

Office アプリケーションで実現するナレッジ管理

坂口 良
松田 有司
榊原 淳

ファイル管理の手法として最も一般的なものはバイナリデータとしてファイルシステムに保存する方法であるが、検索性能の低さやデータの二次活用の困難さという問題を抱えている。当社は Office アプリケーションで作成された文書を登録するだけで属性情報から関連付けデータを取得し、自動的にデータベース化を行うアプリケーションを開発した。このアプリケーションは核となるデータベースに完全にスキーマレスでデータ構造に依存しないネイティブ XML データベースを用いることで実現が可能となった。

Realization of Knowledge Management with Office Application

Ryo Sakaguchi
Yuji Matsuda
Atsushi Sakakibara

The most common file management method is to store binary data in the file system. However, this method contains problems such as complications of secondary utilization of data and low-performance searches. Media Fusion has developed an application that automatically creates a data base and retrieves related data from property information simply by registering sentences created with Office Application. This application, by utilizing a native XML data base which is completely schemaless and does not rely on data structure, has made realization possible.

1. ドキュメント管理の現状について

ドキュメント管理と言っても様々な手法が存在する。最も原始的な方法としては、紙媒

体の物理的なファイルとして格納しておく方法だが、現状ペーパーレス化が進む企業、自治体において、媒体を電子化し記憶装置に格納しているかと思われる。最も一般的な管理方法としては、ファイルシステムに直接ディレクトリを作成し格納する手法だが、この手法だと排他制御やバージョンの管理など細心の注意を払わなければ効率的な管理は難しいと思われる。特に企業内での部署間の情報共有などを行なう場合、極めて困難である。例を挙げると過去作成した稟議書を検索する場合、テキスト文書として格納しておけば **grep** などで検索出来るが、**Office** アプリケーションなどで作成された文書の場合、キーワードとなる名称、文言を指定しない限り探し出すのは困難であるかと考えられる。

近年ドキュメントを管理する為のソフトウェアが多数存在している。これはドキュメント管理の必要性、重要性が高まってきたことの現れだが、ソフトウェアがユーザの要望に対して確実に実装出来ていないのが現状で、機能的に十分だと言い難い物が多々存在している。基本的にはドキュメント管理システムの定義は、ドキュメントを格納でき、検索が行なえればドキュメント管理システムとされているが果たしてそれだけで良いのだろうか。弊社ではそれらの問題を解決しドキュメント管理システムの開発を行なったが、技術的どのような手法で構築されているか解説を行なう。

2. ドキュメント管理ソフトウェアの問題点

前章で述べたように現在利用されているドキュメント管理ソフトウェアでも、ユーザの要求を満たしていないものが多々存在する。そこで考えられる問題を下記に列挙した。

- **Office** 文書や **PDF** などのアプリケーションに依存した電子媒体は、実データに対しての検索を行なえない。
- 登録したドキュメントに対して検索を行なう場合、キーワードとなるデータを別途手動で登録する必要がある。(メタ情報の入力が必要である。)
- ワークフロー機能などドキュメントの公開までの一連の機能に拡張性がない。
- セキュリティ機能が貧弱な為、情報漏洩の危険性がある。
- ドキュメントのプレビューを行なう場合、**Office** アプリケーションなどの機能を使っている為、全てのクライアント側に **Office** を導入する必要がある。
- 登録されたドキュメント **INDEX** などを生成するが基本的にはドキュメントをそのまま格納する為、データベース化されておらず、二次利用が非常に難しい。
- 登録されたデータをデータベースに格納する場合、原本の保証が難しい。
- 登録をドキュメント単位でしか格納出来ない。
- 登録するドキュメントの分類分けなどの自動化は難しい。
- ドキュメント管理システムと他の基幹システムとの連携が非常に難しい。

上記以外にも多数の問題を抱えている。これらの問題を解決する為今後の章で述べる。

3. 弊社ドキュメント管理ソフトの特徴

弊社ドキュメント管理ソフトの特徴としては、文書ファイルにおける定められたセル位置に記憶された実データに基づいて当該文書ファイルの特徴データを前記階層情報に基づいて階層構造を含むタグ情報を生成し、当該タグ情報を取得した実データに付加することによって XML 形式の特徴データの生成が出来ることである。上記の XML 形式の特徴データを生成する為の情報は、当該データの属性情報に所定のデータを記録することにより実現している。(現在特許出願中)

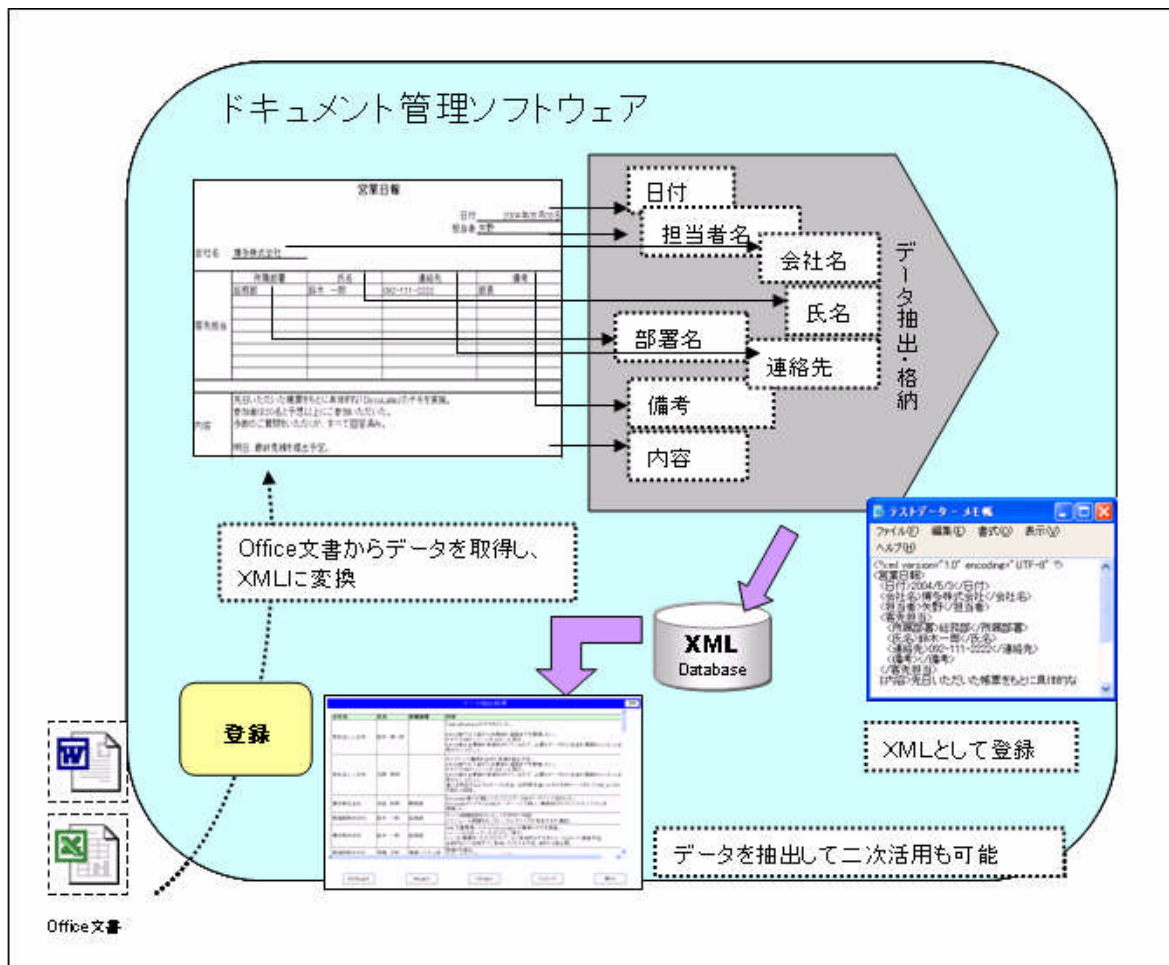
簡単に言うと Office 文書の属性情報 (プロパティ) にシート名、セルの位置などの情報とマッピングする XML のノードを関連付ける情報が記入されている。関連付け定義が Office 文書に内包されていることにより実際に XML へのコンバートを行なう際、関連付けファイルを別途取得する必要がなく Office 文書単体で処理を完結することが可能である。実装上問題となったのが、Office 文書のバージョンの違いによる Office アプリケーション API の差異である。実装としては、Word や Excel の機能をフックし操作しているのだが、最新のバージョンの Office 文書だけの対応であれば、文書全体を XML 形式でエクスポートする機能があり、定義の埋め込みも容易に行なうことが可能である。

だが XML 形式でのエクスポート機能は Word97 など旧バージョンでは実装されていない為、その機能を用いて実装することは不可能である。その他各 Office のバージョン依存 API が多々存在し、使える機能が制限されている。そこで Word95 や Excel 95 などから存在する属性情報を用いて実装を行なった。

他にも完全にスキーマレスであるという特徴がある。作成した文書の登録を行なうと自動的にデータベース化され検索・抽出などが可能となる。上記の関連付け定義がない文書が登録された場合、文書に記述されている内容 (データ) を全て取得し階層を持たせた形で XML 形式にコンバートしネイティブ XML ストレージに格納している。その他 WINAPI を用いて文書のファイル情報 (ファイル名、作成日時、更新日時、作成者、ファイルサイズなど) を取得し、合わせて XML 化しストレージに格納している。

その他の属性として独自にキーワードを設定し XML に属性を追加することにより一般的なキーワード検索機能も実装している。この実装はリレーショナルデータベースで非常に困難である。リレーショナルデータベースは常にスキーマが決まっており定型のデータしか扱うことが出来ない。リレーションを複雑に設定し実装することは不可能ではないが、処理速度の問題なども発生することは確実で現実的ではない。このドキュメント管理システムは、バックエンドにネイティブ XML データベースが存在していることにより実装出来た良い例である。

概念図



Office 文書を登録すると関連付け定義を元に XML を自動的に生成。同時に検索用 INDEX ファイルも作成し XML ネイティブストレージに格納

4. 弊社ドキュメント管理ソフトの機能と実装概要

・ ドキュメントの格納機能

Word、Excel を登録する際、Office 文書の内容から自動的に XML 形式のドキュメントを作成し XML ネイティブデータベースへ登録される。これにより Office 文書のデータなどを二次利用することが可能となる。無論全文検索を行なうことも可能である。

また登録を行なう際、下記の 4 つの手順で登録することが可能である。

● 単体ドキュメントの登録

ドキュメント名を指定し登録。パスから Office 文書のインスタンスを生成しデータを取得し登録

- ドラッグ&ドロップ

複数のドキュメントに対しても対応。ドラッグ&ドロップを行なった時のイベントをフックし複数のドキュメントのパスを取得。その後は単体登録と同一の処理を行なう。
 - ディレクトリからの一括登録

指定ディレクトリにドキュメントを格納しておき自動的に登録。ディレクトリ監視アプリケーションを起動しておきディレクトリにファイルが登録されるとそのドキュメントのパスを取得。その後は単体登録と同一の処理を行なう。
 - メールの添付ファイルからの登録

メールに添付されているドキュメントを登録。予めメールサーバの情報を登録しておきタイマー処理でメールサーバにアクセスする。新着メールが存在した場合、添付の文書を取得し登録する。
- ・ 関連付け定義機能

登録するドキュメント内にデータマッピング定義を埋め込むことによりデータを構造化することが可能である。実装方法など詳細については「3. 弊社ドキュメント管理ソフトの特徴」を参照。
 - ・ ドキュメントの検索機能

登録時に自動的に作成された項目一覧からの検索が可能である。その他全文検索、絞込み検索などが可能である。もちろん登録時にキーワードを登録しておけばそのキーワードで検索を行なうことも可能である。
 - ・ プレビュー機能

登録時に自動的にプレビュー用のファイルが作成される為、プレビューを行なう時に Office の起動が必要ない。もちろん対応する Office アプリケーションなどクライアントにインストールしておく必要はない。
 - ・ 抽出機能

登録されているデータは XML 化されている為、必要なデータだけ取得し結果を整形した形で出力することが可能である。マイニングツールとして使用することが可能である。

抽出を行なう際に検索項目を選択するが、選択された項目には対象となる XML のノードパスが保持されており、それを元に XML に対して検索を行い一覧が作成される。

- ・ ワークフロー機能
登録されたデータをワークフローに流すことが可能である。承認、決済を行い指定日に公開することが可能である。
ドキュメントの属性にステータスを保持しており、そのステータスを変更することによりドキュメントの状態を変更する。

- ・ 階層型管理機能
分類のフィルタリングを予め行なっておくことにより、登録するドキュメントを階層化された分類に自動的に振り分け登録することが可能である。
分類の階層を XML で保持している。
ディレクトリ構成のプロパティなどは属性に値を保持させるなどの処理を行なっている。

- ・ バージョン管理機能
更新したドキュメントに対して再度登録を行なうと自動的に履歴をとる為、バージョン管理を行なうことが可能である。例えば過去のある時点のドキュメントを取得するなどが可能である。
過去のデータに関しても全て原本、XML データを保持しており当該ドキュメントの公開期間も保持している。

- ・ チェックイン・チェックアウト機能
ドキュメントの排他制御を行える。
ドキュメントのロック情報を XML で保持しておりその値を確認して制御を行なっている。

- ・ 原本保証
ドキュメントを登録した際、データを抽出して XML 化しデータベースへ格納するが、原本はそのままの状態を管理している為、原本の保証が可能である。
ドキュメントをバイトレベルで同一のものをドキュメント管理システムで保持している。

- ・ セキュリティ機能
通信時に強固な暗号化を行っている為、セキュリティ上安全である。

5. 弊社ドキュメント管理ソフトの利用用途

通常のドキュメント管理ソフトと違い、XML としてデータを管理している為色々な使

い方が可能となります。

1. ファイリング

基本的なドキュメント管理の機能です。Word、Excelなどのドキュメントをチェックイン・チェックアウト、バージョン管理、ワークフロー等の機能により管理することが可能である。

2. データベース化

先に述べたように登録時に自動的にXMLを生成されXMLネイティブデータベースに登録をしますので、ノンプログラミングでデータベースを構築することが可能である。

3. 既存システムとの連携

ドキュメント管理システムで管理するWord、Excelから必要な項目だけをXML形式など様々な形式で抽出することが可能である為、他の既存システムとの連携が容易である。出力することが可能な形式として、XML、CSV、Excelシートがある。これだけの形式に対応していれば、他のシステムとの連携がスムーズに行なえる。

6. まとめ

企業、自治体で扱う膨大な量のドキュメントを管理・活用するにはやはりドキュメント管理ソフトは必須である。ただ管理出来るだけでなくスキーマレスでありノンプログラミングでデータベース化されることによってより効率化が図れると考えられる。これらを実現するにはリレーショナルデータベースでは困難でありネイティブXMLデータベースが最適であるであろう。

近年デファクトスタンダードとなりつつあるOfficeアプリケーションをインターフェイスとし自動でXML形式に変換される機能を有するドキュメント管理システムを利用し、ナレッジマネジメント化を計り業務効率を向上することが可能であると考えられる。

株式会社メディアフュージョン
Media Fusion Co., Ltd.