

ワークプロセスと文書の緩やかな 制約に基づく管理方法の検討

尾崎正治 林浩一

富士ゼロックス(株) IT 事業開発部

文書の散在性が解消され、多種多様な文書がアクセス可能となっている。これらの文書を非定型のワークプロセスで有効利用するには、文書の表現形態の異質性と、プロセスにおける文書のコンテキストとの分離が問題となる。これを解決するために、文書の表現形式の違いを吸収して同種の情報を取り出す仕組みと、ワークプロセスの自動捕捉を行う仕組みを導入する。これらを前提として、文書要素に与える意味付け、ワークプロセスでの文書に与える意味付けを緩やかに制約し、さらに両者を文書群に対するメタデータと統合的に管理することによって、文書情報の有効利用を促進することを提案する。

A Proposal of Loosely Constrained Management of Work Processes and Documents

Masaharu OZAKI and Koichi HAYASHI

Fuji Xerox Co., Ltd.

We are now able to access a large number of electronic documents through the network. In this circumstance, structural variation in representing information in documents and separated management of work processes and documents prevent from repurposing the information. In order to solve this problem, we propose to introduce a technique that extracts the information represented in various document structures, and an environment where dynamic work processes are captured semi-automatically. Based on these, we also propose a new direction for repurposing information in the documents in which the captured work processes and the extracted information from the documents are unified and managed as metadata to the documents.

1 はじめに

知的生産活動に従事するワーカーの作業環境は、ここ数年の間に大きく変化してきた。パーソナルコンピュータを用いた業務の比重が高まり、文書やメッセージの作成・編集・交換とい

った活動が電子媒体を通じて行えるようになった。

これらの情報を組織的に管理されたりポジトリで共有し、さらにワークフロー管理システムによってワークプロセスに統合化することも可能になっている。しかし、現状では、文書とワ

ークプロセスの統合の対象となっているのは経理業務など定型化しやすいものに限られている。一方、試行錯誤的に進められるダイナミックな知的生産活動はますます重要になってきており、各ワーカーやグループは様々なプロセスから作成・収集された多種多様な情報を扱う必要に迫られている。このような状況下で組織的な活動によるメリットを生かすには、共有されている文書群に内在化している情報を、各ワーカーやチームの活動目的や状態に適合する構造へ組み替え、提示する情報統合を実現することで、文書共有の効果を高めてゆかなければならない。

本論文では、上記を実現するための課題と、それに対する解決の指針を示すことを試みる。まず、電子文書の発展を概観した上で、今後解決すべき課題について示す。次に、文書、ワークプロセスに対して緩やかな制約を課すことを指針とする解決方法を提示する。最後にその緩やかな制約のもとで、ワークプロセス、文書群を統一的に管理することの有用性を示す。

2 電子文書処理のこれまでの発展

2.1 文書処理の発展

文書は長い間、紙を媒体として人の間で情報を交換するためのコンテナとして存在してきた。レーザープリンタ等の印字ハードウェアの進展と、ワードプロセッサ、DTP ソフトウェアにおける、WYSIWYG と呼ばれる紙への出力を前提とした編集メタファーの実現によって、美しい体裁で紙に出力するためのプロセスが大幅に効率化された。

これと並行して、紙文書管理のアナロジー（ドロワ、フォルダ）で電子的な文書管理が実現されるようになった。これを発展させる形で、検索機能などの多くの機能が文書管理システムとして実現されている。文書配布の手段については、電子メールの登場によって、紙を媒介とせ

ずに電子的な表現のまま文書を、同時に多数の人間に対して配布することが可能となった。さらに、インターネット上の WWW は、文書を限られた範囲における情報配信の媒体から、不特定多数に対する情報共有のための媒体へと変化させた。

文書が電子的に作成され、蓄積・配布されるようになると、次第にその内容を機械的に処理する要求が顕在化してきた。この要求に応えるために、文書そのものを論理的な意味を持つデータ構造として表現することが必要となった。SGML [1]を代表とする構造化文書は、文書における各要素に意味付けを与え、それらを階層的に構成した論理構造によって文書を表現し、この構造をもとにさまざまな処理を実現可能とするものである。最近では、インターネット上で構造化文書処理を実現するための規格 XML [2]が制定されている。このように、多くの文書フォーマットがその利用目的に応じて利用されるようになってきている。

2.2 文書の散在性の解消

紙文書管理、およびそのアナロジーに基づく電子文書管理においては、小人数のグループ、企業でいえば部門単位で文書は蓄積・管理されてきた。このために多くの文書は部門ごとに管理され、全社的なネットワークには接続されていないながら、個々の文書管理システムに対するアクセスインタフェースの違いから、同一のアプリケーションからアクセスできないということが生じていた。

このような文書管理システムのアクセスインタフェースの違いによる文書の散在性はこれまでも問題視されており、文書管理システムのベンダーによって、文書管理システムアクセスインタフェースの標準である DMA が制定されている [3]。DMA は、異種の文書リポジトリを統合し、統一的なインタフェースでアクセスする

ことを可能とするアクセスインタフェースの標準である。これが実現されれば、文書管理システムの違いを意識することなく、文書をアクセスすることが可能となる。

DMA とは異なるが、WWW すなわち Web サーバの技術を企業内に展開したイントラネットによって、特別な仕組みを用意することなく、容易に文書を共有することが可能となった。現状の Web サーバは基本的に文書を閲覧するための機能を中心としたものであり、文書管理に必要となるいくつかの機能、格納、バージョン管理、アクセス権管理などは含まれていない。しかし、WebDAV と呼ばれる、Web サーバに対して上記の機能を付加するためのプロトコルの拡張も提案されている [4]。

このような DMA や Web ベースの文書共有の仕組みは、これまでアクセスが困難であった散在している文書群へのアクセスを可能とするインフラストラクチャーとして機能することになる。これらが広く採用され、利用されていくと、散在する文書への到達可能性については問題となくなる。しかしながら、このことは次節で述べる文書の異質性を顕在化することになる。

3 解決すべき課題

組織において、文書の共有の仕組みが整うと、共有されている文書群から、必要となる情報を抽出し、利用するための支援が求められる。特に、本稿で対象としているダイナミックな知的生産活動の中では、必要となる情報の構造は試行錯誤的に、あるいは一時的に規定される。われわれは、このような活動の中で、計算機が主体となって文書群から同種の情報を抽出して、利用者が一覧、比較するというを行う上で、文書の異質性と、文書とワークプロセスの分離という二つの課題に焦点をあてる。

3.1 文書の異質性

文書上に情報を表現する表現形態は、配布対象者によってさまざまである。多くの場面で、文書上の情報の表現形態¹は、作成者と利用者の相互了解のもとに設定される。文書の作成者と直接の利用者以外の第三者が、二次利用しようとした場合には、その相互了解を理解した上で操作をしなければならない。

前述の文書の散在性が解消され、アクセス可能な文書の範囲が広がると、配布目的とは異なる目的に文書を利用する要求が増加することが予想される。

したがって、問題は、人間がみれば同種の情報を表現しているにもかかわらず、表現形態の違いから計算機による抽出が自動化できないという状況である。もっとも単純な例で言えば、企業内の複数の部門における出張報告書を一覧したい場合、少なくとも、「日時」、「出張者」、「出張先」、「サマリ」など定型化された情報が文書中に存在することは了解しているが、これらがどのような表現形態になっているかについてはわからない、といった図 1 に示すような状況がこれにあたる。これを計算機処理から見た文書の異質性と呼ぶこととする。

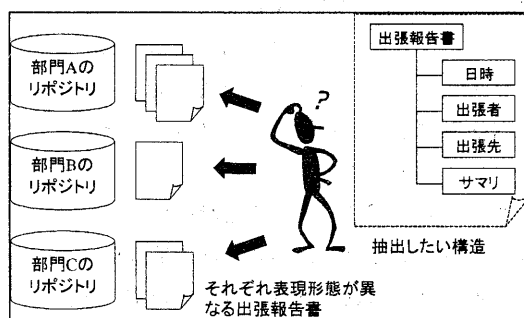


図 1 文書の異質性

¹ 表現形態とは、文書フォーマットとそこにおける情報の構成方法の両者を意味する。

異質性の原因は、文書フォーマットの違いと、各文書フォーマットにおける情報を表現する構造の自由度である。前者についてはフォーマット変換などによって、解決は可能である。しかし、後者については、解決は困難である。たとえば、ワードプロセッサなどの書式指向の文書において、ある項目と対応する内容の組を並べるだけでも、箇条書き、表、フォントを変更した段落など多くの表現方法が存在する。XMLをはじめとする構造化文書では、作成者と利用者の相互了解を、文書型 (Document Type Definition: DTD) として共有し、計算機処理がそれを参照することで解釈の曖昧性をなくすることができる。しかしながら、構造化文書を利用した場合でも、同一のセマンティクスを持つ情報が、文書型によって異なる構造表現がなされる可能性がある。

このような異質性の解消については、以下の2つの方法が考えられる。

■ **作成者が、利用目的ごとに統一した表現形態で文書を作成する (作成者の負担)**

これは、構造化文書が提唱している方法論であるといえる。たとえば、NTTが行っているXMLを利用したカタログ作成サービスの実験では、日本酒のカタログデータベースを作成している[5]。文書の作成者は、文書型に沿った日本酒のデータをHTML文書の中に埋め込む。この文書の利用者は、カタログサービス提供者であって、文書を受け取り自動的にデータベース化して、サービスを提供する。

われわれの想定している活動においては、文書作成時にあらかじめ必要となる論理構造を規定することは困難である。仮に、規定できたとしても、上述の例のように作成者に何らかの利益をもたらさない限り、規定された構造での文書作成は、促進されないといえる。

■ **利用者が、目的に応じて表現形態の違う文書を変換して均質にして利用する (利用者**

の負担)

作成者には負担をかけずに、利用者側で必要な情報を抽出することを行うものである。この例としては、Junglee社が行っている、WWW上で複数の新聞社が提供する求人広告や、オンライン書籍販売の情報を、均質にして仮想データベース化して比較・検討を行うサービスの提供があげられる[6]。これらのサービスでは、対象となる文書群、すなわち個々の求人広告や書籍データを提供する側には何ら負担をかけることなく、その対象ごとにサービス提供者、すなわち文書の利用者が変換ルールを記述して均質化している。しかしながら、この処理は、対象文書の範囲があらかじめ設定できるということが前提となっている。したがって、新しい対象が出現した場合に、それに対する変換ルールを記述しなければならなくなる。表現形態のバリエーションが多くなればなるほど、この方法では対処が困難になっていく。

以上、述べたように作成者、利用者のいずれか一方が、異質性解消のコストを負担するという方法では、必ずしも、われわれが想定している問題点を解決するものとはならない。したがって、作成者、利用者の負担がそれぞれ少なく、対象文書の表現形態の変動に対して個別に対処することなく、指定した構造と同種の情報を抽出することを可能とする機能の実現が必要となる。

3.2 文書とワークプロセスの分離

一般に、文書は何らかのワークプロセスの中で生成され、他のプロセスで参照される情報のコンテナとして位置づけられる。[7]で示されているように、ワークプロセスは、文書にその中での意味合い (コンテキスト) を与える。

文書の内容を正確に把握するために、その文書が作成されたコンテキストを知らなければならないこともある。例えば、仕様書のような文

書の場合、全く同じ文書であっても、承認前であるのか、承認後であるのかによって、記述の内容の扱いは異なる。新たな意図に情報を組み替えるという情報統合を考えると、文書だけでは不足している情報を、活動のコンテキストを手がかりに補うことができる。しかしながら、ワークフロー管理のできる定型的な業務を除いては、文書を共有する段階で、これらのコンテキスト情報は失われてしまう。これを文書とワークプロセスの分離と呼ぶことにする。

文書とワークプロセスの分離の原因は、コンテキストの記述方法とその共有方法にある。今後、テレワーク、SOHO など遠隔地で文書をネットワーク上で共有しながら進めていくようなワークスタイルが広がると予想される。このような活動の日常的なコミュニケーションを補うために、活動を実施するメンバー間で、文書群に対する意味付けを明示的に共有することが求められていく。文書とワークプロセスの分離の解消についても、以下の 2 つの方法が考えられる。

■ 作成者が業務ごとに統一した方法でコンテキストを表現する(作成者の負担)

これは、ワークフローが提唱している方法論であるといえる。定型的なワークプロセスについては、ワークフロー管理などのアプリケーションを導入することによって、プロセスのどのステップで、どのような文書が参照され、新たに生成されたのかを管理することができる。しかし、非定型なプロセスについては、ワークフローとして管理できず、結果として意味的な関連を保持することは困難となっている。非定型なワークプロセスにおいて、作成・参照される文書群に対する意味付けを管理する方法としては、現状ではファイルドロワ、フォルダに意味付けを注意深く与えて保管しておくか、もしくは文書ごとに意味付けを示す属性を定義し、そ

こに状態を入力していくなどが考えられる。いずれにしても活動そのものとは別に意味付けを管理するための手間が発生することになり、活動遂行の妨げとなる。

■ 利用者が文書の作成コンテキストを探索した上で利用する(利用者の負担)

作成者には負担をかけずに、利用者側に必要な情報を抽出することを行うものである。この例としては、筆者らが、Interlocus と呼ぶプロトタイプで提唱した、ワークプロセスと文書との関連を捕捉するための枠組みがある[8]。このシステムは、文書をもとに進められる活動ごとにネットワーク上での文書操作環境を一種のデスクトップとして提供している。デスクトップ上には、二次元の領域を与え、それに意味合いを持たせることができる。すなわち、文書はその領域に置くという操作で文書に対する意味付けを与えることができる。これに加えて、作業者が文書に対して実施した操作(新規作成、領域への移動、更新、削除等)を行った結果生じた、文書群に対する意味付けの変化、状態の変化を時系列で捕捉する。

このことにより、以下のことが実現できる。

1. ある文書に注目したときに、それがどのような活動において作成された、あるいはどのような活動で参照されたかというコンテキストを明確に把握できる。
2. あるワークプロセスに注目したときに、実施された活動を時系列で参照し、どのような進め方を行ったかということを知ることができる。

しかしながら、この方法は対象とする文書のそれぞれに対するコンテキストを確認することを利用者に強いるものであって、少数の文書の再利用の場合は有効でも、大量の文書の目的変更のために利用するのは現実的ではない。このためには、個々のワークプロセスを時間軸で俯

瞰したり、複数の同種のワークプロセスを参照比較して見る必要がある。

4 課題解決のための指針

我々は、文書の異質性および文書とワークプロセスの分離という 2 つの課題を解消するために、それぞれ、緩やかな制約に基づく管理を導入することを提案する。

4.1 異質な文書からの文書内容の抽出

3.1で述べたように、作成者、利用者ともに大きな負担をかけずに、多様な文書から、指定した構造と同種とみなせる部分を抽出する技術を実現する必要がある。このためには、図 2 に示すように、抽出すべき構造が表現されうる形態のバリエーションにシステムが自動的に展開し、それぞれについて、適合するかどうかを判断することが必要となる。バリエーションは、書式指向文書での表現方法と、構造化文書における DTD の構築方法論などを調査する必要がある。

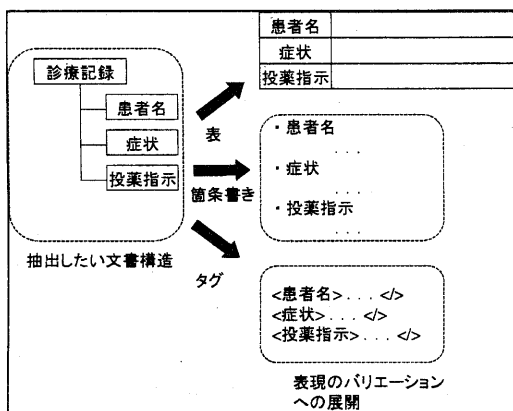


図2 指定構造からの多様な表現形態への展開

上記の技術が実現されたとしても、制約なしに自由に作成された文書に対しては情報の抽出は困難である。したがって、運用面での前提条件として、少なくとも必要となる情報を特定するために、表現形態の最低限の取り決めが必要となる。この取り決めは、文書全体の論理構造

を規定するのではなく、部分ごとにある制約に従っていけばよいものとする。この制約を文書構造における緩やかな制約と呼ぶことにする。

緩やかな制約のもっとも単純な例としては、個々の要素に対しては、その内容を示す見出しをつけることを義務づける。たとえば、出張報告書には、どのような構造で表現してもよいが、かならず「日時」、「出張者」などを明記するという制約になる。これは、構造化文書のタグで表現されてもよいし、書式指向文書の見出しなどを用いてもよい。このような要素に対する意味付けを与える語彙については情報流通の範囲において、統一しておくことが望ましい。また、個々の要素ではなく、要素の集合、あるいは単純な構造として意味を持つものについても、それを単位として制約することも必要である。たとえば、会議レポートの中で「出席会議」は「名称」、「場所」、「参加人数」から成ることを了解としておき、その部分は「出席会議」という単位で均質に扱うことができる。

このような制約は、作成者の負担を必要とするが、文書全体の構造を制約するものではないため、比較的受け入れられやすいと考える。利用者は、部分構造の組み合わせによって必要となる情報を特定することが可能となる。すなわち、構造化文書の文書型のような文書全体に対する強い制約ではなく、個々の要素について、相互了解として認識できるように、それぞれ独立して意味付けを与えておくという緩やかな制約のもとに文書を作成しておくことで、情報の利用性を高めることができる。

4.2 ワークプロセスにおける文書への意味付けの捕捉

3.2で述べたように、作成者、利用者ともに大きな負担をかけずに多様な活動コンテキストで作られた文書を統合管理するには、複数の同種のワークプロセスを参照比較して扱う技術を

実現する必要がある。

そのためには、文書への緩やかな制約と同様に、同種のワークプロセスにおいては文書群に対する意味合いを示すための語彙などを取り決めて運用することが前提となる。たとえば、ソフトウェア開発においては、「要求仕様」、「機能仕様」、「テスト結果」などのキーとなる語彙を統一するということを意味する。それ以外に必要な意味付けは自由に与えればよい。このような制約をワークプロセスに対する緩やかな制約と呼ぶ。

この活動履歴のデータは、ある場面では文書として扱うことが要求される。たとえば、進捗状況の報告などがそれにあたる。活動履歴のうちの、ある期間における文書群の意味付けの変化を提示することによって、進捗報告のひな形が自動で生成できる。このようなワークプロセスの履歴を文書として扱うことができることは、ワークプロセス、文書を統合的に管理できるということを示唆している。

5 ワークプロセスと文書の統合的管理

5.1 メタデータによる統合管理の実現

これまで、ワークプロセスにおける文書への意味付けと文書の要素に対する意味付けを別個に考えてきたが、これらは、どちらも文書外部に存在するメタデータとして統合的に扱うことが可能である。これを模式的に図 3 で示す。この統合化によって実現できる機能の例を示す。

ここでは、新しいプリンタを開発しようとするプロセスで、過去の設計においてどのようなスペックの機器を競合としてあげたのかを参照したいという場合を考える。この例では、これまで行われた製品開発プロセスの中で、「概要設計」というステップで「参照文書」という意味付けを持つ文書群を、まず特定し、その文書群の中から、「メーカー」、「機種名」、「スピード」、

「印字方式」という組から構成される部分のみを抽出して一覧するという処理になる。すなわち、プロセスで利用される文書群に対する意味付けを制約しておくことで、複数の同種の情報を一括して処理することができる。

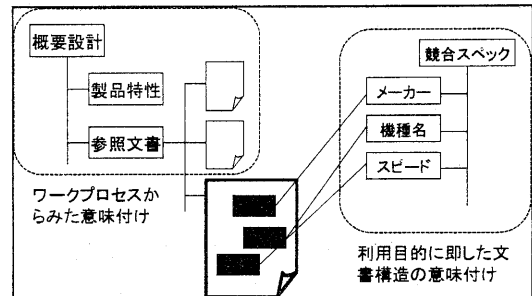


図3 メタデータとしてのワークプロセスと文書の統合管理

このようにワークプロセスにおける文書の意味付けが構造として捕捉されていれば、そのコンテキストに基づいて対象文書群を絞りこむことができる。また、抽出すべきスペックの情報は、箇条書きや表など、その時点で最適とされた表現形態で記述されているが、その表現形態の違いを利用者が意識することなく指定できる。このように、活動の目的に沿った情報のみを計算機の支援によって得ることができる。

5.2 統合管理実現に向けて

上記のようなメタデータによる統合的な管理の実現のためには、まず表現のためのデータモデルを設定する必要がある。これに求められるのは、文書中の要素の構造、あるいはワークプロセスで表現される情報の構造から容易にマッピングできるものであるとともに、柔軟に組み替えを行えるようなものでなければならない。

実際の実現には、XML、そのリンク機構の表現 XLink [9]/XPointer [10]、および XML によるメタデータ記述の標準 RDF [11]などが有効であると考えられる。文書をその元のフォーマットで表

現されている構造を保ったまま XML に変換しておき、文書外からのリンクで文書群、文書要素群を対象とするメタデータを表現することができる。今後は、これらの規格の特質を理解した上で、ワークプロセス、文書の緩やかな制約のもとでの情報統合を実現する枠組みを検討していく。

6 おわりに

アクセス可能な文書の範囲が広まると、多様な文書から目的に即した構造で情報統合を実現することが求められる。構造化文書の文書型や、ワークフローなどのプロセス定義などを用いて、あらかじめ構造を定義することによる解決は、ダイナミックなワークプロセスを支援するには十分といえない。

本論では、情報をより柔軟に目的に即して抽出し、組み替えることが可能とするために、文書とワークプロセスに緩やかな制約を課すことを提案した。さらにこれらを統合化されたメタデータとして管理することで、相乗的な効果が発揮できる可能性を示した。今後は、この考え方のもとに実現方法を具体化していく。

参考文献

- [1] ISO 8879: Standard Generalized Markup Language (SGML), October 1986
- [2] Extensible Markup Language (XML) 1.0, W3C recommendation,
<http://www.w3c.org/TR/REC-xml>
- [3] Document Management Alliance (DMA),
<http://www.aiim.org/dma/index.html>
- [4] WWW Distributed Authoring and Versioning (webdav),
<http://www.ietf.org/html.charters/webdav-charter.html>
- [5] <http://mom.nsc.cae.ntt.co.jp>
- [6] <http://www.junglee.com>
- [7] 若山、Djakovacki、陌間、XML による情報統合とプロセス統合: Enterprise Process Integration Calculus、情処研報 DD15-4、98 年 11 月
- [8] Hayashi, K., Nomura, T. et al., *Temporally Threaded Workspace: A Model for Providing Activity Based Perspectives on Document Spaces*, in Proc. of the Ninth ACM Conference on Hypertext and Hypermedia (Hypertext '98), Pittsburgh, June, 1998
- [9] XML Linking Language (XLink), W3C Working Draft,
<http://www.w3c.org/TR/WD-xlink>
- [10] XML Pointer Language (XPointer), W3C Working Draft,
<http://www.w3c.org/TR/WD-xptr>
- [11] Resource Description Framework (RDF), W3C Working Draft,
<http://www.w3c.org/TR/WD-rdf-syntax/>