

# セマンティック技術に関する 調査活動TF報告

津田宏, 齊藤孝広(富士通研究所),  
佐藤 宏之(NTT),  
高梨益樹(富士通ソフトウェアテクノロジーズ),  
乙守信行(ジャストシステム),  
渡邊 圭輔(三菱電機),  
細見格(NEC)

1

## 1.調査対象

- Semantic Web
  - OWL, RDF, SparQL, ... : W3Cベースの規約を元にした取組み
- Semantic Web的な技術
  - TopicMap, microformats, ... : オントロジー、メタデータを元にした取組み
- Semantic技術
  - Powerset, twine, ... : 上よりも広い、何か知的なことができる技術/サービス
- まずは
  - そもそも「セマンティック技術」について世の中が何を期待しているのか: 本コンファレンスでアンケート
  - おそらくセマンティック技術が大きく関係するであろう「検索技術」に関して世界的にどのような動きがあるのか:
    - 日本: 情報大航海、情報大爆発
    - フランス: クエロ
    - ドイツ: テセウス
    - ノルウェー: iAD
    - EU: ファロス計画

2

## 2. セマンティック技術への期待

- そもそも「セマンティック技術」に対して、世の中が何を期待しているのか、本コンファレンス出席者に事前アンケート。結果はコンファレンス当日に発表します。
- 質問1: 所属されている組織ではどのようにWebを利用していますか? (複数選択可)
  - 正式な社内情報の連絡
  - 個人ホームページ, ブログ
  - SNS
  - 検索インタフェース
  - スケジューラ・グループウェア
- 質問2: あなたの職種と、どんな業務でWebや検索技術を使っていますか?
- 質問3: 仕事でWeb技術を使う上での不満は? (最大3つ選択可)
  - 量が多すぎて情報が探せない
  - 同義語のゆれなどにより、キーワードでうまく探せない
  - 情報の信頼性を自分で判断する必要がある
  - 新しい情報と古い情報が混在している
  - 情報の変化、経過がわからない
  - 一度見た情報を後で探そうとすると大変

3

- 質問4: 「セマンティック技術」として期待するものは? (最大3つ選択可)
  - 大量に作られるログ的な情報を次々と整理して溜めてくれる
  - わざわざ検索しなくても必要な情報を勝手に教えてくれる
  - 自然言語や文章から、情報を検索できる
  - 様々な言語の情報を探して日本語で結果をくれる
  - 複数の情報源を一度で意味的に横断して検索してくれる
  - 製品名などろ覚えで入力しても適切に近そうなものを検索してくれる
  - 専門的な内容でも易しい言葉で探することができる
  - 状況(場所, 時間, etc.)にあわせた検索
  - 内容の信頼性によって区別して返してくれる
  - 新しい情報が追加されると古い情報は自動で消してくれる
  - 検索結果をリストだけでなく、表とか色々な形で見える化してくれる
  - 文書だけでなく、人とかモノ、ノウハウ(動画)が検索できる
  - その他
- 質問5: 企業内でセマンティック技術の導入が難しい要因は何でしょう? (複数選択可)
  - 技術的な問題
  - 金額的な問題 ( )万円くらいでないと厳しい
  - 運用面の問題
  - 効果が不明

4

## 3.1 情報大航海 (日本)

- プロジェクト名称
  - 情報大航海プロジェクト(Infomation Von Voyage)
- 期間、予算
  - 2007年～2009年
  - 150億円(3年間)
- 内容
  - 個人の生活、ビジネスなどあらゆる局面で必要な時に必要な情報を提供するための「次世代検索・解析技術」をとサービス事業について実証。経産省主導。
  - 「次世代検索・解析技術」は、実際に役に立つセマンティック技術の開発を狙いとしている。沖電気工業(株)のラダリング型検索サービス「ラダサーチ」は、深いドメイン知識と、自然文で入力されるユーザの特性や現在の状況を自然文による対話戦略でマッチングさせて適切な情報を提供するサービス

5

## 3.2 情報爆発 (日本)

- プロジェクト名称
  - 情報爆発時代に向けた新しいIT基盤技術の研究
- 期間
  - 2006-2010
- 内容
  - 情報爆発時代に向けた先進的なIT基盤技術の構築。大航海プロジェクトと比べて、学術的な取組み。文科省主導
  - A01 情報爆発時代における情報管理・融合・活用基盤
    - Semantic Web関係: 人工物オントロジーに基づく製造業における技術ドキュメントの融合(阪大)、メタデータの創造的流通(NII)
  - A02 情報爆発時代における安全・安心ITシステム基盤
  - A03 情報爆発時代におけるヒューマンコミュニケーション基盤
  - B01 情報爆発時代における知識社会形成ガバナンス

6

## 3.3 クエロ (フランス)

- プロジェクト名称
  - Quaero(クエロ)
- 期間、予算
  - 2006.4～準備、2008.3～(5年間)
  - 2.5億ユーロ
- 担当する主な組織
  - Thomson社(フランスのマルチメディア・家電企業)他23社(仏独)
  - ただし、ドイツは2006年12月脱退し、独自にTheseusプロジェクトを開始
- 内容
  - 対Google的な性格で仏独政府主導で始まった検索に関するプロジェクト。5つのアプリケーションプロジェクト
    - デジタル化とコンテンツ強化
    - デジタル権利管理
    - パーソナライズされた動画
    - マルチメディア検索エンジン
    - PC、モバイルポータルからの動画等へのアクセス

7

## 3.4 テセウス(ドイツ)

- プロジェクト名称
  - テセウス(THESEUS)
- 期間、予算
  - 2006.12～(基礎技術)、2009～(5年間、TESEUS SME)
  - 200万ユーロ(2009-)
- 担当する主な組織
  - フラウンホーファー協会、DFKIをはじめとする研究機関と、ドイツ国立図書館、empolis社、ライコスヨーロッパ、SAP、シーメンスなどを中心にアプリケーションシナリオを担当
- 内容
  - 「Web3.0=Web2.0+意味」「サービスとしてのインターネット」として、インターネットベースでの知識処理技術をインフラとした6つのシナリオ
  - 1. ALEXANDRIA インターネットにおけるナレッジプラットフォームの開発
  - 2. CONTENTUS 文化遺産保護およびマルチメディアライブラリに向けての技術開発
  - 3. MEDICO 医療分野でのスケーラブルなイメージ検索
  - 4. ORDO デジタル情報処理としてテキストに対するセマンティックアノテーションと利用、情報の可視化アプリの開発を行う
  - 5. PROCESUSS ビジネスプロセスの最適化をセマンティック技術を用いて行う
  - 6. TEXO 「サービスとしてのインターネット」によるビジネスWEBのインフラ構築

8

## 3.5 PHAROS(EU)

- プロジェクト名称
  - The PHAROS Project (Platform for search of Audio-visual Resources across Online Spaces)
- 期間、予算
  - 2007年1月1日から3年間 (EUの第6次フレームワークFP6での計画)
  - 1,425万ユーロ (ファロス計画全体の総額)
- 担当する主な組織
  - Engineering Ingegneria Infomatica S.p.A. (イタリア), Fast Search & Transfer (ノルウェー), France Telecom (フランス), Fraunhofer IDMT (ドイツ) 他、計9カ国12団体で PHAROS Consortium
- 内容
  - Audio/Visualデータを含む次世代マルチメディア検索プラットフォームの開発
  - 従来のプリミティブな情報検索エンジンから総合的な検索プラットフォームへのパラダイムシフトを狙っているという。マルチメディア・コンテンツからセマンティックな情報を抽出し、メタデータとしてコンテンツに付与する自動アンノテーション技術の実現を、主要な目標の1つとしている。これまでの主な成果の1つとしてAV-RSSというAVコンテンツ向けRSSがある。

9

## 3.6 iAD (ノルウェー)

- プロジェクト名称
  - iAD (information Access Disruptions)
- 期間、予算
  - 2007年から8年間
  - 240百万NOK (参考 約32.6億円 2/16時点)
- 担当する主な組織
  - Fast Search & Transferがホスト
- 内容
  - 情報アクセス分野における、次世代の精度 (precision)、分析 (analytics)、スケール (scale) のためのコア研究
    - 1. 構造化、非構造化、マルチメディアなどの混在データに対する、情報アクセスサービスに最適化した索引生成
    - 2. スケーラブルかつ対障害性の高いシステムのアーキテクチャの開発。高速なデータストリームからの情報のキャプチャ、クリーニング、抽出のためのデータ処理およびデータマイニングのプラットフォームを含む。
    - 3. 分散情報アクセスのための次世代インフラの開発と実環境での評価
    - 4. マルチメディア情報アクセスのための高精度なソリューションの開発
    - 5. 情報アクセスそのものや情報アクセスソリューションによって可能となる破壊的過程 (disruptive processes) を明らかにすること

10