

企業の情報活用でのオントロジー技術の適用

Ontology Technologies Application for Enterprise Information Integration

企業ではアンケート分析や顧客の声分析などへのテキストマイニングの適用が進んでいる。しかし、意味解析技術は適用が難しい。ヘビーウェイト・オントロジーを設計・製造や品質保証に適用した成功事例がある一方、業務上の判断を支援するために大量の情報から判断の裏づけ情報を抽出するためのテキストマイニングからのアプローチはライトウェイト・オントロジーを扱うこととなり、企業で活用できるオントロジーの構築が課題である。そこで筆者らは、業務にオントロジー技術を最適化することを狙いとしてユーザニーズを収集し、日常業務を通じて半自動的にオントロジーの構築を可能とする情報共有システムを開発した。試行の結果、組織の判断スピードの向上が示唆された。

高梨 勝敏	Takanashi Katsutoshi
塚原 朋哉	Tsukahara Tomoya
渡邊 まり子	Watanabe Mariko
佐藤 俊也	Sato Shunya

1. はじめに

企業では、共有している情報にオントロジー技術を適用する取り組みがある¹⁾。これら取り組みにより、数値データやテキストデータを、意味を持つ情報として機械処理することが可能となる。ここでオントロジー技術はヘビーウェイト・オントロジーとライトウェイト・オントロジーに大別される。両者について考察した研究事例²⁾では、前者は対象世界を適切に捉えることを重視して強い公理で構成されるのに対し、後者では情報論的な利用効率を重視して統一された語彙集合と簡単なスキーマで構成されると述べている。ヘビーウェイト・オントロジーは、企業が持つ知識体系を厳密に定義することにより、機械処理の精度が高く、設計支援システムの事例で設計知識の共有や不具合分析に効果を上げている¹⁾³⁾。しかし、関係者間の整合性を保ったオントロジー構築のコストと、情報の増加・変化に従ったメンテナンスのコストが大きくかかる。ライトウェイト・オントロジーは、オントロジーの構成が単純であることと自由度が高いことにより構築コストが低いが、業務上の判断の裏づけとなりうる情報の抽出は難しい。現在では、情報の分類・検索や、情報作成者間の関係の視覚化などに適用されるのが中心である。

そこで筆者らは、ライトウェイト・オントロジーを業務に適用するための枠組みを作ることを推進している⁴⁾。このためには、オントロジーを自由度の高い状態で使用

するのではなく、業務課題に即して効率のよい構築と活用のモデルをつくることが重要と考える。また、企業の目的は業務課題の解決であり、オントロジー構築のソリューションを求めているわけではない。このため、業務の目的を分析して必要なソリューションを提案している。本論文では、企業でのユーザの業務課題解決のニーズ収集、活用モデルの構築および適用・評価の一連の流れを説明し、研究技術とソリューションを効果的に結びつけるための施策を述べる。

2. 研究の概要

本事例は、ユーザ企業での情報共有のための研究プロジェクトである。対象とする情報は、文書管理システムの蓄積情報、不具合情報および営業日報の3種類である。研究の目的は、「活用対象とする情報の特性に応じ、情報を効果的に活用できるモデルを考案する」こととした。モデルがオントロジーを利用する際、先行研究での利用形態とオントロジーの種類²⁾との比較をおこなった。利用形態は、辞書としての利用から体系化まで各種ある中で、「知識共有」と位置づけた。企業での情報共有のテーマのひとつである、ナレッジマネジメントの実現を狙っている。また、オントロジーの種類は、語彙や is-a の階層構造の扱いから出発し、情報の活用を進めるにつれてより深いオントロジーへ進化させることを狙った。

プロジェクト遂行にあたり、ニーズを効率よく分析す

ることを狙いとして、ユーザ企業と(株)日立東日本ソリューションズ(日立 TO)が協同でプロジェクトに従事した。プロジェクトのメンバ構成を図 1 に示す。ユーザ企業はユーザ、SE および両者を取りまとめるリーダーで構成した。日立 TO はコンサルティング、研究開発およびリーダーで構成した。ユーザと研究開発者は、対象とする情報に応じて異なる者が担当した。ユーザ企業のリーダーは各ユーザが持つ複数のニーズに共通する根本的なニーズを見出す汎化など、ニーズの整理をおこなう。日立 TO のリーダーは技術の整理をおこなう。各リーダーがハブとなることにより、ニーズと技術のマッチングの効率化を図った。

プロジェクトの進行での構成は、情報活用のニーズ分析、活用のためのモデル構築、プロトタイプ構築および適用・評価の 4 フェーズである。研究開発プロセスの各フェーズでのメンバ間の情報の流れを図 2 に示す。各フェーズで起点となるメンバが異なっている。ニーズ収集ではユーザを起点に情報が流れており、これはユーザに対しヒアリングを行なったことを表している。活用モデルの構築では、コンサルタントと研究開発者を起点として、情報活用の仕組みについて提案している。プロトタイプの構築では、SE と研究開発者が調整しながら開発およびシステム構築をおこなっている。適用・評価では、ユーザによる試行を通して、ユーザ企業と日立 TO がプロトタイプを評価・改善する。

プロジェクト全体を通して、打合せは月 2 回の頻度でおこない、ニーズ収集フェーズで分析したユーザニーズが後のフェーズで対応できているかを確認しながら進めた。また、各フェーズでの情報の流れ(図 2)は厳密で

はなく、参加メンバが打合せを通して得た気づきを適宜取り上げて検討内容を改善した。以降の章では、三つの事例でのニーズ収集、活用モデルの構築フェーズでの内容(3章から5章)と、モデルに基づいたプロトタイプの開発、適用の結果得られた評価結果(6章)について述べる。

3. 事例 1 : 文書管理システムの蓄積情報の活用

3.1 ニーズ収集

マーケット情報の共有業務では、Web や報道機関から収集した情報を文書管理システムに蓄積し、設計者が検索することにより市場に適合した設計をおこなうための判断を支援している。

ニーズ収集フェーズでは、まず情報活用の目的設定をおこなった。次にユーザに対し、情報共有の現状と情報活用に対するニーズをヒアリングした。ヒアリング結果を元に、目的を実現するためのニーズ構造の分析をおこなった。研究の対象とした各情報の概要と情報活用の目的は、以下のとおりである。

(1) 情報の概要

散在する情報源ではそれぞれの分類体系でコンテンツを分類・整理しているが、すべての情報源の分類を統一的に表現できるオントロジーを構築することは困難である。各情報にコメントやキーワードを付与し、階層構造を持ったフォルダ群の中の該当箇所に登録している。

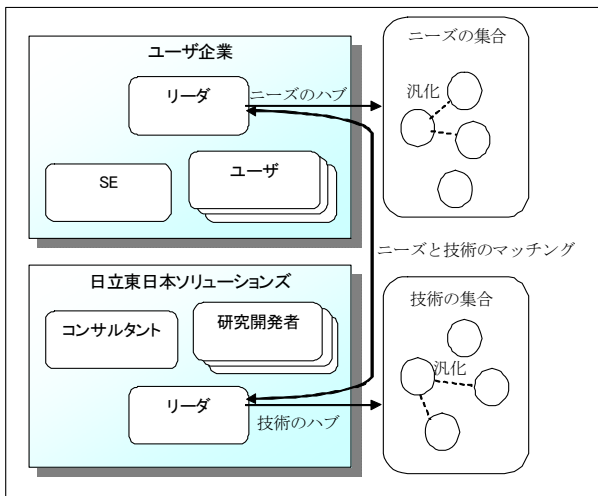


図 1 メンバ構成

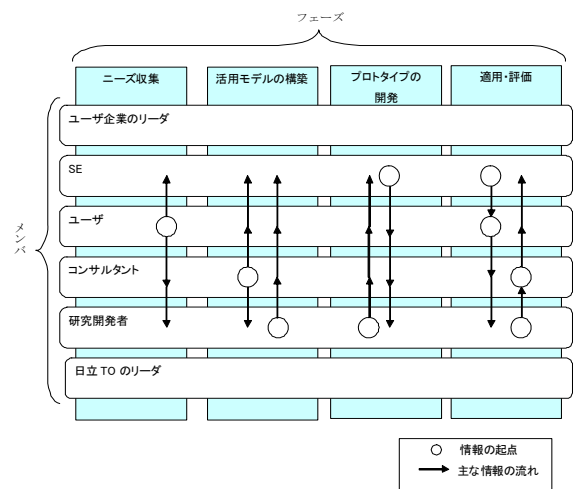


図 2 研究開発プロセス

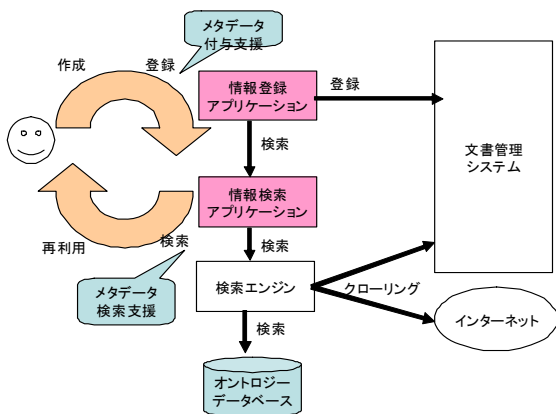


図 3 文書管理システムの情報活用モデル

(2) 情報活用の目的

情報の登録者が付与するキーワードや登録先フォルダが、情報の参照者にとって探しやすく適切なものにする。

これら情報は、大量に蓄積されるため管理しきれなくなっている問題と、登録者間で情報整理の基準が共有されていないために必要な情報を探せなくなっているという問題があった。前者の問題はテキストマイニング技術で対応可能であるが、後者の問題は情報自身の量だけでなく情報に付けられたメタデータに起因するものであると考えた。そこで、情報整理の基準をユーザ間で整合性をとるための、情報活用モデルの構築を検討した。

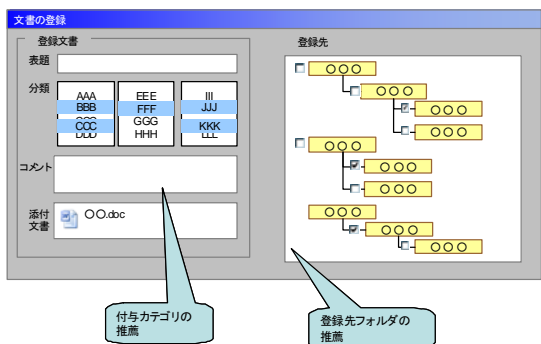


図 4 情報登録アプリケーションの画面例 (1)

3.2 活用モデルの構築

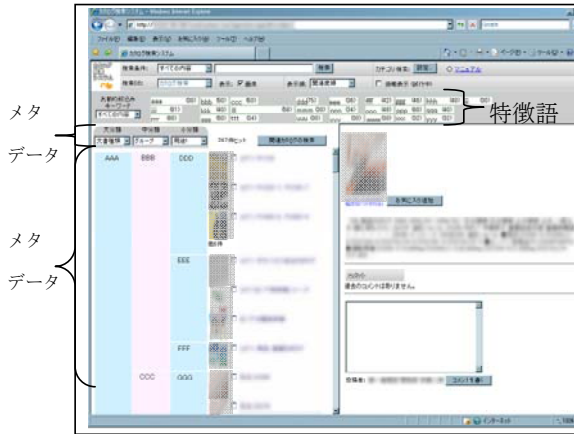
情報を整理して共有するにあたり、分類カテゴリの作

成・文書分類のコスト低減と、分類の質の向上を狙って情報活用モデルを作成した。モデルの概要を図 3 に示す。このモデルは、ユーザによる文書作成→登録→検索→再利用の一連の活動を支援するものである。文書の作成・登録では情報登録アプリケーションが情報へのメタデータ付与を支援する。文書の検索・再利用では、情報検索アプリケーションがメタデータを利用した検索および検索結果の表示をおこなう。

既存の文書管理システムでは、情報の登録時に自由にキーワードを付与できたため、関連する文書であってもユーザ毎に異なる基準でキーワードが付与されていた。そこでキーワードの基準を共有するため、オントロジーデータベースに格納されているノード一覧を提示し、その中から選ぶことにより、メタデータとして文書に付与したうえで文書管理システムに登録する。ここで、オントロジーデータベースはメタデータ名と値のペアを保持するとともに、メタデータ間の階層を定義しており、ライトウェイト・オントロジーの中でもタクソノミー(情報を分類するための階層構造)に位置づけられる。また、ユーザが過去に登録した文書や、関連する文書に付与されているメタデータを提示することによって、関連する文書に同じメタデータが付与されることが多くなることが期待される。ここで、関連する文書の判断には tf-idf 法による特徴語の抽出と、特徴語ベクトルによるベクトル空間法を用いた。前者は TF(term frequency : 単語出現頻度)と IDF(inverse document frequency : 出現文書数の逆数)の積であり、特定の文書によく現れる単語を抽出することができる。後者は各文書で単語集合の tf-idf スコアのベクトルを求めておき、ベクトル間の内積を求め



図 5 情報登録アプリケーションの画面例 (2)



(図はイメージであり具体値は見えないように加工している。)

図 6 情報検索アプリケーションの画面例

ることにより、類似する文書を検索することができる。

情報登録アプリケーションの画面例を図 4 と図 5 に示す。図 4 では、文書を登録する際にタイトルやコメントだけでなく、分類用のメタデータを選択することができる。また、画面右のフォルダ階層の中で 1 個以上のフォルダに文書を登録することができる。この場合も、フォルダに対応するメタデータを文書に付与したと見なす。これらメタデータは、既存の関連文書に付与されているものを推薦することによって、ユーザはメタデータの選択を効率よくおこなうことができる。図 5 では、フォルダ階層は用意していないが、関連文書のメタデータを一覧で表示し、必要なものをユーザが選択することにより、容易にメタデータを付与することができる。登録された文書を検索する場合の画面例を図 6 に示す。

文書をメタデータで分類して表示している。分類は 3 階層としており、各階層に任意のメタデータを指定可能

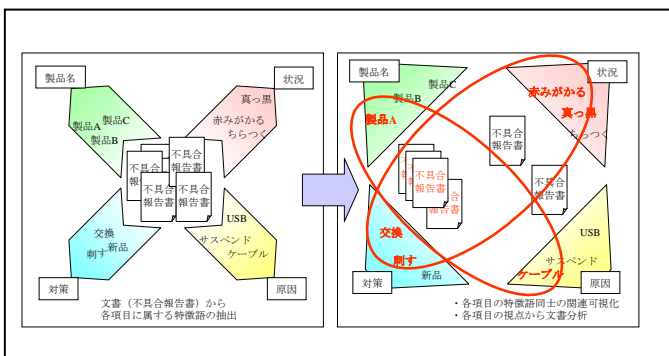


図 7 不具合情報の活用モデル

である。画面上部には検索結果の文書群が持つ特徴語を表示している。ユーザは特徴語を選択することにより、検索結果を絞り込むことができる。

4. 事例 2：不具合情報分析

4.1 ニーズ収集

品質保証の業務では、製品の不具合情報を蓄積し、クレーム対応に役立てている。

(1) 情報の概要：

製品に発生した不具合現象、調査の結果特定した原因および対策内容をデータベースに蓄積している。

(2) 情報活用の目的：

- ・不具合・クレーム対策の迅速化

日々収集・蓄積される不具合情報を分類・体系化することにより、不具合やクレームの原因追求・対策を迅速化する。

- ・設計品質の向上

既存製品の不具合の分析により、類似の新製品での不具合を未然に防止する。

これら目的でのニーズの構造を分析した。

- ・不具合・クレーム対策の迅速化

(i) ある製品に特有の不具合傾向の発見

(ii) 未知の不具合原因の発見

- ・設計品質の向上

(i) 特定の部品を共通に使う複数の製品での不具合の予測

(ii) 既存機種での不具合の分析結果を用いた、新製品開発時のチェックリストの作成

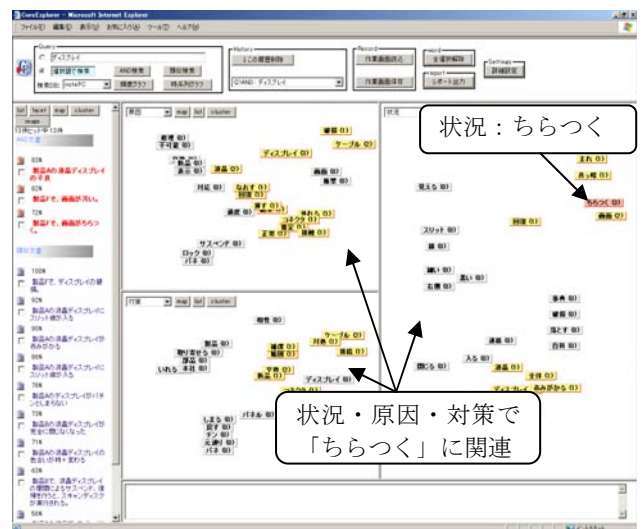


図 8 不具合情報分析の画面例

4.2 活用モデルの構築

上記ニーズから、製品、パーツ、部品および不具合事象間の因果関係を分析するモデルを定義した(図7)。モデルに従い、プロトタイプを開発した。対象とするデータは不具合報告書であり、不具合の現象、原因および対策の記入欄がある。これらを分析する画面の例を図8に示す。文書から特徴語を抽出する際、状況・原因・対策のそれぞれについてインデックスを作成している。これら特徴語を比較検討できるように特徴語表示画面(以下特徴語パネル)を複数配置した。1つの特徴語パネルで特徴語を選択したとき、その特徴語に関連する文書が絞り込まれるだけでなく、選択された特徴語に関連する特徴語がわかるように各特徴語パネルの特徴語の色を動的に変更した。

特徴語同士の関連は、絞り込まれた文書から求めている。ある特徴語を選択したときに絞り込まれる文書群があるとき、その文書群に紐づく特徴語は関連があると考える。例えば、「状況」項目の特徴語「ちらつく」を選択すると、ディスプレイ関連の文書が絞り込まれるため、「原因」項目の特徴語「ディスプレイ」が関連を表す色で表現される。これにより、ユーザが着目した特徴語に関連する原因を発見することができる。また、原因に着目して関連する状況を発見するなど、3つの特徴語パネル間で相互に関連を表示することができる。

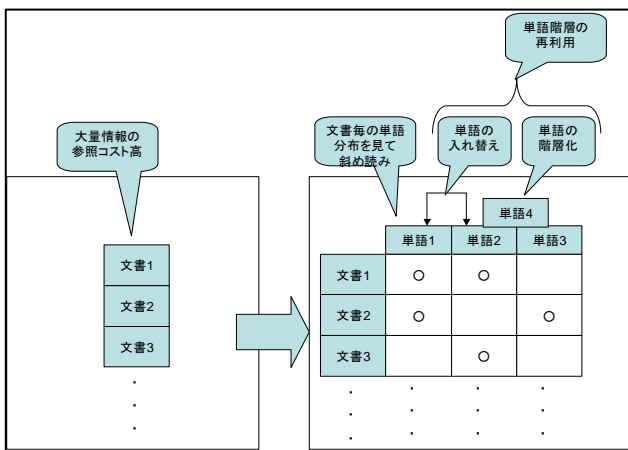


図9 営業情報の活用モデル

5. 事例3：営業日報の活用

5.1 ニーズ収集

営業の業務では、市場や顧客の動向に対応するために、営業日報を活用している。

(1) 情報の概要

営業担当者が日々の営業活動の内容、進捗および問題点・対策を営業支援システムに登録している。

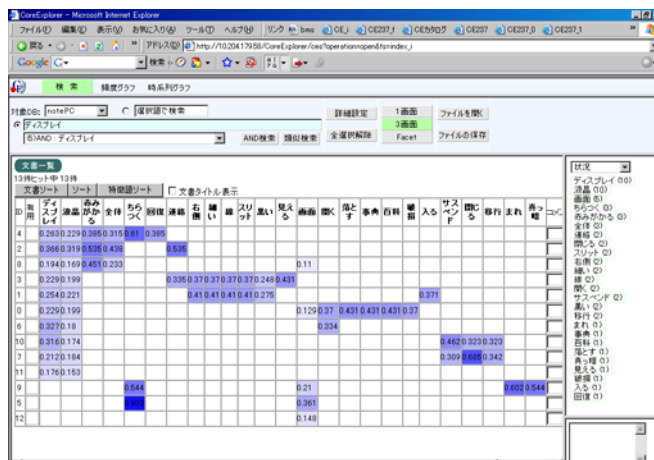
(2) 情報活用の目的

営業部署の管理者が、管轄する地域の営業状況や問題点を発見する効率を向上させる。

営業部門の管理者は、管轄エリアの営業所から日々登録される営業日報をチェックし、問題を示す報告を発見して対策を打っている。しかし、すべての日報をもれなく確認するのはコストの問題がある。そこで、大量の情報を効率よく読むための仕組みを検討した。

5.2 活用モデルの構築

ユーザのニーズには、斜め読みの効率化があった。ここで斜め読みとは、すべての文や単語に同じ力点をおいて読むのではなく、注目すべき単語を探しながらすばやく読み進める行為である。斜め読みを支援するために、図9の活用モデルを検討した。テキストマイニングによって特徴語を抽出することはできるが、それらは必ずしも注目すべき言葉ではない。ユーザは特徴語の中で注目した単語を選択し、並べ替えやグルーピングの操作を通して、自分にとって見やすい配置に修正する。これにより、注目すべき単語と単語間の関係がシステムに記憶される。これら記憶された情報は、同じユーザが次回以降



(図はイメージであり具体値は見えないように加工している。)

図10 斜め読みの画面例



(図はイメージであり具体値は見えないように加工している。)

図 11 特徴語の階層化の例

の使用で活用し、また、他のユーザが再利用することも可能である。

モデルを検証するため、斜め読みのプロトタイプを開発した。画面例を図 10 に示す。文章を行に、特徴語を列にとった表の自動生成機能を実装している。文章／特徴語の一覧をマップ表示することにより、可読性の向上を狙っている。文章の行と特徴語の列が交差するセルに、当該文書での特徴語の tf-idf スコアを表示する。これにより、文書群に対する単語の分布を視覚化している。ユーザの好みの順番で特徴語を並び替えることができる。並び替えは、特徴語のセルをマウスでドラッグすることによりおこなう。並び替えた特徴語に関連する順に文書を自動ソートすることができる。これにより、関連する特徴語と文書が近傍に配置され、tf-idf スコアが高いセルがグループを形成する。これを見て、どのような話題のグループがあるかを俯瞰することができる。

また、特徴語の階層を自動で生成する機能を開発した。画面例を図 11 に示す。階層生成のアルゴリズムは以下

の 3 種類から選ぶことができる。

(1) AND

グループのうち、すべての文書に含まれる特徴語を出す。文書との関連度が 1 つの単語について、すべての文書のうち最小値の値をその単語の関連度とし、その関連度の降順で表示する。

(2) OR

グループのうち 1 つの文書だけでも含まれる特徴語を出す。各文書との関連度の最大値を求め、降順で表示する。

(3) Average

グループのうち 1 つの文書だけでも含まれる特徴語を出す。各文書との関連度の平均値を求め、降順で表示する。

(4) 差分

Average で単語の関連度が他のグループでの平均値よりも高い場合表示する。親の階層は計算対象外とする。

これらの中で、絞り込まれた情報を参照する目的では AND を使用し、関連する情報を発見する目的では OR を参照するという使い分けを想定した。差分は Average に比べてよりグループ毎に特化した情報を発見することを狙っている。階層生成により、それまで 1 次元上に配置していた特徴語間の親子関係を視覚化し、文書を分類して参照することができる。

6. 評価

以上の対象とした各情報について、情報を利用している業務に対し活用モデルとプロトタイプを適用した。適

表 1 適用結果の評価

#	対象	内容	対象文書数	利用者	業務	改善点	効果
1	文書管理システムの蓄積情報	マーケット情報	約 3,000 文書	設計者	市場に適合した設計の企画	検索漏れの低減	従来:企画業務者は個人の経験に頼っていた。 ↓ 適用後:データによる裏づけをもとに判断する人が増えた。
2	不具合情報	クレーム分析	1,000 文書以上	設計者	品質保証体制の改善企画 設計の改善点の判断	不具合傾向の視覚化	↓ 裏づけが他人に分かりやすいため、組織の判断が早くなった。 (会議の中で判断ができる事例が出てきた。)
3	営業日報	営業日報分析	1,000 文書以上	営業企画	市場動向に適合した営業改善の判断	日報確認の迅速化	

用結果の評価を表 1 に示す。以下、評価内容について述べる。

6.1 業務毎の改善点

(1) 文書管理システムの蓄積情報

本研究適用前の検索システムでは、情報に付与するメタデータを表形式の台帳やフォルダ階層で管理していた。このため、既存のシステムではメタデータの表記が統一されていない問題があった。本研究適用後は、一覧表示されている中からメタデータを選択することにより“表記ゆれ”を防止でき、また最近登録した、あるいは関連するとシステムが判断したカタログ文書を表示し、これら文書のメタデータを選択することにより、文書間のメタデータを統一することができた。

利用者はメタデータを用いて情報を絞り込むことにより、検索もれを低減できることを確認した。

(2) 不具合情報分析

蓄積しているクレーム情報の分析に適用した。機種や部品毎の不具合の傾向を視覚化できることを確認した。

(3) 営業日報の活用

営業支援システムに格納されている営業日報をプロトタイプに取り込み、斜め読みの効果を検証した。日報を速く読む効果を確認した。

6.2 効果

個々のプロトタイプでの改善点に加えて、組織の業務の進め方が変わったことが重要な効果であった。従来は企画業務で判断をする場合、個人の経験に頼っていた。プロトタイプの適用後は、データによる裏づけをもとに判断する人が増えた。客観的な裏づけのほうが個人の経験よりも他人にとって分かりやすいため、個人の判断を通して組織の判断が早くなった。例えば、従来は会議で挙げられた仮定はその後の経過によって判断していたのに対し、会議の中で判断する事例が出てきた。

7. 今後の課題

本研究では、オントロジー技術の活用による情報分析の改善により、業務上の判断を支援できることが示唆された。今後は、大きな判断に対応できる人を増やし、人と組織の成長を継続して支援し、オントロジーの成長を検証してゆく。

8. 謝辞

本研究は、日立建機株式会社殿の研究プロジェクトとして実施された。研究開発では同社の多大なご指導とご協力をいただいている。関係者各位には、ここに記して感謝の意を表する。

9. おわりに

本研究ではユーザ企業と日立 TO が協働し、またユーザ、SE、コンサルタントおよび研究開発者の情報共有をおこなったことにより、ニーズから出発してソリューションを構築することができた。各案件は研究プロジェクトとして実行することにより、業務適用前のモデル構築段階から研究の内容をユーザと共有し、オントロジー技術を適用することができた。適用の結果、判断業務の裏づけとなるデータが他人に分かりやすく、組織の判断が早くなるという効果が得られた。今後は大きな判断を多くこなす人が増え、人が育ち、企業が強くなるソリューションを展開していく所存である。

参考文献

- 1) 武内雅宇, 小路悠介, 來村徳信, 林雄介, 池田満, 溝口理一郎: 知識成長過程を指向した設計意図知識管理システムの構築, 人工知能学会誌, Vol. 22, No. 3, pp. 263-275 (2007)
- 2) 古崎晃司, 來村徳信, 溝口理一郎: Web2.0 時代のオントロジー利用雑感—ライトウェイトからヘビーウェイトまで—, 人工知能学会研究会資料(2007)
- 3) 垂見晋也, 古崎晃司, 來村徳信, 溝口理一郎: 機能發揮・製造プロセス知識統合的記述枠組みに基づくナノテク材料設計支援システムの開発, 人工知能学会誌, Vol. 23, No. 1, pp.36-49 (2008)
- 4) 高梨勝敏, 佐藤俊也, 原島一郎: Web コンテンツからのオントロジーの再構成方法の提案と試作, 人工知能学会論文誌, Vol. 20, No. 6, pp. 417-425 (2005)



高梨 勝敏 1995 年入社
ナレッジソリューショングループ
知識管理システムの開発,
CoreExplorer の開発, 知識処理ツールの研究開発
takana@hitachi-to.co.jp



塚原 朋哉 1997 年入社
ナレッジソリューショングループ
CoreExplorer, テキストマイニングツールの研究開発
tomo@hitachi-to.co.jp



渡邊 まり子 2004 年入社
ナレッジソリューショングループ
CoreExplorer を利用したテキスト分析・コンサルティング
w_mariko@hitachi-to.co.jp



佐藤 俊也 1993 年入社
ナレッジソリューショングループ
CoreExplorer, テキストマイニングツール, 知識管理システムの拡販, コンサルティング
shu_sato@hitachi-to.co.jp