セマンティックWebの課題と研究の方向性
～セマンティックメタデータの観点から～

（財）情報処理相互運用技術協会　セマンティックWeb委員会委員
日本電信電話株式会社　ＮＴＴ情報流通プラットフォーム研究所
佐藤宏之

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

セマンティックWebの現状

- リソースの意味，知識表現（セマンティックメタデータ）のXMLによる記述形式の標準化が進んだ
- 基本的なツール（エディタ，DBなど）が揃った
- 仕様に基づいたボキャブラリなどが公開され始めた
- さあ本当に使える？
  - キラーアプリはまだ？
  - 高度な検索，Webサービス，EAI，電子政府，ナレッジマネジメント，コミュニティ，なんでもあり？

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

リソースの意味，知識表現のための仕様

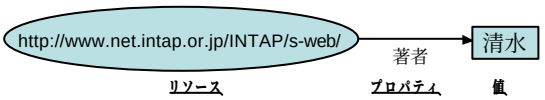
- モデルと文法
  - RDF
- ボキャブラリの定義方法（ボキャブラリの階層など）
  - RDF Schema
- ボキャブラリの概念定義方法（概念間の階層，同等性，非重複性，制約，集合などの関係）
  - DAML+OIL
  - OWL (Web Ontology Language)

知識表現へ

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

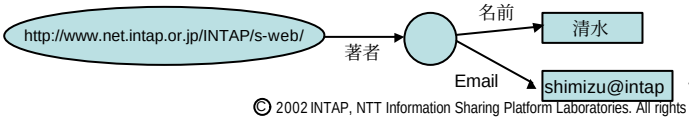
Resource Description Framework (RDF)

- リソース，プロパティ，値の三つ組によるメタデータ，リソース間の関係表現のモデル
  - 有向グラフによる表現が可能
- W3CでXMLによる記述形式が標準化




```
<?xml version="1.0"?>
<rdf:RDF  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
  <rdf:Description  rdf:about="http://www.net.intap.or.jp/INTAP/s-web/">
    <著者>清水</著者>
  </rdf:Description>
</rdf:RDF>
```

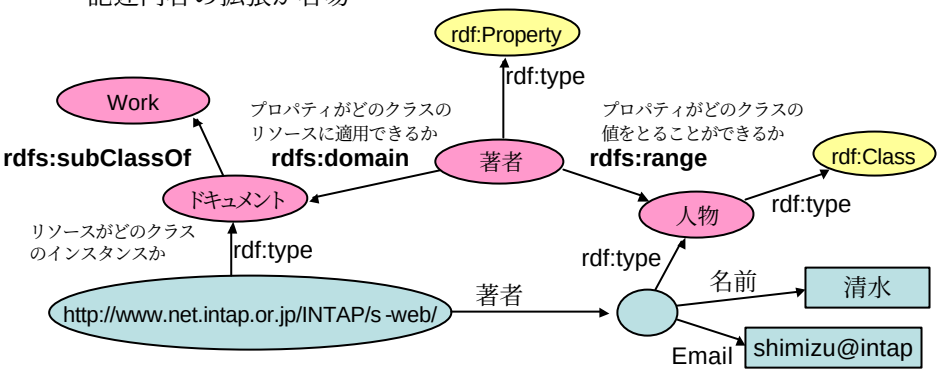
右のような表現も可能



© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

RDF Vocabulary Description Language: RDF Schema

- オブジェクト指向のクラスの考えを利用
- リソースのクラスとクラスの階層関係を定義
- プロパティの階層関係を定義
- プロパティを適用できるリソースやプロパティがとり得る値の制約を定義
- 記述内容の拡張が容易



OWL Web Ontology Language

- オントロジ記述言語
- 概念間の関係記述ができる
  - 記述できる例
    - 同等性 sameClassAs
      - 「人物」と「人間」は同じ概念であることなどを表現できる
    - 非重複性 disjointWith
      - 「男性」と「女性」が互いに素であることなどを表現できる
    - 集合 unionOf
      - 「男性」と「女性」の集合を表現できる．（これを利用して，その集合によって「人間」を定義するといった記述ができる．）
    - この他，階層，制約など
- 利用者やツール構築者の利便性を考慮してOWL Lite (OWLのサブセット)を規定

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

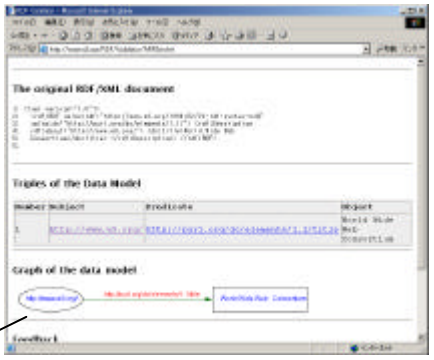
セマンティックWebのツール

- さまざまな組織がプロトタイプレベルだが実用的なツールを公開
  - RDFのパーサ
  - RDFのDB
    - リポジトリ，クエリープロセッサ
  - メタデータ（RDF）エディタ
  - オントロジエディタ
- 次ページ以降に代表的なものを紹介
  - ツールについては，森田，津田，清水，布目，来間，佐藤，「セマンティックWebのツール」，情報処理Vol.43 No.7 July 2002のセマンティックWeb特集に解説記事
    - http://www.net.intap.or.jp/INTAP/s-web/data/ipsj_vol43-no7/ipsj4.pdf からも参照可能

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

RDF Validation Service

- W3CのWebページから利用可能
  - http://www.w3.org/RDF/Validator/
- 直接入力もしくは指定したURIにあるRDFファイルを解析し結果を三つ組のリストやグラフで表示
- 内部で以下のツールを利用
  - パーサ：HP-Labs in BristolのAnother RDF Parser (ARP)
  - ビジュアライゼーション：AT&T LabsのGraphViz



© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

RDFのデータベース

- 基本的には，既存のXMLのデータベース管理システムの流用が可能
  - 要素などをRDBにマッピングしても良い
  - DOMのインタフェースを用いてツリー構造をそのままバイナリ化する方が高速性が期待できるという点などはいっしょ
- 三つ組（トリプル）の検索を効率よくできるクエリー言語とともに提供されることが求められる

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

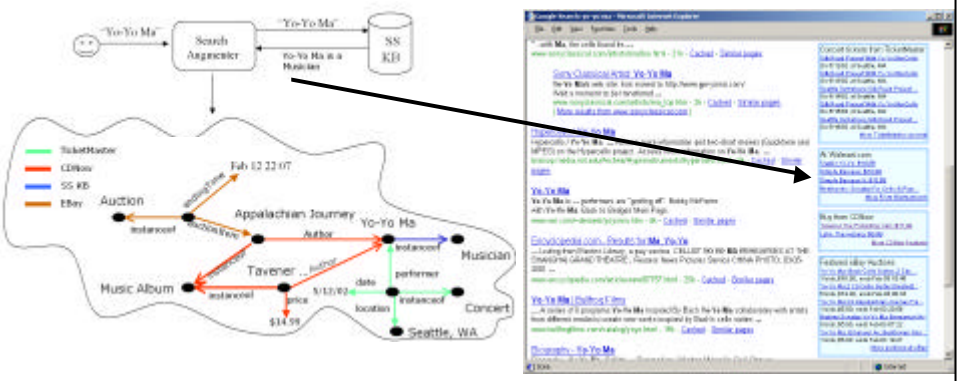
利用可能なRDFのDB

- HP Labsの**Jena**
  - http://www.hpl.hp.com/semweb/
  - JavaのRDF APIとtoolkit
  - persistent storage
  - クエリー言語としてRDQLを準備
  - オープンソースでBSDのようなライセンスで利用可能
- ICS-FORTHの**RDFSuite**
  - http://139.91.183.30:9090/RDF/
  - Validating RDF Parser (VRP)を備える
  - RDF Schema Specific DataBase (RSSDB)
  - クエリー言語としてRDF Query Language (RQL)
  - オープンソース
- R.V.Guhaの**rdfDB**
  - http://www.guha.com/rdfdb/
  - RDFのデータモデルを直接サポートするデータベース
  - SQLのようなクエリー言語をサポート

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

TAPのSemantic Search

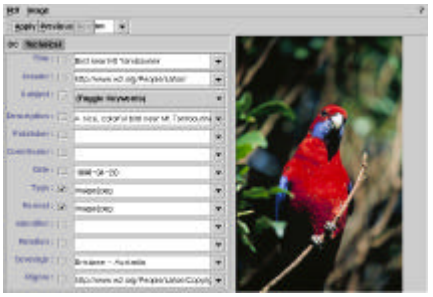
- http://tap.stanford.edu/tap/ss.html
- メタデータがRDFのデータモデルでTAP Knowledge Baseに格納されている
- “Yo-Yo Ma”という文字列で検索をかけると，それがミュージシャンであることがわかり， TicketMasterのデータからコンサートのスケジュール， EBayのデータから彼の音楽アルバムがオークションにかけられていること， CDNow のデータから彼の音楽アルバムが$14.99 で売られていることなどが分かるという例



TAP: Towards the Semantic Web http://tap.stanford.edu/www2002.ppt より

RDFPic

- 写真のメタデータ登録ツール
- RDFのメタデータ記述をデジタル写真のイメージそのものに埋め込むことが可能

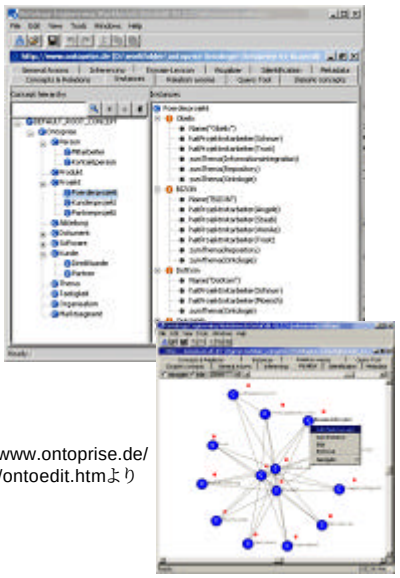


http://jigsaw.w3.org/rdfpic/より

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

オントロジエディタ

- OntopriseのOntoEdit
  - カールスルーエ大のAIFBが開発
  - 階層，背反などの概念の編集などが可能
  - RDFやDAML+OILの入出力が可能
  - 述語論理の処理系が組みこまれ，推論規則に従った記述の導出が可能
- スタンフォード大のProtégé
  - http://protege.stanford.edu/
  - RDFやRDF Schemaを最近はサポート



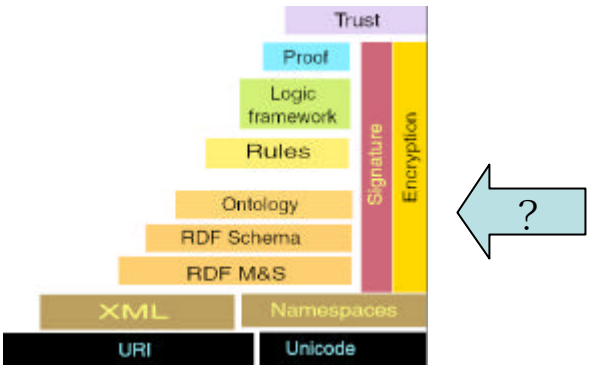
http://www.ontoprise.de/com/ontoedit.htmより

ボキャブラリをRDFの規定に従って定義する試みがあるもの

- Dublin Core
  - http://dublincore.org/
  - 情報資源の１５の基本的なメタデータ要素（Title，Creator，Subjectなど）を規定
- RDF Site Summary (RSS)
  - http://purl.org/rss/1.0/spec
  - Web コンテンツ（特にニュース，日記など）の同時配信と集約のためのメタデータ（要約，公表日など）を規定
- MusicBrainz
  - http://www.musicbrainz.org/MM/
  - デジタルオーディオやビデオトラックのメタデータ
- W3Cのデジタル写真のメタデータ
  - http://www.w3.org/TR/photo-rdf/
  - どんな機材を使ったか，何が写っているかなどを規定
- vCard
  - http://www.w3.org/TR/vcard-rdf
  - ビジネスカードのプロファイル
- Annotea
  - http://www.w3.org/2001/Annotea/
  - Webドキュメントへの注釈
- eXtensible Metadata Platform (XMP)
  - Adobeのアプリケーションが生成したファイルにはRDFベースのメタデータが含まれている

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

現状の再整理：どこまでできているのか



Enabling Standards & Technologies - Layer Cake

2002-04, The Semantic Web, Tim Berners-Lee Academic discussion, Japan Prize 2002

http://www.w3.org/2002/Talks/04-sweb/slide12-0.htmlより

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

現状の再整理：
セマンティックWeb以前と変わったところ

- RDFのXMLによる記述方式の標準化がW3Cでできた
  - 今まで意味や知識は異なるコミュニティで異なった表現によって管理されていた
    - RDFの背景にAppleのMCF (Meta Content Framework)があるが，メタデータや意味表現のための仕組みはこれまでにもさまざまなものが提案されてきた
  - 今後，さまざまなコミュニティがもつ意味情報や知識を融合して扱える可能性に対する期待感
    - 相互運用のための課題の一つ目はクリア

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

現状の再整理：ユーザの環境の変化

- Webによって，知識表現の相互運用の可能性が現実のものとなった
  - URI (Uniform Resource Identifiers)によってあらゆる情報資源を示すことが可能になり，さらに誰もがインターネットを利用してその情報資源の取得が可能
  - 誰かが新たに拡張し，URIによって示したボキャブラリ定義，オントロジも簡易に取得可能

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

課題：誰がメタデータを付けるのか？

- 見返りやモチベーションなく誰が貢献するのか？
- 既にメタデータがたくさんあるならサービスは利用するけど…
- ニワトリが先か卵が先か？
  - メタデータ少ないから良いアプリが登場しない
  - アプリが使われないからメタデータが増えない

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：既存の情報資源やシステムからのメタデータの生成

- HTMLからXSLTを利用してRDFを生成したり，RDBからRDFを生成できる
- XHTMLからRSSを生成したり，Emailのリプライ状況などからRDFを生成する試みがある
  - W3CのSemantic Web Advanced Development
    - http://www.w3.org/2000/01/sw/
- Webリンク解析，ページ内容の自然言語処理からページのメタデータの自動生成
  - 津田他，Webディレクトリのためのページメタデータの自動付与の試み，情報学シンポジウム2002
    - http://www.ipsj.or.jp/katsudou/sig/sighp/fi/sympo/2002/proceedings/papers/tsuda.pdf

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：XMLによる既存のメタデータの活用

- マルチメディア
  - MPEG7
- 教育
  - LOM (Learning Object Metadata)，SCORM
- 放送
  - TV Anytime
- レーティング，フィルタリング
  - PICS (Platform for Internet Content Selection)
- ユーザプロファイル
  - P3P (Platform for Privacy Preference Project)
- 電子政府
  - e-GMS，MIReG，AGLS…

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：既存スキーマ，オントロジの活用（1）

- これまでさまざまなモデルで表現された意味，知識の資産を，RDFを中間形式として活用できないか？
  - Knowledge Interchange Format (KIF) as an RDF Schema
    - http://www.w3.org/2000/07/hs78/KIF
  - Topic MapsとRDF間の変換
    - かなり議論されている
  - XLinkで表現されたリンクの意味もRDFに変換可能
  - UMLからRDF
  - SemTalk
    - MS-Visioをベースとしたグラフィカルなエディタから RDF Schemaを生成
    - http://www2002.org/CDROM/refereed/549/

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：既存スキーマ，オントロジの活用（2）

- NetscapeのOpen Directory Project
  - http://dmoz.org/
  - ボランティア編集者によって分かり易いWebのディレクトリを生成
  - ディレクトリ階層などのデータをRDFでダウンロード可能
- Princeton大のWordNetの語彙体系をRDF化する試み
  - WordNetでは語彙間の階層関係を管理
  - Wordnet for the Web
    - http://xmlns.com/2001/08/wordnet/
  - http://www.kanzaki.com/docs/sw/jwebont.htmlに解説

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：モチベーション向上

- ユーザが自身のために行うアクティビティからメタデータを抽出させてもらう
  - MITのHaystack
    - http://haystack.lcs.mit.edu/
    - e-mailやカレンダー情報や個人が扱うドキュメントから収集したメタデータをRDFのデータモデルで管理してパーソナライゼーションなどに利用
  - NTTのHyperclipと知識流通プラットフォーム
    - ブックマークのような個人のアクティビティからセマンティックメタデータを生成（後述）
    - http://www.net.intap.or.jp/INTAP/s-web/data/ipsj_vol43-no7/ipsj5.pdf
    - http://www.cs.rutgers.edu/~shklar/www11/final_submissions/paper12.pdf
- 集客力があがるような仕掛け作り
  - 商業的なサイトのRDFによるメタデータ
- 公共的な使命として付けてもらう
  - 電子政府，行政サービス

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

課題：メタデータの収集・流通

- メタデータ収集の仕組みが確立していない
  - Webロボット？
  - メタデータのURIをサーバに登録？
- メタデータ公開の仕組みが確立していない
  - 既存のHTMLにRDFメタデータを付与する方法は示されている
    - http://www.w3.org/TR/REC-rdf-syntax/#transport
  - しかし，ほとんどの人が行なっていない

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

方向性：流通の段階的な進展

- レベル1：inB
  - 信頼できるコミュニティ内でメタデータ利用
    - 企業内部
  - ボキャブラリもある程度統一
- レベル2：+BtoB
  - 信頼できる他のコミュニティとメタデータ交換
    - 特定のパートナー会社
  - 簡単なオントロジ（階層や同義など）が必要
    - ここを1対1のXML変換でもよしとするとレベル3に進まない
- レベル3：+BtoC, CtoB, CtoC
  - 不特定のコミュニティや個人とメタデータ交換
  - メタデータはRDFによる表現が必要
    - 多対多の自由なメタデータの交換，利用を保障
    - ボキャブラリの自由な拡張を保障
  - 高度なオントロジが必要

RDFでメタデータを記述したり，世界的にコンセンサスのとれたボキャブラリ（Dublin Coreなど）を利用したりしておくと，他のシステムとの接続や将来のシステム更改が低コストとなるかの検証が必要

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

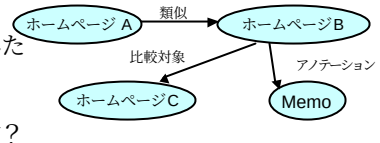
方向性：セマンティックWebサービス

- セマンティックWeb
  - 意味や知識の表現性を重視
- Webサービス
  - 今使えるもの，プロトコル，サービスのインタフェースの規定を重視
    - SOAP，WSDL
- 両者の融合はメリットがある
  - セマンティックWebではプロトコルは特に規定していない
  - Webサービスでは柔軟なサービスの発見が求められている
  - Using RDF with SOAP
    - http://www-106.ibm.com/developerworks/webservices/library/ws-soaprdf/
  - Supercharging WSDL with RDF
    - http://www-106.ibm.com/developerworks/library/ws-rdf/

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

知識流通プラットフォーム（NTTの研究）の出発点

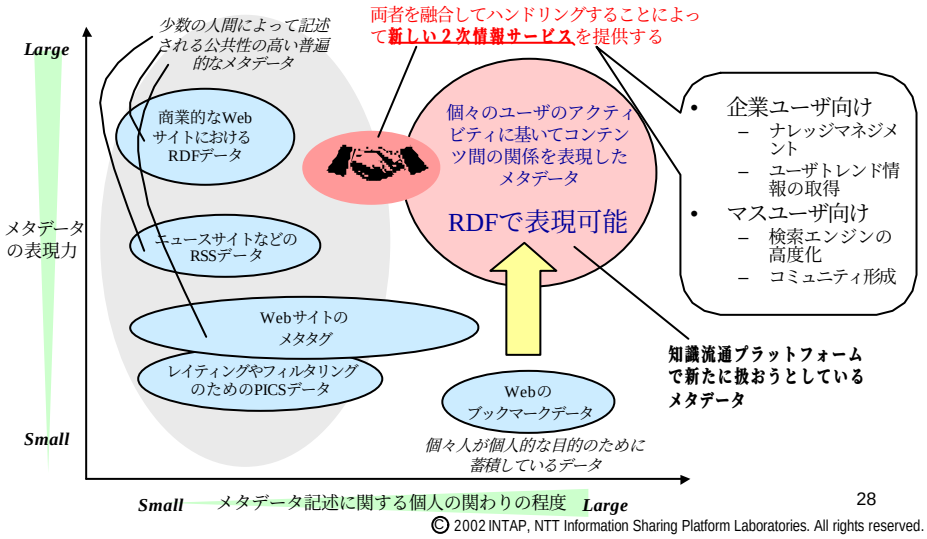
- インターネット人口の中で自ら情報発信を行う人の割合の極端な低下（商業的なサイトの氾濫とそのブラウジングで手一杯のユーザ）
  - 「インターネットする　自分のホームページを持つ」は昔の話
  - もっと多くのユーザから情報を引き出すインフラレベルのフレームワークができないか？
- 人間しか付与できないメタデータがおもしろい
  - ホームページAとページBって似てる
  - ページBの商品を買うのにページCと比較した
  - ページBの商品には言いたいことがある



- セマンティックメタデータは手軽な情報発信？
  - URL(URI)を持つリソースは既に大量に存在，それらの関係付けは誰でもできるし，それほどコストはかからない
  - フォーマットはRDFで良い
- 個々のセマンティックメタデータの集合は知識とならないか？
  - またコミュニティのきっかけとならないか？

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

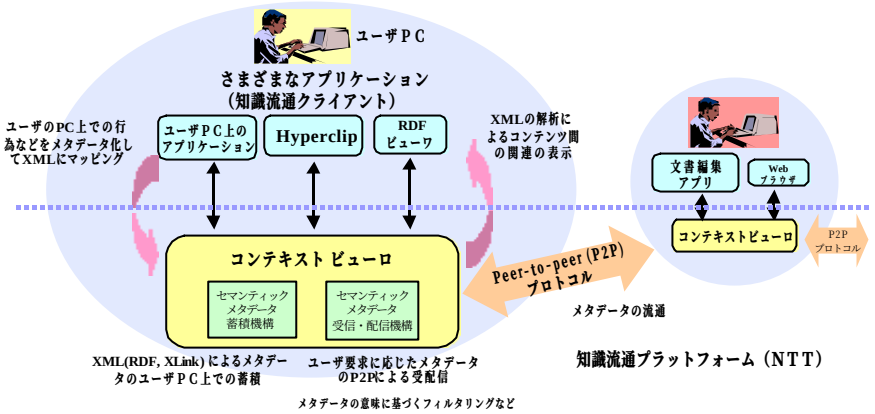
メタデータの現状と開拓したい領域



© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

知識流通プラットフォーム

- RDFなどのコンテンツ間の関係を表現するメタデータをP2Pで流通
- コンテンツに関する個々のユーザのアクティビテイに基づいたメタデータを必要に応じて共有可能



Hyperclip

知識流通プラットフォーム上のアプリの一例

- ユーザはさまざまなコンテンツをブックマークのようにクリップ．その際，コンテンツ間の意味を指定できる
  - セマンティックメタデータが生成される

- メタデータを媒介に意味的に関連するコンテンツを芋づる式に探せる
  - 他ユーザの保持するセマンティックメタデータをP2Pプロトコルで検索できる

比較対象となるコンテンツには次のようなコンテンツがあるということを，hiroyukiというリモートのユーザが個人的に蓄積しているメタデータを検索して表示した

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

課題：その他

- 信頼性
  - メタデータスパム
    - 推論ができなくなる
    - Layer CakeのTrustをどうするか？
  - セマンティックWebを用いたエージェントが提示した情報やサービスの根拠
    - Layer CakeのProofをどうするか？
  - 古いメタデータ
    - メタデータのexpire dateの設定
- オントロジ
  - オントロジでどこまで記述する必要があるか？
  - オントロジのメンテナンス

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.

おわりに

- RDFのXML表現の標準化とURIによって実現する意味や知識表現の相互運用性に対する期待感が高まっている
- 既存のリソースやユーザアクティビティなどの活用によるメタデータの生成が重要
- ボトムアップアプローチの継続が重要
- メタデータの本格的な流通に向けて，まずはビジネス領域での適用の評価が必要

© 2002 INTAP, NTT Information Sharing Platform Laboratories. All rights reserved.