

# セマンティックWebは次世代Webの 主役になれるか！？



発表者

株式会社ジェー・アイ・イー・シー

大泉 英之

# セマンティックWebの背景



セマンティックWebは1999年Tim Berners-Leeによって提唱された次世代のWebビジョンである。

国際会議「XML 2000」において、Tim Berners-Leeの講演をきっかけに、セマンティックWebの関心が強まった。

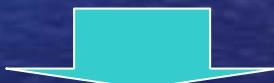
2001年からW3Cの活動として承認されて以来、活動が活発に始まった。

# 現状のWebの問題点

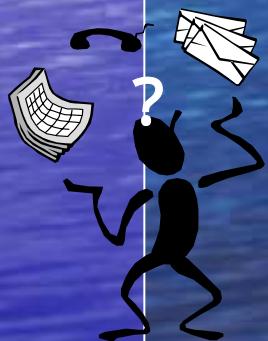
HTMLは世界中の情報にアクセスできる、  
WWWを生み出した



WWWは爆発的に普及し、情報が氾濫している



- ・自分の欲しい情報を手に入れる手間が次第に増加
- ・手にした情報は本当に一番良いもの？



# 次世代Webで何ができる？



検索エンジンが変わる！

Web上にある意味情報を機械が理解することにより、利用者の要求にあった商品検索や情報配信が可能

Webサービスとのコラボレーション

WSDLのようにインターフェースを記述するファイルはデータ型などの定義はできるが、「どういう目的で、どういう意味を持って、どういう処理が行われるか」ということは定義できない。意味情報を付与することによって柔軟なサービスが期待できる

# 次世代Webの現状は？



欧米では活発に議論されている

昨年のWWW10 (*Tenth International World Wide Web Conference*) ではXMLからセマンティックWebに議論の対象がシフトした。

今年5月に開かれたWWW11でもWebサービスとほぼ同数のセミナーなどが見受けられた。

日本ではほとんど議論されていない

日本での議論はセマンティックWebよりもWebサービスに大きく傾いている。

# セマンティックWebとは？

人間が読み、理解することが前提のWebに対して、意味情報(Semantic)を付加することにより、Webを知的にし、有用性を高めるWebビジョン

意味情報を表現するものは...？



オントロジ  
(ontology)



# オントロジ

意味情報を表現する方法の一つ。



哲学分野  
「ある」

WebやAI分野  
「概念間の関係の明確な定義の集まり」

セマンティックWeb  
「概念どうしの関係や、それらを解釈するための論理的なルールを定義する文章の集合」

# データは何を表しているでしょう？

1、3、4、6、8、10、12

正解：東京都のTVチャンネル



# セマンティックWebの概念

人間対人間のやり取りを人間対機械、もしくは機械対機械とのやり取りにし、自動処理を可能にする。

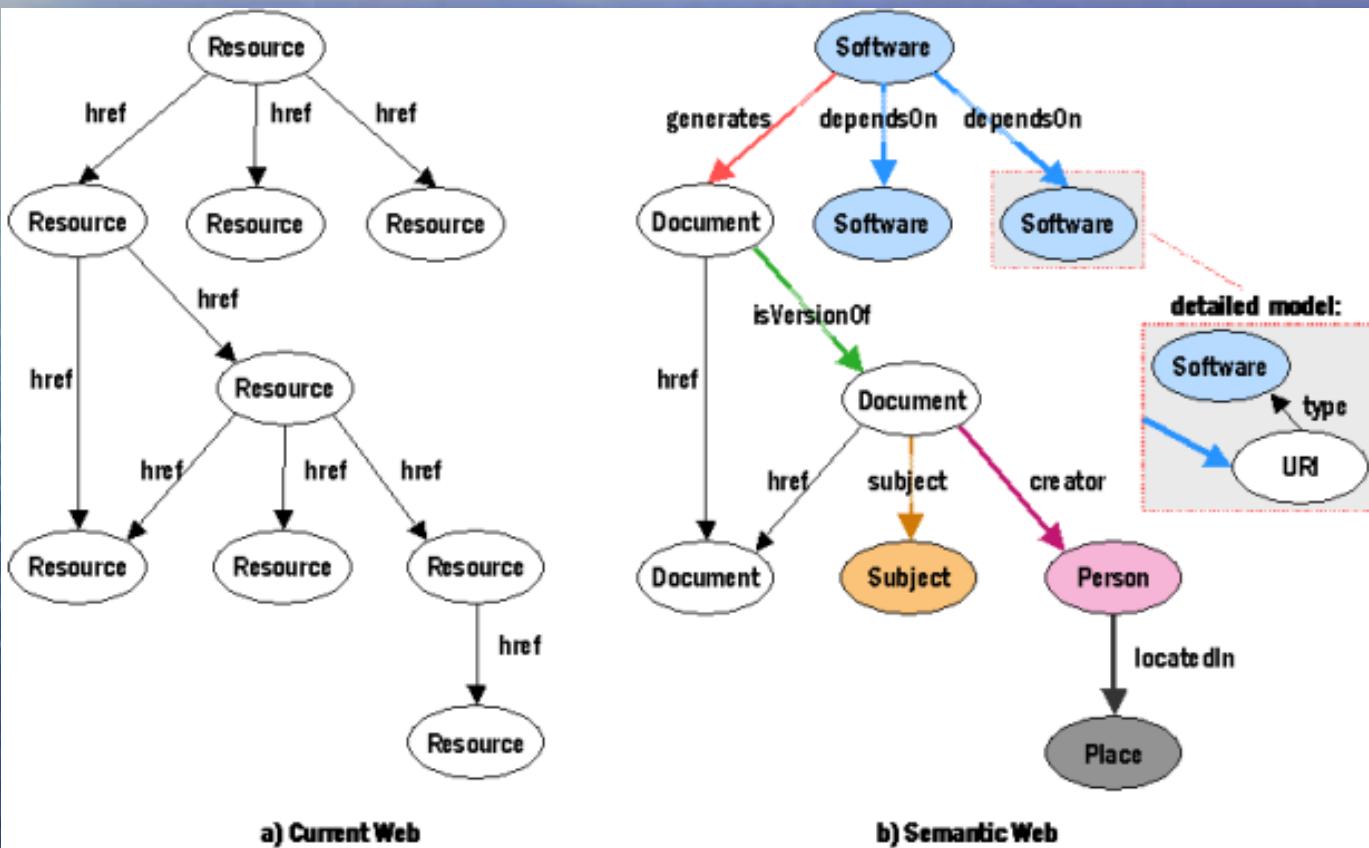


図: Resources and links can have types in the Semantic Web  
<http://www.w3.org/2001/12/semweb-fin/w3csw>

# セマンティックWebが有用な分野

## 検索

検索の手間が削減、必要な情報が得られる

## Webサービス

サービスの連携など

## ナレッジマネジメント

蓄積されたナレッジの検索、人事の適材適所への配置

## リアルタイムなサイトマップ

デットリンクを回避する

## その他、人間の思考が必要な分野

Webの意味情報増加によって、さまざまな可能性



# (例) 従来の検索



博多にあるラーメン屋を探す場合

「博多」と「ラーメン屋」で検索

博多ラーメンの店が見つかる

“博多”という名前のラーメン屋も見つかる

大量の検索結果から人間が判断



コンテンツの内容から住所や地図を調べる手間が増加

最近、開店した店を探す場合

「最近」というキーワードでは検索は難しい

# (例) セマンティックWebの検索

博多にあるラーメン屋を探す場合

住所のメタデータを元に判断



最近、開店したラーメン屋を探す場合

「最近」など時間に関するメタデータを元に判断

行列ができるラーメン屋を探す場合

オンライン雑誌などの評判に関するメタデータをもとに判断。

# (例) Webサービスでは

例: スtring型の「name」に対してint型の数値データを返す  
「getInt」メソッドが存在する場合。

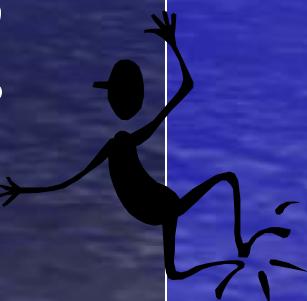
これだけでは関連付けがわからない。

オントロジを加えると…

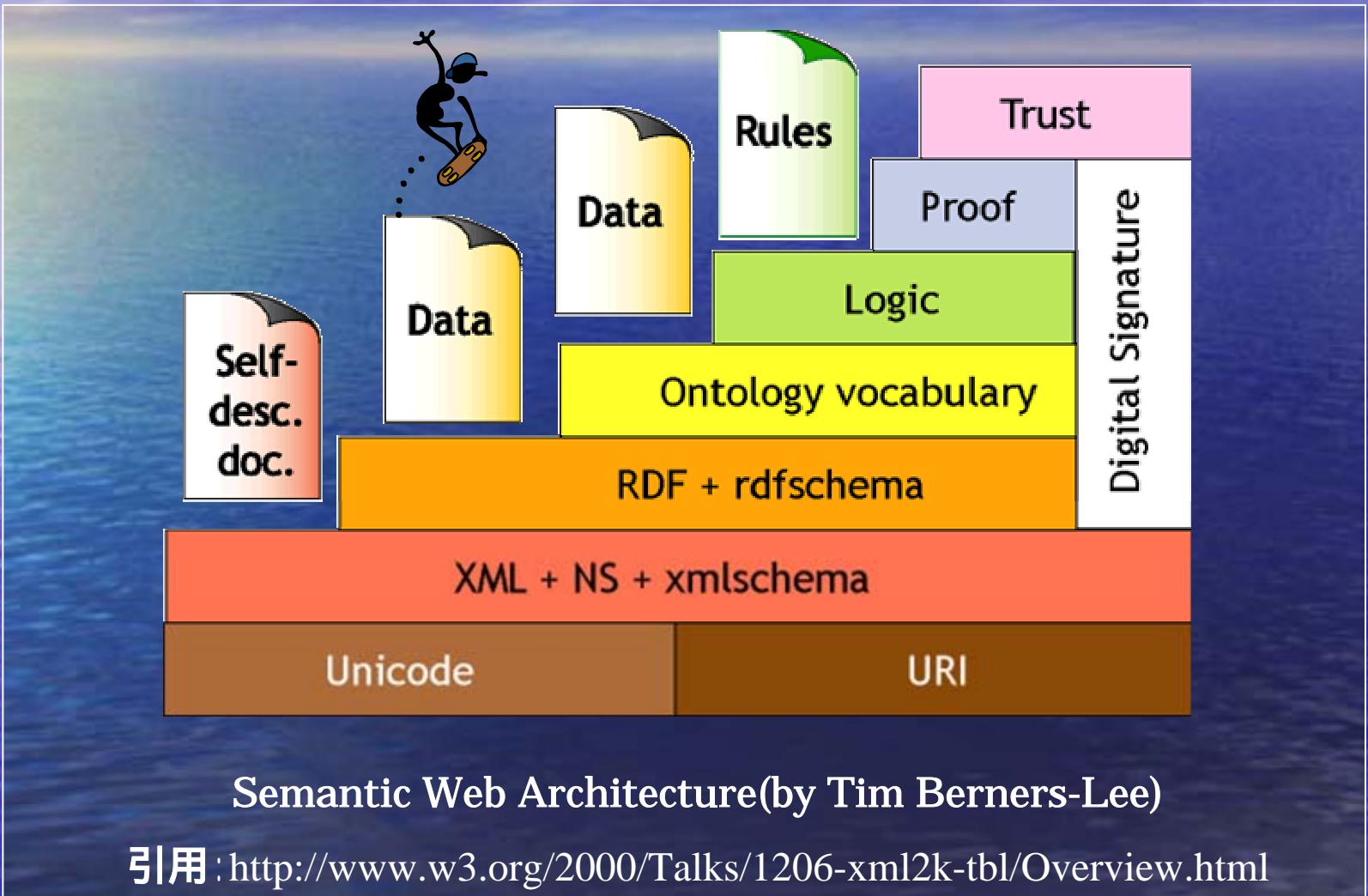
回答1: 「お見合いサービス」に相手の名前を渡す際、財産に関するオントロジを付加することによって、その人の年収が返ってくる。

回答2: 名前「赤ちゃん」をサービスに渡したら、「九九の答え」が返ってくる。

これが「かけあわせた」  
結果です。



# セマンティックWebアーキテクチャ



# オントロジ言語



- RDF  
*(Resource Description Framework)*
- RDFS  
*(RDF Schema)*
- DAML + OIL  
*(DARPA Agent Markup Language + Ontology Inference Layer)*
- DAML-S  
*(DAML-Service)*

# オントロジ言語(RDF)



*Ora Lassila is the creator of the resource <http://www.w3.org/Home/Lassila>*

「Ora Lassilaさんはリソース<http://www.w3.org/Home/Lassila>の作者です。」

```
<rdf:RDF>  
  <rdf:Description about="http://www.w3.org/Home/Lassila">  
    <s:Creator>Ora Lassila</s:Creator>  
  </rdf:Description>  
</rdf:RDF>
```





# オントロジ言語(RDFS)

<!--例: クラスManの定義-->

```
<rdfs:Class rdf:ID="Man">
    <rdfs:subClassOf rdf:resource="#Person"/>
    <rdfs:subClassOf rdf:resource="#Male">
</rdfs:Class>
```

<!--例: プロパティhasParentの定義-->

```
<rdfs:Property rdf:ID="hasParent">
    <rdfs:Domain rdf:resource="#Person"/>
    <rdfs:range rdf:resource="#Person">
</rdfs:Property>
```

<!--例: プロパティhasFatherの定義-->

```
<rdfs:Property rdf:ID="hasFather">
    <rdfs:subPropertyOf rdf:resource="#hasParent"/>
    <rdfs:range rdf:resource="#Man">
</rdfs:Property>
```

# オントロジ言語(DAML+OIL)



```
<daml:Class rdf:ID="MarriedPerson">
  <daml:intersectionOf rdf:parseType="daml:collection">
    <daml:Class rdf:about="#Person"/>
    <daml:Restriction daml:cardinality="1">
      <daml:onProperty rdf:resource="#hasSpouse"/>
    </daml:Restriction>
  </daml:intersectionOf>
</daml:Class>
```

<http://www.daml.org/2001/03/daml+oil-ex#MarriedPerson>

# オントロジは推論



オントロジは個々にロジックがあり、個のつながりにロジックがあり、個のつながりの流れにロジックがなければならぬ。

演繹法

```
<daml:Class rdf:ID="ソクラテス">  
  <daml:intersectionOf rdf:parseType="daml:collection">  
    <daml:Class rdf:about="#Person"/>  
    <daml:Restriction>  
      <daml:onProperty rdf:resource="#人間は必ず死ぬ"/>  
    </daml:Restriction>  
  </daml:intersectionOf>  
</daml:Class>
```

# エージェント



人間の代理としてエージェントを組み込むことにより、機械同士の会話が可能。

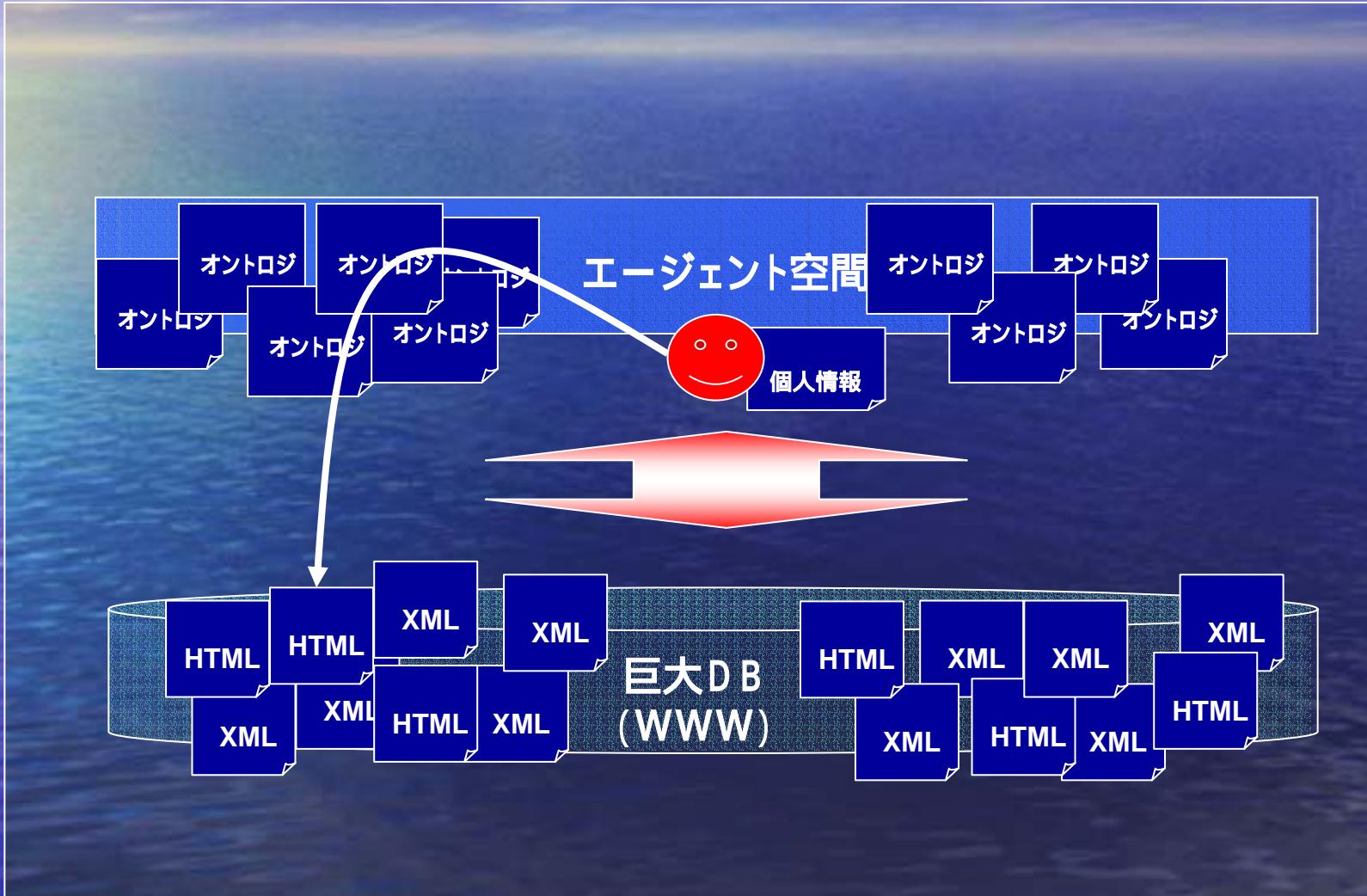
エージェントは複数存在し、各ジャンルに特化する。

エージェントが伝っていくオントロジの深さに制限が必要。

エージェントに個人情報を持たせる。



# エージェント空間とWebの関係



# セマンティックWebの問題点



オントロジ言語があまりにも複雑で、個人の価値観によって書き方が大きくことなってしまう。オントロジ自動生成ツールなどが必須。

セマンティックWeb普及のためのキラーアプリケーションがない。

人間が意味情報を付加するため、確実な証拠や信頼性の証明が難しい。

# RDF記述は信頼できるか？

## 信頼できる情報

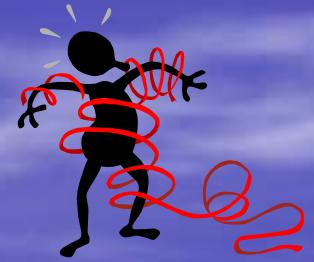
```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Lassila</s:Creator>      ここ正直
  </rdf:Description>
</rdf:RDF>
```

## 嘘の情報

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:s="http://description.org/schema/">
  <rdf:Description
    about="http://www.w3.org/Home/Lassila">
    <s:Creator>Ora Gokuu</s:Creator>      ここうそつき
  </rdf:Description>
</rdf:RDF>
```



# ではどうするの？



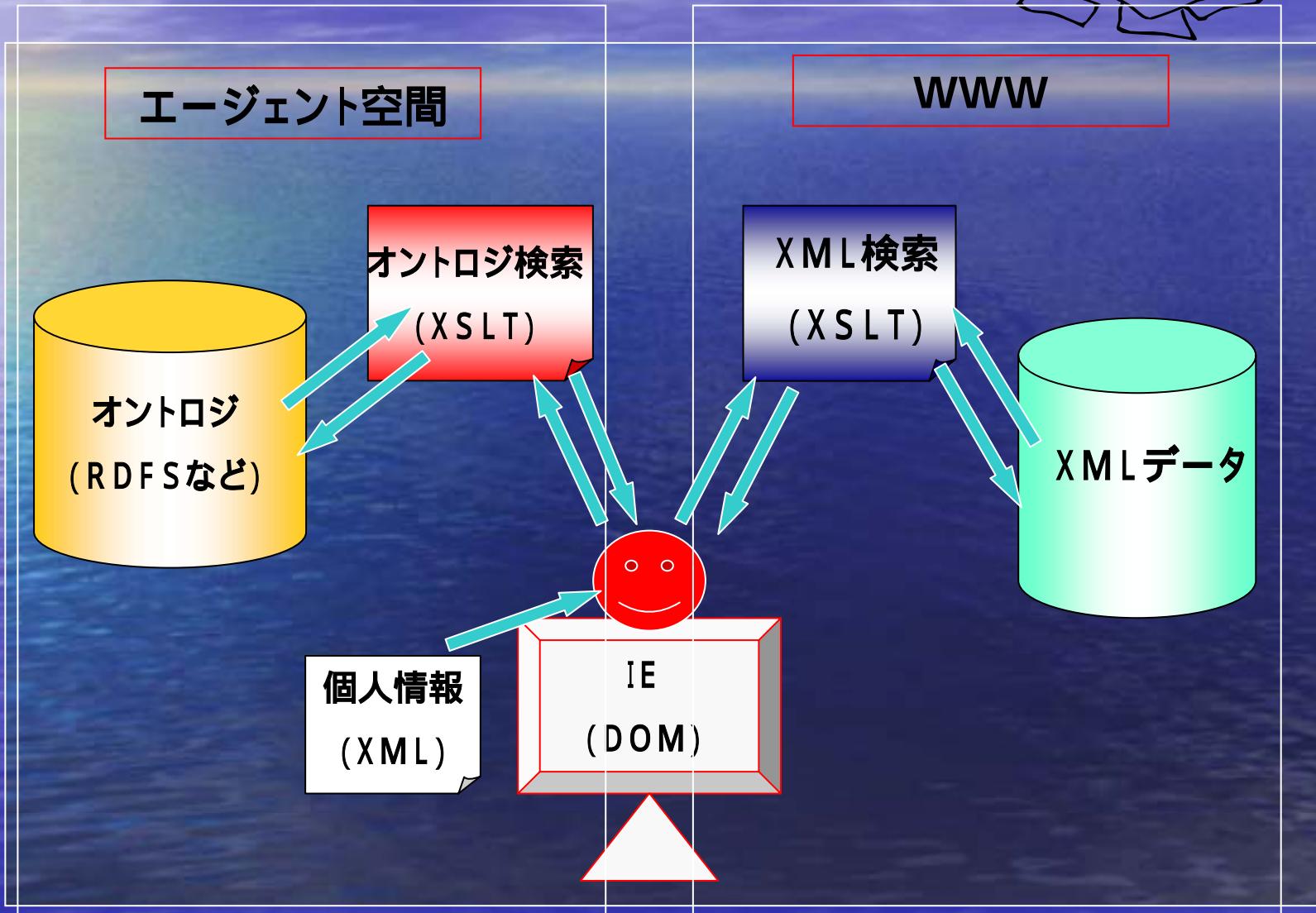
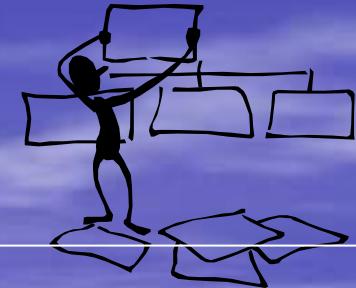
はじめは信頼できる場所から普及させていく。

イントラネット、エクストラネットなど最初から信頼が確立されている場所から普及する。

オントロジ作成にあたり何か付加価値をつける。

オントロジを一般向けに公開もしくは作成するためには付加価値が必要。

# DEMOの構成



# デモファイルの説明



インターフェース : **search.html**

オントロジ検索エンジン : **rdfsearch.xsl**

Web検索エンジン : **search.xsl**

個人情報ファイル : **personinfo.xml**

オントロジ記述ファイル : **kensaku.rdf**

音楽データファイル : **data1** (歌、演歌、氷川きよし、きよしのズンドコ節 )

: **data2** (歌、ポップス、SMAP、freebird、)

: **data3** (歌、演歌、水森かおり、東尋坊 )

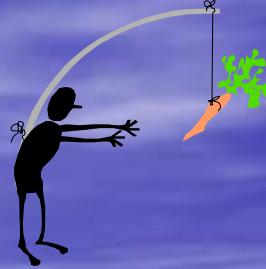
: **data4** (歌、ポップス、宇多田ヒカル、SAKURAドロップス)

: **data5** (歌、洋楽、マライアキャリー、恋人たちのクリスマス)

: **data6** (歌、童謡、不明、うさぎとかめ )

: **data7** (歌、童謡、不明、線路はつづくよどこまでも )

# 全文検索とオントロジ検索



## 全文検索

テキスト内にキーワードとすべて等しい、もしくは一部が等しい場合のみ結果を返す



検索する人間のボキャブラリーに影響される。

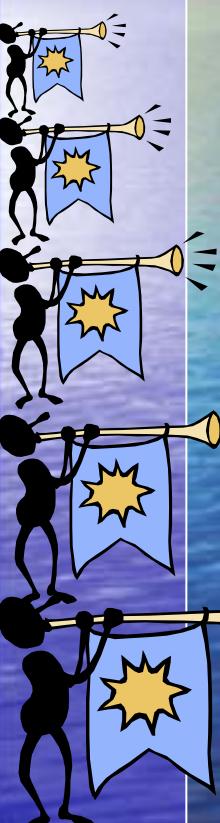
## オントロジ検索

キーワードと個人情報を元に、ある程度推論した結果を返す



検索する人間のボキャブラリーの影響は少ない。

# 今後のセマンティックWeb

- 
- 現状では業者に任せるほうがよい  
個人レベルで作成するよりも、正確なオントロジを作成できる。
  - 付加価値としてデータの蓄積  
データマインニングに利用可能
  - モバイル環境とセマンティックWebの融合  
携帯電話の個人認識機能とCRMのメタデータを関連つける。

# まとめ



## 直感的意見

セマンティックWebは大きな可能性を秘めているが、実現には時間がかかりそうだ。(アプローチ: ピックピクチャーからパズルピースへ)

## 想像力を逞しくしたときの意見

セマンティックがWebを巨大なDBにすることが可能。

+

エージェントは人間よりも多くの知識を持ち、考えることができる。

+

コンピューター(ロボット)にエージェントを搭載し、WWWに常時接続



は



に



は



に