

XML ベースのコンテンツ管理システムにおける現状と課題

樋川恭平[†]

サービスマニュアルなどの分野では、効率的なマニュアル作成のために XML ベースの DITA (Darwin Information Typing Architecture) の利用が普及しつつある。また、様々な XML ベースのツールが利用できるようになってきておりユーザにとっては、選択の幅が広がり目的に合致したシステム構築ができるようになっていく。

しかし、DITA を利用した場合に得られるメリットは示されるものの導入プロセスについて具体的に示されることはほとんど無いのが現状である。本稿では、XML ベースのシステム導入にあたりユーザの懸念事項であるツール選定と既存データの移行方法について考察する。

The current state and issues on XML-based CMS

Kyohei Hikawa[†]

Darwin Information Typing Architecture (DITA) has been applied to creating technical manuals. Recently many XML-based applications are available and help us develop an XML-based computer system which fits our purposes.

While we appreciate many benefits of DITA, we know little about its incorporation. In this paper, we discuss of what users feel trouble when they introduce DITA. That is, we show how we can choose a proper one of various applications and transfer an existing data to DITA.

1. はじめに

近年、製品の多品種化およびグローバルな展開が進む中で、サービスマニュアルなどの技術資料を効率的に作成し、正確な情報を迅速にユーザに提供することが必要になってきている。また、印刷物から web、さらには様々な携帯端末へとメディアが多様化していく中で、技術資料は新しい表現方法をとっていくと考えられる。これは、新しいメディアに対しては、イラストと文字だけでなく、アニメーション、音声などを駆使していかに技術情報を正確に伝えていくかが、サービスの質を高める上でも必要になってくるためである。

こうした新しい表現方法がもめられる状況において技術情報は、メディアの種類に依存しない形式であること、それによって、高い再利用性と動的な出力ができることが重要な要素になってくる。

テキストについては、長い間、主に印刷物という形で提供されてきた。しかし、XML を利用して内容と体裁を分離することによって、コンテンツ作成の生産性と質向上を図りながら同一インスタンスから web などへ動的に展開することが可能になっている。最近では、さらに再利用性の高い DITA (Darwin Information Typing Architecture) が登場しているが広く普及するには至っていない。

現在、様々なベンダーが XML、DITA に対応したツールを提供しており、高い機能を持ったシステムが構成可能になっている。こうした環境下で XML あるいは DITA を普及させていくには、採用後のメリットを強調するだけでなく、システム導入時にユーザが持つ不安を低減する具体的な方法を提示していく必要がある。本稿では、DITA 導入に際してユーザが不安を持つ既存データの移行、および、ツール選定時の問題について検討する。

まず、DITA について簡単に説明し、DITA の導入にはいくつかのレベルが存在することを示す。次に、導入時に必要となる既存データ移行について考察する。また、現在利用できるツールを紹介しシステム構築に際して注意すべき点を指摘する。最後にまとめと今後の課題について述べる。

2. DITA について

ここでは DITA について簡単に説明し、DITA の導入にはいくつかのレベルがあることを示す。

DITA は、1 つのテーマに関して記述した topic とその topic を参照する map からなる。topic は、1 つのテーマについて記述された自己完結的なそれぞれが独立した情報の単位である。また、map は複数の topic を参照し、提供する情報の目的に合わせてそ

[†] 株式会社 リナレッジ
ReKnowledge Corporation.

の順序や階層構造を定義する。topic も map も XML インスタンスであるがそれぞれ別の DTD (Document Type Definition) に従っている。map が topic とは別の DTD に従うために、PDF、web コンテンツなどの出力メディアでの表現は、map とスタイルファイルによって制御されることになる。そのため、topic は独立した情報単位であると同時に最終的に出力されるメディアからも独立している。この構成により、topic レベルでの情報の再利用性が高まり、技術情報の生産性向上と質の向上が実現できる。

まず、DITA の導入について考察する。DITA 導入には 2 つの側面がある。

(1) コンテンツの書き方としての側面

DITA では、通常、topic から派生した concept、task、reference の三種類を使って情報を作成する。concept は概念を、task は具体的な手順を、そして reference は関連情報を記述する。しかし、既存の技術文書は、一般的に印刷物の構成要素である章、節などの単位からなっており、意味的な concept、task、reference の単位で記述されているわけではない。そのため、既存データを再利用可能な適切な単位に分解できたとしても、そのままでは concept、task、reference と同一視することはできない。そもそも構造が異なるためである。concept、task、reference を作成していくにはそれに合ったライティングの技術が必要である。

(2) コンテンツを利用する側面

DITA に従ったコンテンツがあれば、DITA 対応のシステムを導入することでコンテンツの再利用性が高まり、効率的なコンテンツ作成や質の高いコンテンツ提供が可能になる。

ここでは、map、conref、conditional processing を取り上げて DITA 対応システムで何が実現できるのかを簡単に説明する。

map

map は参照する topic の順序、階層構造を指定する。出力として PDF を作成する場合には目次構成に相当する。

conref

conref を使うと topic の一部分だけを参照することができる。

たとえば、すべてのドキュメントにわたって共通して使われている「注意」、「警告」などの項目を一元管理し、その中の「注意」や「警告」を様々なドキュメントの中で参照して利用することができる。こうすることで、すべてのドキュメントの「注意」、「警告」などを同一内容、同一表現に保つことが可能になる。

conditional processing

同一の topic から指定した条件に合致したコンテンツだけを処理することができる。

たとえば、1 つの topic に異なるモデルの固有の情報を追加しておき、出力時 (PDF あるいは Web コンテンツ生成時) に特定のモデル固有の情報のみを抽出し情報を生成することができる。この topic に修正が発生した場合、異なるモデルの固有情報が全て

含まれているため整合性のとれた修正ができることは大きなメリットになる。以下、図を使って具体的に説明する。

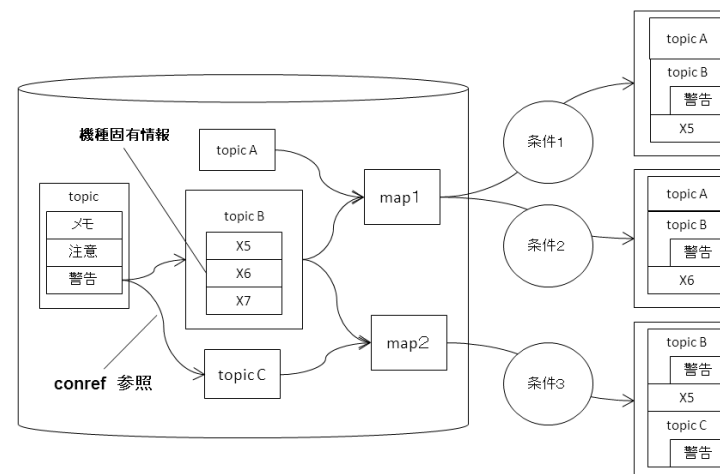


図 1 DITA を使った処理の概念図

左端にすべてのドキュメントにわたって共通して使われる「メモ」、「注意」、「警告」をまとめている topic がある。中央には 3 つの topic (topic A から C) があり、topic B には 3 機種 (X5、X6、X7) の固有情報が含まれている。また、topic B と topic C は、左端の「警告」を参照 (conref) している。map は、2 種類あり map1 は、topic A と topic B から、また、map2 は topic B と topic C から構成されている。右端には、条件 1 から 3 によって処理 (conditional processing) された結果 (たとえば PDF) が示されている。条件 1、3 は、X5 の固有情報を、また、条件 2 は X6 の固有情報を指定している。

この図から、内容の整合性がとれた topic からいかに多様な出力結果が得られるかが分かる。ここで、注意すべき点は、この仕組みは各 topic の詳細な記述規則にはほとんど関係していないということである。つまり、既存コンテンツを concept、task、reference に厳密に変換できなくても、既存コンテンツにわたって共通する部分を抽出できればこの仕組みを活用し、メリットを享受できるということである。もちろん concept、task、reference を使い、更に特殊化を行うことでより自由度の高い map の活用と編集環境の改善を行うことができる。

このように、DITA 導入の効果は、コンテンツを concept、task、reference で記述できなければ得られないわけではなく、いくつかのレベルがあることが分かる。つまり、

ユーザの現状にあったレベルでの導入をすることによって十分にその効果を得ることができる。

3. データ移行

DITA の導入に際して、ユーザが心配する問題として既存データの移行がある。

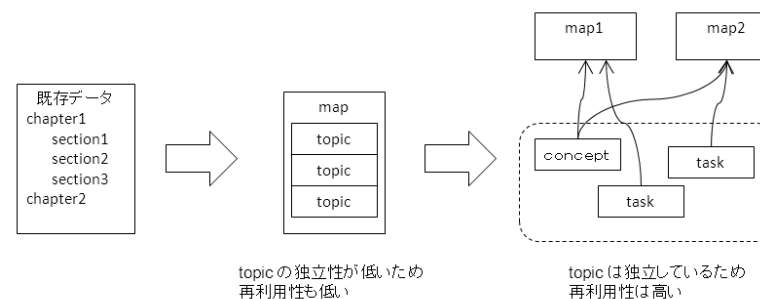
一般的に既存データはそのまま concept ,task ,reference に変換できる構造にはなっておらず、厳密にデータ移行を行おうとすると技術情報全体に対する見直しが必要になる。これは大きなプロジェクトであり、現場に混乱を生じる恐れがある。一方、すでに述べたように、既存データを厳密に concept ,task ,reference で表現できなくても、DITA システムを採用することで、コンテンツ作成の生産性向上や質の向上などのメリットを享受することができる。

DITA 導入に関しては、Michael Priestley と Amber Swope による「DITA Maturity Model」がある¹。これは、費用対効果を考えて6つの導入レベルのモデルを提案している。簡単に費用がかからないレベルから複雑で費用がかかるレベルまで6つの段階に分けているが、導入費用が低いレベルはその効果も限定的になる。例えば最も低いレベルでは、既存データを topic レベルにすら分解せず、topic のネストを使って1ファイルに変換する。この場合の効果は、1インスタンスから PDF と紙へ自動出力が可能になる程度の限定されたものであるが、DITA のタグセットを用いているため conditional processing を使えるような変換を行い効果を高めることもできる。

ここでは、現場の混乱を最小限にとどめ DITA へ移行するための既存データの移行方法について検討する。

まず、既存データの変換先として、詳細に定義されている concept ,task ,reference ではなく、その派生元の topic (generic topic) を使う。既存データが、手順中心である場合には task を利用することも考えられる。いずれにしても、章、節単位で作成されたコンテンツを無理に意味的な concept ,task ,reference などに変換しようとせず、詳細な意味付けは無視してでも DITA 形式に変換することを優先させる。この際、既存データを分析する過程で、conditional processing が利用できるようにデータを整理することもできる。generic topic へのデータ移行により DITA 対応システムの機能を利用することが可能になる。ただし、変換した topic の再利用性に関して制限が存在することは認識しておく必要がある。

DITA は、段階的に導入を行うことが可能であるようにできている。そのため、一度、DITA 対応システムでの運用が開始すれば、制限はあるもののコンテンツの再利用、PDF と web コンテンツの動的出力などのメリットを享受しながら、技術情報全体に対する見直しを進めて DITA 対応を深めていくことができる。



なお、データ移行の具体的な方法については、プログラムでの変換の他、既存データの種類 (Word ,FrameMaker) によってはプログラミングを必要としない変換ツールも存在している。いずれも、変換後の最終調整の手作業が必要であるが、こうした修正には XML エディタが有効である。

4. XML 対応ツール

現在では XML や DITA に関して、日本語環境で利用可能なインスタンス作成ツール (XML エディタ)、CMS、出力ツールが数多く提供されている。比較的知られているツールを図2にまとめる。図には、関連する作業と処理するインスタンスの流れも併せて示す。

図2では「編集」、「管理」、「利用」に分けてツールをまとめてある。「編集」はインスタンス作成を行う部分であり、XML エディタを利用する。「管理」は、ワークフローやバージョン管理を行う重要な部分であるが、コンテンツ作成者からは見えにくい機能が多い。専ら管理者が利用する機能を持ち CMS を利用する。「利用」は様々なメディアへの出力部分である。出力先としては PDF や web コンテンツなどがあるが、ツールに関しては PDF 出力に限定している。図2には翻訳支援ツールとチェッカーツールも「編集」に記載している。これらツールは、データ形式が XML に限定されるものではないが、情報の質向上および多言語展開という必要性から今後急速に連携が進むと考えられ、システムの一要素として重要であるためここに含めてある。なお、チェッカーツールは、単にスペルチェックを行うだけでなく、表記の揺れのチェックや用語統一を支援するツールである。「編集」、「管理」、「利用」のツールについて特徴を簡単に説明する。

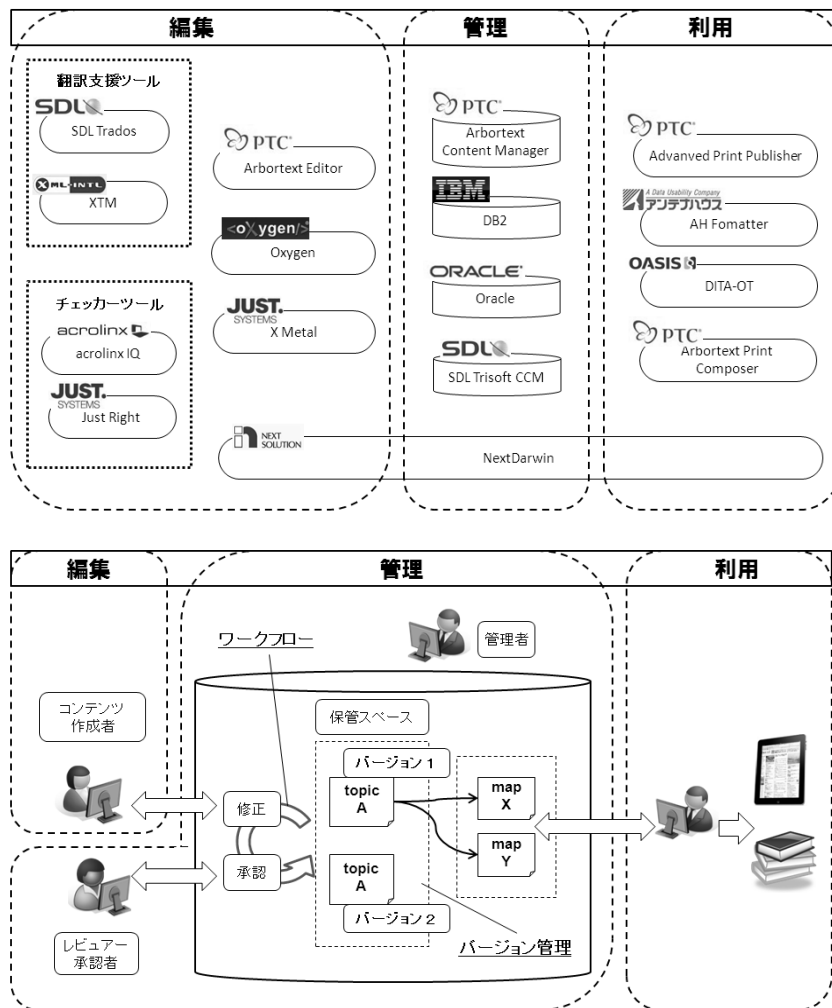


図 2 XML 関連ツール一覧および作業概要[a]

編集

a) 図中の会社名、製品名は一般に各社の登録商標または商標である。

基本的な機能は備えているため、エディタの機能として比較すると差異はあまりない。ただし、エディタは単体ではなく、CMS との連携における一機能として評価すべきである。CMS と連携して使用する場合には、連携の度合いがそれぞれ異なるため使いやすさ、カスタマイズのしやすさなどに違いが出てくる。

管理

CMS は得意とする分野が異なり、CAD 管理との統合に特徴があるもの、翻訳システムとの連携に特徴があるものなどがある。バージョン管理、ワークフロー等基本機能に違いは少ないが、ユーザインターフェイスなどは異なる。ワークフローの設定など詳細にできるほどユーザインターフェイスは複雑になる傾向がある。また、エディタや出力ツールとの連携の実績にも違いがある。さらに、ライセンス形態が異なるため価格比較も注意を要する。

出力

ここでは PDF 出力するものをまとめている。DITA-OT はフリーのツールであるが、出力設定など有償のツールに比べて手間がかかる。DITA-OT は利用価値は高いが、商用ツールに比べて手間もかかるために導入に当たっては、関連するコストと商用のツールを利用した場合のコストを比較する必要がある。また、PDF 出力の場合、どのツールを採用してもスタイル設定や修正にコストがかかるためこの点も考慮する必要がある。

この図からわかるように XML あるいは DITA を使ったシステム構築を考える場合にはいろいろなツールの選択肢が存在する。ユーザにとっては、好ましい一面、製品選択にはアドバイスが必要な状況にあると言える。

XML, DITA を使う場合、データ形式が標準化されているためにこれらツールの選択は特定のベンダーが提供するツールに縛られることなく自由な組み合わせが可能なのである。標準化されたデータ形式であるため、ベンダーが連携を保証している組み合わせ、あるいは、実績のある組み合わせであればデータ処理そのものに問題は起らない。しかし、編集支援機能あるいはユーザインターフェイスの使いやすさなど注意すべき点がある。

(1) 機能の重複

DITA 対応のツールの場合 map 作成の機能は通常、エディタにも CMS 側にも存在している。同一ベンダーのエディタ、CMS であれば統合されているので問題ないが、異なるベンダーツールを利用する場合は一般的に CMS 側の機能を利用する。このように、機能そのものに問題が発生するわけではないが、導入時に注意しないとユーザが混乱する可能性がある。

(2) ツールの連携

エディタ単体での機能はツール連携を行った場合、利用できなくなる場合がある。例えば、エディタにプレビュー画面があり編集画面と同期する機能があった場合、シス

テム構築時にエディタと出力ツールが異なるベンダーの製品である場合には編集画面との同期はとれなくなる。ユーザは、業務に必要な機能を明確にして確認する必要がある。

(3) DITA のバージョン

ツールが対応している DITA バージョンは一般的には異なっている。特に、DITA は、現在発展途上であるため異なっていると認識すべきである。最新バージョンの機能を実装しているエディタを採用したとしても、出力ツールが対応していない場合にはその機能は利用できないことになる。動作保証されているツール群であればツールのバージョン保証情報に従えば問題はない。また、specialization を行う場合にも注意が必要である。エディタ、CMS、出力ツールに対してそれぞれに DTD あるいはスキーマの設定が必要な場合がある。

(4) カスタマイズの必要性

DITA ベースのシステムは XML ベースのシステムに比べるとカスタマイズの必要性は低くなっている。これは、DITA 以前の XML では、map に相当する部分が標準化されていなかったために設計、開発の余地が大きかったことが理由である。DITA システムでは、異なるベンダーのツールを組み合わせたととしても、そのベンダーがサポートしている組み合わせであれば、ツールに対するカスタマイズなしでも十分利用できる。ただし、編集支援や進捗管理用のための機能についてはユーザによる差異が大きく使いやすくするために一部カスタマイズが必要かもしれない。ツールの選択は、単にその価格だけでなく、カスタマイズした場合の費用も考慮して選択すべきである。それぞれのツールには、一長一短がある。そのため、DITA 導入の目的を明確にして、それに合った組み合わせを選択することが重要になる。次にツールを選択する際のポイントをまとめておく。

確認する項目	詳細
CMS に期待するものはなにか	将来的に CAD のデータ管理も行うのか、多言語展開が必須なのか、あるいは、どのようなデータと連携していくのか
カスタマイズの有無は	標準機能でよいのか、あるいは、いくつかカスタマイズした機能が必須なのか

最後に、DITA の導入に際してユーザが DITA の規格についてどれくらい理解している必要があるのかという点について説明する。DITA 対応のエディタを使う限り、タグ

名やタグ間の階層関係、あるいは、属性名など詳細知識は必要ない。例えば、map 上で topic を参照する場合のマークアップを次に示す。

```
<topicgroup collection-type="family">  
<topicref format="dita" href="Topic%本体梱包箱の開梱.dita" navtitle="本体梱包箱の開梱" scope="local" type="task">  
</topicref>
```

ここでは、いくつかのタグと属性が現れているが、このマークアップを行うためのエディタ上の操作は、新規 map 上で参照する topic ファイルを選択し、適切な属性値を選択するだけである。基本的には Word 上の画像挿入と同じ操作である。DITA に対応したエディタを採用するのであれば、ユーザとしては DITA の記述方法の詳細知識は必要なく、ただ、DITA では何ができるかを理解していれば十分である。

5. まとめ

本稿では、XML、DITA のシステム導入に際してユーザが心配するデータ移行とツール選定について考察した。データ移行に関しては、DITA 導入にいくつかのレベルが存在することを示し、ユーザの実情に応じた DITA の導入モデルを簡単に説明した。また、ツールの選定に関しては、異なるベンダーのツールを組み合わせる場合の注意点を説明し、選定する時のポイントをまとめた。

6. 今後の課題

XML、DITA はコンテンツの枠組みを決めているだけであって内容を保証するものではない。タイトルタグで囲まれている文字列は、間違いなくタイトルであることを前提にシステムは処理を行う。従って、XML、DITA を使うことで提供するコンテンツの質が向上すると言った場合、同じ表現が使われるので表現が統一されることによる質の向上が保証されているのであって、その内容の質が向上することとは別のことである。

情報の質を向上させるには、当然のことながら内容を向上させる必要がある。DITA の concept、task、reference という分類の意味は本来こうした点にある。XML や DITA のシステムを導入し作業を標準化した後には、情報の内容を高める方向に進むことが望まれる。

本稿では XML、DITA のシステムと翻訳支援ツールとの連携については触れなかった。コンテンツの多言語展開を考える場合、翻訳支援ツールの機能を最大限利用するために原文を翻訳しやすい表現にするなどの作業も重要になってくる。このような構

文のチェックを行うツールも利用可能になっている。XML, DITA に適した情報の多くが複数言語への展開を必要としていることから、チェッカーツールや翻訳支援ツールと XML, DITA システムの連携, および, それに伴い既存の翻訳メモリーなどの翻訳資産を活用する具体的方法の提案が今後の課題である。

参考文献

- 1) M. Priestley and A. Swope, DITA maturity model. A JustSystems whitepaper. 2008 ;
http://na.justsystems.com/files/Whitepaper-DITA_MM.pdf.