

E-019

日常の文書管理操作を用いた企業オントロジーの自動構築[†]高梨勝敏^{††} 佐藤俊也^{†††}(株)日立東日本ソリューションズ ナレッジソリューショングループ^{††††}

1. はじめに

近年、多くの企業では、情報共有システムによる業務効率・生産性向上に大きな関心が寄せられている。しかし、断片化された情報の共有だけでは業務効率の向上は難しく、情報間の関係を体系立てて共有することが課題である。このためには、システムが文書の情報だけでなく企業が共有する知識体系(企業オントロジー)を扱えるようにすることが有効な手段と考えられる。

そこで、文書群の内容および文書に対する管理操作を通して企業オントロジーを自動構築するシステムを考案した。これにより、情報間の関係の把握が容易となり、情報共有システムの利用効率向上を通して業務効率の向上が期待できる。

2. 背景

我々は2001年より、組織の情報共有を目的として、組織の情報を容易に探索できるツールの研究²⁾を行ってきた。次のステップでは、情報そのものだけでなく、その背景となる知識も併せて共有することにより、情報作成者の意図を理解して協働できるようにすることが重要と認識した。

そこで、本研究の目的を、背景知識を表現するための知識体系の定義と、知識体系の可視化・共有ができるツールの開発とした。

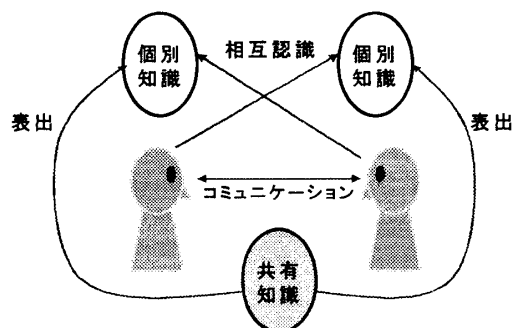


図1 共有-個別知識交流モデル

3. 知識交流モデル

知識体系構築についての既存の研究では、オントロジーをグローバルオントロジーとローカルオントロジーに分け、これらの対応付けを行なうための手法が提案されている³⁾。本研究では、オントロジーのメンテナンスは組織メンバの文書作成・共有を通じて行ない、メンバ間でオントロジーに不整合が起きた場合は交流によって解消することを狙いとする(図1)。

組織が生み出した全ての情報を以下のように定義する。

$$K = \{C_1, C_2, \dots, C_l\} \quad \dots\dots\dots(1)$$

ここで、Cは組織の共通用語、共有文書、共有文書が含む単語やメタデータである。共有知識Sと、個人iの個別知識E_iを以下のように定義する。

$$S = \{s(C_1, C_2), s(C_1, C_3), \dots, s(C_m, C_n)\}, \\ E_i = \{e_i(C_1, C_2), e_i(C_1, C_3), \dots, e_i(C_m, C_n)\} \quad \dots\dots(2)$$

ここで、sおよびe_iは、各知識でのC_mとC_nの関連度である。これにより、以下のような関連を統一的に表現できる。

- (1) 用語間の関係(〇〇課がマーケティング業務に関わっている等)
- (2) 各用語に関連する文書(食品業界調査資料はマーケティングに関わる等)
- (3) 各文書とそれが含む単語・メタデータの関係(食品業界調査資料に単語「水産物」が多く現れる、作成者は鈴木一郎、等)

以上のモデルに基づき、文書管理ツールCoreLibrary(以下CoreLib)を開発した。CoreLibは、SとE_iを知識マップとして視覚的に表示し、共有知識とメンバ毎の個別知識を認識できるよう

[†] Ontology Construction through Document Management System Operation

^{††} TAKANASHI Katsutoshi (takana@hitachi-to.co.jp)

^{†††} SATOU Shunya (shu_sato@hitachi-to.co.jp)

^{††††} Hitachi East Japan Solutions Ltd., Knowledge Solution Group

にする。また、自然言語処理によってCの関連度を自動的に求め、視覚化および通知することによって、互いの知識の整合性を取るための交流を促す。

4. CoreLibraryの機能

CoreLibは、パソコン上のファイル管理ツールとして実装し、ユーザが容易に構成要素を取り扱えるようにする。CoreLibの画面例を図2に、各機能を以下に示す。

(1) 知識マップ表示機能

共有知識Sを、2次元平面上のフォルダ階層で表示する。これらフォルダの定義や配置をメンテナンスすることにより、式(2)のsが求められる。

(2) 文書登録機能

ユーザは、いずれかのフォルダに文書を登録できる。文書の近傍に、文書の内容を表す単語である特徴語を表示するとともに、文書に関連する他の文書も表示する。ここで特徴語の抽出にはtf-idf法を、関連文書の抽出にはベクトル空間法を用いた4)。文書を登録する操作と特徴語および関連文書の自動抽出によって、式(2)の e_i が求められる。

(3) マップ編集機能

個人毎の画面では、ユーザは2次元平面上でフォルダおよび文書の配置を変更することが可能である。フォルダや文書はマウスを近づけることにより拡大し、概要を表示する。ユーザは概要を容易に確認し、配置の決定に役立てることができる。

5. 期待される効果

CoreLibの利用によって、以下の効果が期待される。

- (1) 知識マップ表示機能によって、各メンバーの活動が共通知識のどこに位置付けられるか、他メンバーの活動とどのように関係付けられるかを互いに理解しあうことができる。
 - (2) 文書登録機能およびマップ編集機能によって、組織の共通知識に基づいて、各メンバーが自分の文書を整理することができる。
 - (3) 特徴語と関連文書の抽出によって、関係のある活動をしているメンバー間で、情報がタイムリーに流通し、メンバーの活動を活性化させる。
- 以上の手順により構築された企業オントロジーは共有知識を基本としているが、共有知識のみを

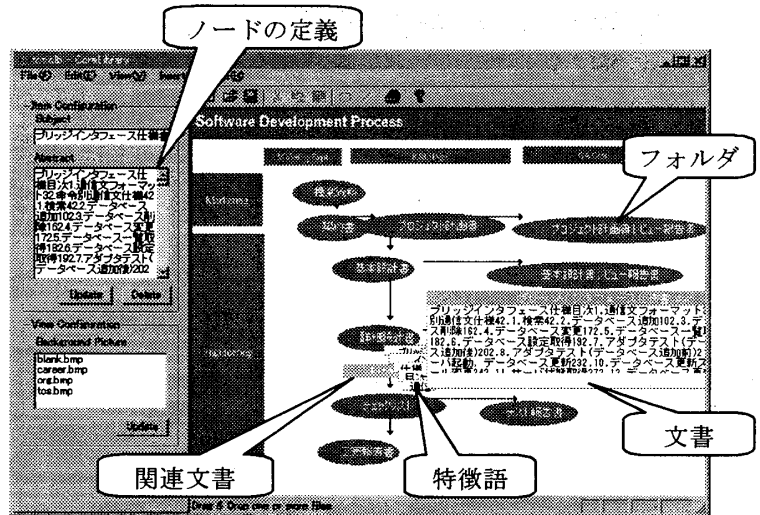


図2 CoreLibraryの画面

用いた構築に比べて各ユーザの個別知識がより反映されていると考えられる。本オントロジーを用いて、個人が作成した文書を組織共通の分類体系で閲覧・検索することが可能となり、情報共有システムの有効利用に役立つことが期待される。今後は自社内でのプロジェクト管理で利用することにより、情報の流通と企業オントロジーの構築を定量評価する予定である。

6. おわりに

組織の知識を共通知識とメンバーの個別知識に分け、これらを統合して相互理解するためのモデルと、組織の情報を知識として表現するためのモデルを構築した。このモデルに基づき文書管理ツールを開発した。本ツールを使用することにより、組織共通の知識マップによる文書の整理を通して、自然に個人の知識マップが構築される。また、個人の知識を反映した企業オントロジーを構築することが可能である。

参考文献

- 1) 塚原朋哉：情報の視覚的検索方法，fit2003 一般講演論文集 第二分冊 pp.179-180
- 2) 中挟知延子：Web オントロジーのメンテナンスにおける語の多義性解消からのアプローチ，情報処理学会 研究報告「デジタル・ドキュメント」No.036, pp.006-013
- 3) 廣田啓一，佐々木裕，加藤恒昭：文書情報からの分野オントロジー構築の支援，情報処理学会研究報告「自然言語処理」，Vol. 2000, No. 107, pp.55-60