

# Semantic Web 利用による次世代グループウェア WorkWare++

小櫻文彦、片山佳則、井形伸之、渡部勇、津田宏

富士通研究所 ITメディア研究所 ドキュメント研究部

{kozac, katayama.yoshin, igata, watanabe.isamu, htsuda}@jp.fujitsu.com

## 概要

情報統合型グループウェア WorkWare++は、RDF を利用したイントラネット情報統合基盤である。情報抽出技術によるメタデータ半自動生成機能を持ち、高速全文検索およびグラフ視覚化技術によるメタデータ検索活用が可能である。本論文では、WorkWare++の概要と KM における Know Who 検索への適用について述べる。

## 1. はじめに

Semantic Web は、Web 規模の情報整理/統合の枠組みであるが、現実に全世界規模での実現には、上位の Logic, Rule, Trust 層の規格整備が必要となりまだ時間がかかると思われる。一方、Semantic Web におけるメタデータ (RDF:Resource Description Framework) <sup>1)</sup> やオントロジーといった基盤技術は、個人情報管理 (Haystack) <sup>2)</sup> や、特定サイト内情報管理 (TAP-KB Semantic Search) <sup>3)</sup> 等への応用で有効性が示されている。本論文では、こうした Semantic Web の基盤技術をイントラネットにおける KM (Knowledge Management) に適用した、情報統合型グループウェア WorkWare++の概要と適用事例について述べる。

情報統合型グループウェア WorkWare++は、時間情報自動抽出による異種情報統合 WorkWare<sup>4)</sup> をメタデータにより汎用的なグラフ形式に発展させたものである。メタデータのモデルとしては、単に属性情報だけでなく複雑な関係情報を記述できることが求められるため、RDF を採用した。これにより各アプリケーションや文書などのリソースの連携が可能となり、単なるグループ間の共有だけでなく、情報共有の構成要素として重要な人、イベント、文書、場など従来バラバラに存在していたものを統合する。また WorkWare++は異種情報をグループ内で共有しシームレスに各種アプリケーションが連携できる

EAI(Enterprise Application Integration) <sup>5)</sup> ツールでもある。異種データの統合には RDF とオントロジー技術を利用している。

WorkWare++では Semantic Web の一つの問題として上げられるメタデータ作成コスト <sup>6)</sup> を抑えるため、自然言語処理(情報抽出)技術を利用している。これは、グループウェアの主機能である個人スケジュール管理を中心に、イベント(会議など)における文書(コンテンツ)作成など通常の業務活動を行いながら半自動で RDF メタデータを抽出する機能である。

本論文では、以下、WorkWare++の概要、およびメタデータ抽出、メタデータ検索活用機能について、KM における Know Who 検索への適用事例を交えて述べる。

## 2. 統合型グループウェア : WorkWare++

ここでは、WorkWare++の、目的と基本システム構成と特徴について述べる。

### 2.1 ねらい

グループウェアは、スケジューラや文書共有システムの組み合わせ、個人の情報をグループで共有することで業務効率向上を目指している。しかし、一般のグループウェアでは全員が同一のシステムを用いることが前提とされ、また異種アプリにおける情報の連携も手動で行なうものが多い。

WorkWare++では、メタデータ標準である RDF を中心データとすることで異種のアプリにおけるコンテンツ連携を容易にする。また、日常業務における作業過程や、自然言語処理技術により、人と情報の間の様々な関連付けをメタデータとして自動的に抽出し新たな情報活用を考察することを目的として開発している。

### 2.2 システム構成

図1に示すように、WorkWare++では、活用、管理、抽出と3階層で構成される。

抽出層は、各種コンテンツ/アプリケーション

WorkWare++: An Integrated Groupware Using Semantic Web  
Fumihiko Kozakura, Yoshinori Katayama, Nobuyuki Igata,  
Isamu Watanabe, Hiroshi Tsuda

から自然言語処理などを使用してメタデータの自動抽出を行なう。

管理層は、抽出したメタデータを RDF として管理し、さらにオントロジーによるメタデータ間の関連付けも行う。

活用層は、全文検索技術およびグラフ情報の視覚化技術により、メタデータから、ユーザーに様々なビューを対話的に提供する。

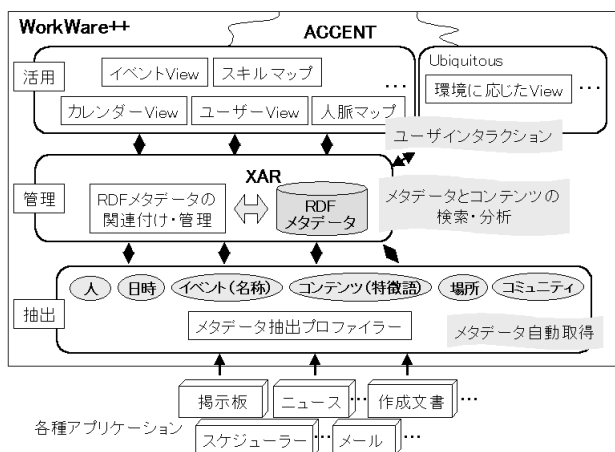


図1 WorkWare++の構成図

### 2.3 WorkWare++の特徴機能

一般に Semantic Web では、メタデータの作成コストが一つの問題である。メタデータが大量にあればそれを活用した有用なアプリが作れるものの、肝心のメタデータを作るコストが大きすぎて、結局の所アプリもできないというデッドロックになりがちである。

WorkWare++では、自然言語処理(情報抽出技術)よりメタデータにおける属性情報の自動抽出を行い、メタデータ間の関連についてはユーザーインタフェースやログ分析により関連性を抽出しメタデータに付与することで、メタデータ作成コストを低くしている。

#### 1) メタデータ半自動生成

イベント、イベントの配布資料などの文書作成、スケジュール管理、メールの送受信、情報検索ログなどの日常作業において、個人に関わる情報が溢れている。これらのデータからメタデータを抽出することで、個人や文書情報に付随する大量のメタデータを自動抽出できる。

例えば、資料やメール、論文など個人がアウトプットした情報は、個人の作成コンテンツ情報というだけでなく、コンテンツから特徴キーワ

ードを取り出すことで個人の所有するスキル情報とすることが可能である。各種サービスを利用したログからは個人のインプット系のメタデータとして、単に履歴情報だけでなく、時系列における個人の興味情報とすることが可能である。また、スケジュールの参加者情報からは、同じ打合せに良く出る人達の間の人脈情報を得ることが可能である。

#### 2) 検索・可視化によるビュー

半自動生成で得られる大量の RDF によるメタデータの活用において、まず高速な検索機能が必要となる。WorkWare++では大規模な XML 高速全文検索技術(XAR)<sup>7)</sup>を、RDF 検索に適用することで処理の高速化をはかっている。

また、メタデータを検索により絞りこんだとしても単純なリスト表示ではグラフ状の情報の活用は難しい。例えばリンクを辿るなどの操作は可能であるが、視点を変えることによる情報の2次利用を行なうことができない。そこでテキストマイニング技術である ACCENT<sup>8)</sup>を使用し、メタデータ間の関連をグラフにより可視化することで、視点を変え表示することを可能とした。これにより分析などの2次利用を支援するビューを構築することが可能となる。例えば、人と人との関連を可視化することで人脈マップを簡単に見ることができる。

#### 3) 異種メタデータ間マッチング(オントロジー)

大量のメタデータをさらに活用するには、異なるアプリケーションを通して抽出されたメタデータ間の関連付けが必要となる。個人に付随するメタデータが大量にあったとしても、あるアプリと別のアプリで、同じ個人を示す名称が違っては、メタデータの活用場面は限られてしまう。

そこで WorkWare++では異種メタデータ間の関連付けとして、抽出された関連情報やオントロジーを利用したマッチング、ユーザーインタフェースによる情報間の自然な関連付けを支援することで、半自動で異種メタデータ間の関連付けを行なうことができる。これにより一段と多様な情報の2次利用が可能となる。

関連付けの例として、異なるグループウェアで管理されていたり、人によって呼び方が異なるミーティング情報の関連付けや、打合せ情報と議事録や打合せ資料などの文書の関連付けなどがあげられる。

### 3. WorkWare++におけるメタデータ抽出

本節では WorkWare++における個人のメタデータ(ユーザプロフィール)抽出の実例について述べる。

ユーザプロフィールの属性には、所属部署やメールアドレス、電話番号などの単値で扱えるものから、技術スキルなど単値では表現が難しいものまである。Know Who 検索の実現において必要となる技術スキル情報については、これまで、自己申告情報を元にした人手による入力があった。しかし、一方人手による入力は、情報操作の悪用や情報の陳腐化、入力コストの高さなどの問題をかかえている。

WorkWare++では、技術スキル属性を、単値の属性として扱うのではなくすでに入力されている情報(個人の作成コンテンツなど)から抽出した、特徴キーワードやそれを作成した年月日間のグラフ構造として扱っている。個人は通常業務を行なう上で、さまざまな電子化された文書を作成している。その中には、個人が所有する技術のキーワードが散りばめられている。また、キーワード間にも関連性がある。文書から自然言語処理を使用して技術キーワードを抽出し、キーワードの共出現情報からキーワード間の関連度を計算することにより図2に示すような個人のスキル情報を得ることができる。

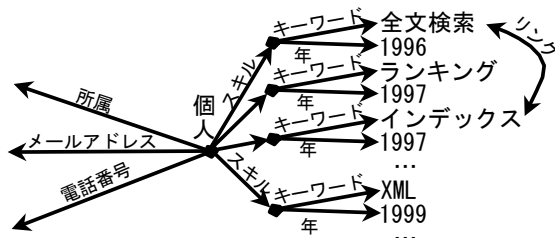


図2 個人メタデータ(ユーザプロフィール)

こうしたグラフ形式によるプロフィールの表示においては、キーワードが膨大になるため、表形式で表示しても人間には把握できない。そこで WorkWare++では、テキストマイニング技術である ACCENT を使用しメタデータにある大量のスキルの表示を行なうことで、技術スキルとしての表示を可能にしている。このようにメタデータの自動抽出を行ない用途にあったビューで表示させることにより、従来では難しかった情報管理を容易にすることができる。

### 4. WorkWare++によるメタデータ検索活用

WorkWare++では、検索により絞り込んだ同一のメタデータから、時間を中心に見たり、人を中心に見たりと様々なビューによる情報提示が可能である。さらに、こうした複数のビューをリンクにより組み合わせることで一連のタスクを行なうこともできる。本節では、その活用例として企業内における Know How 検索を紹介する。

企業内 KM においては、知識は最終的には人に付随することが多い。しかし、単なる文書検索ではなかなかノウハウを持った人に辿りつくのは困難である。このように特定のスキルを持った人やグループ情報を検索するのが Know Who 検索である。Know Who 検索においては、単に人を提示するだけでなく、その人への連絡手段やその人の背景情報も得られなければならない。

Know Who 検索の実現にあたって、まず前述したユーザプロフィール取得の困難さがある。さらに、検索の困難さとして、文書から抽出したスキルのメタデータを単に使用しても全文検索で著者を求めるのと同様に平面的になり、人脈などの背景情報を含めた情報を得ることは難しい。個人の背景情報を得るためには、1) 検索したいスキル(キーワード)の関係、2) 当該スキルの関連するグループ、3) グループ内の人脈(背景情報)といった一連の情報検索を相互に連携させながらマップ化する必要がある。

#### 4.1 グループスキルマップ

3節で述べたように、WorkWare++では、個人の作成コンテンツなどから、時系列のスキル情報を自動で取得している。

この情報を活用し、まず検索したいスキルを特定することを考えよう。文書検索と同じく、人のスキル検索においてもスキルを特定するうまいキーワードを見つけることは重要であるものの難しい。図2の個人スキル情報は、所属でまとめることで、(所属部署、キーワード)のグラフと見ることができる。

まず、スキルキーワード検索の第一歩として、まず初期キーワード(例えば「XML」)を与えると、そのキーワードの周辺関連キーワードや関連部署を得ることができる。

次にその結果から ACCENT を使用して技術間、技術と所属間の関連を可視化したものが図3である。これは技術を中心とした関連技術を含めた企業内部署のスキルマップである。なお、太い線や近い配置は、関連性が高いことを表している。

