

# 情報技術演習

## 第6回 「情報流通」(2)

2006/11/7

久保田秀和

文学部／情報学研究科

kubota@ii.ist.i.kyoto-u.ac.jp

<http://www.ii.ist.i.kyoto-u.ac.jp/~kubota/>

# 本日の講義・演習

- 情報流通(1)フォーカソノミー
  - 前回提出されたレポートを題材に
- 情報流通(2)メタ情報
  - セマンティックWeb
- 中間報告会

# 情報流通(1) フォークソノミー (前回の復習)

- 課題
  - あるコンテンツに対してフォークソノミーに基づいて付与されたタグと、自然言語処理的な手法によって求まる重要語とを比較し、その違いについて論じよ。

# 比較の例(1)

- YM, 「ライフログ」を調査
  - 重要語抽出
    - ライフログ, 携帯, KDDI, バーコード, ブログ
    - 「記事の要約」
    - 「何だろう? という興味を引く要素は薄くなっている」
  - フォークソノミー
    - 携帯, **ネタ**, gps, ライフハック, ログ, lifelog, **ケータイ**, blog
    - 「見た人の印象がより濃く投影されている。」
    - 「ケータイのようなくだけた表現」

## 比較の例(2)

- YH(i), 「PR会社」を調査
  - 重要語抽出
    - メディア, 株式会社, マーケティング, コンサルティング, 会社, 仕事, 広報, ジャパン, 情報
    - 「会社名や仕事内容といった具体的なコンテンツ」
  - フォークソノミー
    - 政治, 気になる, メディアリテラシー, ジャーナリズム, PR会社, 情報操作, 愚民化計画, 広告代理店, マーケティング, マスコミmedia, ネタ, 政治, 危機管理
    - 「PR会社と私たちの社会の関わり方」「より抽象的」

# 比較の例(3)

- YH(a), 「asahi.com」を調査
  - 重要語抽出(言選 <http://gensen.dl.itc.u-tokyo.ac.jp/gensenweb.html> を利用)
    - 朝日新聞社, 日本, 社会, 記事, 特集, 等
    - 「オンラインでニュースを配信するニュースサイトであることは多少分かりにくい」
  - フォーカソノミー
    - news, 新聞, **いつも見る**, ニュースサイト, ニュース
    - 「サイトのコンテンツ自体の情報ではなく、ユーザの中でそのサイトが占めている位置についての情報」

# フォークソノミーとメタ情報

- 「メタ情報」 一般に……
  - 情報それ自身について記述する情報
  - より抽象化された情報(≒意味)
- フォークソノミーを用いてつけられたタグ(メタ情報)
  - ある情報について、個人による意味づけや印象を記述した情報
  - フラットな連想のネットワーク(抽象化も部分的に含む)
  - とくに文字以外の情報(写真、音楽、映像など)に対するタグは、情報を言語的に分類あるいは検索する上で有用
    - <http://www.flickr.com/>
  - 必ずしもタグとして記述する必要はなく、人が書いた感想文などから自然言語処理を用いて抽出することも可能

# 本日の講義・演習

- 情報流通(1)フォーカソノミー
  - 前回提出されたレポートを題材に
- 情報流通(2)メタ情報
  - セマンティックWeb
- 中間報告会

# セマンティックWeb

- メタ情報を記述, 整理, 活用する手法の歴史
  - Webが生まれ, 発展した歴史と重なる
- Webの目的(「カッコ」内はTim Berners-Lee, “Webの創成”より)
  - 「人々の共同作業の手助け」
  - 「世界中に散らばっている私たちが織りなしている網の目のような存在を支援し, 改善すること」
    - 家族, 友人の輪, 会社などへの帰属と, 信頼, 推薦, 同意のような社会的相互作用から成る網の目
  - セマンティックWeb(1998頃-) <http://www.w3.org/2001/sw/>
    - 計算機を用いた支援を可能とするための, 意味論的(Semantic) Web
    - 意味論的?
      - 「マシン処理可能な相対的な“meaning(意味)”」
      - 相対的に計算機可読で, 計算機による推論や判断を可能とするメタ情報としての意味と, 意味処理のための枠組み

# SGMLとHTML

- SGML(Standard Generalized Markup Language)
  - 1986年, ISO(国際標準化機構)
  - 文書の構造を記述するためのメタマークアップ言語
  - アメリカ国防総省の公文書, CERN (欧洲原子核研究機構)における文書, 航空機産業におけるマニュアル等
- HTML(HyperText Markup Language)
  - WWWにおいて情報を頒布するためのマークアップ言語
  - 1990頃, Tim Berners-Lee (CERN) . 文法はSGMLを参考.
  - 1993年, IETF Internet Draft
  - 1999年, W3C Recommendation, HTML4.01

# XML

- XML(Extensive Markup Language)
  - SGMLのサブセット
    - 複雑すぎて実装が難しい、あるいは冗長な機能のそぎ落とし
  - 1996年, W3C Working Draft
  - 1998年, W3C Recommendation, XML1.0

## 練習: XML文書の例(1)

- RSS(RDF site summary)
  - Webサイトの見出しや要約, 構造などを表現するメタ情報
  - 例) 政府インターネットテレビより
    - <http://nettv.gov-online.go.jp/>
    -  RSS をクリック

# サイトの概要やヘッドライン, 日付, リンクなどが記述されている

```
<?xml version="1.0" encoding="UTF-8" ?>
- <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:sy="http://purl.org/rss/modules/syndication/" xml:lang="ja">
- <channel rdf:about="http://nettv.gov-online.go.jp/">
  <title>政府インターネットテレビ</title>
  <link>http://nettv.gov-online.go.jp/index.html</link>
  <description>総理の一週間・政府からのお知らせを番組形式で配信</description>
  <language>ja</language>
  <dc:date>2006-11-07T01:00:02+09:00</dc:date>
- <items>
  - <rdf:Seq>
    <rdf:li rdf:resource="http://nettv.gov-online.go.jp/mag_mov.php?n=813" />
  </rdf:Seq>
</items>
</channel>
- <item rdf:about="http://nettv.gov-online.go.jp/mag_mov.php?n=813">
  <title>平成18年11月2日(木)午前-内閣官房長官記者会見-</title>
  <link>http://nettv.gov-online.go.jp/mag_mov.php?n=813</link>
  <description>閣議の概要について経済連携促進に関する主要6閣僚による打ち合わせについて</description>
  <category>官房長官記者会見</category>
  <dc:subject />
  <dc:creator />
  <dc:date>2006-11-02T11:00:00+09:00</dc:date>
</item>
</rdf:RDF>
```

IE6上の表示例

(各種RSSリーダーでブラウズ可能)

## 練習: XML文書の例(2)

- FOAF(Friend of a Friend)
  - 自己紹介, 知人関係を記述
  - 例)はてなのプロフィール欄より(ただし知人関係の記述はなし)
    - プロフィール:  
<http://d.hatena.ne.jp/?????/about>
    - FOAF:  
<http://d.hatena.ne.jp/?????/foaf>  
(はてな社長のjkondo氏など, 有名人, 知人で試してみてください)

# HTMLから見たXML

- “Extensive”
  - HTMLはマークアップ言語, XMLはメタマークアップ言語
- 誰もが独自のタグないしはマークアップ言語を制作できる
  - XML Application Language(XML応用言語)
    - 先に見たRSS, FOAFもその一例
  - HTMLのタグセットで表現できない情報(カレンダー, アドレス帳, 銀行取引明細など)をHTMLに変換する際に起こる情報の紛失を回避
- XHTML
  - XMLに適合するよう定義しなおしたHTML

# Webの思想

- 非集権化
  - 局所的な合意の集合と、合意間の関係記述
  - 実践のなかで漸進的に合意形成
    - IETF (Internet Engineering Task Force) における、“Rough consensus and running code” という思想の影響
    - 「部分的理解」に基づいて活動するということが基本であり、同様のことを私たちは非電子的な日常の世界で行っている (Tim Berners-Lee, “Webの創成”より)
- XML名前空間、RDFスキーマによる実現

# XML名前空間

- それぞれ独自に作られたXML応用言語を区別するための命名法

```
<?xml version="1.0" encoding="UTF-8" ?>
- <rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-
    rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:sy="http://purl.org/rss/modules/syndication/
  xml:lang="ja">
- <channel rdf:about="http://nettv.gov-
    online.go.jp/">
  <title>政府インターネットテレビ </title>
  <link>http://nettv.gov-
    online.go.jp/index.html</link>
  <description>総理の一週間・政府からのお知らせ
    を番組形式で配信 </description>
  <language>ja</language>
  <dc:date>2006-11-
    07T01:00:02+09:00</dc:date>
```

xmlns:rdf

…rdfという名前空間

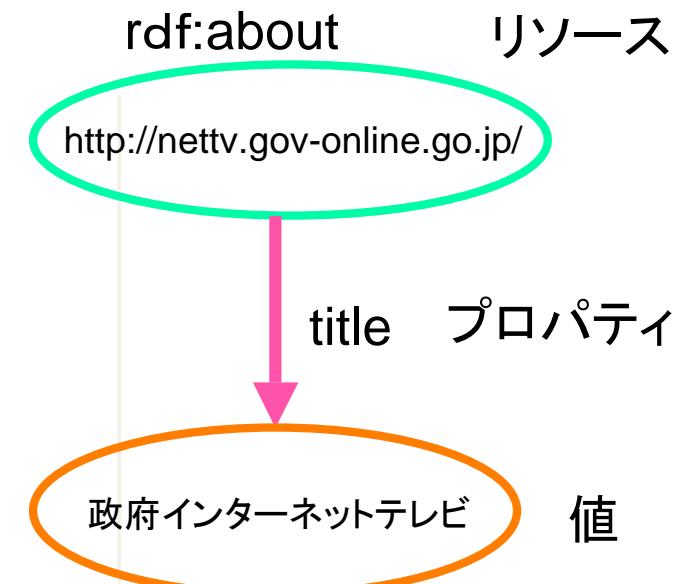
<http://www.w3.org/1999/02/22-rdf-syntax-ns#> というURIによって一意に識別可能

rdfという名前空間におけるaboutという語彙

# RDF (Resource Description Framework)

- セマンティックWebの枠組みにおいて、意味を記述するための規約
- 「リソース」「プロパティ」「値」から構成されるステートメント(文)を用いて記述

```
- <channel rdf:about="http://nettv.gov-online.go.jp/">
  <title>政府インターネットテレビ</title>
  <link>http://nettv.gov-online.go.jp/index.html</link>
  <description>総理の一週間・政府からのお知らせを番組形式で配信</description>
  <language>ja</language>
  <dc:date>2006-11-07T01:00:02+09:00</dc:date>
```



# RDFスキーマ

- RDFの語彙を定義する基本的な仕組み
  - 型(人名は文字列である, 年齢は数字である)
  - 分類(鳥は動物に分類される, トラックは自動車に分類される)
- など
- 単一の権威的な定義付けではない
  - 個々のサイトがそれぞれに異なる独自の「自動車」についての定義を行うことを許す(Tim Berners-Lee, “Webの創成”より)
  - 個々のRDFスキーマのリンクとオントロジーを用いた推論により, 計算機的に共通の理解を作り上げてゆく

# フォークソノミーとセマンティックWeb

- フォークソノミーによるタグ
  - 暗黙的に決定, ゆるやかな定義
  - 推論処理は連想的, 統計的(クラスタリングなど)
  - フラットに形成される集合知(Web2.0的)
- RDF／オントロジー
  - 明示的に決定, 明確な定義
  - 推論処理は論理的
  - セマンティックWebにおけるオントロジーは, 集合知ではないが, 古典的な中央集権的オントロジー(Cycなど)でもない. 興味あるコミュニティにおいて局所的に定義され, 推論処理に基づいて, 必要に応じて別の場所で再利用, 拡張, 共有可能とした知識.
- 参考: Nigel Shadbolt, Wendy Hall, Tim Berners-Lee, “The Semantic Web Revisited”, IEEE INTELLIGENT SYSTEMS 21(3) pp. 96-101, 2006. (<http://eprints.ecs.soton.ac.uk/12614/>)

# 参考文献

- Tim Berners-Lee, “Webの創成 World Wide Web  
はいかにして生まれどこに向かうのか”, 毎日コミュニケーションズ
- Elliotte Rusty Harold, “XMLクイックリファレンス”, オライリー・ジャパン
- 萩野ら, “特集 セマンティックWeb”, 情報処理, Vol. 43, No. 7, 2002.