



## 次世代Web情報流通エンジンを用いた 人材マッチングシステム

NTT情報流通プラットフォーム研究所<sup>†1</sup> / NTTソフトウェア<sup>†2</sup>

おおとも けんじ<sup>†1</sup>   むらやま たかひこ<sup>†1</sup>   せのお まさみ<sup>†2</sup>   こくぼ ゆういち<sup>†2</sup>  
大友 健治 / 村山 隆彦 / 妹尾 正身 / 小久保 祐一

複数の情報源から抽出した多種多様なデータ群をRDF (Resource Description Framework) のグラフ構造を持つような表現に統合、解析し新たな知識の発見を支援する次世代Web情報流通エンジンを企業内データ群に適用した共同実験を実施し、人材マッチングの分野において良好な結果が得られました。

### セマンティックWebの広がり

セマンティックWebとは、Tim Berners-Lee氏によって提唱されているWebページに対して付与されるメタ情報（作成者、作成日、キーワード等）、意味を活用することにより、WWW (World Wide Web) の利便性を向上させようというプロジェクトのことです。当初メタ情報は誰が付与するのかといったことがあいまいで普及については懐疑的にみられていました。しかし最近のブログやSNSでは自動的に作成者等のメタ情報が付与され、またタグ付けということでユーザが積極的にメタ情報を付与し始めており、セマンティックWebが本格的に展開されることが期待されています。RSS (Rich Site Summary) による記事の自動配信等もセマンティックWebのサービスの1つとみることができます。我々もセマンティックWebで利用されるメタ情報に着目し、新たな知識・情報の抽出、発見を行う次世代Web情報流通エンジン (iMage: information Mixable graph explorer: アイメージ)<sup>(1)</sup> の研究開発を行っています。

### 次世代Web情報流通エンジン (iMage) とは

セマンティックWebにおいてメタ情報を記述するフレームワークとしてRDF

(Resource Description Framework)<sup>(2)</sup> という規格がW3C (World Wide Web Consortium) で策定されています。メタ情報を記述するためにRDFではさまざまなことを規定していますが、その特徴の1つにRDFで記述されたものはグラフとして表現できるということがあります。RDFにおける記述の基本単位はトリプルと呼ばれ、「ある論文の著者はXさんです」といったことが「ある論文」と「Xさん」という2つのノードを「著者」というアークが結ぶグラフとして表現されます。このような記述が多数集まり、メタ情報は大きなグラフとして表現されることになります。

我々はこのグラフの「構造」に着目し、メタ情報を表現する大きなグラフの中で特徴的に現れるパターンはきつと人間にとって意味のある有用な知識・情報を表しているに違いない、という仮説をたてました。コンテンツ (情報) そのものに対し自然言語処理等を行って有用な知識・情報 (評判やトレンド) を得ようという試みも多くなされていますが、そのような複雑な知識処理を行わなくても「構造」に着目すれば有用な知識・情報が得られるであろうという考えです。

図1にiMageの技術概要を示します。この例では研究者や、研究者が利用するSNSのデータを対象としています。メタ情報がノード間をアークが結ぶ大きなグラフとして表現されていることが分

かります。このグラフ構造を解析することにより、例にあるような「ある技術について詳しい人」や「ある技術についてSNSにおいて積極的に交流している人」等の情報を得ることができます。

### 人材マッチングへの適用

今回NTT情報流通プラットフォーム研究所とNTTソフトウェアは企業内のデータを対象とすることで、社内における人材マッチングにiMageの技術が適用できるのではないかと考えました。すなわち、社内には存在する管理データをiMageの対象とすることで、管理データの本来の利用目的を越えて人材マッチングという新たな利用、既存管理データの再活用が期待できると考えました。

#### ■共同実験

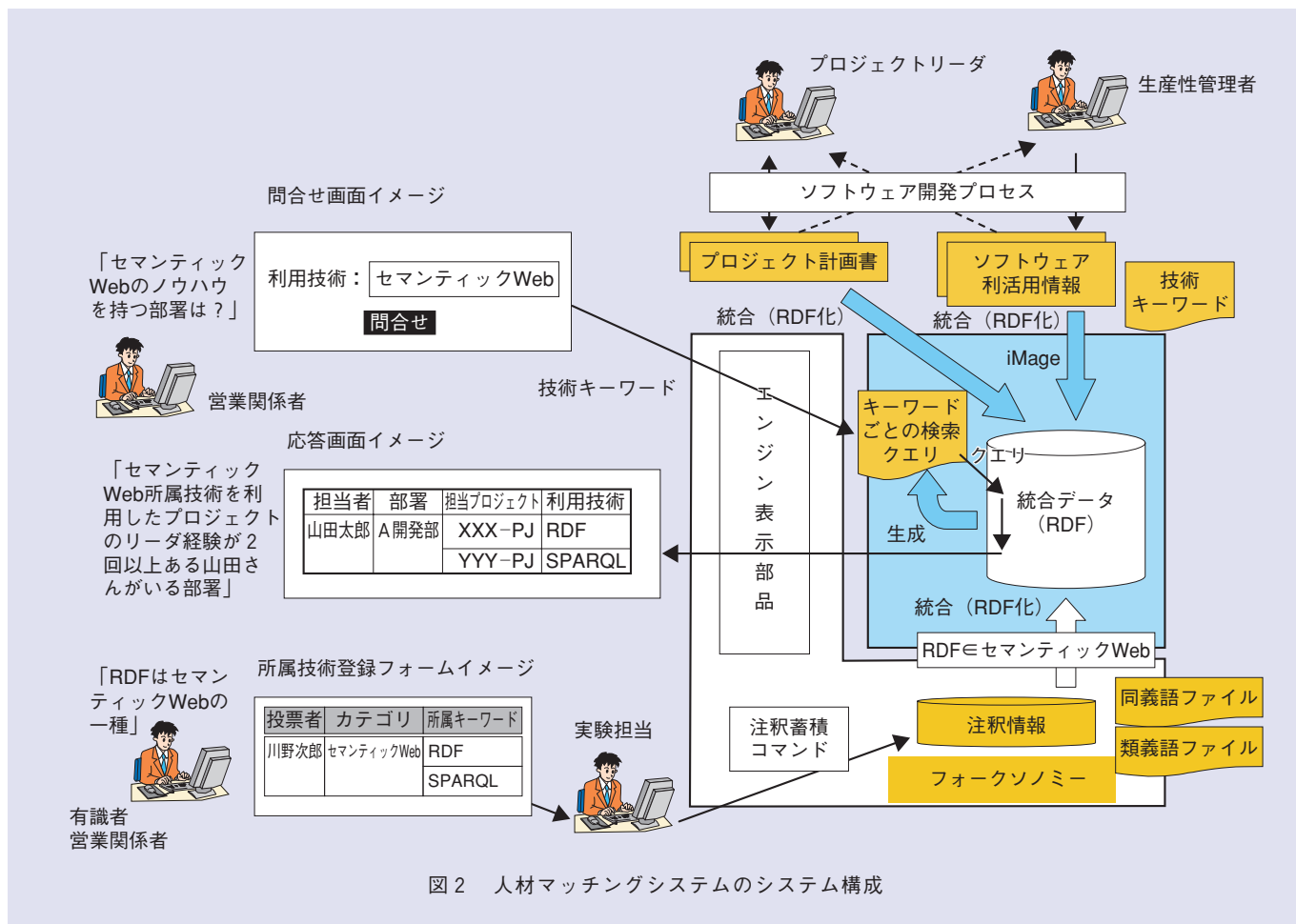
NTTソフトウェアの企業内データを利用した人材マッチングシステムの構築を行い、利用者へのアンケート等を通してそのサービス性の検証を行いました。

#### ■共同実験の概要

- (1) 営業向け技術・案件マッチングサービス

NTTソフトウェアなどのシステムインテグレータにおいては営業担当が営業先にて見聞きした技術に関し自社で対応可能なのか、より詳しい情報が自社内に存在するのを知りたいといった状況が発生します。そこで顧客から得た技術要件から問い合わせる部署が分からないと





します。人材マッチングシステムとして完成させるためには図中にある「エンジン表示部品」といったユーザインタフェース部分のアプリケーションをつくり込む必要があります。

#### (4) 特徴的パターン

iImageではグラフ構造として表現された対象データを解析し特徴的なパターンの抽出を行い、それをクエリとすることで知識・情報を獲得します。今回の人材マッチングシステムにおいてiImageによって抽出され実験担当者により選択された特徴的パターンを図3に示します。通常の検索では技術キーワードにマッチした人が一覧として得られるだけですが、特徴的パターンを用いて検索することにより、例えば技術キーワードに関連するプロジェクトに対して、上位管理者として(パターン1)、プロジェクトリーダーとして

(パターン2)、技術担当者として(パターン3)等、どのような立場で関係していた人なのかが分類されて得られます。

#### (5) 有効性を高める工夫

今回の人材マッチングシステムでは技術キーワードを与えたとき、その技術に関する有識者が漏れなく正しく検索されることがその有効性を決定することになります。この有効性を高めるために以下の2点の対応を行いました。

- ・フォークソノミー<sup>(3)</sup>による類義語対応
- ・表記のゆれ対応

フォークソノミーによる類義語対応とは、与えた技術キーワードが特徴的過ぎて有識者が見つからないとき、もう少し抽象度を上げて同じ技術に類する類義語とともに検索を行うことによりマッチする有識者を増やそうというものです。

図2の「所属技術登録フォームイメージ」では、「RDF」「SPARQL」が「セマンティックWeb」という技術分野における類義語として登録されています。これにより「SPARQL」という技術キーワードで有識者を検索してもマッチする有識者がいない場合、「SPARQL」とともに「RDF」という技術キーワードでも検索してマッチする有識者を探すことになります。こうした類義語の知識をユーザ自ら登録できるような仕組みをつくり、フォークソノミー的に有効性が高められるようにしました。

表記のゆれ対応というのは同義語に対する対応です。同義を表す言葉でも、漢字表記、平仮名表記、英文表記、さらには英文表記においても大文字、小文字の区別等さまざまな表記のゆれが存在します。表記のゆれに対しては代表語辞

書を用意し、対象となるRDFデータおよびユーザーが入力する技術キーワードの双方を代表語で正規化することで対応

しました。

■ヒアリングによる評価

評価は一部の利用者にヒアリングを行

うことで実施しました。図4にヒアリング対象者のプロフィールを示します。これらヒアリング対象者に顧客から得た情

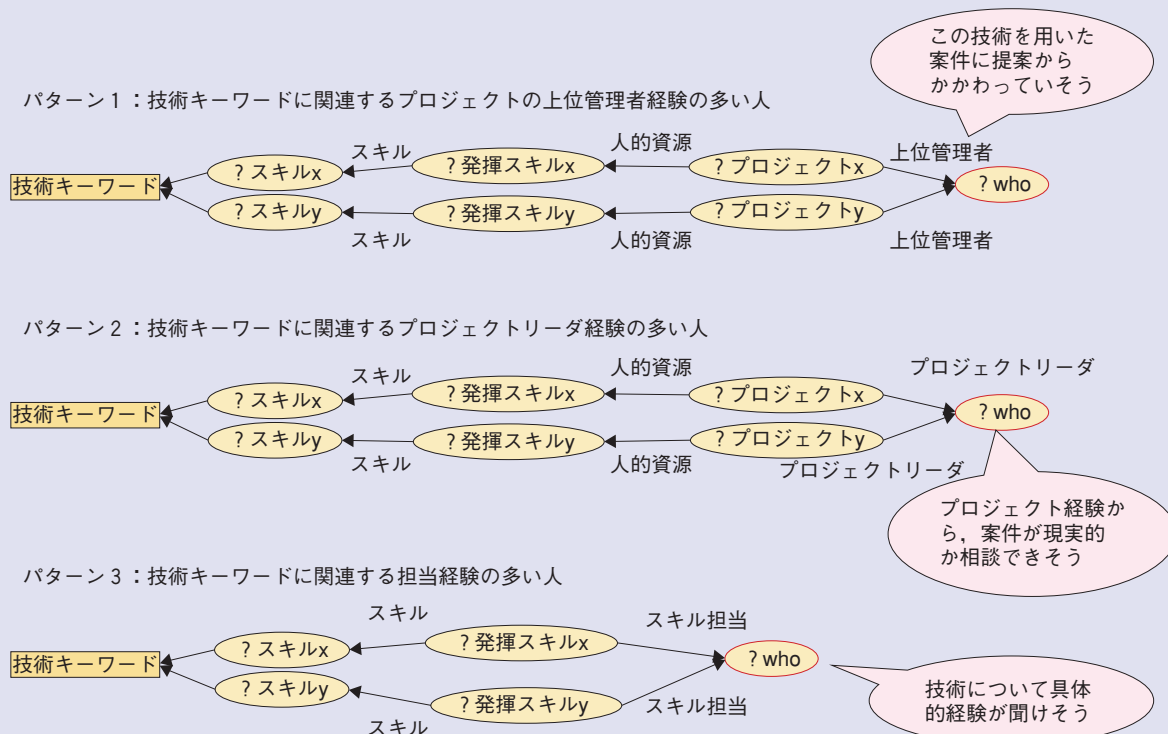
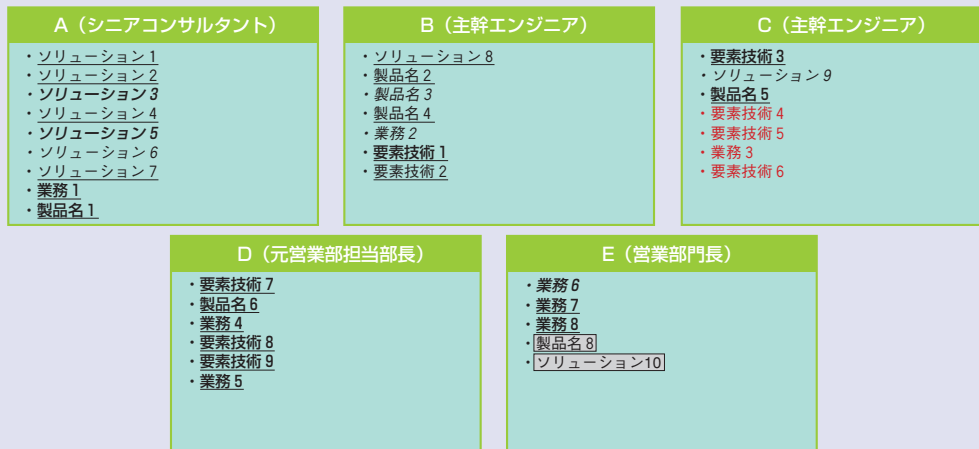


図3 人材マッチングシステムで用いた特徴的パターン

A (シニアコンサルタント)	B (主幹エンジニア)	C (主幹エンジニア)
<ul style="list-style-type: none"> <li>・NTTグループ歴：約15年</li> <li>・NTTソフトウェア歴：約15年</li> <li>・営業歴：約1年</li> <li>・今までの探し方：関係するソリューションの営業に声を掛ける、PL・営業担当に声を掛ける、上長の情報網、人脈</li> </ul>	<ul style="list-style-type: none"> <li>・NTTグループ歴：約7年</li> <li>・NTTソフトウェア歴：約7年</li> <li>・営業歴：約6年</li> <li>・今までの探し方：身近な有識者に聞く、ハイアラーキーをたどる</li> </ul>	<ul style="list-style-type: none"> <li>・NTTグループ歴：約26年</li> <li>・NTTソフトウェア歴：約4年</li> <li>・営業歴：約3年</li> <li>・今までの探し方：人づて</li> </ul>
D (元営業部担当部長)	E (営業部門長)	
<ul style="list-style-type: none"> <li>・NTTグループ歴：約34年</li> <li>・NTTソフトウェア歴：約20年</li> <li>・営業歴：約5年（営業支援）</li> <li>・今までの探し方：人づて</li> </ul>	<ul style="list-style-type: none"> <li>・NTTグループ歴：約30年</li> <li>・NTTソフトウェア歴：約15年</li> <li>・営業歴：約3年</li> <li>・今までの探し方：各種会議の情報に基に</li> </ul>	

図4 ヒアリング対象者のプロフィール



- ・キーワードは、ソリューション・要素技術・製品名・業務という種別ごとに連番で匿名化されている
- ・(A) 下線太字は有識者情報のみが得られ、技術があると知らなかった人が含まれていたキーワード
- ・(B) 斜字体太字は有識者情報と有識者ではない情報が混在し、技術があると知らなかった人が含まれていたキーワード
- ・(C) 下線は有識者情報のみが得られ、技術があると知っている人のみが含まれていたキーワード
- ・(D) 斜字体は有識者情報と有識者ではない情報が混在し、技術があると知っている人のみが含まれていたキーワード
- ・(E) 赤字は有識者情報が得られなかったキーワード
- ・(F) 灰色網掛け文字は検索結果が多すぎる等分類不能なキーワード

図5 ヒアリングの結果

報から問合せ先を探すときに利用した技術キーワードで検索をしてもらい、結果を実験担当者としてヒアリング対象者で確認することで評価を行いました。図5にヒアリングの結果を示します。下線太字で示された技術キーワードで検索をしたときにはその技術キーワードに関する有識者のみがマッチし、かつ検索を行ったヒアリング対象者の知らない有識者もマッチしたということで、もっとも有効な検索が行えたケースだといえます。検索結果のうち80%で有識者が得られていることと合わせ、人材マッチングシステムの想定する利用対象者が技術や技術者を熟知していない営業担当者であり、営業職であるヒアリング対象者D氏、E氏において下線太字の技術キーワードが多いこと等から本システムが業務上有効であるとの評価を得ることができました。

## 今後の展開

共同実験を通し、次世代Web流通エ

ンジン (iMage) の技術が企業内データを対象とした人材マッチングの分野において業務上有効であることが検証できました。今回は約1000人規模のシステムインテグレータの4年間ほどの企業内データを対象としましたが、より大規模な企業のデータや、より長期にわたる期間のデータを対象にしようとした場合にはデータ量に対するスケーラビリティが求められます。さらに対象となるデータの効率的な収集、検索結果の質のさらなる向上等に関する検討を進め、さまざまな分野にも適用可能となるよう研究開発および実用化を進めていく予定です。

### 参考文献

- (1) 佐藤・飯塚・プラムディオノ・大友・村山：“iMage ～情報の再構成が生み出す意外な発見～,” SWC2007.
- (2) <http://www.w3.org/TR/rdf-concepts/>
- (3) <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>



(上段左から) 大友 健治/ 村山 隆彦  
(下段左から) 妹尾 正身/ 小久保 祐一

どのような情報を対象とするかによってさまざまな適用が可能な技術だと考えています。このような情報で試してみたいということがあればぜひご連絡ください。

### ◆問い合わせ先

NTT情報流通プラットフォーム研究所  
ソフトウェアアーキテクチャプロジェクト  
TEL 0422-59-6349  
FAX 0422-59-2699  
E-mail csm@lab.ntt.co.jp