

# DITA を用いたソフトウェアドキュメンテーションにおける トピックの意味検索システムの提案

坂井 麻里恵<sup>†</sup> 大場 みち子<sup>‡</sup> 伊藤 恵<sup>‡</sup> 奥野 拓<sup>‡</sup>  
 公立ほこだて未来大学大学院<sup>†</sup> 公立ほこだて未来大学<sup>‡</sup>

## 1 はじめに

ソフトウェアドキュメントの多くは一般的に汎用のワードプロセッサやスプレッドシートを用いて作成されており、記載された情報の横断的な閲覧や再利用が難しい。それにより、トレーサビリティの管理や、変更管理などのメンテナンスにかかるコストが増大し、生産性の低下やソフトウェアの品質低下を招く恐れがある。

本研究では、これらの問題を解決するために、技術文書向けの XML アーキテクチャ標準である DITA (Darwin Information Typing Architecture) を用いる。DITA の構造に則ってソフトウェアドキュメントを記述し、コンテンツと文書構造を分離させることで、関連する情報の横断的な参照や、記述内容の再利用が容易になる。これは、トレーサビリティの向上やメンテナンスコストの削減に繋がる。

一方で、ソフトウェアドキュメントをトピックに分割した場合、蓄積されるトピックの量は膨大になる。そこで、トピック間に意味的な関係情報を設定し、その情報を元にトピックを検索できるような仕組みの構築を目指す。

## 2 ソフトウェアドキュメントへの DITA 適用の利点

DITA とは、技術文書向けの XML ベースのアーキテクチャ標準である [1]。DITA に基づくドキュメントは、トピックとマップの二つの要素で構成される。トピックとは、異なるコンテキスト間での再利用を想定した比較的小さな単位の情報を指し、文章や図などのコンテンツは全てトピックとして記述される。これに対し、マップはコンテンツそのものを含まず、トピックの組み合わせによって表される文書の構造のみを定義したものである。これらの二つの要素でドキュメントを記述することにより、必要に応じてマップ上でトピックを組み替え、同じトピック群から様々な形のドキュメントを出力する

ことができる構造になっている。

この仕組みに基づいてドキュメントを作成すると、目的に応じてマップを作成することで求める情報のみを取り出すことができるため、従来のドキュメントに比べてトレーサビリティが向上する。また、トピックは複数のマップから参照し、ドキュメント間で共有することができる。この構造により、元トピックの変更が全ての共有先に反映されるようになるため、ドキュメント修正時の修正回数や範囲が小さくなり、メンテナンスコストの削減が見込める。これらの利点から、本研究では DITA に基づいたソフトウェアドキュメンテーション手法の構築を目指す。

## 3 トピックの意味検索手法の構築

DITA に則ってソフトウェアドキュメントを記述した場合、蓄積されるトピックの量は膨大になる。そのため、トピック群の中から必要なトピックを検索するための有効な手段を考える。

ある目的を持ってソフトウェアドキュメントを閲覧する際、読み手は主にソフトウェアの機能、それに関わる要求、業務など、ドキュメントに記載されている内容そのものの概念や関係性から必要な情報を辿っていく。従って、トピックを検索する際も、トピックに記載された内容やそれらの間の意味的な関係から、必要なトピックを取得できることが望ましい。

一方、利用者の持つ概念体系に合わせて情報を分類・整理し、情報を見つけやすくするトピックマップという国際規格がある [2]。トピックマップの構造は本の索引の概念に基づいており、複数のドキュメントにまたがるインデックスの役割を果たすことができる。また、セマンティック Web の技術を利用して、意味に基づいた検索を行う手法も多く研究されている。庭山らは、ソースコードのメタデータを RDF で記述することで必要とする部品を意味的に検索する手法を提案している [3]。本研究ではこれらの考え方を取り入れ、トピック間の関係を RDF で表現し、それをインデックスとして用いることで、求めるトピックを検索できるような仕組みの構築を目指す。RDF でトピック間の関係を表現すると、SPARQL クエリを用いて条件に該当するトピック

A Proposal of a Semantic Topic Search System for Software Documentation Using DITA

<sup>†</sup> Marie Sakai, Graduate School of Future University Hakodate

<sup>‡</sup> Michiko Oba, Kei Ito, Taku Okuno, Future University Hakodate

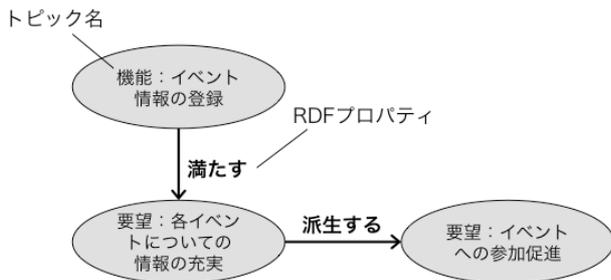


図 1 グラフによるトピックの関係の可視化

クを意味的に一括検索することができるため、マップを作成する際のトピック検索の手間を削減することができると思われる。

また、トピック間の関係を一覧できるようにするために、ラベル付き有向グラフ（以下、グラフ）による可視化を行う（図 1）。これまでも、SysML の要求図などによって要求や機能の関係をモデル化する手法が用いられてきた。しかし、従来のモデルはドキュメントとは乖離しており、モデルに表されている要求や機能の関係が正しくドキュメントに記載されているかは別途確認する必要がある。一方、DITA と RDF を組み合わせた場合、グラフによって可視化された要求や機能の関係はそのままドキュメントに記載された項目間の関係を表している。そのため、グラフ上の各要求や機能について漏れなく文書化されていることが保証されるだけでなく、ドキュメント読解時に補助としてグラフを閲覧することで、書き手の意図する論理構造が誤解無く読み手に伝わることも期待できる。

#### 4 グラフ表示の効果に関する予備実験

トピック間の関係のグラフ表示がドキュメントの解釈に与える効果を検証するために、予備実験を行った。被験者は 2 名で、いずれもシステム開発 PBL においてソフトウェアドキュメントの作成・閲覧経験のある学生である。

まず、DITA トピックから生成した要件定義書を用意した。これに 15 分間目を通した後、要件定義書の内容に関する質問に回答するよう被験者に指示した。続いて、トピック間の関係を表したグラフを見せ、再び同じ質問に回答するよう指示した。質問内容は、「この要望に対応している機能はどれか?」「この機能に含まれる部分機能はどれか?」など、主にドキュメントに記載されている項目間の関係を問うものである。設問は全部で 8 問用意した。

実験の結果を表 1 に示す。2 名とも、一回目の回答に比べ、二回目の回答の方が、こちらの想定回答と一致した設問数が増加している。

表 1 被験者回答が想定回答と一致した設問数（全 8 問中）

	一回目	二回目
被験者 1	6 問	7 問
被験者 2	4 問	7 問

#### 5 考察

実験結果から、トピック間の関係を RDF で記述し、それをグラフで可視化することで、書き手の意図と読み手の解釈が一致しやすくなると考えられる。

その他の傾向として、複数回答が可能な設問では、一回目に比べて二回目の回答で回答数が増加する傾向がみられた。例えば、「この機能が実現している要望はどれか?」という設問に対して被験者 2 が回答した要望の数は、一回目が 1 つ、二回目が 2 つである。これは、グラフ上で機能と直接繋がっている要望だけではなく、さらに上位の要望も回答として挙げているためである。このことから、グラフが要求トレーサビリティ理解の補助になっていると考えられる。

また、「この機能に含まれる部分機能はどれか?」といった単純な包含・親子関係を問う設問の場合、一回目の回答から正答率が高かった。これは、包含関係や親子関係は DITA のみで表現が可能なため、これらの関係の理解にはグラフを必要としなかったことが理由と考えられる。

#### 6 おわりに

ソフトウェアドキュメントを DITA で記述し、トピック間を RDF で意味的に関連付ける仕組みを提案した。予備実験の結果、トピック間の意味的関係をラベル付き有向グラフで可視化することで、ドキュメント理解の補助となる可能性が示唆された。

今後は、被験者数を増やして評価実験を行うとともに、SPARQL クエリを用いたトピックの一括検索の有効性についても評価を行う。

#### 謝辞

本研究は科研費(23500127)の助成を受けたものである。

#### 参考文献

- [1] J. Hackos, Introduction to DITA: A User Guide to the Darwin Information Typing Architecture, 2nd ed. Comtech Services, Inc., 2011.
- [2] 内藤求: トピックマップ入門(セマンティック技術シリーズ), 東京電機大学出版局, 12 2006.
- [3] 庭山他: セマンティック Web を利用したソフトウェア検索システムの提案, 信学技報. SS, Vol. 102, No. 704(2003), pp. 27-31.