

企業活動における情報量の飛躍的な増加に伴い、これらの情報を蓄積、共有化し、更に、蓄積された情報を新たな知識として活用させるナレッジマネジメントが注目されている。

デジタルドキュメント プラットフォームは、OCR(Optical Character Reader)認識、自然言語処理などの当社独自のコア技術をベースに、情報共有 e 知識共有 e 知識創造のステップで、ドキュメント管理とナレッジマネジメントを支援する製品群、及びサポートサービスである。情報共有では、ドキュメント管理及びドキュメントフロー、知識共有では、自然言語検索、自動要約及び自動分類、知識創造では、テキストマイニングを提供し、蓄積された情報から知識を創造し、より価値を生み出す仕組みを実現する。

With the rapid expansion of information in corporate activities, attention is being focused on knowledge management for the storage, sharing, and optimal use of information. The Digital Document Platform is composed of products and support services which support document management and knowledge management in the steps of information sharing, knowledge sharing, and then knowledge creation. Information sharing provides document management and flow; knowledge sharing provides natural language retrieval, automatic summarization, and automatic categorization; and knowledge creation provides text mining.

These technologies realize a structure for creating knowledge with stored information and producing significant value.

## 1 まえがき




グループウェア、インターネット/イントラネットの普及により、企業活動に伴う情報量が飛躍的に増加している。企業活動に存在する情報は、建物や不動産同様の資産価値があると見なされ、これらの情報を共有化、再利用することが企業価値を向上させるポイントとなる。蓄積された情報を新たな知識として活用させる手段として、ナレッジマネジメントが注目されている。当社は、これまで培ってきたドキュメント管理の技術と製品をベースとして、ナレッジマネジメントを実践する。

以下に、ナレッジマネジメントの考え方を示す。

- (1) 情報共有 企業内に存在する情報(紙、電子ファイル、メール、Web情報など)を効率よく管理する。
- (2) 知識共有 管理された情報の中から役立つ情報を抽出し、それを知識として共有する。
- (3) 知識創造 更に、知識を共有するだけでなく、共有された知識を分析し、蓄積された知識の中から新たな知識を創造する。

デジタルドキュメント プラットフォームは、情報共有 e 知識共有 e 知識創造のステップで、ドキュメント管理とナレッジマネジメントを支援する製品群、及びサポートサービスである。また、ネットビジネスプラットフォームと連携し、インターネット経由でもドキュメント管理やナレッジマネジメントを提供する。

ここでは、図1に示すデジタルドキュメントプラットフォームの製品群、及び、それらを支えるコア技術について、情報共有 e 知識共有 e 知識創造のステップに沿って述べる。

活用レベル	製品	機能
情報共有	 全文検索オプション ドキュメントフローオプション	ドキュメント管理 入力・蓄積・検索・表示、 アクセス管理、バージョン管理、 暗号化保存、改ざん検出 全文検索 ワークフロー
	OCRシリーズ DVD, RAID デジタル複合機	文字認識、テキスト抽出 大容量ストレージ イメージ登録、変換
知識共有	 ナレッジマネジメントサーバ LotusNotes® 連携オプション Webサイト情報収集オプション	自然言語検索 全文検索 自動要約 自動分類 LotusNotes®の情報を検索 Webサイトの情報を検索
知識創造	 ナレッジマネジメントサーバ テキストマイニングオプション (開発中)	分析、マイニング 業務ノウハウと組み合わせ 新しい価値を創造

RAID: Redundant Array of Inexpensive(Independent)Disks

図1. デジタルドキュメント プラットフォームの製品群 情報共有 e 知識共有 e 知識創造を実現する。  
Products of Digital Document Platform

## 2 情報共有

eFilingMeister™は、Webベースで情報共有を実現するドキュメント管理システムである。ドキュメントフローと組み合わせることで、申請、承認、保管、活用といった業務を電子化し、これまで紙や人手で実施していた作業を大幅に軽減することができる。

### 2.1 ドキュメント管理

eFilingMeister™は、全文検索、アクセス管理、バージョン管理といった通常のドキュメント管理機能のほかに、以下の特長を持っている。

- (1) 簡単な情報入力手段の提供 デジタル複合機からコピー感覚で簡単に登録できる。また、優れたOCR機能により、紙、イメージからテキストを抽出し、知識として活用できる。
- (2) 高度なセキュリティ機能 文書の暗号化保存やデジタル署名による改ざん検出など、セキュリティ機能も充実している。融資約定書や行政文書など、高度なセキュリティが要求されるドキュメント管理にも対応できる。
- (3) 大容量ストレージのサポート 1.5 T(テラ)バイトを超えるDVD-RAMオートチェンジャや高信頼で高速なアクセスが可能なディスクアレイ装置により、大容量ストレージシステムの構築が可能である。

### 2.2 ドキュメントフロー

ドキュメントフローは、紙ベースで実施していた机上での作業をWebベースで実現するワークフローシステムである。ドキュメントフローの特長を以下に示す。

- (1) 実感豊かな作業空間 入力箱、出力箱、差戻し箱といった実感豊かなアイコンに対し、クリックやドラッグ&ドロップするだけで文書进行操作することができる。
- (2) 簡単な押印機能と付箋(ふせん)機能 日付印の押印による承認や付箋によるコメント付けなど、直感的でわかりやすい操作を提供している。
- (3) エージェントの指定 フローの宛先に保管者や配布先といったエージェントを指定しておくことで、フロー完了後に文書を保管したり、保管完了を他の人に通知したりすることができる。

## 3 知識共有

知識を共有するためには、目的とする知識をいかにすばやく探し出せるかが重要となる。eFilingMeister™ナレッジマネジメントサーバ(以下、ナレッジマネジメントサーバと略記)では、高度な自然言語処理技術をベースとした自然言語検索や自動分類によって、統一されたインターフェースですばやく知識を探し出せる。知識共有の構成を図2に示す。

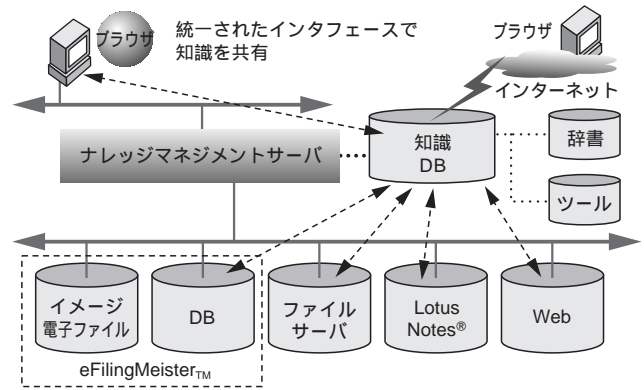


図2. 知識共有 管理された情報の中から役だつ情報を抽出し、知識として共有する。

Knowledge sharing

### 3.1 自然言語検索

自然言語検索は、会話感覚で質問文を入力すると、知識データベースの中でもっとも質問文に近い知識を探し出す機能である。自然言語検索の画面例を図3に示す。検索結果一覧には、自動的に作成される要約も合わせて表示されるため、一目で探したい文書を見つけられる。また、検索された文書をそのまま質問文として利用できる関連文書検索も提供している。この関連文書検索を繰り返すことによって、容易に目的の知識や新たな知識にたどり着くことができる。

### 3.2 自動分類

ナレッジマネジメントサーバは検索だけでなく、あらかじめ定められたカテゴリーに情報を分類する自動分類



図3. 自然言語検索の画面例 ごく自然な会話感覚で質問文を入力すれば、質問文に近い知識が得られる。

Typical display of natural language retrieval

機能も備えている。カテゴリーごとに代表的な情報を登録しておくだけで、以降、新しく登録された情報は、適切なカテゴリーに自動的に振り分けられる。そのため、自然と情報が整理され、系統的に情報にアクセスできるようになる。

### 3.3 既存情報の活用

知識共有では、新たに蓄積されたものだけでなく、既存情報を活用することも重要である。ナレッジマネジメントサーバでは、eFilingMeister™、ファイルサーバ上のイメージや電子ファイル、LotusNotes®(注1)のデータベース情報、Webサイト上のHTML(HyperText Markup Language)コンテンツなどを自動的に収集し、知識として活用することができる。特に、Webサイト情報収集は、自然言語検索と組み合わせることによって、ニュース、技術動向、事例などの情報を知識として共有できる。ポータルサイトやコンテンツプロバイダへの適用が可能である。

## 4 知識創造

### 4.1 テキストマイニング

電子化されたデータの中には、数値やキーワードのような定型データと報告書やコメントなどの非定型データがある。定型データについては、データマイニングなどの手法で活用されているが、非定型のテキストデータについては、まだ十分に活用されていない。しかし、営業日報やコールセンターへの問合せなど、大量のテキストデータを分析し、内容を迅速に把握したいというニーズが高まりつつある。そのため的手段として、テキストマイニングが注目されている。

ナレッジマネジメントサーバでは、ルールを基にテキストデータに含まれる要因を抽出し、定型データへ変換することによって、テキストマイニングを実現している。

### 4.2 テキストマイニングの実践例

営業日報から成功事例、及び機会損失事例の要因抽出を行うシステムにテキストマイニングを適用した。今回実践したテキストマイニングの概略を図4に示す。

まず、営業日報のテキストデータから、業務ごとにカスタマイズされたルール/辞書に従い、テキストに含まれる要因を抽出する。要因抽出では、自立語だけでなく、付属語も考慮しているため、“売れている”、“売れていない”という表現を区別することができる。図4の入力データ①では、納期、確保という組合せから、該当日報の営業活動では、要因として、納期が問題となっていることを抽出している。実践例では、営業日報から、競合、価格、仕様、納期、対応・保守、提案の六つの要因を抽出した。次に、抽出した要因を数値化し、データベース(DB)に出力する。数値化した要因データと受注/失注などの定型データを組み合わせることによって、例えば、納期の問題で失注した案件はN件あり、納期を

(注1) LotusNotesは、Lotus Development社の商標。

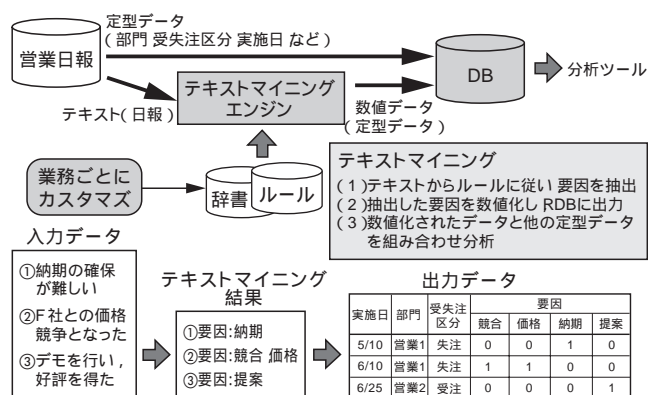


図4. テキストマイニング テキストデータに含まれる要因を抽出し、定型データへ変換する。

Text mining

強化すれば、売上げアップにつながるというような分析が可能となる。

## 5 デジタルドキュメント プラットフォームを支えるコア技術

デジタルドキュメント プラットフォームを支えるコア技術として、OCR 認識技術、文書の改ざん検出技術、自然言語処理技術について述べる。

### 5.1 OCR 認識技術

既存文書の入力において、OCR 認識技術は非常に重要である。帳票OCRの新シリーズ(OCR2000iシリーズ)では、帳票のレイアウトを自動的に理解する技術により、扱える帳票の種類を大幅に拡大した。

- (1) 既存帳票の読取りが可能 フォームを形成する罫線(けいせん)情報を認識し、あらかじめ登録されたフォームモデルと突き合わせることによって、帳票レイアウトを自動的に理解しているので、OCR専用帳票以外の帳票も読み取ることができる。また、レイアウト情報や漢字タイトルなどから帳票の種類を自動的に識別できるので、複数種類の既存帳票が混在した状態でも読み取ることが可能である。
- (2) 新認識方式による高精度読取りを実現 複数の認識アルゴリズムを統合的に処理することで、文字認識において誤読率を従来機種に比べ1/15に改良した。
- (3) アプリケーションソフトウェアとの親和性を向上 認識機能をソフトウェア化し、更に各モジュールのインタフェースは、Microsoft®(注2)COM(Component Object Model)で提供している。

文書OCR(ドキュメントリーダー)では、一般文書の文字認

(注2) Microsoftは、米国Microsoft Corporationの米国及びその他の国における登録商標。



識だけではなく、文書の自動タグ付けを応用したHTML、RTF(Rich Text Format)など、事実上の標準(DFS)の構造化文書形式への変換機能も提供している。今後の応用として、文書論理構造の認識技術を更に高度化し、XML(eXtensible Markup Language)タグ付け支援などを実現していく。

### 5.2 文書の改ざん検出技術

文書の改ざん検出は、ハッシュアルゴリズムを用いたデジタル署名によって実現している。ハッシュアルゴリズムが持っている「異なるデータから同じ特性値が作成される可能性は極めて低い」という特長を利用して、改ざん検出を行う。以下に、改ざん検出の仕組みを示す。

- (1) 文書の登録時に、ハッシュアルゴリズムを用いて、文書の特性値を作成する。
- (2) 更に、求めた特性値を公開鍵(かぎ)方式によって暗号化し、デジタル署名として、文書とともに保管する。
- (3) 改ざん検出の際には、再度、文書をハッシュアルゴリズムによって特性値を求め、保存してあるデジタル署名を復号化した特性値と比較する。二つの値が一致していれば、改ざんがなかったことになる。

### 5.3 自然言語処理技術

知識共有、知識創造を支える技術として、優れた自然言

語処理技術がある。ナレッジマネジメントサーバで利用している自然言語処理技術を図5に示す。他の自然言語検索では、一般的に(1)の単語情報だけが使われているが、ナレッジマネジメントサーバでは、(2)の意味情報を加えることによって精度が高い自然言語検索を実現している<sup>(1)</sup>。

- (1) 単語情報 入力された文書を形態素解析で単語に展開し、単語情報を登録する。検索の際は、単語の出現頻度などによって重み付けを行い、重みが大い単語で該当した文書ほど、質問文に対する類似度(スコア)が高くなるようにする。
- (2) 意味情報 形態素パターンルールを用いて意味単位に分割し、構造抽出を行い、意味情報を登録する。検索の際は、質問文の意味情報と登録された意味情報が一致する文書の類似度が高くなるようにする。図5の例では、意味情報(症状)の一致している文書(事例1)の方が一致していない文書(事例2)より、類似度が高くなっている。これにより、単語だけでなく、文脈を考慮した検索が可能となっている。

## 6 あとがき

デジタルドキュメントプラットフォームは、情報の管理から知識の管理へ進化させることによって、共有・創造した知識を更に価値ある知識として、蓄積・活用できる仕組みを提供する。また、ルール作成支援など、ナレッジマネジメント構築支援を中心としたサポートサービスも提供する。今後も、知識共有、知識創造の分野に注力し、業務ノウハウと組み合わせる新しい価値の創造を行う分析・マイニング機能の提供、コミュニケーション機能と組み合わせる新しい価値を創造するコラボレーション機能の提供に取り組んでいく予定である。

## 文献

- (1) Nakayama, Y., et al. "Know-how sharing Using a knowledge sharing system KIDS." Proc. of 3rd Practical Aspects of knowledge Management. 2000-10, SGAICO. 2000, p.21.1 - 7.

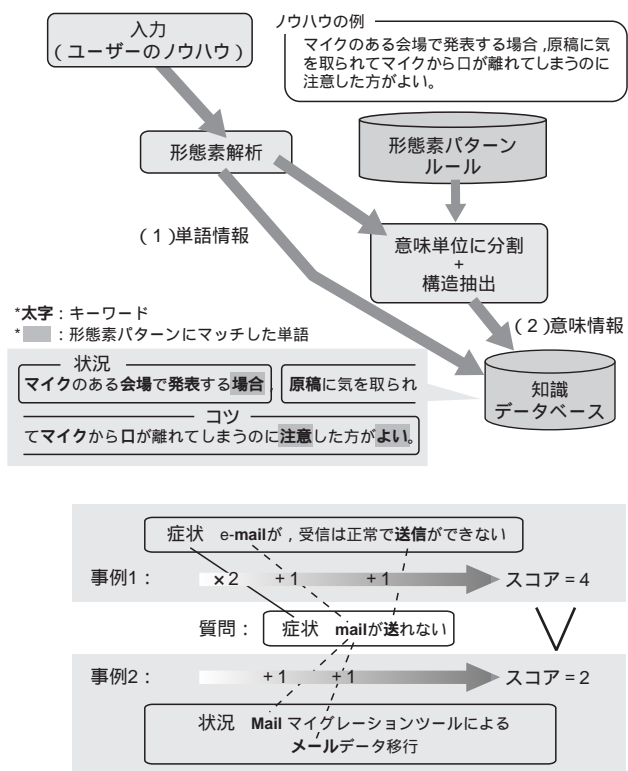


図5 自然言語処理技術 形態素パターンルールを用いて構造/意味抽出を行う。

Natural language processing technology



齋藤 稔 SAITO Minoru  
デジタルメディアネットワーク社 府中デジタルメディア工場 ミドルウェア部グループ長。コンピュータミドルウェアの開発・設計に従事。  
Fuchu Operations - Digital Media Equipment



永村 栄治 NAGAMURA Eiji  
デジタルメディアネットワーク社 府中デジタルメディア工場 ミドルウェア部主務。コンピュータミドルウェアの開発・設計に従事。  
Fuchu Operations - Digital Media Equipment



豊田 真代 TOYOTA Mayo  
デジタルメディアネットワーク社 府中デジタルメディア工場 ミドルウェア部主務。コンピュータミドルウェアの開発・設計に従事。  
Fuchu Operations - Digital Media Equipment