Web 2.0からセマンティック・ウェブへ

人工知能学会　セマンティックウェブとオントロジー研究会
2006-11-21　神崎正英

“Web 2.0”の主張

1

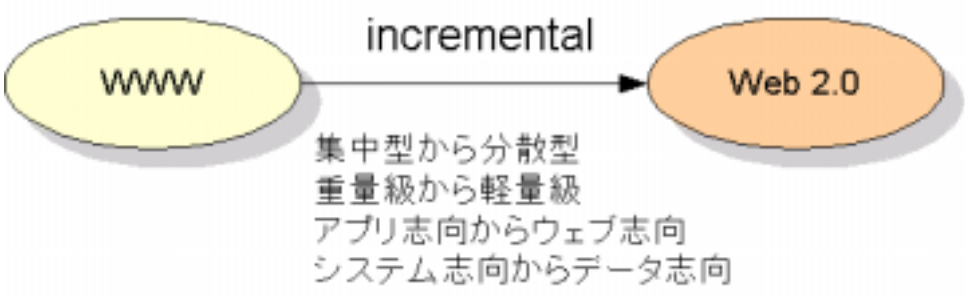
- **Tim O'Reillyによれば**
  - プラットフォームとしてのウェブ (The Web As Platform)
  - 集合知の利用 (Harnessing Collective Intelligence)
  - データが重要 (Data is the Next Intel Inside)
  - 商品としてのソフトからサービスとしてのソフトへ (End of the Software Release Cycle)
  - ハッキングと連動が容易な軽量プログラミングLightweight Programming Models
  - PCに限定されない利用 (Software Above the Level of a Single Device)
  - Ajaxに代表されるリッチなUI (Rich User Experiences)

(*What Is Web 2.0[1] Design Patterns and Business Models for the Next Generation of Software*)

Web 1.0からWeb 2.0へ

2

- **既存の技術を組み合わせたアプリケーション、サービスの連動**

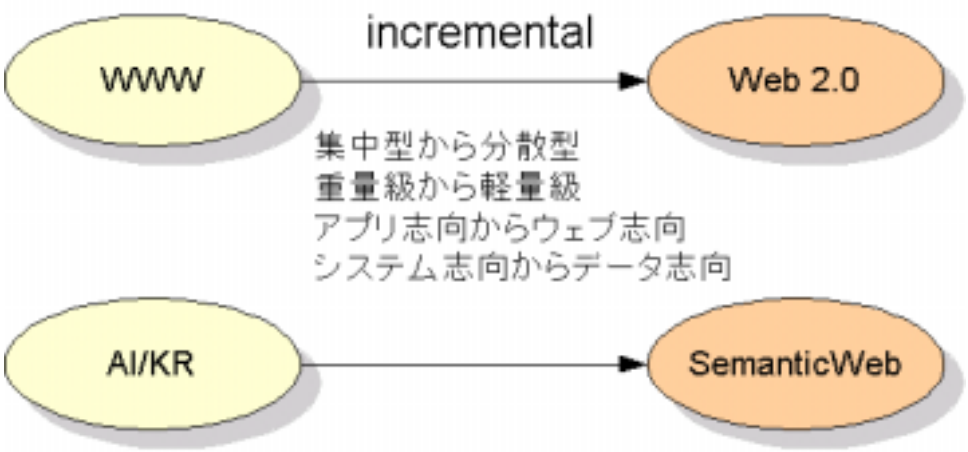


  - 既存ウェブからの移行、組み込みが容易な**漸次的な発展**

AI/KRとセマンティック・ウェブ

3

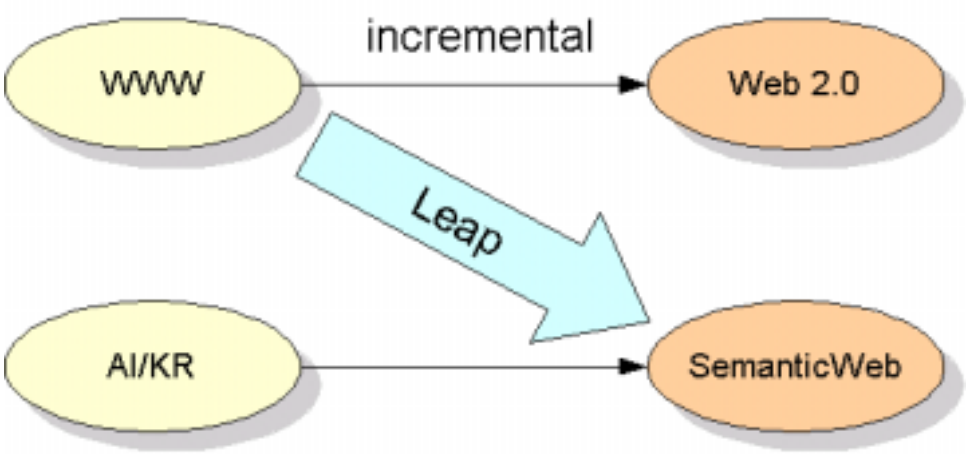
- **W10とW20の関係はAI/KRとSWの関係に類似**



セマンティック・ウェブへの跳躍の問題

4

- **従来のウェブからSWへは大きな跳躍が必要**

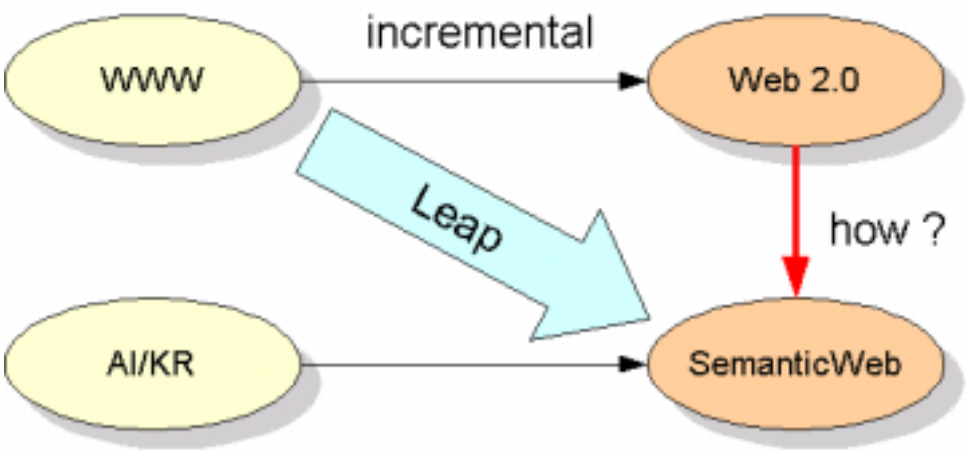


  - The attempt to get the world to switch to XML, including quotes around attribute values and slashes in empty tags and namespaces all at once didn't work -- Tim Berners-Lee

Web 2.0による橋渡し

5

- **W20を利用したセマンティック・ウェブへの漸次発展**



  - うまく機能すれば、ドメインを越えた「データのウェブ」が出現
  - W20が橋渡しとして機能するには何が必要か

Web 2.0とセマンティック・ウェブ

6

- **比較**

	Web 2.0	セマンティック・ウェブ
データモデル	XHTML/microformats	RDF(RDFa？)
データアクセス	個別API	SPARQL
スコープ	コミュニティ	グローバル
語彙	フォークソノミー（タグ）	オントロジー
推論	統計・確率的	論理的
ユーザインターフェイス	Ajax	SWUIWG…

- **W20からSWへ**
  - 軽量、導入しやすいというW20の長所を生かしながら
  - コミュニティ/APIレベルからグローバルへの跳躍を橋渡しする

橋渡しの3つのテーマ

7

- **テーマ1：microformats**
  - 既存のXHTML要素、属性を利用し、CSSと結びつけて扱いやすく
  - Wikiなどをベースにしたルールの公開
  - 語彙を混在や再利用が困難
  - →microformatsのグローバル化
- **テーマ2：個別API**
  - データを重視し、外部からでも利用できるAPIを公開
  - サービスごとに異なるAPIではスケーラビリティがない
  - →個別APIから共通のクエリへ
- **テーマ3：フォークソノミー・タグ**
  - 「タグ」は便利で有益だが、語彙の揺れや同義語をはじめとして問題は多い
  - 「タグ」の統計的な処理から論理的処理への手立ては
  - →タグとオントロジーの連携

テーマ1：microformatsのグローバル化

8

- **名前空間のないローカル名のみのmicroformats**

```
<div class="vevent">
 <abbr class="dtstart" title="2006-11-21">11月21日</abbr>に
 <span class="summary">セマンティックウェブとオントロジー研究会</span>を実施します。
</div>
```

- **グローバル名とデータモデル**
  - 一定のルールでmicroformatsの語彙をURIにマッピング
  - 合わせてRDFのモデルを与える

```
<ical:Vevent>
 <ical:dtstart rdf:parseType="Resource">
  <ical:date>2006-11-21</ical:date>
 </ical:dtstart>
 <ical:summary>セマンティックウェブとオントロジー研究会</ical:summary>
</ical:Vevent>
```

変換ルールの汎用フレームワークGRDDL

9

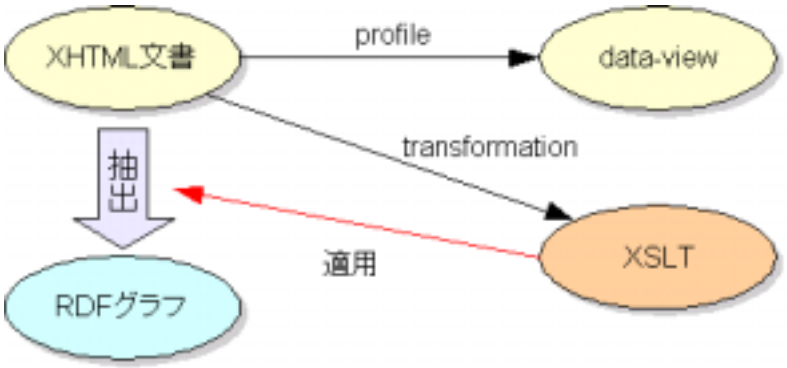
- **サービスに依存せずに変換ルールを適用する**
  - 変換アルゴリズムをXSLT（など）で表現
  - 文書とXSLTを結びつけるルールを定めることで汎用的な変換を実現
- **仕様策定中**
  - 2006年10月に初の公開草案
  - 基本仕様Gleaning Resource Descriptions from Dialects of Languages[2]
  - 入門GRDDL Primer[3]とユースケースGRDDL Usecases[4]

GRDDLとhtmlプロファイル

10

- **profile属性とリンクタイプの利用**
  - XHTML文書のprofile属性値が`http://www.w3.org/2003/g/data-view`を持つとき
  - `rel="transformation"`であるlink要素で示されるリソースは文書のGRDDL変換（＝XSLTスクリプト）

```
<head profile="http://www.w3.org/2003/g/data-view">
 <link rel="transformation" href="(XSLTのURI)" />
```



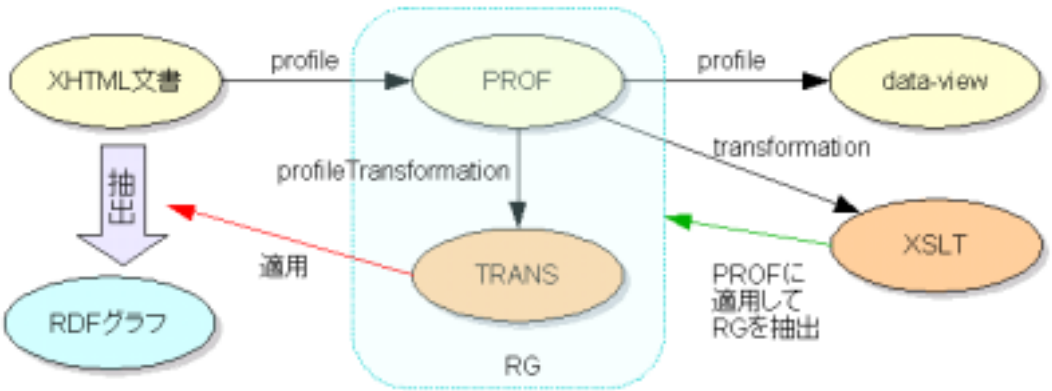
  - GRDDL変換結果グラフが複数あるときは、結果グラフを併合したものが文書のGRDDL変換結果

GRDDL変換の間接指定

11

- **プロファイル文書経由でスクリプトを指定**
  - XHTML文書Dのprofile属性で示される文書PROFがGRDDL変換結果グラフRGを持つとき、RGはDの変換結果でもある
  - さらRGが `PROF grddl:profileTransformation TRANS` となるトリプルを持つ場合、TRANSは文書DのGRDDL変換（スクリプト）を示す

```
<head profile="(独自のプロファイル:PROF)">
<!-- 文書Dにはlink要素は不要 -->
```



プロファイル文書でのXSLT指定

12

- **プロファイル文書自身にGRDDLを適用して参照XSLTを得る**
  - PROFからTRANSを得るにはW3Cの用意するXSLTが利用できる
  - プロファイル文書にはrel="profileTransformation"としたリンクを埋め込む

```
<head profile="http://www.w3.org/2003/g/data-view">
 <link rel="transformation" href="http://www.w3.org/2003/g/glean-profile" />
 ...
<a rel="profileTransformation" href="(XSLTのURI)">...
```

- **メリットと課題**
  - link要素すら敷居が高い場合、microformatsほかで既にプロファイルを定義している場合などに有効
  - 処理ステップが一つ増える
  - 現段階ではサポートしていないツールもある

もう少し進んだXHTMLとRDFの連携

13

- **Embedded RDF**
  - Dublin CoreをHTMLに埋め込む方法を利用して、接頭辞とスキーマURIをマッピング
  - クラスのローカル名の前に接頭辞をハイフン(-)もしくはピリオド(.)でつなぐ
  - プロファイル指定で汎用的にRDFを抽出

```
<head profile="http://purl.org/NET/erdf/profile">
 <link rel="schema.dc" href="http://purl.org/dc/elements/1.1/" />
 ...
 <p><span class="dc.date">2006-11-21</span> by
  <span class="dc.creator">神崎正英</span>…</p>
```

  - 詳しくはEmbedded RDF Wiki[5]参照

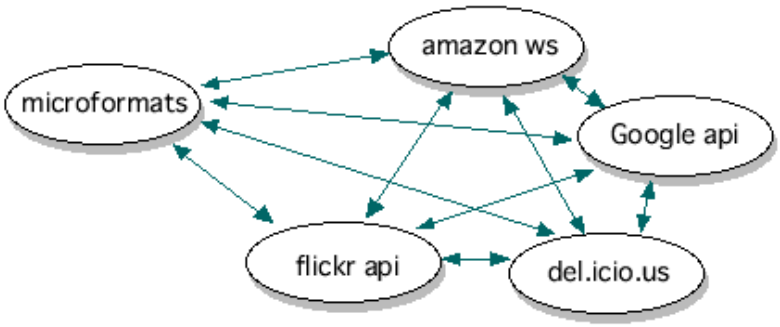
- **より手軽なプロファイル**
  - Embedded RDFは汎用性が高く現在のXHTMLでそのまま利用できるが、スキーマ指定などはまだ煩雑
  - created、topicなど、頻繁に使われる(と思われる)クラスをあらかじめDC、FOAFにマップしたプロファイルXHTML metainformation profile[6]を使うのも一案

GRDDLの利用

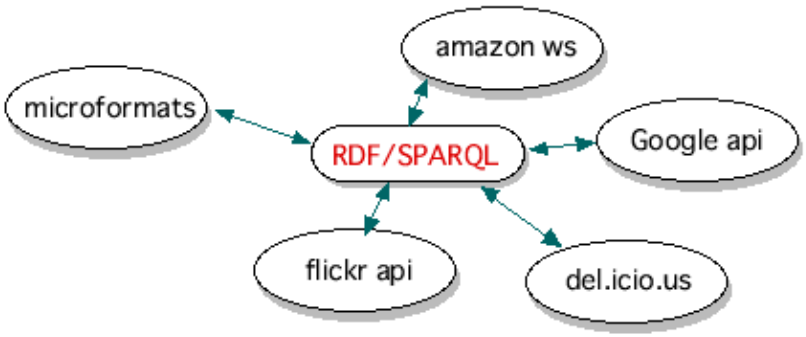
- **GRDDL経由でRDFを取り出す**
  - GRDDL demo at W3C[7]
  - Raptor RDF parser[8] がGRDDLパースを実装
  - Tabulator、Piggy Bankも（部分的に）GRDDLをサポート
- **microformats変換のためのXSLT**
  - ESW Wiki MicroModels[9] のリストなど
- **肝心なのは規則的な記述**
  - XPathでパターンが指定できれば、クラス名がなくても抽出可能
  - XHTMLでなくても、たとえばxsltprocの `--html`オプションなどで抽出できる
  - 厄介なのは意図しない不正な実体参照（特にリンクURIの`&`）

テーマ2：個別APIはスケールさせにくい

- **N対N問題とSPARQL**
  - N個のサービスを利用するにはN個のプログラムが必要→スケーラビリティがない



- **データのハブとしてのRDF**
  - RDFを共通項とすることでN対1に



RDFグラフへのクエリSPARQL

16

- **“データのウェブ”のクエリ**
  - クエリ言語：SPARQL Query Language[10]
  - プロトコル：SPARQL Protocol[11]
  - 結果フォーマット：SPARQL Results XML Format[12]、SPARQL Results in JSON[13]
- **WHERE句にグラフパターンを用いるSQL**

```
PREFIX rss: <http://purl.org/rss/1.0/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
SELECT ?item ?title
WHERE {
  ?item dc:date ?date .
  ?item rss:title ?title .
  FILTER regex(str(?date), "^2006-11-21")
}
```

ウェブの共通クエリとしてのSPARQL

17

- **RDFデータへのクエリ**
  - RSS、FOAFなどのグラフを直接検索
- **XHTML内のデータへのクエリ**
  - GRDDLを介した検索ができる
  - RedlandのRasqal RDF Query Library[14]は、GRDDLを利用してSPARQLでXHTML文書を直接検索可能
- **ウェブサービスのDBへのクエリ**
  - SPARQLとSQLのマッピングを利用して既存DBにも共通クエリを
    - 例：マッピング言語D2RQ[15]を用いてSPARQLからRDBにアクセスできるD2R Server[16]
    - 例：Eric Prud'hommeauxによる**SPASQL**（Notes on Adding SPARQL to MySQL[17]）
  - 個々のサービスはデータをRDFに置き換える必要はなく、マッピングさえあればよい

テーマ3：フォークソノミー・タグとオントロジー

18

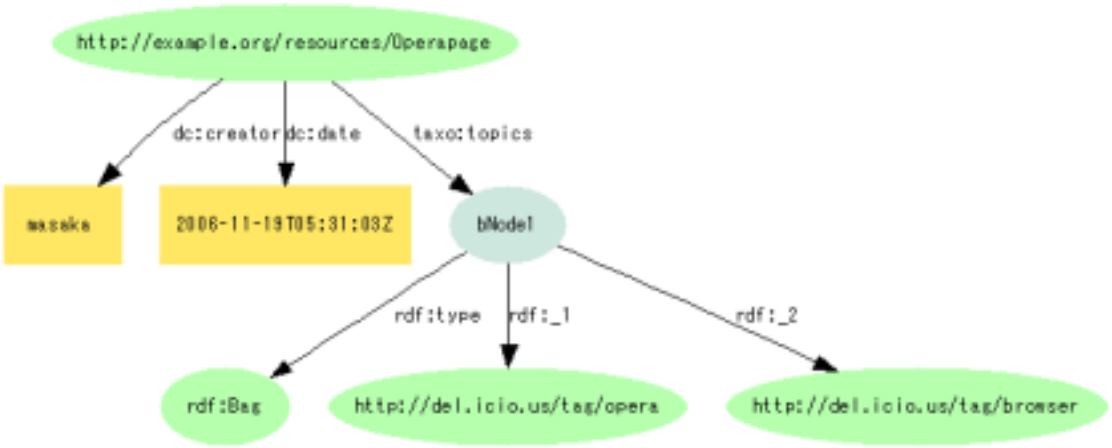
- **タグは現実的で有益**
  - 実用レベルの精度、再現率の検索ができる
  - 多くのタグはURIすら持っている！（ex. http://del.icio.us/tag/opera）
- **だが、任意のテキストラベルでは**
  - 言葉のゆれやあいまいさ、多義語を解決しない
  - システムを越えた共有が（一般に）できない
- **タグを統計的に処理する現実解**
  - タグをクラスタリングして同義語などをまとめることはある程度可能
    - 多くの研究があり、タグ・システムでも（部分的に）導入されている
  - →できればこれを論理処理に結び付けたい

タグの論理処理の第一歩：データモデル

19

- **既存のタグシステムによるRDF**
  - del.icio.usのRSSフィードはタグをtaxo:語彙を使って記述している

```
<item rdf:about="http://example.org/resources/Operapage">
 <dc:creator>masaka</dc:creator>
 <dc:date>2006-11-19T05:31:03Z</dc:date>
 <taxo:topics>
  <rdf:Bag>
   <rdf:li rdf:resource="http://del.icio.us/tag/opera"/>
   <rdf:li rdf:resource="http://del.icio.us/tag/browser"/>
  </rdf:Bag>
 </taxo:topics>
</item>
```



  - 同じリソース（文書）に複数の人がタグを付けているとき、itemを併合したときに、どのタグセットが誰によるものか識別できなくなる

次の一歩：タグ・オントロジー

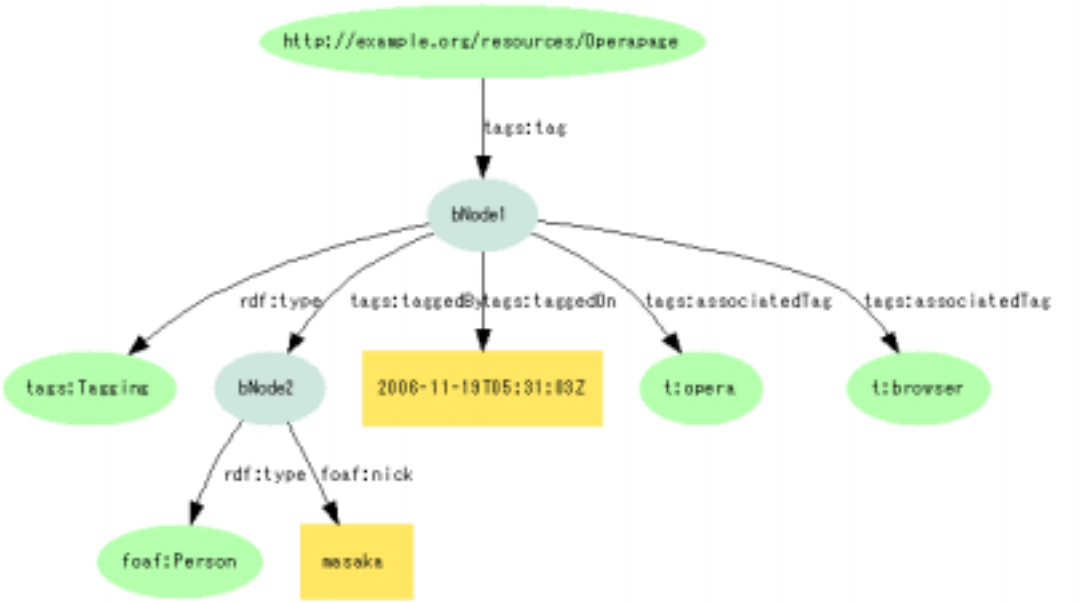
20

- **Tom GruberのTagOntology**
  - `Tagging(Object, Tag, Tagger, Source [, +/-])`
  - Ontology of Folksonomy: A Mash-up of Apples and Oranges[18]
- **Richard NewmanらのTag Ontology**
  - `Tagged(Resource, Tagging(Tag, Agent, DateTime))`
  - Tag ontology design[19]
- **基本要素はリソース、タグ、人**
  - ひとつのタグ付け（タグセット）に対して文書、人（エージェント）を結びつける
  - Gruberはソース（タグシステム）を与えているが、タグ自身をURIとして記述すれば吸収できそう
  - Newmanのモデルは、「ある人にとってのあるタグの意味は変化しうる」ことを前提に、タグ付けをプロセスもしくはイベントと捉えて日付を加える
  - タグとタグ文字列は区別して扱う
    - タグはタグラベル（タグ文字列）を持つ。もしくは文字列からの関数として表現される。
    - これによって大小文字、単複数、スペースなどの表記のゆれを吸収する。（多国語対応も可能？）
    - タグそのものは概念を直接表現しない（Operaというタグが何を表すかは人によって異なる）

タグ・オントロジーのグラフ

21

- **Newmanのモデルをグラフで表現すると**



  - 日付、タグ付与者がタグセット（Taggingクラスのインスタンス）に結び付けられる

タグのグループ化

22

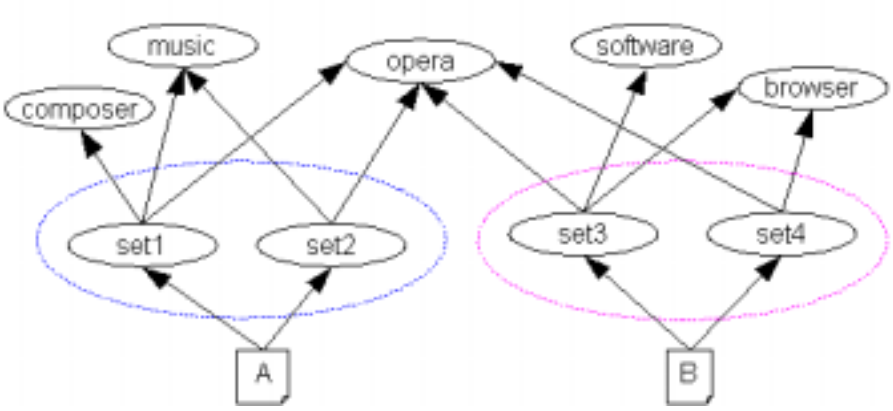
- **クラスタリングと検索**
  - タグの使われ方（主として共起）のパターンから同義語、類義語のクラスタを見出す
  - 一般に検索の再現率を高めることが可能
- **タグ・クラスタ経由のクエリ**
  - タグのクラスタをグラフに加え、（検索文字列をラベルとして持つ）タグの属するクラスタを探し、そのクラスタに含まれるタグの検索を行う

```
SELECT ?resource ?tag
WHERE {
  ?cluster t:includes ?a_tag .
  ?a_tag rdfs:label "検索語句" .
  ?cluster t:includes ?tag .
  ?resource t:tagged ?tagset .
  ?tagset t:associatedTag ?tag .
}
```

タグのグループと曖昧さ

23

- **タグそのもののクラスタは曖昧さに対応しにくい**
  - operaをbrowser, softwareなどとクラスタ化した場合：
    - Operaブラウザの検索は可能だが精度は低い（音楽もヒットする）
    - 歌劇をoperaのクラスタで検索しても、再現率、精度ともに低い
- **タグセットのクラスタ**
  - タグごとに共起関係によって「タグセット」をクラスタリングする
  - Flickrのように、タグごとにクラスタを見せてユーザに選択させるなど
    - 注目するタグが変わるとクラスタも変化する



タグとWiki

24

- **Wikitag**
  - タグのセマンティクスを利用者がWikiで定義、討議、変遷を記録
  - たとえばDave BeckettによるSemantics Through the Tag[20]、GruberのTagCommons[21]

```
<a href="http://en.wikitag.org/wiki/opera" rel="tag">オペラ</a>
```

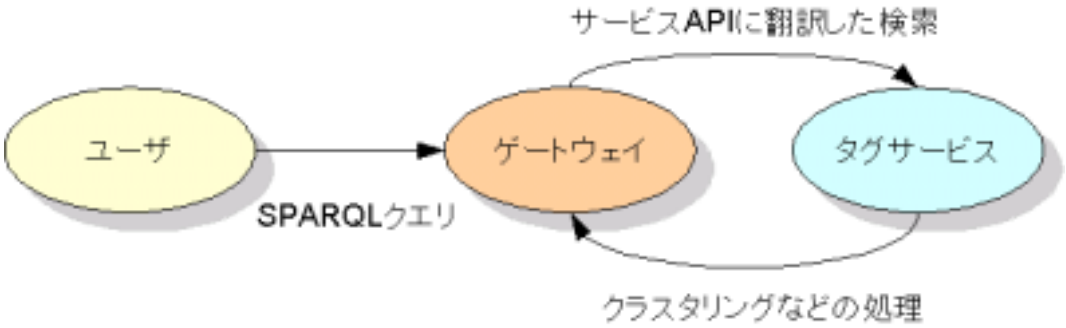
- **Wikipediaの利用**
  - 現実には、新たに何万ものタグをWikiで登録、討議するのは容易ではない
  - Wikipediaが持つ機能をタグの定義、共有に利用できないか
    - disambiguationを利用した曖昧さの解決
    - カテゴリを用いたグループ化
  - Semantic Wikipedia[22]のように、MediaWikiのプラグインを用いることでWikipediaと連動させられるか

タグのグローバルな利用のためのサービス

25

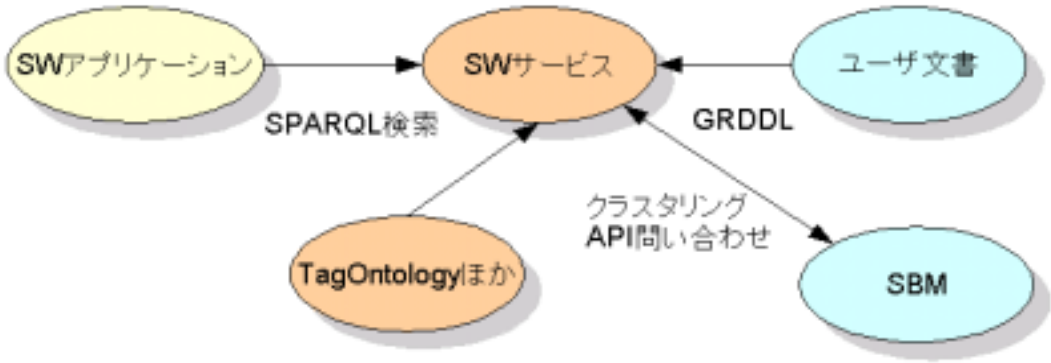
- **タグ・オントロジーの現実性**
  - タグ対応ウェブサービスが共通のタグオントロジーやクラスタリングモデルを用意するとは限らない
  - タグオントロジーのモデルによる検索はユーザにとっても複雑

- **ゲートウェイサービス**
  - ウェブサービスのタグの統計処理、オントロジー化などを提供するゲートウェイサービスもありうる
  - ゲートウェイとWikipediaなどを連動させることで曖昧さの解決も視野に入る
  - ユーザからは共通のSPARQLクエリを受け取り、それぞれのサービスのAPIに翻訳して結果を取り出す



W20からSWへの道筋

- **現実的に**
  - 既存の仕組みやデータをできる限り利用する
  - 大きな跳躍を求めず、段階的（漸進的）に
  - できれば利用者にメリットがよく分かるように
- **橋渡しをする**
  - 既存のデータをできるだけURIを使ってグローバル化する（ex. GRDDL）
  - できることならオントロジーと結び付けて体系化する。ニーズがあればサービス化する（ex.タグゲートウェイ）
  - RDBをRDFにマッピングして共通のクエリ（SPARQL）で検索できるようにする。RDFへのマッピングは自前でもゲートウェイでもよい。



参照先

- **このスライド**
  - <http://www.kanzaki.com/works/2006/pub/1121swo.html>

- **参照したページ**

1. What Is Web 2.0
   <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
2. Gleaning Resource Descriptions from Dialects of Languages
   <http://www.w3.org/TR/grddl/>
3. GRDDL Primer
   <http://www.w3.org/TR/grddl-primer/>
4. GRDDL Usecases
   <http://www.w3.org/TR/grddl-scenarios/>
5. Embedded RDF Wiki
   <http://research.talis.com/2005/erdf/wiki>
6. XHTML metainformation profile
   <http://purl.org/net/ns/metaprof>
7. GRDDL demo at W3C
   <http://www.w3.org/2003/11/rdf-in-xhtml-demo>
8. Raptor RDF parser
   <http://librdf.org/raptor/>
9. ESW Wiki MicroModels
   <http://esw.w3.org/topic/MicroModels>
10. SPARQL Query Language
    <http://www.w3.org/TR/rdf-sparql-query/>
11. SPARQL Protocol
    <http://www.w3.org/TR/rdf-sparql-protocol/>
12. SPARQL Results XML Format
    <http://www.w3.org/TR/rdf-sparql-XMLres/>
13. SPARQL Results in JSON
    <http://www.w3.org/2001/sw/DataAccess/json-sparql/>
14. Rasqal RDF Query Library
    <http://librdf.org/rasqal/>
15. D2RQ
    <http://sites.wiwiss.fu-berlin.de/suhl/bizer/D2RQ/spec/>
16. D2R Server
    <http://sites.wiwiss.fu-berlin.de/suhl/bizer/d2r-server/>
17. Notes on Adding SPARQL to MySQL
    <http://www.w3.org/2005/05/22-SPARQL-MySQL/>
18. Ontology of Folksonomy: A Mash-up of Apples and Oranges
    <http://tomgruber.org/writing/ontology-of-folksonomy.htm>
19. Tag ontology design
    <http://www.holygoat.co.uk/projects/tags/>
20. Semantics Through the Tag
    <http://www.dajobe.org/talks/200605-semantics-tag/>
21. TagCommons
    <http://www.tagcommons.org/>
22. Semantic Wikipedia
    <http://wiki.ontoworld.org/>