

MIIDAS: 情報の選別的共有のためのオントロジ構築とその増進的学習

5 Q-8

奥村明俊 池田 崇博 村木 一至

NEC C&C メディア研究所

1 はじめに

オフィス業務など知的生産活動において、必要な情報を選別・提供しアイデアの効率的生産を促進するため、情報活用・共有プラットフォーム MIIDAS (Multiple-Indexing Information Dissemination and Acquisition Service) を開発している。MIIDAS は、オントロジベース情報集配信、5W1H分類・ナビゲーション、ノウハウ活用エージェントから構成される[1; 2; 3]。本論文では、オントロジベース情報集配信のオントロジについて述べる。

必要な情報をユーザに提供する技術として、ユーザの必要とする情報をキーワードとして具現化し言語知識を用いて情報内容と照合する CBF(content-based filtering) と、ユーザの情報に対する評価結果のような情報メタ知識を用いて関心を共有する他のユーザに情報を配信する SIF(social information filtering) が提案されている。CBF では、ユーザの興味対象を示すキーワードの準備・更新の手間がかかり、意外な情報の発見も期待しにくい。また、SIF では、情報に対する関心の類似度計算で分野の違いといった多次元的な要素を扱うことや、情報配信理由を明示するのが困難である[4; 5]。

ビジネスイントラネットにおいては、上記のようなキーワードや評価結果だけでなく、その個人が属するグループや組織の中での位置付けという利用者知識が情報選別の手がかりとなる。実際、キーワードや評価結果がなくとも、利用者知識を用いて情報が有効に配信されることが多い。本論文では、ユーザの所属や業務テーマなどの利用者知識、ユーザのドキュメントから抽出したキーワードとキーワードに関するシソーラスといった言語知識、ユーザの情報に対するアクセス履歴や評価という情報メタ知識の3種類の知識を複合オントロジとして管理し、それぞれの視点から情報を選別・配信するモデルを提案する。

2 複合オントロジとその活用

複合オントロジは、利用者知識、言語知識、情報メタ知識の3軸の知識を利用者にリンクした階層型知識ベースである(図1)。利用者知識は、所属部門、業務の種類、業務テーマ、業務経験、専門分野、顧客リスト、発行文書、趣味などの利用者属性情報であり、それぞれの属性の階層関係を記述する。言語知識は、ユーザが発

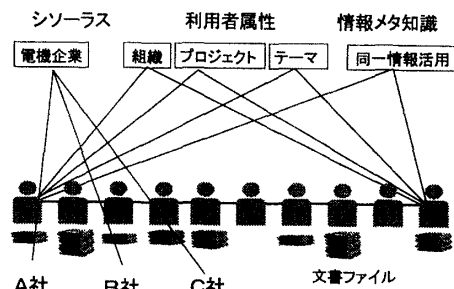


図1 複合オントロジの例

行したドキュメントから抽出したキーワードを記述したキーワードとキーワードに関するシソーラスをキーワード属性として記述する。情報メタ知識は、利用者の情報活用履歴と情報に対する評価結果を情報メタ属性として記述する。

情報配信は、その情報を必要とする可能性の高いユーザ(positive user)の集合をオントロジの言語知識と情報メタ知識に CBF や SIF を適用して求める。次に、positive user の共通点をオントロジの属性の組み合わせで表現し配信条件とする。オントロジの多次元的な知識を用いることで、個人の関心の多次元性に対応し、オントロジ属性によって配信条件を定めることで、言語知識と情報メタ知識のないユーザ(unprofiled user)への情報配信と配信理由の明確化が可能となる。さらに、ユーザの情報に対する要・不要の情報と配信条件の対応を情報メタ知識のフィードバック属性としてオントロジに記述し、ユーザのドキュメントからキーワード属性を更新することで、増進的に配信条件を適合することができる。

3 配信条件探索アルゴリズム

まず、profiled user の中からプロフィールや情報活用履歴を用いて、その情報に関する positive user を検索する。ここでは、オントロジの言語知識と情報メタ知識を用い CBF ベースで求める[6]。次に、positive user に共通する属性と属性の組み合わせを条件として探索し、その条件を満たす全ユーザの中で、ある割合 R_0 を越えるユーザが positive user である時、その条件を満たすグループを配信先とする。これは、イントラネット環境下においては、グループメンバは何らかの形で情報の必要性を共有しているので、ある割合を越えるメンバにとって価値のある情報は、そのグループ全員にとって有用であるという経験に基づいている。positive user

MIIDAS: Ontology-based information organization and sharing

Akitoshi Okumura, Takahiro Ikeda, Kazunori Muraki
NEC C&C Media Research Laboratory

が共有する属性集合は以下のように抽出する。

1) 配信の1次条件として、 $G^{(1)}$ を求める。

$$G^{(1)} = \{G_i \mid \text{positive user 1名以上が } G_i \text{ に属する}\}$$

$$G_i = \{u_n \mid \text{ユーザ } u_n \text{ は属性 } a_i \text{ を持つ}\}$$

2) 配信の k 次条件として、 $G^{(k)}$ を求める。

$$G^{(k)} = \left\{ G_n^{(k)} = \bigcap_{i=1}^k G_{n_i} \mid |G_n^{(k)}| \geq m_0, \right.$$

positive user 1名以上が $G_n^{(k)}$ に属する,
 $G_{n_i} \not\subset G_{n_j}$ and $G_{n_i} \not\supset G_{n_j}$ for all i, j

$$m_0 = (\text{予め定義されたグループの最小ユーザ数})$$

3) 配信条件 G_{dis} が、閾値 R_0 によって決定される。

$$G_{\text{dis}} = \left\{ G \in G^{(k)} \mid \frac{|P(G)|}{|G|} \geq R_0 \right\}$$

$$P(G) = (\text{グループ } G \text{ における positive user の集合})$$

例えば、図2のオントロジによって、A B X Y のキーワードを含む情報を配信する場合、キーワード属性を共有するユーザ1,2,3を positive user とする。この時、配信候補は表1のようになる。 R_0 が0.7の場

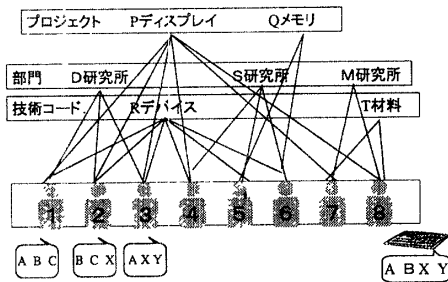


図2 オントロジによる配信例

合、グループ $G_1, G_1^{(2)}, G_2^{(2)}, G_3^{(2)}, G_1^{(3)}$ が配信グループとして決定され、ユーザ1,2,3,4が配信対象者となる。ユーザ4は、キーワードを持たない unprofiled

表1: 配信先候補

グループ	属性(部門、技術コード、プロジェクト)	positive user / 全ユーザ
G_1	D 研究所	3 / 3
G_2	R デバイス	3 / 6
G_3	P ディスプレイ	3 / 6
$G_1^{(2)}$	D 研究所 & R デバイス	3 / 3
$G_2^{(2)}$	D 研究所 & P ディスプレイ	3 / 3
$G_3^{(2)}$	R デバイス & P ディスプレイ	3 / 4
$G_1^{(3)}$	D 研究所 & P ディスプレイ & R デバイス	3 / 3

user であるが、技術コード属性“R デバイス”とプロジェクト属性“P ディスプレイ”を持つユーザの R_0 以上が positive user であったので情報が配信される。

4 実験と課題

オントロジによる多次元情報集配信システムは、企業イントラネットユーザ1,500名を対象として構築中である。今回、表2のような利用者属性と利用者の特許から抽出したキーワード属性およびキーワードのシソーラス情報を含むオントロジを用いて900名に特許配信の実験を行ない、unprofiled user 100名にも配信理由を沿えて情報を配信できることを確認した。イントラネット

オントロジ	分類構造*	インスタンス数
組織	9	26
技術コード	7	35
テーマ	7	(プロフィールなし100名)
プロジェクト	15	
キーワード**	28	137
		2450団体

* 分類構造が2カラムあるのはインスタンスの上位概念が2層になっていることを示す。
 ** キーワードのオントロジは団体に関するシソーラスである。

表2 実験オントロジ

環境では、スタッフ部門が管理する情報がオントロジ属性として有効であるが、1) それぞれのオントロジの軸から配信情報を分類し配信情報の分野と有効なオントロジ属性を明らかにすること、2) 閾値はオントロジ属性に依存しており、属性ごとに重みとして定義するか、配信条件によって、positive user である確度とともに最適化することが必要である。

今後、閾値、配信条件と属性数、配信情報の内容などの関係を分析し、利用者が適合的に配信条件を選択できる機構と、フィードバック機構を実装する。さらに、positive user を求める手法の改良とともにシステムの改善を図る。

謝辞 本研究に関して NEC 情報システムズの落合尚良氏の協力に感謝致します。

参考文献

- [1] 奥村明俊, 他, “オントロジによる多次元情報集配信,” 人工知能学会第11回全国大会, pp.368-339,1997
- [2] 佐藤研治, 他, “文書参照 / 引用履歴を利用したノウハウ活用エージェント,” 人工知能学会第11回全国大会, pp.366-336,1997
- [3] 池田崇博, 他, “5W1H 情報を利用する情報 分類・ナビゲーション,” 人工知能学会第11回全国大会, pp.370-371,1997.
- [4] Scott Deerweester, et all, “Indexing by latent semantic analysis,” Journal of the American Society for Information Science, pp.391-407, 1990.
- [5] Upendra Shardanand and Pattie Maes, Social information filtering: Algorithms for automating word of mouth, Proceedings of CHI'95, 1995.
- [6] 石川開, 他, “MIIDAS: 情報の適合的選別による文書フィルタリング,” 情報処理学会第55回全国大会, 5Q-09,1997.