

セマンティックWebコンファレンス 2011

オントロジーを用いたセマンティック検索の試み： 気象庁防災情報XMLの例

2011年3月4日 慶應義塾大学 三田キャンパス

先端IT活用推進コンソーシアム(AITC)
コンテキスト・コンピューティング(CC)研究部会
湯本 正典

本発表の共同研究者

- AITC 田原春美
- AITC CC研究部会 牧野友紀、和泉憲明、小林茂

セマンティックWebの研究の背景・目的



- 背景 「クラウド」「データ爆発」
 - 豊富なITリソース、爆発するデータ
 - オープンガバメントへの流れ到来
 - 「人工知能」の研究が活きる環境が到来？
- 目的 「情報価値の創造」
 - 同じデータを違うユーザーに見せ方を変えて提供し、価値ある情報とする。
- 課題 「ユーザーのコンテキストを反映した情報提供」
 - ユーザーのコンテキストを理解して、情報提供、特に検索の結果に反映させる。

「いつでも、どこでも、だれにでも」使えるITで、
「いまだけ、ここだけ、あなただけ」のサービスを提供しよう！

- アプローチ 「セマンティックからコンテキストへ」
 - XMLで記述されるメタデータ間の関係を明確にする。
 - それにより、「意味」を付加して処理するセマンティック技術
 - 個人個人にとつての「意味」を表現できるか？
 - コンテキスト・コンピューティングへ。
- 実証実験例「気象庁防災情報XML」
 - 住民が「わがことと思える防災情報」の提供を目指して！



XML Consortium

2009年度XMLコンソーシアムWeek

次世代Web活用部会

2010年度 第3回

セマンティックWeb委員会

雨量情報のセマンティック処理

2010-06-25

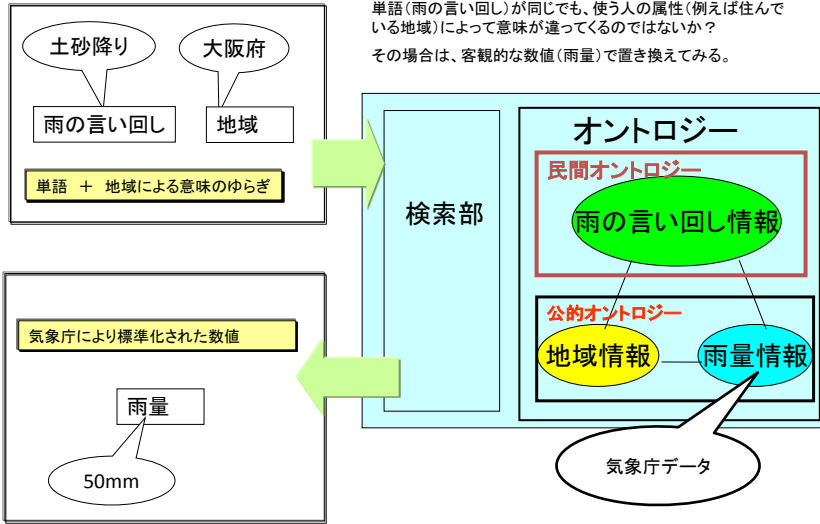
XMLコンソーシアム コミュニティ
田原春美 荒本道隆 芦田尚人

次世代Web活用部会
小林茂 湯本正典 西一嘉

雨量情報のセマンティック検索

仮設(仮説):

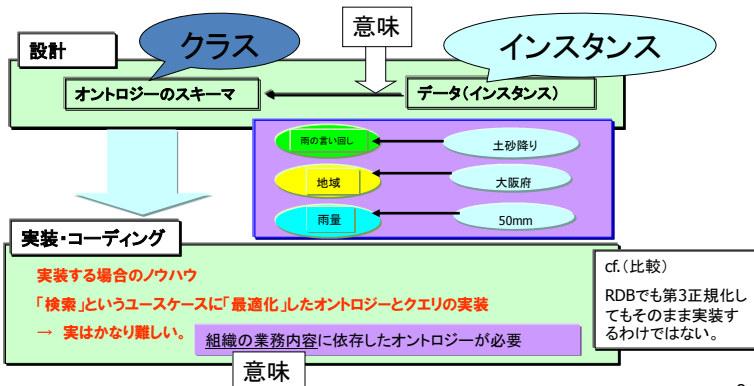
単語(雨の言い回し)が同じでも、使う人の属性(例えば住んでいる地域)によって意味が違ってくるのではないかと。その場合は、客観的な数値(雨量)で置き換えてみる。



セマンティック検索機能の設計と実装

- セマンティックシステムを実現するための技術要素

	クエリ	リソースを記述する枠組み
記述言語	SPARQL	RDF, RDFS, OWL
Cf. 従来の構造化の手法	SQL	RDB



cf.(比較)
RDBでも第3正規化してもそのまま実装するわけではない。

セマンティック検索の実装の一例

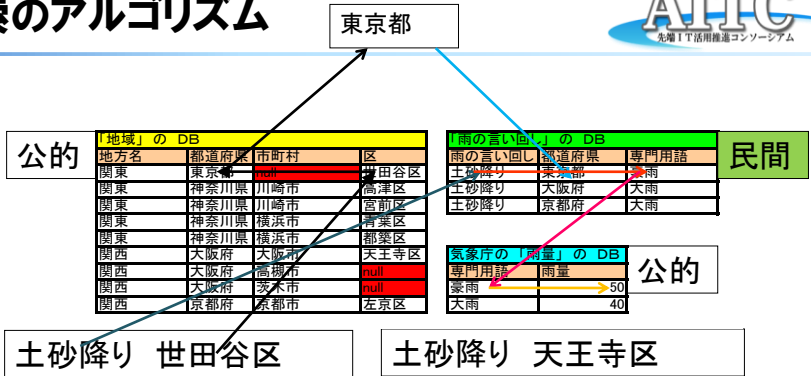


公的	「地域」の DB			「雨の言い回し」の DB			民間	
	地名	都道府県	市町村	区	雨の言い回し	都道府県		専門用語
	関東	東京都	世田谷区	世田谷区	土砂降り	東京都		豪雨
	関東	神奈川県	川崎市	高津区	土砂降り	大阪府		大雨
	関東	神奈川県	横浜市	青葉区	土砂降り	京都府		大雨
	関東	神奈川県	横浜市	都築区				
	関西	大阪府	大阪市	天王寺区				
	関西	大阪府	高槻市	null				
	関西	大阪府	茨木市	null				
	関西	京都府	京都市	左京区				
	気象庁の「雨量」の DB			公的				
	専門用語	雨量		豪雨	50			
				大雨	40			

土砂降り 世田谷区 土砂降り 天王寺区

- 「雨の言い回し」(文字列入力) + 「地域名」(文字列入力) から 「雨量」 を導く検索。
- 「世田谷区」から直接「東京都」が計算機にわかるように、また、「天王寺区」から直接「大阪府」が計算機にわかるように、「地域」オントロジーで、<owl:TransitiveProperty> (推移)を用いる。
 - 推移とは[A→B、かつ、B→C、ならば、A→C]である。
 - 地域オントロジーで推移を実装すると
世田谷区→東京都
天王寺区→大阪市、大阪市→大阪府、よって、天王寺区→大阪府
と計算機が処理できるようになる。

検索のアルゴリズム



- の過程の検索を実現させるための手段
 - RDBとSQLで実装する。SQLのロジックが複雑になる。
 - XML文書と呼び出し言語で実装する。呼び出し言語が非常に複雑になる。
 - RDF/RDFS/OWLとSPARQLで実装する。OWLにより、SPARQLの実装を簡素化できる。
- 検索プロセスを、ロジックに実装するか？それとも、DBに実装するか？

「検索」の実現方法の考察



- 検索を、ロジックに実装するか？DBに実装するか？
 1. RDB:ロジック、DB両方に硬く実装する。
 2. XML文書:DBは非常に柔らかく、ロジックは非常に硬く実装する。この場合にDBに硬い構造を入れる試みとして、XMLスキーマがある。
 - 例外:keyValue
 3. セマンティック技術:
 - A) XML文書に新たに「ロジック」のタグを追加する。(今回の発表のOWLに相当する。)
 - B) XMLの本文のメタデータの間関係を明記するような、メタファイルを用意する。(今回の発表の「民間オントロジー」に相当する。)
- 検索(参照)の比較

検索サービス	検索のアルゴリズム	対象データ	データの関連性のルール
google	PAGEランク アルゴリズム	統計処理 非構造化	すべてのデータ ルールなし
Yahoo!の将来系? ツリー構造検索など	SPARQL言語	論理処理 脱構造化	オントロジー LOD 専門家によるルール
Oracle	SQL言語	論理処理 構造化	RDB 特定業務のルール

結論



- 業務用のRDBは、業務知識を反映し、特化した、業務特有の構造を持っていて、複数の知識を結合したデータを初めから格納しておく。検索ロジックは、RDBが変わるたびに修正する。「硬いDB」と「硬い検索」で正確性が非常に高い。
- XML文書は、メタデータによりその単語の属性は定義できるが、単語と単語と間関係は、RDBのように簡単には実装できない。「柔らかいDB」すぎて検索が困難な場合がある。
- 「柔らかいDB」は、RDFとRDFSとOWLをつかって、単語間関係を表記して、オントロジーにて構築する。この場合、検索のロジックが簡素化され、しかも、オントロジーが変わっても、比較的柔軟に対応できる。適度な柔軟さとしなやかさがある。ロバストなDBと検索システムの実現を目指す。

DBの構造	格納構造	要素間関係意味付け	検索	変更作業	実装手順
RDB	硬い	硬い	正確	複雑・困難	第3正規化
XML文書	柔らかい	ほとんどできない	不正確	容易	定型的には無い
RDF/OWL	柔らかい	柔らかい	正確	容易	定型的には無い

考察① 意味検索の実装

公的	【地域】の DB				【雨の言い回し】の DB			民間
	地方名	都道府県	市町村	区	雨の言い回し	都道府県	専門用語	
	関東	東京都	autil	世田谷区	土砂降り	東京都	豪雨	
	関東	神奈川県	川崎市	高津区	土砂降り	大阪府	大雨	
	関東	神奈川県	川崎市	宮前区	土砂降り	京都府	大雨	
	関東	神奈川県	横浜市	青葉区				
	関東	神奈川県	横浜市	都築区				
	関西	大阪府	大阪市	天王寺区				
	関西	大阪府	高槻市	autil				
	関西	大阪府	茨木市	autil				
	関西	京都府	京都市	左京区				

気象庁の「雨量」の DB		公的
専門用語	雨量	
豪雨	50	
大雨	40	

- RDBは、「論理構造(意味構造)」を「行列の物理構造」に射影している(第3正規化)。これを用いて、“いろいろなユースケース”(非定型業務)の意味検索をすることには、無理がある。
- 無理があるので、検索のロジックが複雑になる。
- ロジックが複雑になると、運用・保守に手間がかかる。

11

考察② 論理構造(意味構造)

- 論理構造(意味構造)は、物理構造としては柔らかい「ネットワーク構造」で示すのが便利。特に、階層構造が使えると嬉しい。
- 「ネットワーク構造」は、ノード(要素)とリンク(関係)で表現できる。
- XMLではノード(要素)の意味は、メタデータ(タグ)で正確に表現・実装できる。しかし、現在のXMLでは、リンク(関係)を実装し、さらにその意味を表現・実装することが難しい。
- XMLスキーマ言語の標準化の停滞を招いた。
- ユーザがDBに合わせなければならないような、“硬い構造”はもうたくさん。“柔らかい構造”は使えないのか？

12

考察③ 柔らかい物理構造

- ネットワーク
 - “柔らかすぎず”、“硬すぎず”が大切と思われる。
 - したがって、Linking Open Data (LOD) が台頭してきた。
 - やはり、Googleの考え方が活きている。すなわち、スケールメリットが有効になる場合の考え方である。
 - スケールメリットが有効になる場合の考え方とは
 - » リンクの数でランキングを決める(ページランク)、これは、Citation Indexと同じ考え方である。
 - » リンクのネットワーク構造は「集合知」で決まる。
 - しかし、専門家の知恵を反映した、賢いネットワークの実現を目指す。
- セマンティックWeb
 - W3Cで標準化が進む。基本的に従来のXMLにタグを追加した感じ。
 - RDF 主語、述語、目的語の関係を表現する。
 - RDFS スキーマ(構造)をオブジェクト指向で定義する。
 - OWL 関係性を定義・表現する。

13

結果

- 非常に簡単ではあるが、オントロジーと検索処理を実装した。

土砂降り 世田谷区	→	50mm
土砂降り 天王寺区	→	40mm

- セマンティック技術、特に、OWLの活用により、オントロジーの場合にも、RDBと同様のピンポイント検索ができるだろう。
- ただし、OWLを導入するフィールドには任意性がある。また、それによってSPARQLの記述が変わってくる。

14

今後の展望



- システム設計
 - オントロジー実装の「デザインパターン」と、それを簡単に実装するための(RDBにおける第3正規化のような)簡単な手順を追究する。
- サービス設計
 - 各ユーザ(ステークホルダー)の「コンテキスト」(文脈、背景など)を意識した検索の実装を目指す。
 - 気象情報の場合
 - 東京都の「大雪」と新潟県の「大雪」は住民の経験により違う。
 - 千葉の南部と北部で気候が同じか？行政区分と気候区分の違い。
 - オープンガバメントによる公共情報を個人化する。

先端IT活用推進コンソーシアム

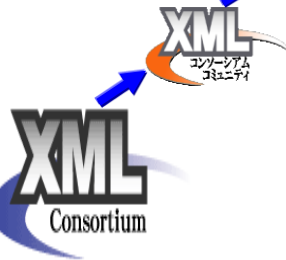


2010年9月8日設立
 企業における先端ITの利活用と技術育成の推進
 社会への貢献を目指して、
 企業の枠を超えた活動
 一社ではできない活動(実証実験等)
 利活用推進のための提案・提言
 情報と知見の提供
 成果物の公開



- 2010年度の活動対象:
- クラウド・テクノロジー分野
(Hadoop, NoSQLなど)
 - インターネットデバイス・アプリケーション
開発環境分野
(iPhone, iPad, Androidなど)
 - リッチ・インターネット・アプリケーション
(RIA) 分野
(HTML5, Ajaxなど)
 - コンテキスト・コンピューティング分野
(Semantic Web, Service Engineeringなど)
 - ソーシャル・コミュニケーション分野
(Twitter, Facebook, OpenSocialなど)
 - 実世界と情報世界の重ね合わせ分野
(ARなど)

- Windows
コンソーシアム
- Java
コンソーシアム
XML 部会
- 日本経営協会
XML フェスタ



2000/07 設立宣言 2001/06~2010/03 実活動 2010/03~2010/09 2010/09/08 設立 16

ご案内



- 近々の催事情報

2月23日(水) 13:00~18:00

「AITC Day 2011 (第一回中間活動報告会)」

@日立ソリューションズ 本社別館 20F セミナールーム

4月14日(木) 13:30~18:00

「ARセミナー」

@日立ソリューションズ 本社別館 20F セミナールーム

- 月例の勉強会や部会活動、その他の情報は、

<http://aitc.jp/>

- ニュース配信MLへのご登録は、

<http://aitc.jp> トップページでのメーリングリストから。