person>
<foaf:name>Masahide Kanzaki</foaf:name>
<foaf:mbox rdf:resource="mailto:webmaster@kanzaki.com"/>
<foaf:homepage rdf:resource="http://www.kanzaki.com"/>
<foaf:depiction rdf:resource="http://.../masaka.jpg"/>
<foaf:knows>
  <foaf:Person>
    <foaf:name>Dan Brickley</foaf:name>
    <foaf:mbox rdf:resource="mailto:dانبri@w3.org"/>
    <rdfs:seeAlso rdf:resource="http://.../danbri-foaf.rdf"/>
  </foaf:Person>
</foaf:knows>
</foaf:Person>

```

FOAFの語彙

• FOAFのクラス

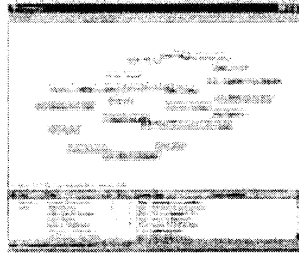
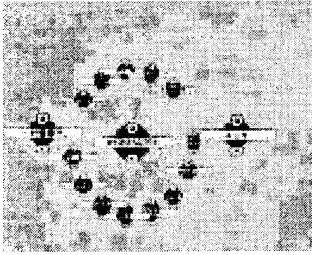
クラス	説明
foaf:Agent	人間、グループ、ソフトウェアなどを総称するクラス
foaf:Person	人物を表すクラス。foaf:Agent のサブクラス
foaf:Group	グループ一般を表すクラス。foaf:Agent のサブクラス
foaf:Organization	組織形態をもつグループを表すクラス。foaf:Agent のサブクラス
foaf:Document	文書を表すクラス
foaf:Image	画像を表すクラス。foaf:Document のサブクラス

• FOAFの主要プロパティ

プロパティ	説明
foaf:mbox	メールボックスURI(mailto)で、家族などと共有していないもの
foaf:name	名前(人に限らず使える): ニックネームはfoaf:nick
foaf:knows	目的語の人を知っている
foaf:interest	関心を持っていることに関するページ
foaf:publications	出版物へのリンク
foaf:made	主語が作ったもの
foaf:schoolHomepage	母校のホームページ: 仕事関連ならfoaf:workplaceHomepage
foaf:currentProject	現在手がけているプロジェクト: 過去のものならfoaf:pastProject
foaf:homepage	ホームページ
foaf:weblog	ウェブログ
foaf:depiction	主語を描いたもの。写真やイラストなど
foaf:topic	主語ページ、文書のトピック(foaf:pageの逆)
foaf:member	目的語は主語グループのメンバーである

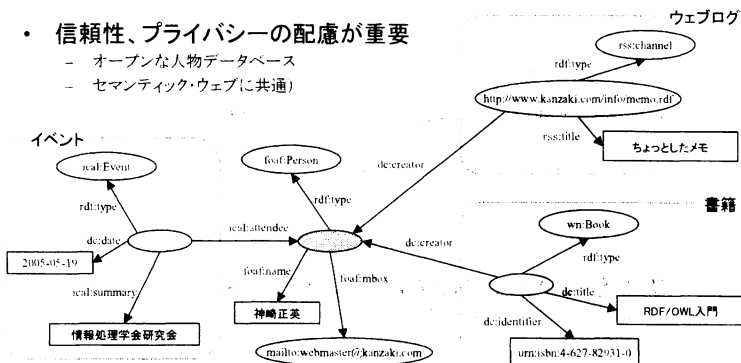
FOAF:人に関する汎用語彙

- 人: 最も関心のあるメタデータ
 - ある人に関する関連情報を知りたい
 - 「この本の著者が出席するイベント」を検索
- 様々な関係の視覚化
 - 交友関係やを同じ関心を持つ人をグラフ化して新しいつながりを発見
 - 携帯ツールと組み合わせた「モバイル人物名鑑」



FOAFの利用

- 人を軸にした様々なメタデータの連動
 - 多くのメタデータは、人に関する情報を含む(コンテンツには基本的に作者がいる)
 - 書籍の著者、そのウェブログ、出席するイベント...
 - RDFによる語彙/データ連動のハブとなりうる
- 信頼性、プライバシーの配慮が重要
 - オープンな人物データベース
 - セマンティック・ウェブに共通)



FOAFとRDFのクエリ

- FOAFデータの収集とデータベース

- 各人が分散して記述したFOAFをロボットが収集、データベースに蓄積
- 複数の語彙が混在していても、トリプルからリソース、述語は明確
- データベースは語彙に関する知識を必要としない
 - スキーマやオントロジーを用いてsubsumptionしてもよい

- RDFのクエリ

- 集積したデータから、グラフパターンを利用して検索
 - 異なるサービス、ソースによっても検索ができる
- RDFクエリ言語SPARQLがW3Cで検討中
 - 例: メールアドレスがwebmaster@kanzaki.comである人物の出席するイベントを調べる

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX ical: <http://www.w3.org/2002/12/cal/ical#>
SELECT ?name ?evdate ?evsummary
WHERE
  (?person foaf:mbox <webmaster@kanzaki.com>)
  (?person foaf:name ?name)
  (?event ical:attendee ?person)
  (?event dc:date ?evdate)
  (?event ical:summary ?evsummary)
```

メタデータをどうやって与えるのか

- ユーザは面倒なことはいらない

- 普通のユーザは自分のページのメタデータなど提供しない
- meta要素によるメタデータは検索エンジンシステムで破綻状態
- 書いたとしても、メンテナンスするのが困難



- ユーザに意識させない(ツールによる)メタデータ生成

- ウェブログツールによるRSS生成
- ウェブログの「プラグイン」によるメタデータ生成
- ウェブログサービスの「プロフィール」からのFOAF生成
- デジタル写真のExifやPDFなどのXMP
- はてなキーワードのようなメタデータ自動抽出

- ユーザが便利と感じるシステムからのメタデータ生成

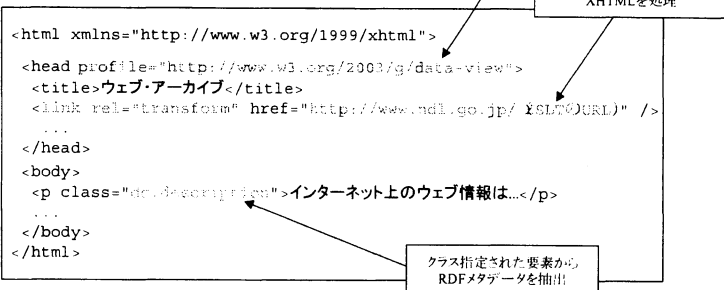
- スタイルシート(クラス)との連動と自動抽出
- del.icio.us, flickrなどの「ソーシャルサービス」のタグシステム
- folksonomy v.s. 統制語彙

作者自身によるメタデータと自動抽出

- XHTML文書のセマンティクス
 - XHTMLにも、link、address、dfnなどの、「メタデータ」を伝える要素がある
 - 適切に記述されたXHTML文書からのメタデータ抽出はある程度可能
 - W3CのSemantic data extractor
 - <http://www.w3.org/2003/12/semantic-extractor.html>
 - マイクロコンテンツ
 - 例: XFN (rel="friend"などのように、a要素のrel属性を利用)
- XHTMLからメタデータを抽出
 - 普通的手段を応用 → スタイルシートなどでclass属性を付与する作者は多い
 - 各人がclass属性値に意味を与える対応表を作れば、class属性からメタデータを生成可能
 - GRDDLによる抽出(後述)
- 作者自身に依存する限界
 - 作者の意識が高くないと現実的には広まりにくい
 - meta要素と同様の「自称メタデータ」の問題

GRDDL

- XSLTを用いて、XHTMLからメタデータを抽出
 - Gleaning Resource Descriptions from Dialects of Languages (<http://www.w3.org/TR/grddl/>)
 - クラス属性など、一定のマーク付けパターンからXSLTを使ってRDFを生成する
 - 文書ルールのある組織内やツールでのメタデータ生成・維持には有効
- XSLTはどんなものでもよい
 - 一般的なルールに従うなら、汎用のXSLTが利用できる
 - 独自ルールを定めるなら、専用のXSLTを記述すればよい
 - profile属性とlink要素でGRDDLを利用



Folksonomy

- コンテンツ作成者とメタデータ作成者

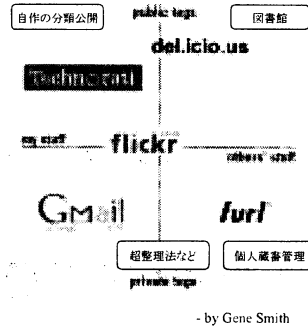
- 誰が、誰のためにメタデータを付与するか
 - X軸: 自分の作品 vs 他人の作品
 - Y軸: 自分用 vs 公開用
- 自分のためのタグがコミュニティでも機能する

- メタデータの適切さの問題

- 同義語や重複数によるゆれ
- 階層を持たない単語
 - operaは音楽かブラウザか
- メタデータ付与の基準なし
- 将来的には「タグスバム」の可能性

- 多数の「タグ」が集積する効果

- Power Law
- 入力補助機能
- クラスタによるタグ、キーワードの関連付け
- 「投票」としてのタグ→信頼性のメタデータ



del.icio.usのタグとRDF

- 同じ「タグ」と関連タグによる連動

- 自分のブックマークエントリにつけた「タグ」から同じタグを持つブックマークへ
- 「related tags」による関連メタデータとの紐付け、揺れの吸収
- 同じタグをつけている人=同じ関心を持つ人のブックマークを参照できる

- RSSのtaxo:topicsとしてメタデータを取得できる

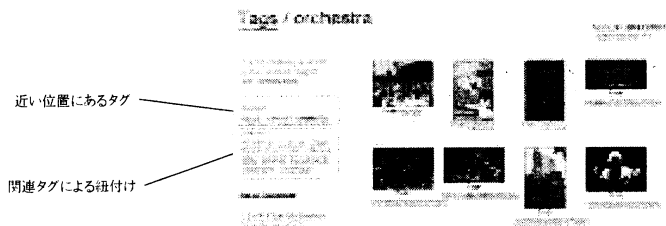
- 「http://del.icio.us/tag/semanticweb」といったURIでトピックを特定

```

<taxo:topics>
<rdf:Bag>
  <rdf:li resource="http://del.icio.us/tag/rdf"/>
  <rdf:li resource="http://del.icio.us/tag/semanticweb"/>
  <rdf:li resource="http://del.icio.us/tag/foaf"/>
  <rdf:li resource="http://del.icio.us/tag/xml"/>
  <rdf:li resource="http://del.icio.us/tag/socialsoftware"/>
  <rdf:li resource="http://del.icio.us/tag/semanticweb"/>
  <rdf:li resource="http://del.icio.us/tag/semanticweb"/>
</rdf:Bag>
</taxo:topics>
    
```

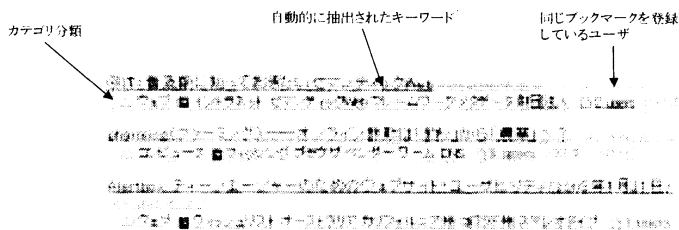
flickr.comとRDF

- 同じタグを持つ写真の検索や一覧表示
 - related tag, see also による関連付け、揺れの吸収
 - 「<http://www.flickr.com/tags/orchestra/>」などのURI
- Note機能による画像のアノテーション
 - 画像の一部分にテキストで注釈を加える
- flickr からのRDFメタデータの抽出
 - APIを使ったタグ、ノート抽出とRDFへの変換
 - プロフィールページからのFOAFの抽出



自動的なメタデータの付与

- 図書館などでの自動メタデータ作成の試み
 - 一般に、日付、フォーマットなど形式的なものは積極的
 - タイトルでも検証必要、主題分類などの自動付与は懐疑的
- はてなブックマーク
 - ユーザがブックマークを登録すると、カテゴリ分類とキーワードが自動的に付与される
 - 見当はずれのものもあるが、ユーザの負荷は軽く、確実
 - 「<http://b.hatena.ne.jp/keyword/セマンティックWb/>」のような形でリンク
 - 今のところRSSへのキーワード出力はないが、del.icio.usと同様の形は可能



とりあえずのまとめ:メタデータの現在と近未来

- **高度なリソース探索**
 - RSSを中心としたウェブ文書リソースのメタデータ
 - CMSツールによるメタデータ付与やフォークソノミーによる自発的なメタデータ付与
 - GRDDLなどによる自動抽出
- **語彙の共有とデータの多角的利用**
 - メタデータの標準コンテナとしてのRSS
 - FOAF:人情報を介したさまざまなデータの連動
- **サービスの自動発見と連動**
 - セマンティック・ウェブ・サービス?
 - 公開APIとRDF語彙を用いたサービスの相互運用
- **信頼のウェブ**
 - リンクのタイプを利用した「信頼度」の表現
 - ブックマークなどの登録数:Power Law
 - スパムとの戦いと個人情報保護