



セマンティックWEBとW3C

慶應義塾大学 環境情報学部

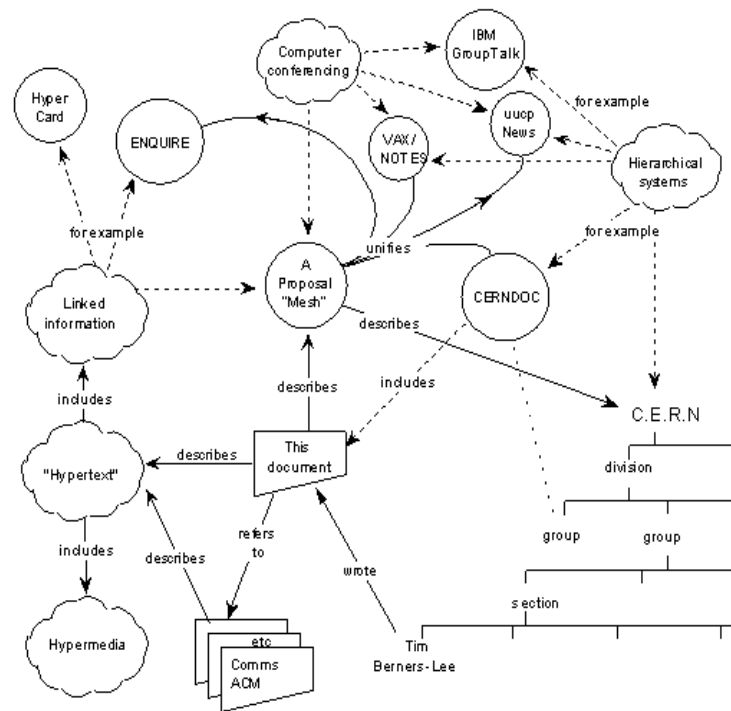
World Wide Web Consortium

萩野 達也

1

セマンティックWEB

- Tim Berners-Leeが1998年ごろに提唱しはじめる
- 機械的処理可能なメタデータ空間
- Timの1989年のWebの提案書に始まる



セマンティックWEBで実現したいこと

- Webを知識空間とみなしさまざまな問題を解決したい
 - 検索エンジンはWebページを見つけるだけ
 - 旅行予約などの特定のサービスはあるが、ユーザが必要とするサービスが見つからないこともある
- Web上にはたくさんの知識がある
 - 検索エンジンを使って日々利用
 - あらゆることに関する情報がある
 - 人が読むためのページがほとんど
 - HTMLはあまりに自由な形式で機械的な処理が難しい



機械処理可能なデータ

時刻表

辻堂駅 東海道本線 横浜・東京方面 (上り)		▶ 土曜・休日の時刻表を表示																											
時	平日																												
5	00 27 50																												
6	快高	02	11	19	27	34	40	44	47	52	55	58	湘ラ																
7	快高	01	04	07	10	13	16	19	22	25	29	31	34	37	40	46	49	51	54	58	快高								
8	快籠	01	04	09	12	15	19	22	27	30	35	43	51	54	湘ラ														
9	快籠	01	10	23	33	45	50	55																					
10	快籠	05	15	26	38	49	59																						
11	快籠	14	32	50	53																								
12	快籠	02	14	35	44	50																							
13	快籠	03	15	26	32	39	50	56																					
14	快籠	07	15	25	38	50	54																						
15	快籠	04	17	32	40	50	58																						
16	快籠	02	16	26	29	42	50	55																					
17	快高	07	14	19	24	34	46	51	54	快籠																			
18	品	快前	品	快籠	05	11	22	26	30	35	45	50	57																
19	快高	08	20	23	32	42	48	54	58	快籠																			
20	快籠	09	21	25	37	42	50																						
21	快高	00	14	21	25	39	51																						
22	快籠	04	19	25	33	39	45	58																					
23	品	品	品	12	27	39																							

列車種別-列車名: 無印=普通 快=快速 湘ラ=湘南ライナー
行き先-経由: 無印=東京 籠=籠原 高=高崎 品=品川 前=前橋

HTML

```
<table>
  <tr>
    <th>時</th>
    <th colspan="19">平日</th>
  </tr>
  ...
  <tr>
    <th>6</th>
    <td>快高<br />02</td>
    <td>11</td>
    <td>19</td>
    <td>快籠<br />27</td>
    <td>34</td>
    ...
  </tr>
  ...
</table>
```

XMLによる記述

```
<?xml version="1.0" encoding="Shift_JIS"?>
<timetable>
  <station name="辻堂">
    <line name="東海道" dir="上り" week="平日">
      <train at="6:02" dest="高崎" kind="快速" />
      <train at="6:11" />
      <train at="6:19" />
      <train at="6:27" dest="籠原" kind="快速" />
      ...
      <train at="6:62" kind="湘南ライナー" />
      ...
    </line>
    ...
  </station>
  ...
</timetable>
```

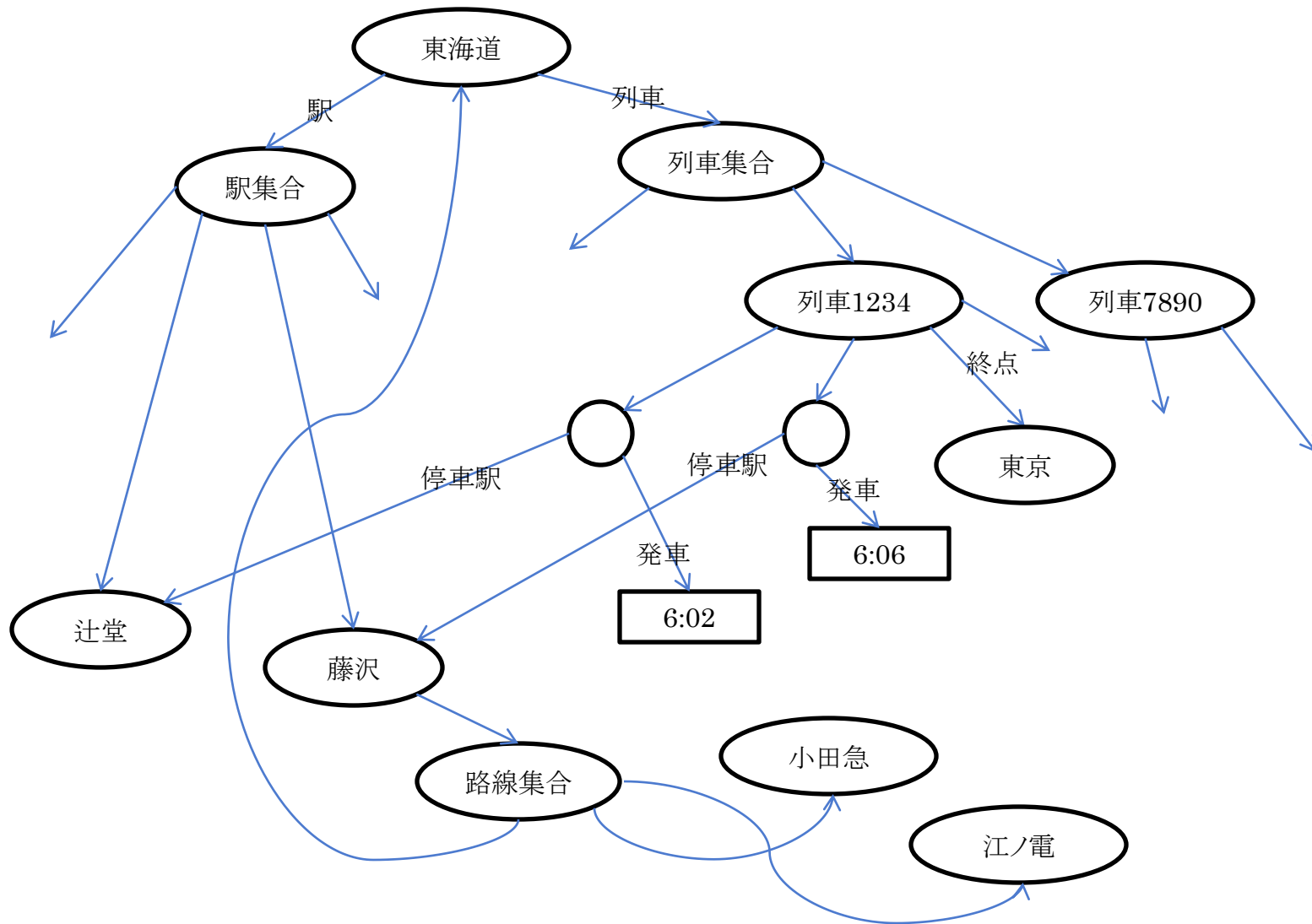
- 機械的な処理が可能
- DTDを決めておく必要がある
- 複数の形式で記述可能
 - 駅を軸に記述
 - 列車を軸に記述

辻堂駅 東海道本線 横浜・東京方面 (上り)	
平日	
時	
5	00 27 50
6	02 11 19 27 34 40 44 47 52 55 58
7	01 04 07 10 13 16 19 22 25 29 31 34 37 40 46 49 51 54 58
8	01 04 09 12 15 19 22 27 30 35 43 51 54
9	01 10 23 33 45 50 55
10	05 15 26 38 49 59
11	14 32 50 53
12	02 14 35 44 50
13	03 15 28 32 39 50 56
14	07 15 25 38 50 54
15	04 17 32 40 50 58
16	02 16 26 29 42 49 55
17	02 14 19 24 34 46 51 54
18	05 11 22 28 39 35 45 50 57
19	08 20 23 32 42 49 54 58
20	09 21 25 37 42 50
21	00 14 21 25 39 51
22	04 19 25 33 39 45 58
23	12 27 32

列車種別: 快速 普通 湘南ライナー
行先: 高崎 横浜 東京 品川 横浜

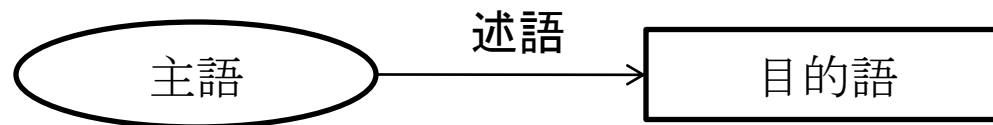


より柔軟な記述



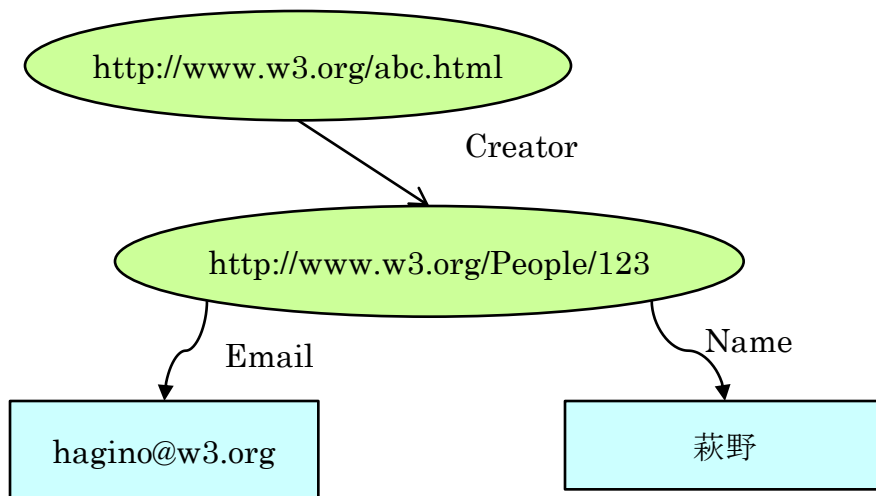
RDF

- Resource Description Framework
- データの記述形式
- ラベル付き有効グラフ



- 三つ組を基本とする
- 主語, 述語はURI
- 目的語は文字列またはURI

RDFの例



RDF/XML

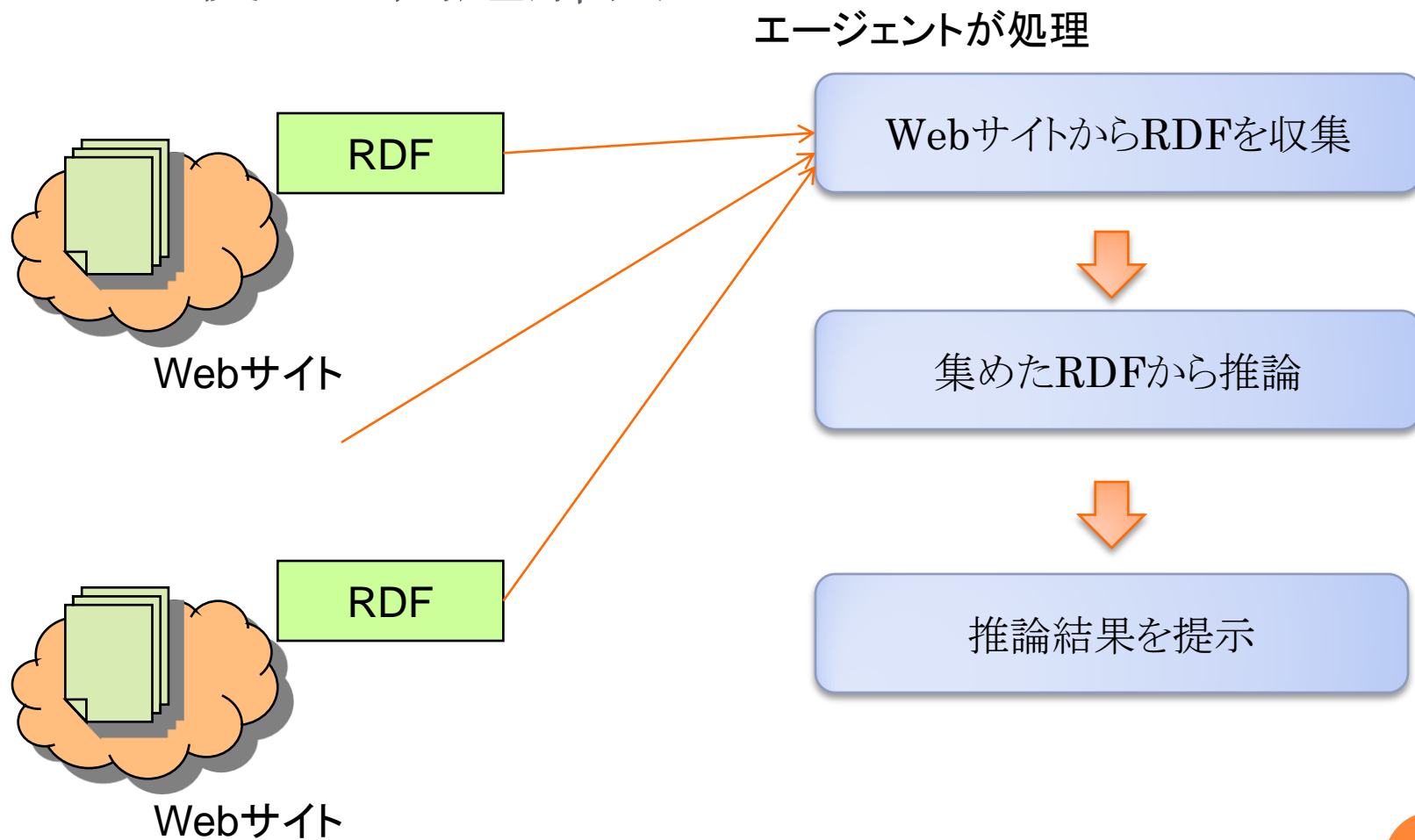
```
<rdf:RDF>
  <rdf:Description about="http://www.w3.org/abc.html">
    <s:Creator rdf:resource="http://www.w3.org/People/123"/>
  </rdf:Description>

  <rdf:Description about="http://www.w3.org/People/123">
    <v:Name>萩野</v:Name>
    <v:Email>hagino@w3.org</v:Email>
  </rdf:Description>
</rdf:RDF>
```


RDFとXMLの違い

	XML	RDF
表現	テキスト	グラフ RDF/XML Triple N3
構造	木構造	任意のグラフ
構造の定義	DTDまたはXML Schema	RDF Schema
要素順序	重要	無視

RDFを使った問題解決



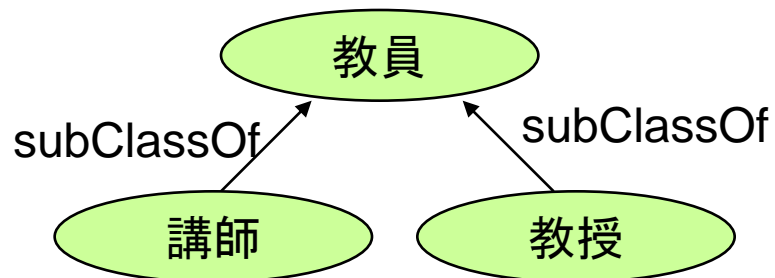
時刻表がRDFで提供されていれば

- 乗り換え案内を自分で作ることができる
 - 複数の電車会社からの時刻表を組み合わせる
 - 好みの電車会社を利用
 - 乗り換えの時間を好みに合わせて調整可能
- 電車と他のサービスを連携することができる
 - バスの発車時刻との連携
 - スケジュールとの連携
 - 出張精算との連携
- コピー・ペーストで別の用途に利用可能
 - スケジュール帳に張る
 - 必要な時間帯のみの平日・土・日の時刻表を切り出す

RDFによるデータ提供

- どのような述語を使うかを決める必要がある
 - 語彙の決定
 - RDF Schema
 - オントロジー
- データベースやテーブルからの変換は比較的容易

RDF Schema



```
<rdfs:Class rdf:ID="教員"/>
<rdfs:Class rdf:ID="教授">
  <rdfs:subClassOf
    rdf:resource="#教員"/>
</rdfs:Class>
<rdfs:Class rdf:ID="講師">
  <rdfs:subClassOf
    rdf:resource="#教員"/>
</rdfs:Class>
```

RDFデータの埋め込み

- HTMLにRDFデータを埋め込む
 - RDFa
 - Microformat

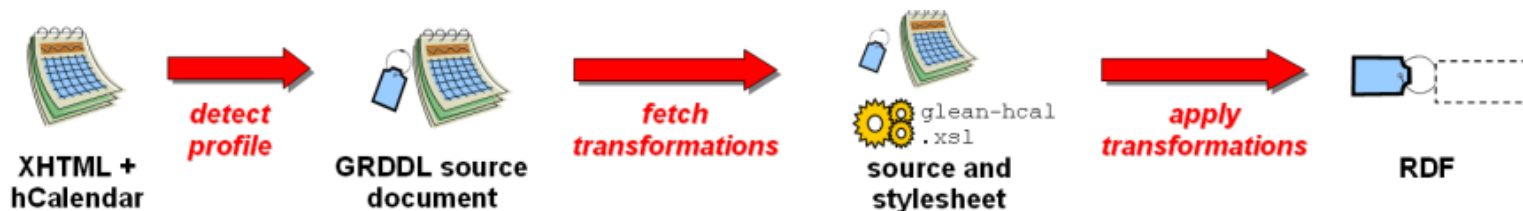
RDFa

```
<p role="cal:Vevent">  
  I'm giving  
  <meta property="cal:summary">  
    a talk at the XTech Conference about web widgets  
  </meta>,  
  on  
  <meta property="cal:dtstart" content="20060508T1000-0500">  
    May 8th at 10am  
  </meta>.  
</p>
```

埋め込まれたRDFデータの抽出

○ GRDDL (Gleaning Resource Descriptions from Dialects of Languages)

- XML文書からRDFデータを抽出する方法の記述
- 動的にRDFデータを生成する
- XSLTを使った変換
- microformatからの変換
- RDFaからの変換

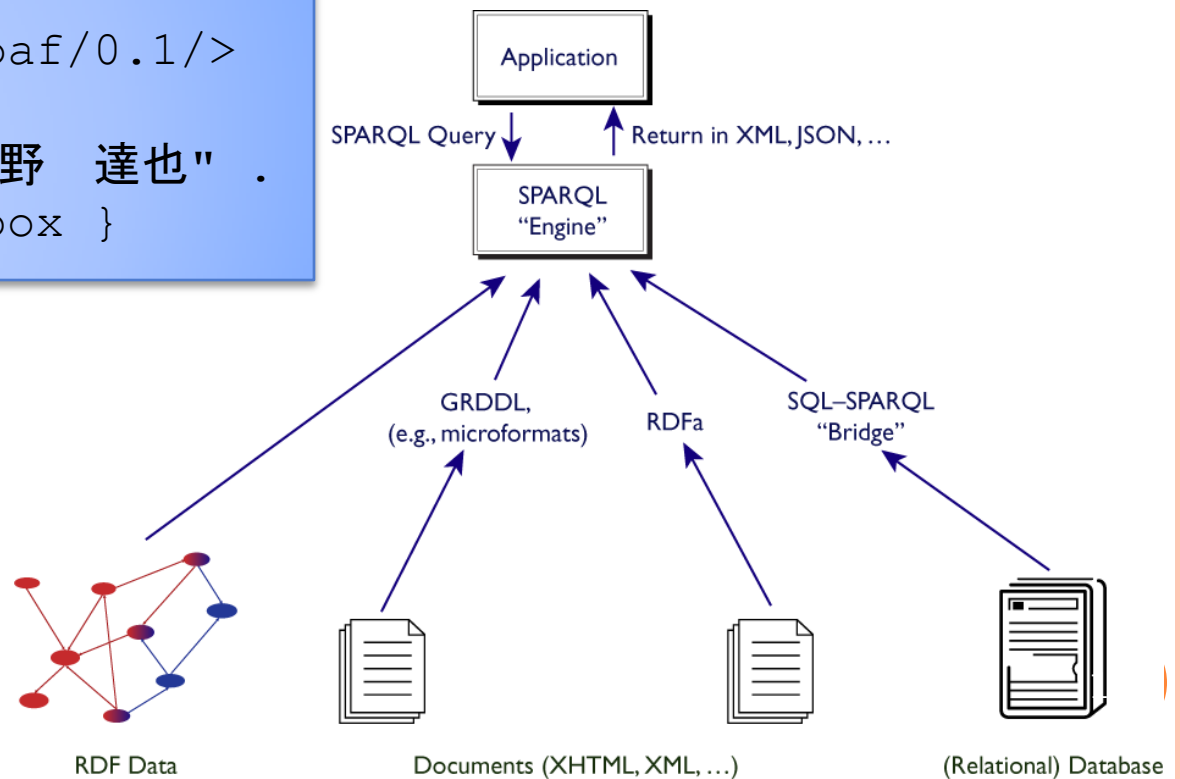


RDFデータの検索

○ SPARQL

- Simple Protocol And RDF Query Language

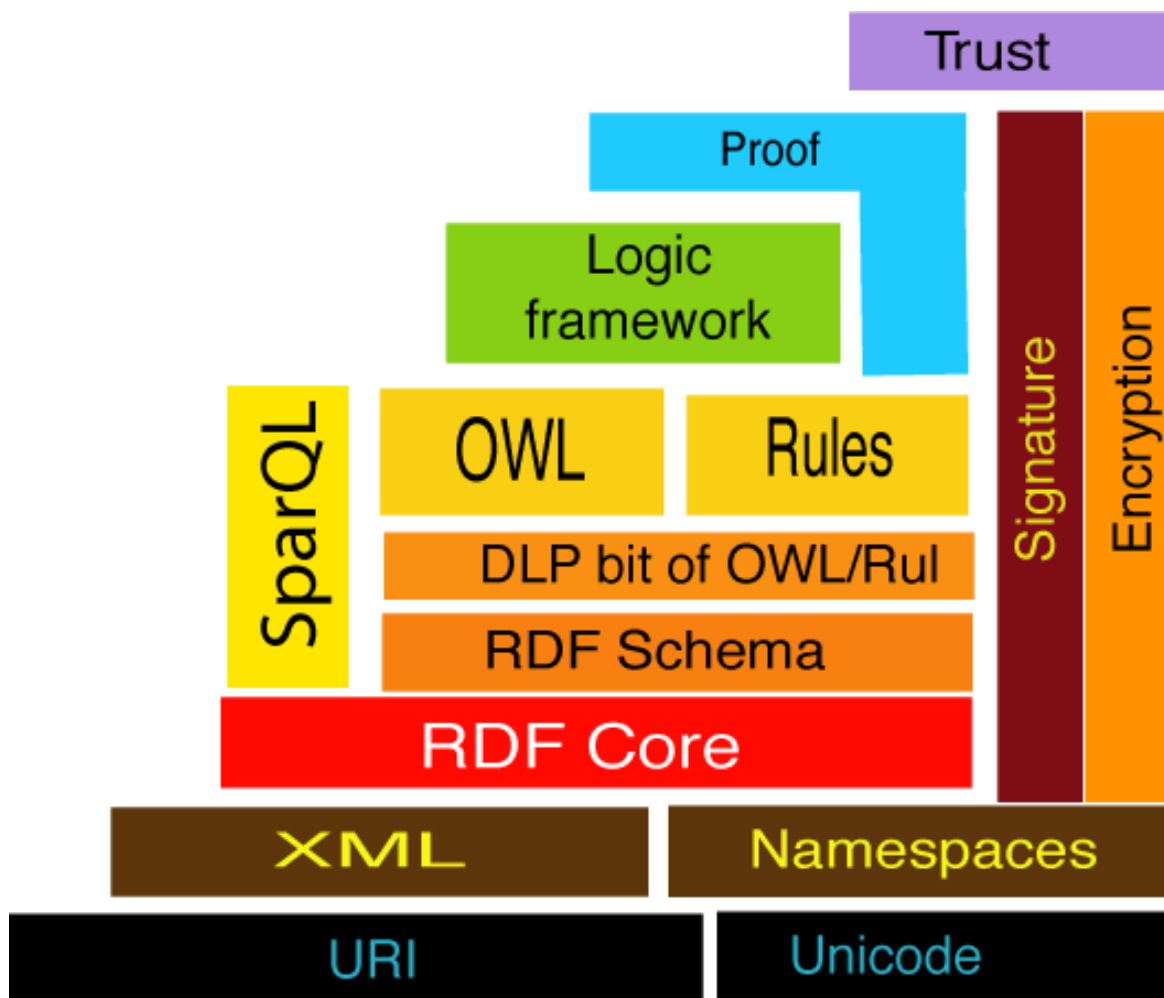
```
PREFIX foaf:  
<http://xmlns.com/foaf/0.1/>  
SELECT ?mbox WHERE  
{ ?x foaf:name "萩野 達也" .  
  ?x foaf:mbox ?mbox }
```



まだできていない部分

- 規則の記述
 - 「A ならば B」
- 論理フレームワーク
 - 単調論理
- 証明の記述

セマンティックWEBの技術階層



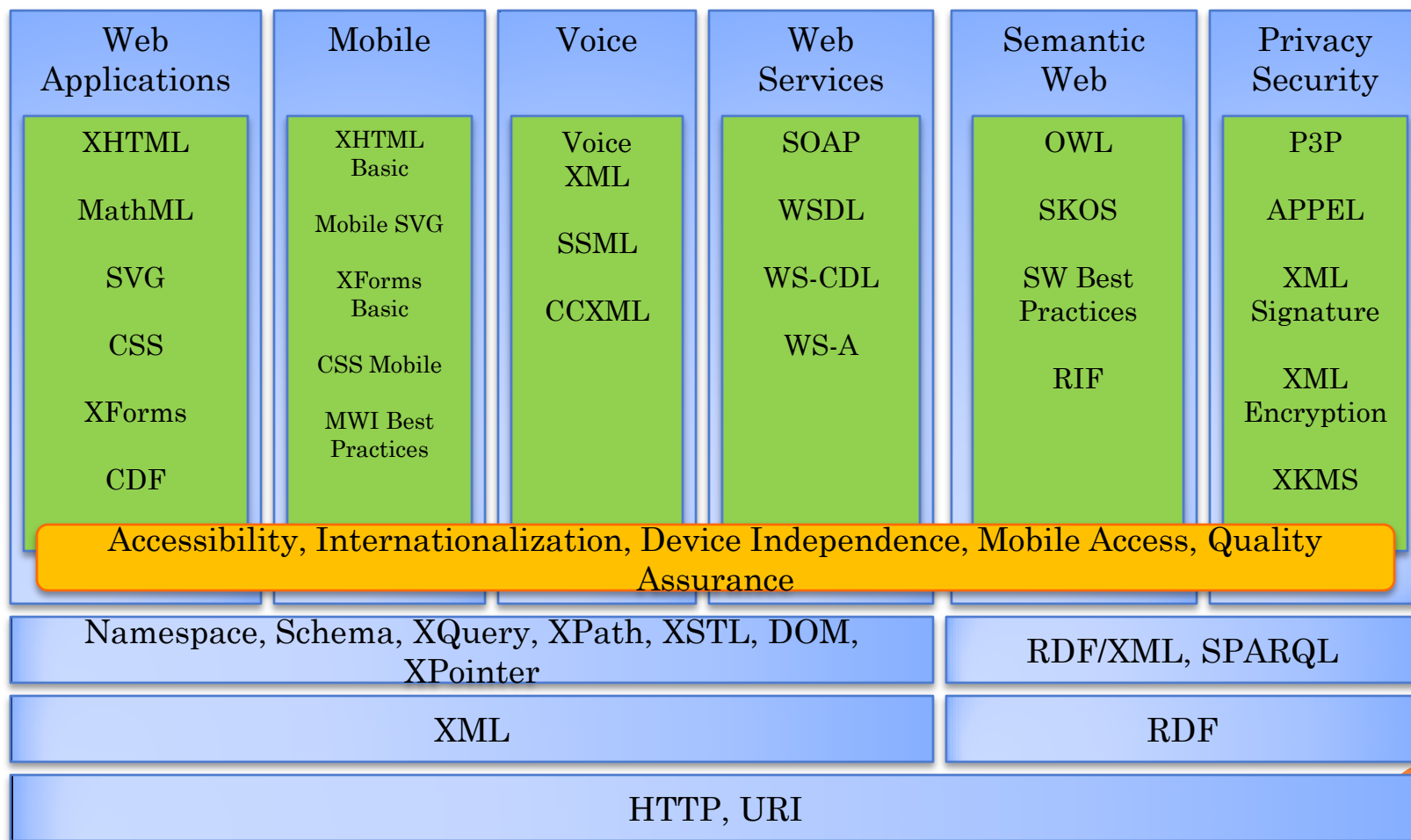
WEB 2.0とセマンティックWEB

	Web 2.0	セマンティックWeb
共通点	ユーザを巻き込む ロングテール Forksonomy Mashup	エージェントがユーザ要求にした がって処理する 自由なオントロジーの利用 複数サイトからデータを収集
相違点	さまざまな技術を利用 Ajax XMLHttpRequest RSS	RDFによるデータ記述 SPARQLによるデータ獲得 RIFによる規則
問題点	ユーザから集めたデータは サイト内で閉じて利用 複数サイトの連携ができにく い	RDFデータが少ない オントロジーが分かりにくい ユーザを巻き込むことができ ない

W3CにおけるセマンティックWEB

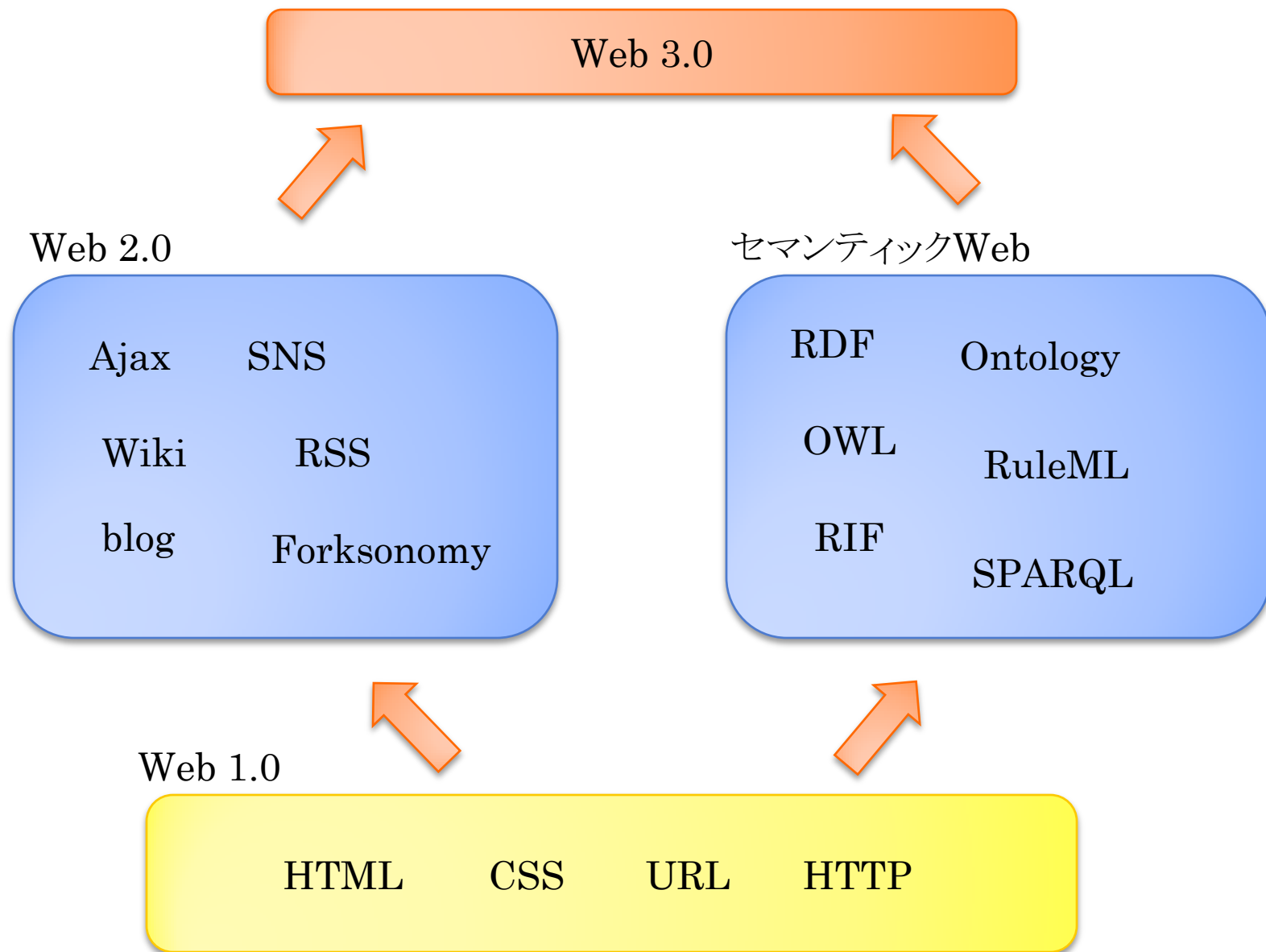
- W3Cの目的
 - Webの可能性を最大限に引き出す
 - Web基盤技術の標準化
 - Web上でのオープンな活動を支援
- 参加組織
 - 437組織(日本30組織)2007年11月18日現在
- 活動分野
 - Architecture Domain
 - Interaction Domain
 - Quality Assurance
 - Technology and Society Domain
 - Ubiquitous Domain
 - Web Accessibility Initiative

WEBの技術階層



SEMANTIC WEB ACTIVITY

- W3C内のSemantic Web関連の活動グループ
 - Semantic Web Coordination Group
 - RDF Data Access Working Group
 - Rules Interchange Working Group
 - Web Ontology (OWL) Working Group
 - GRDDL Working Group
 - Semantic Web Deployment Working Group
 - Semantic Web Interest Group
 - Semantic Web Health Care and Life Sciences Interest Group
 - Semantic Web Education and Outreach (SWEO) Interest Group



WEB 3.0

○ Web 2.0 + セマンティックWeb

- あらゆるデータがRDFとしてWeb上で利用できる
- エージェントがRDFを処理し問題解決をサポート
- RDFの処理を指示するスクリプト言語
- ユーザがRDFデータを提供する環境