

セマンティックWebにおける知識表現

武田 英明



国立情報学研究所



総合研究大学院大学



東京大学

人工物工学研究センター

<http://www-kasm.nii.ac.jp/>

takeda@nii.ac.jp

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}   

あらまし

- セマンティックWebの枠組み
- オントロジーの世界
- Linked Dataの世界
- メタデータの創造的流通

セマンティックWebの目的

- “The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation.”
(セマンティックWebとは現在のWebの拡張であり、そこでは情報はちゃんと定義された意味を与えられていてコンピュータと人のよりよい協調が可能となる)。
The Semantic Web, Scientific American, May 2001, Tim Berners-Lee, James Hendler and Ora Lassila
- The Semantic Web is a vision: the idea of having data on the web defined and linked in a way that it can be used by machines not just for display purposes, but for **automation**, **integration** and **reuse of data across various applications**.
(セマンティックWebとはビジョンである。データはきちんと定義されリンクされており、単に表示用ではなく自動化、統合、アプリケーションを超えたデータの再利用などに使える)

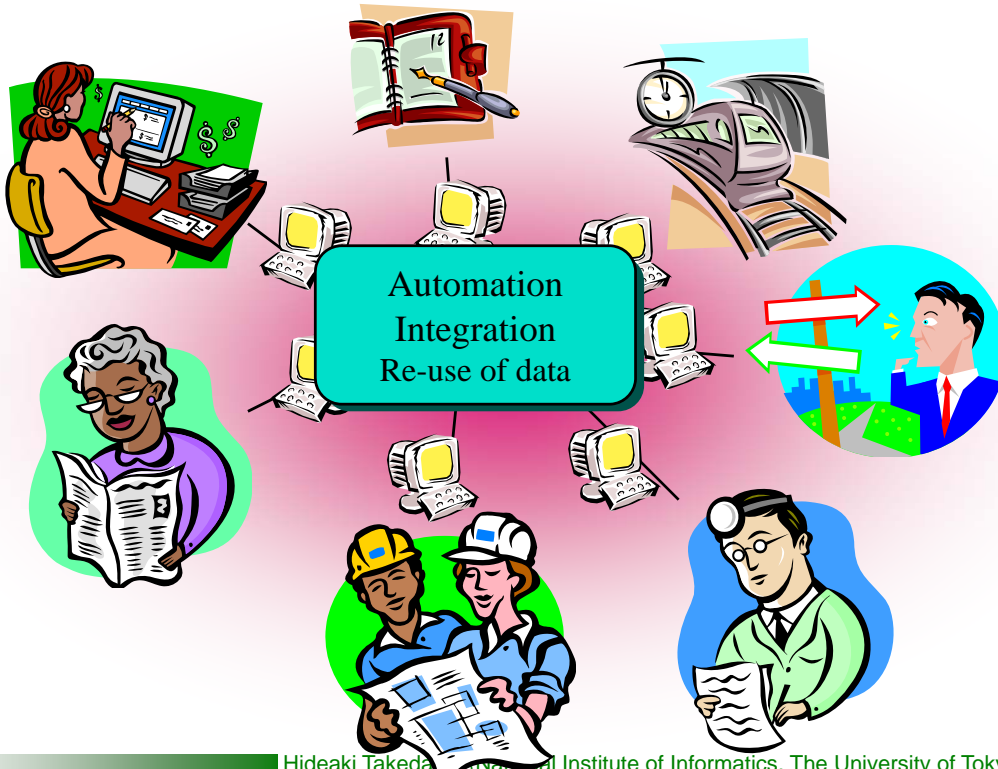
<http://www.w3.org/2001/sw/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



Semantic Web

- Realization of various information exchanging via Web



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



Next Generation Web

- Webの進化
 - HTML: 表示のためのWeb
 - XML: シンタックスをもったWeb
 - ?? : セマンティックスをもったWeb
 - なぜセマンティックスをWebのメカニズムの中に組み込めなければならないか
 - 人間のためのWeb
から
 - 人間と機械のためのWeb
へ
- cf. 機械ためだけのWeb

どうやって意味を記述するか

- 情報に関する情報を記述する仕組みが必要
 - メタデータ
 - ◆ データに関するデータ
- 共通に理解し合える仕組みが必要
 - シンタックス (言語)
 - 語彙 (オントロジー, クラス)
 - データ (インスタンス)

メタデータ

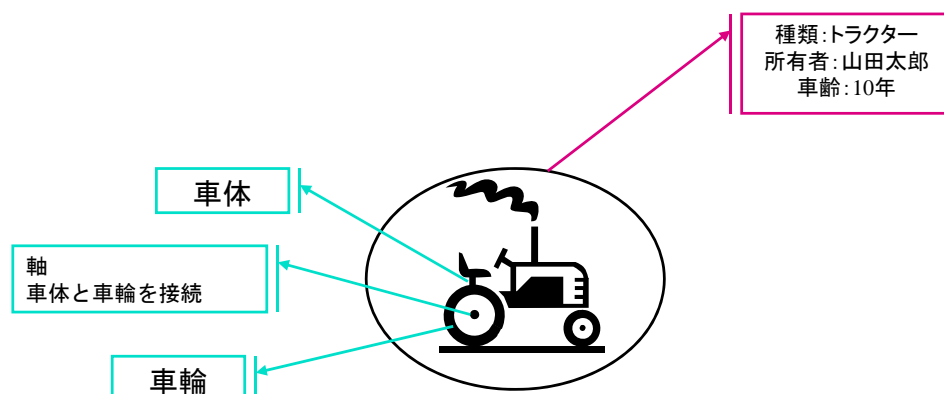
- メタデータとは
 - データに関するデータ
 - さまざまなまとまりとしての情報オブジェクトに関わる言明の集まり
- メタデータの中身は
 - 内容：オブジェクト自身にかかわる情報. 内在性
 - コンテキスト：オブジェクトの生成などに関わる人, 場所, 理由など. 外在性
 - 構造：オブジェクト内部あるいはオブジェクト間の関係

Setting the State, Anne J.Gilliand-Swetland, Introduction to Metadata – Pathways to Digital Information, Murthsa Baca (ed.), Getty Information Institute.

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

メタデータ

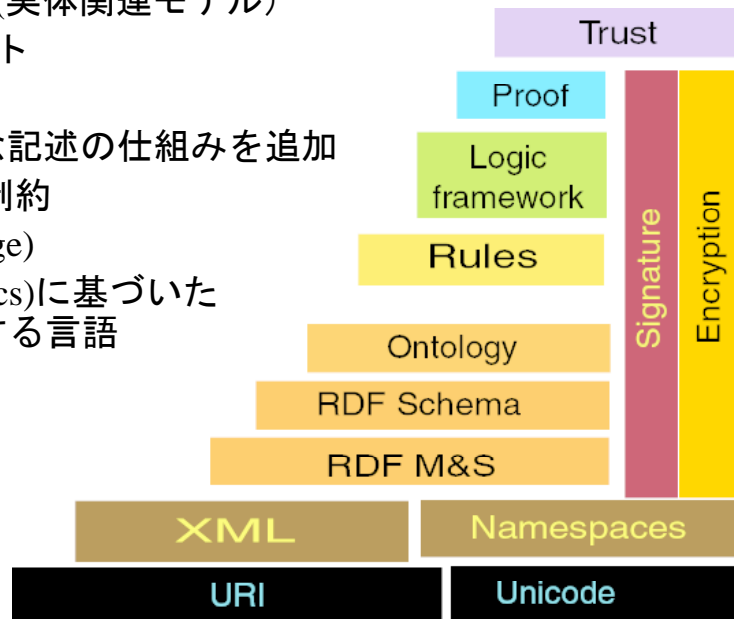
- オブジェクト全体へのメタデータ
 - 書誌情報, Dublin Core
- オブジェクトの部分あるいは構造に対するメタデータ
 - 設計図, RDF, RDFS, DAML+OIL



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

Semantic Webの階梯

- RDF (Resource Description Framework)
 - 最も原始的な意味記述の枠組みを提供一>SVOモデル
 - ◆ Entity-Relation Model(実体関連モデル)
 - ◆ セマンティックネット
- RDF Schema
 - RDFに最も原始的な概念記述の仕組みを追加
 - ◆ class-subclass関係, 制約
- OWL (Web Ontology Language)
 - 記述論理(description logics)に基づいた概念とその関係を記述する言語

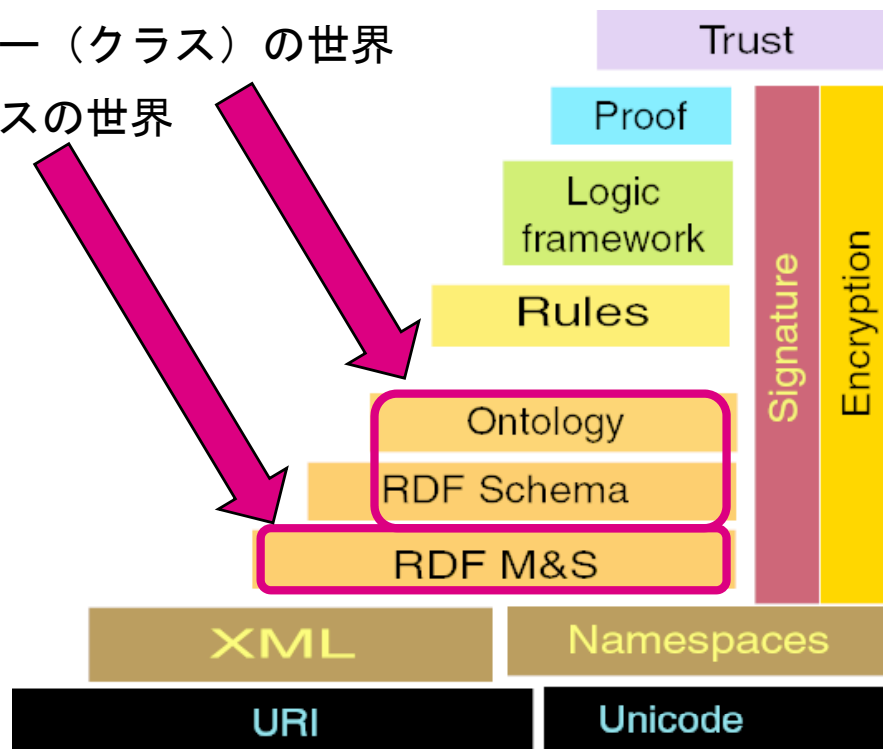


Tim Berners-Lee <http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

Architecture for the Semantic Web

- オントロジー (クラス) の世界
- インスタンスの世界

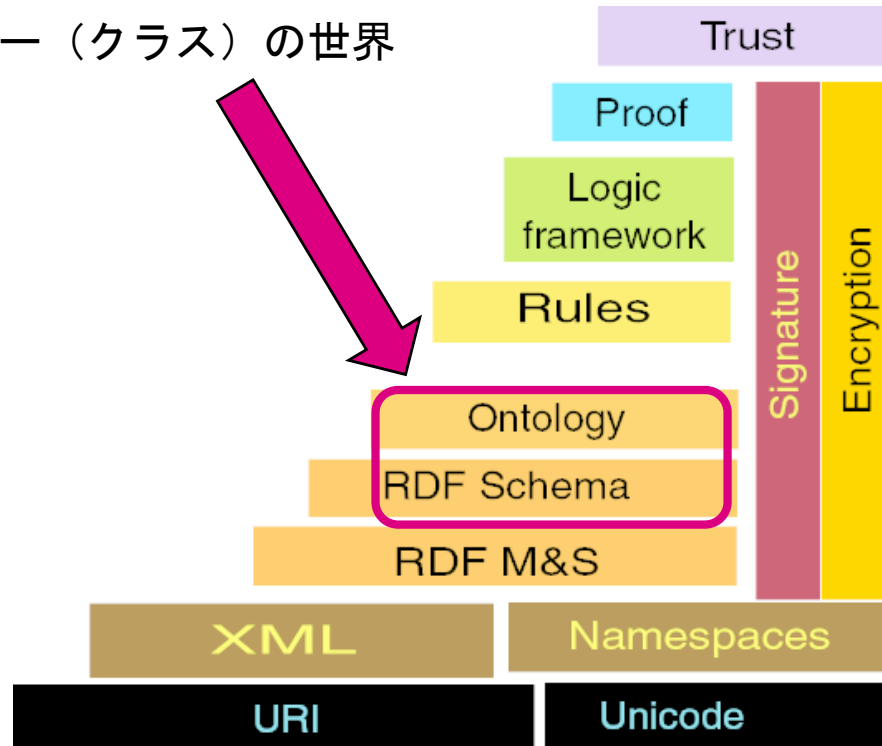


Tim Berners-Lee <http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

Architecture for the Semantic Web

- オントロジー（クラス）の世界



Tim Berners-Lee <http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



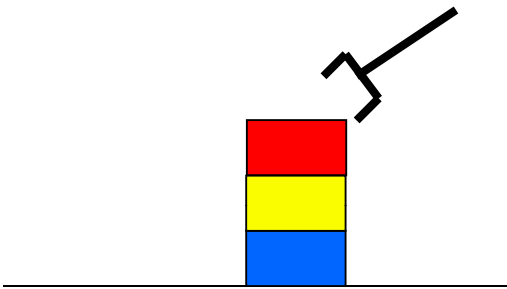
セマンティックWebにおける知識表現

- 一般の知識表現
 - 意味は知識表現言語によって定義される
 - ◆ 例：一階述語論理による公理
 - 知識表現言語で表現されたものがすべて⇒閉じた系
 - ◆ 外部に依存しない
 - 推論：表現されたものの一貫性
 - 定義間の矛盾
 - 定理の導出
 - 外部のデータを表現して、既存の表現との一貫性を見る

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



知識表現の例

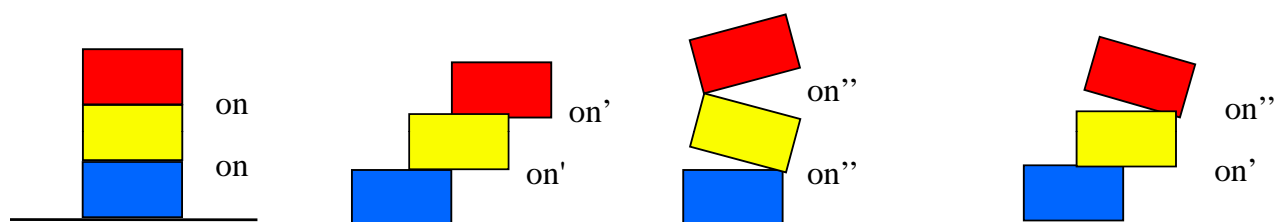


- $\text{above}(X,Z) \leftarrow \text{on}(X, Y) \wedge \text{on}(Y,Z)$
- $\text{on}(\text{ybox}, \text{bbox}), \text{on}(\text{rbox}, \text{ybox})$
- $\text{above}(\text{rbox}, \text{bbox})$

セマンティックWebにおける知識表現

- セマンティックWeb
 - 開いた系：Webのデータは常に不完全
 - ◆ 概念の定義は不完全
 - ◆ 概念の列挙も不完全
 - おおよその定義で満足しざるをえない
 - ◆ 様々な定義
 - ◆ 発散してしまう可能性

開いた系



- $\text{above}(X,Z) \leftarrow \text{on}(X, Y) \wedge \text{on}(Y,Z)$
- $\text{on}(\text{ybox}, \text{bbox}), \text{on}(\text{rbox}, \text{ybox})$
- $\text{above}(\text{rbox}, \text{bbox})?$

URI共有による意味の共有

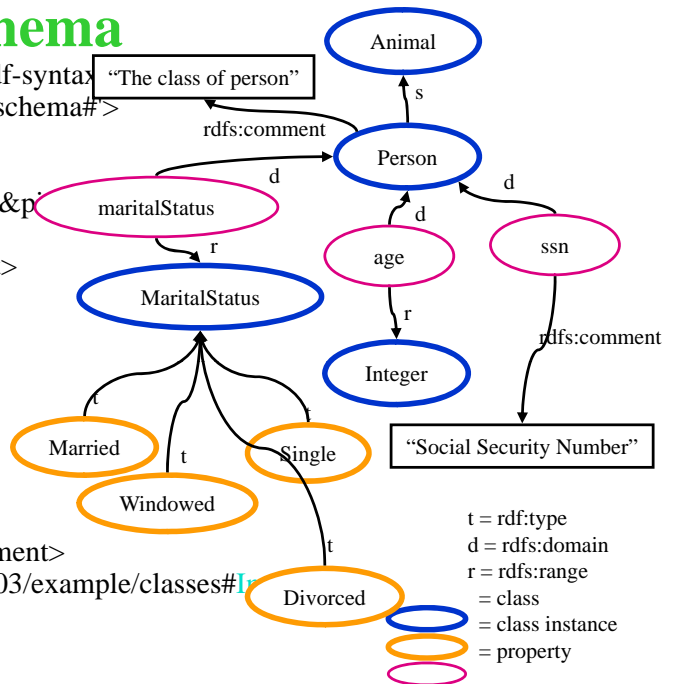
- セマンティックWebにおける開いた系の取り扱い方
 - **URIの共有による意味の共有**
- URI (Uniform Resource Identifier)
 - 概念やデータの識別子として使う
- URIの共有
 - ある概念, データを指し示すときに一つのURIを使う
 - このURIを使うということが意味を共有するという意思表示

RDF Schema

```

<!DOCTYPE rdf:RDF [
  <!ENTITY pi 'http://ex.org/personinfo#'>
  <!ENTITY rdf 'http://www.w3.org/1999/02/22-rdf-syntax#>
  <!ENTITY rdfs 'http://www.w3.org/2000/01/rdf-schema#'>
]>
<rdf:RDF xml:lang="en"
  xmlns:rdf="&rdf;" xmlns:rdfs="&rdfs" xmlns:pi="&pi;">
<rdfs:Class rdf:ID="&pi;Person">
  <rdfs:comment>The class of people.</rdfs:comment>
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/03/example/classes#Animal"/>
</rdfs:Class>
<rdf:Property ID="&pi;maritalStatus">
  <rdfs:range rdf:resource="&pi;MaritalStatus"/>
  <rdfs:domain rdf:resource="&pi;Person"/>
</rdf:Property>
<rdf:Property ID="&pi;ssn">
  <rdfs:comment>Social Security Number</rdfs:comment>
  <rdfs:range rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
  <rdfs:domain rdf:resource="&pi;Person"/>
</rdf:Property>
<rdf:Property ID="&pi;age">
  <rdfs:range rdf:resource="http://www.w3.org/2000/03/example/classes#Integer"/>
  <rdfs:domain rdf:resource="&pi;Person"/>
</rdf:Property>
<rdfs:Class rdf:ID="&pi;MaritalStatus">
<ex:MaritalStatus rdf:ID="&pi;Married"/>
<ex:MaritalStatus rdf:ID="&pi;Divorced"/>
<ex:MaritalStatus rdf:ID="&pi;Single"/>
<ex:MaritalStatus rdf:ID="&pi;Widowed"/>
</rdfs:Class>
</rdf:RDF>

```



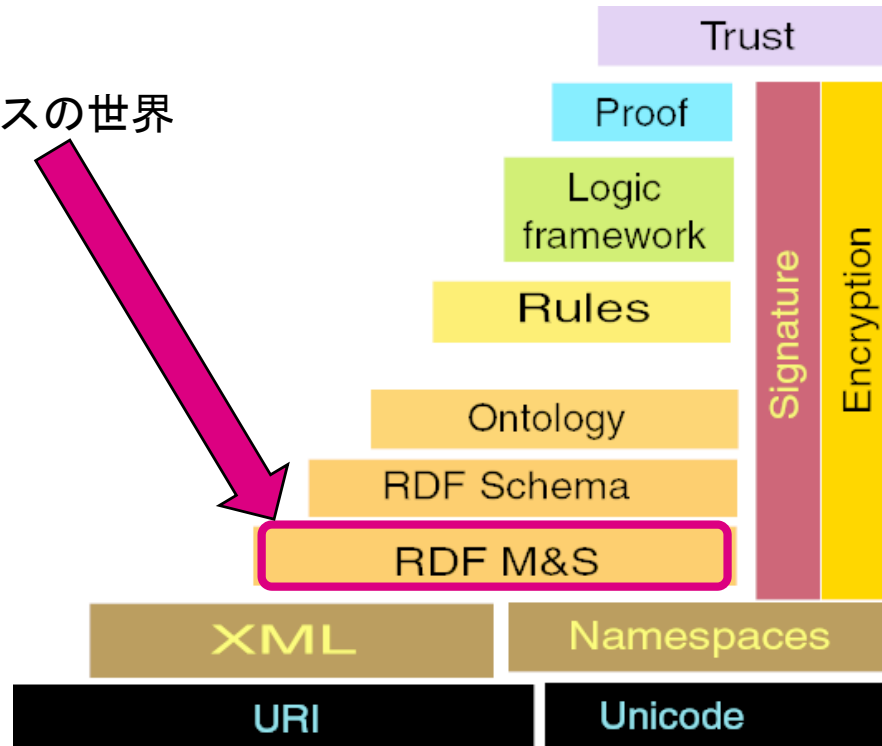
Resource Description Framework(RDF) Schema Specification 1.0
<http://www.w3.org/TR/2000/CR-rdf-schema-20000327/>

RDF Schema

- <http://ex.org/personinfo#Person> -> 「人(の情報)」の定義
- <http://ex.org/personinfo#age> -> 「年齢」の定義
- `<http://ex.org/personalinfo#Person rdf:ID="Takeda">`
`<http://ex.org/personalinfo#age rdf:resource=20 />`
`<http://ex.org/personalinfo#maritalstatus`
`rdf:resource="Married" />`
`</>`

Architecture for the Semantic Web

- インスタンスの世界



Tim Berners-Lee <http://www.w3.org/2002/Talks/09-lcs-sweb-tbl/>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



Linked Data

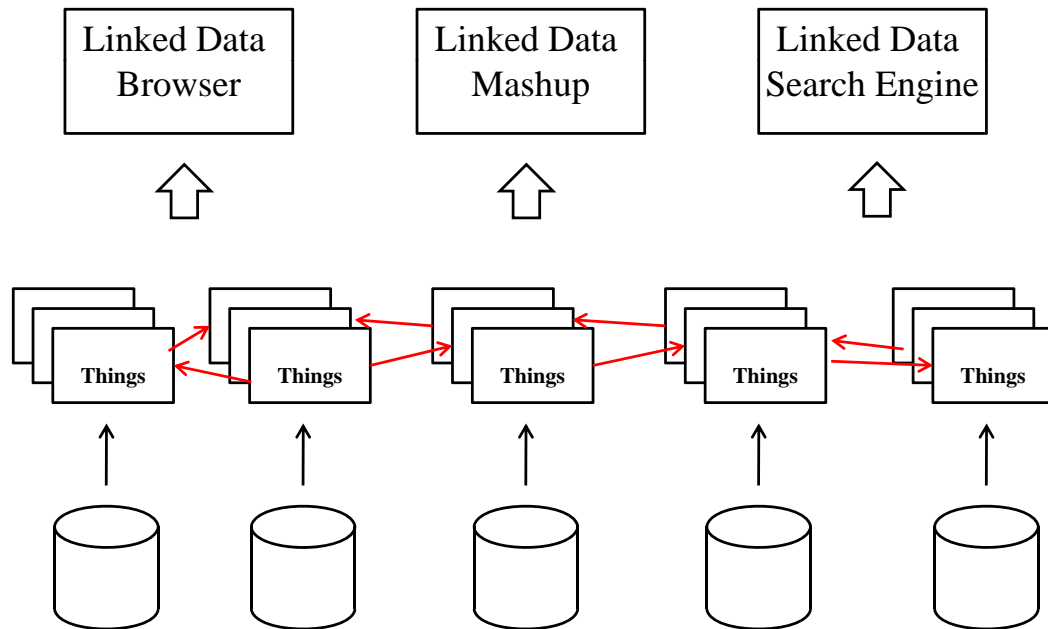
- Linked Dataとは **“Web of Data”**
 - RDFで公開されるデータ
 - 外部から参照可能
- Linked Dataのための4条件
 - 事柄の名前にURIを使うこと
 - ◆ *すべてのモノ、コトにURIを!*
 - 名前の参照がHTTP URIでできること
 - ◆ *DOIとかいったURNは使わないでね*
 - URIを参照したときに関連情報が手に入るように
 - ◆ *理解可能なデータを提供してね.*
 - 外部へのリンクも含めよう
 - ◆ *Webのようにリンクでつながるデータを作ろう*

Linked Data, TBL, <http://www.w3.org/DesignIssues/LinkedData.html>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



LODをどう使うか



Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



LODをどう使うか

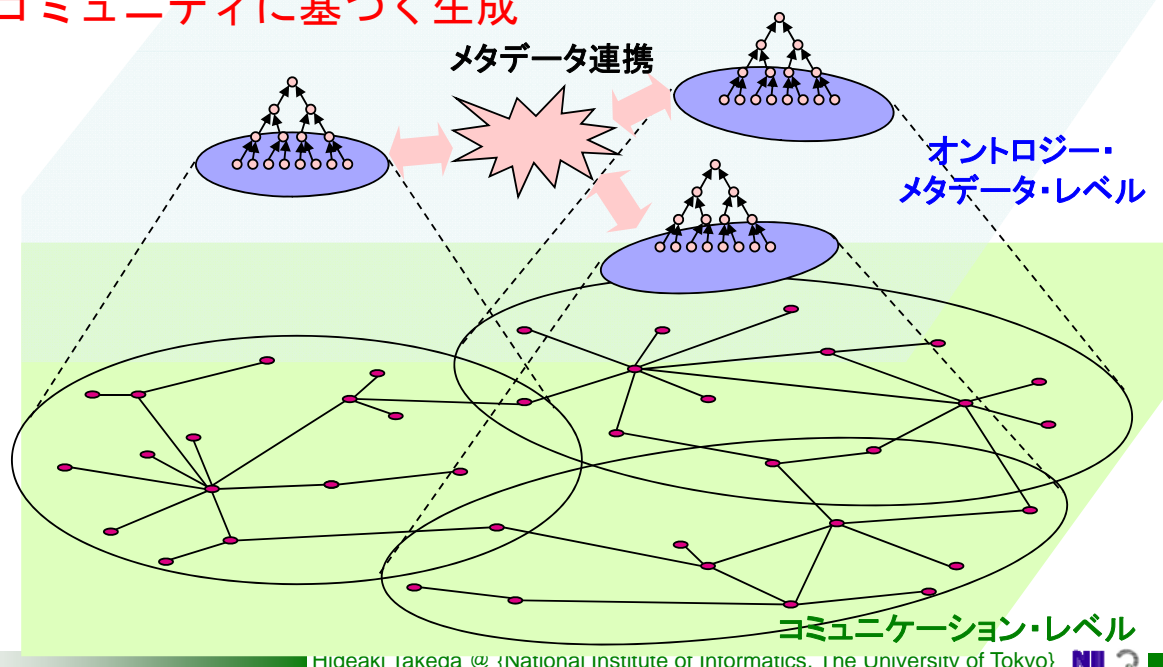
- Semantic Data Mash-up Applications
様々なサイトの様々なデータをつなげて見せる
 - SemaPlover
 - ◆ <http://btc.isweb.uni-koblenz.de/>
 - Dbpedia Mobile
 - ◆ <http://wiki.dbpedia.org/DBpediaMobile>
 - Bio2RDF
 - ◆ <http://bio2rdf.org/>


Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}



コミュニティベースのメタデータ流通の枠組み

- メタデータの自由で柔軟な生成と流通
 - コミュニティに基づく流通
 - コミュニティに基づく生成



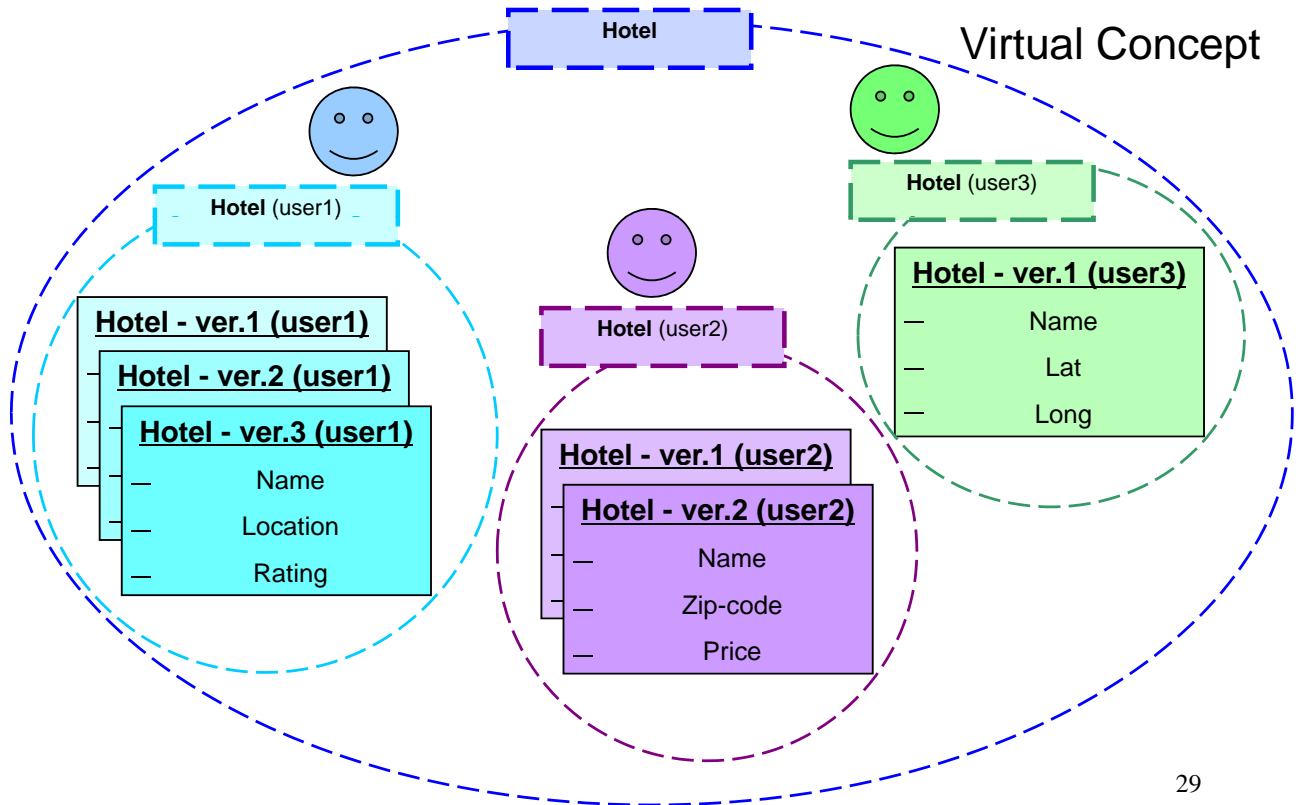
Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

Stylid: 社会Webによる構造的メタデータの共有

- 目的
 - 構造的メタデータの共有基盤の構築
 - ◆ 簡単な作成, 共有
 - ◆ 多様性の実現
 - 人々間の違い, 時間の違い
 - ◆ 統合する仕組み
- Stylid: **S**tructure **y**our own **l**inked **d**ata
 - <http://www.stylid.org>

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}  

多様性のある定義を書く



29

StYLiD

Structure Your own Linked Data

概念を定義する

Published Data Upcoming Data Submit new data Submit a new Concept Modify existing Concept

StYLiD Home » Modify existing Concept

Sort news by: Recently Popu

Submit a new concept, step 2 of 3

Creating a new Concept

"Project" concept

Attributes of the Concept "project"

Attribute labels

Label: <input type="text" value="project_title"/> Suggest range for values	Description: <input type="text"/>
Label: <input type="text" value="organization"/> Suggest range for values Organization remove	Description: <input type="text"/>
Label: <input type="text" value="project_leader"/> Suggest range for values faculty remove Organization remove	Description: <input type="text" value="Leader of the project"/>
Label: <input type="text" value="url"/> Suggest range for values	Description: <input type="text" value="URL of the project page"/>

Suggested Value Range

インスタンスを書く

- Published Data
- Upcoming Data
- Submit new data
- Submit a new Concept
- Modify existing Concept

Submit new data, step 2 of 3

Enter Instance Data

project Data

Entry title:
Please enter the title for your entry. (max 120 characters)
The StYLID project

project title: Literal value
StYLID enter URI
add more..

organization: Suggested range concepts
Suggested range of values: Organization
NII -
http://www.stylid.org/story.php/rdf/id/233 URI Resource URI
add more..

project_leader: Suggested range of values: faculty Organization
(Leader of the project) Hideaki Takeda enter URI
KASM enter URI
add more.. Multiple Values

url:
(URL of the project page)
enter URI

Concept cloud

aeroplane airlines album
chocolate computer C
COURSE event f
lecture movie museum nov
project realstatea
sweets talk test vac

Top Today

Published Data

- 1 International Weblogs and 2008)
- 1 My favourite I
- 1 Determining the infrastructure diversity through analysis
- 1 Information re -Infotrustics
- 1 Digital Silk R

インスタンスの表示

Table View

1
Vote

Electronic entanglement security technology

Posted by [niiscraper](#) 71 days ago (Editorial)
Category: [root](#) | Tags: security
URI: <http://www.stylid.org/story.php/rdf/id/510>
Concept: [project](#) (created by [niiscraper](#))

project_title: Electronic entanglement security technology
project_leader: [YAMAMOTO, Yoshihisa](#)
url: http://www.qis.ex.nii.ac.jp/ex_project2005.html

attributes

outline: This project involves research into the physical implementation of electronic entanglement network technology and associated applications. Physical implementation refers to theoretical and practical validation of core technology elements ? network creation, operation, storage, and observation. Theoretical research focuses on amplifying small non-linearities in order to manipulate the quantum state and facilitate observation, while experimental research encompasses quantum dot photon sources, photon detection using wavelength converters, EIT non-linear operators, and nuclear spin quantum memory. Meanwhile, applications under study include examination of fundamental theory, complete algorithms for future use, and simple protocols designed for immediate implementation.

project_category: Informatics for future value creation

[read more »](#)
[Discuss](#) [Add this link to...](#) [Tell a friend](#) [Bury](#)

Project concept
created by [Username](#) (version no.)

- Machine readable data (RDF, RDFa)
- Scripts can provide useful operations

1
Vote

Science Grid

Posted by [niiscraper](#) 71 days ago (Editorial)
Category: [root](#) | Tags: science grid

Published Data

Upcoming Data

Submit new data

Submit a new Concept

Modify existing Concept

StYLID Home » Concept cloud

Sort news by: Recently Popular

Concept Cloud

All Concepts

My Concept Collection

Concepts Created by Me

The Concept Cloud shows a central node 'faculty' labeled as a 'Consolidated concept'. Below it, a 'Sub-Cloud' is enclosed in a dashed red oval, containing terms like 'faculty', 'god', and 'aman'. Surrounding these are various other concepts such as 'airlines', 'album', 'article', 'band', 'book', 'camera', 'car', 'chocolate', 'composer', 'computer', 'concert', 'conference', 'country', 'course', 'festival', 'guitar', 'harddisk', 'hotel', 'laptop', 'movie', 'museum', 'newconcept', 'novel', 'organization', 'paper', 'person', 'restaurant', 'seminar', 'song', 'sweets', and 'vacancy'.

Published Data

Upcoming Data

Submit new data

Submit a new Concept

Modify existing Concept

StYLID Home

Sort news by: Recently Popular

Concept versions

Align Concepts

The 'Align Concepts' interface shows two source concepts: 'Book created by amazonscraper (ver. 0)' and 'Book created by aman (ver. 0)'. The first concept has attributes: title, author, price, book_type, and publish_date. The second concept has attributes: title, author, editor, ISBN_code, price, and date. These are being aligned to a 'Combined attribute name' list on the right, which includes: title, author, editor, ISBN_code, price, book type, and publish_date. Red dashed boxes and arrows highlight the alignment of 'title' and 'publish_date' from the source concepts to their respective entries in the combined list.

インスタンスの統合表示

Published Data Upcoming Data Submit new data Submit a new Concept Modify existing Concept

StYLiD Home » Instances of Concept book Sort news by: Recently Popular

Search results for book concept

Uniform Table View

Record View Edit combined attributes filter

title	author	editor	ISBN_code	price	book_type	publish_date
Speech and language processing	Speech and language processing	Daniel Jurafsky , James H. Martin	0-13-095069-6	3000		
A Semantic Web Primer (Cooperative Information Systems)	A Semantic Web Primer (Cooperative Information Systems)	Grigoris Antoniou and Frank van Harmelen		\$33.90	Hardcover	April 1, 2004
Explorer's Guide to the Semantic Web	Explorer's Guide to the Semantic Web	Thomas B. Passin		\$26.37	Paperback	Mar 1, 2004
Technologies: Trends and Research in Ontology-based Systems	Technologies: Trends and Research in Ontology-based Systems	John Davies, Rudi Studer, and Paul Warren		\$106.60	Hardcover	Jul 11, 2006
Introduction to the	Introduction to the					

Unified attributes

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}

構造的検索

StYLiD

Structure Your own Linked Data

Published Data Upcoming Data Submit new data Submit a new Concept Modify existing Concept

StYLiD Home » Advanced Search

Advanced Search

Concept name:

Attribute: Value:

[Add more..](#)

Search

Search on consolidated concept

Search on Consolidated Concept

1 search results returned:

[Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential](#)

Enter your own SPARQL Query

SPARQL

Hideaki Takeda @ {National Institute of Informatics, The University of Tokyo}

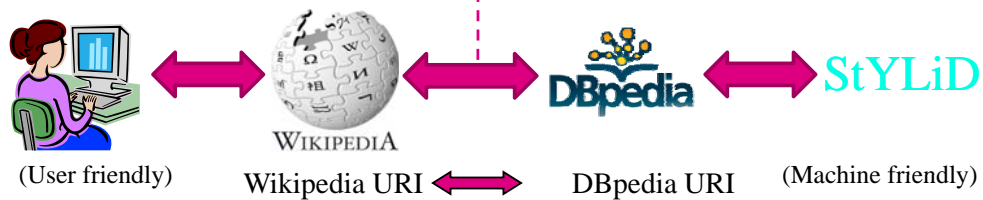
Linked Data

- Link data: URIを通じたインスタンスデータの関連づけ
例:



Dbpedia: Wikipedia (のinfobox) の情報をRDF化して公開

- Wikipediaコンテンツとの結びつけ
 - ユーザは通常のwikipediaのページを指示
 - StlyidはDBpediaのインスタスに関連付け



37

まとめ

- Web of Data としてのLinked Dataは普及しつつある
 - 新しいレイヤー
- 新しいサービス, 新しいアプリケーション
 - Semantic Mashup
- 多様な可能性

