

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 26 日現在

機関番号：20103

研究種目：基盤研究(C)

研究期間：2011～2013

課題番号：23500127

研究課題名(和文)ソフトウェアライフサイクルを支援するコラボラティブDITAドキュメンテーション

研究課題名(英文)Collaborative DITA Documentation for Supporting Software Life Cycle

研究代表者

奥野 拓 (Okuno, Taku)

公立はこだて未来大学・システム情報科学部・准教授

研究者番号：30360936

交付決定額(研究期間全体)：(直接経費) 3,800,000円、(間接経費) 1,140,000円

研究成果の概要(和文)：ソフトウェア開発においては多くのドキュメントが作成される。しかし、ワードプロセッサ等で作成した場合、トレーサビリティの低下、メンテナンスコストの増大という問題が発生する。この問題を技術文書向けのXML標準であるDITAを導入することにより解決している。ソフトウェアドキュメントをDITAの構成要素であるトピックとして記述するために、標準的なソフトウェアドキュメントの構造化モデルを作成し、トピック粒度の検討を行った。また、意味的な関連性によりトピックを検索する手法を構築し、DITAに基づいた統合ソフトウェアドキュメンテーション環境を汎用ウェブCMSを拡張して構築した。

研究成果の概要(英文)：General software documents have some problems with traceability and maintenance. This study aims to solve those problems by using DITA, an XML-based architecture for authoring technical documents. In DITA architecture, documents consist of topics and maps. Topics have small amount of information that can be reused, and maps specify contents of each output by organizing topics. It helps to browse relative contents together and reuse topics, and then DITA improves document traceability and reduces maintenance costs. For managing software documents as DITA contents, an integrated documentation environment that assists in authoring topics and maps, and other specific works for DITA was developed. As a result of an analysis of software documents, it was found that there are three types of structures in software documents, and those need to be broken down into different-sized topics. Besides, the semantic topic search system is realized by using RDF with DITA.

研究分野：総合領域

科研費の分科・細目：情報学，メディア情報学・データベース

キーワード：ソフトウェア開発 ドキュメンテーション トレーサビリティ XML DITA

1. 研究開始当初の背景

ソフトウェアの開発プロセスにおいて、提案書、要件定義書、設計書といった多くのドキュメントが作成されている。これらのドキュメントは、一般的に大きく三つの問題を抱えている。まず、ドキュメントの増加に伴うトレーサビリティの低下、メンテナンスコストの増大の二つである。これら二つの問題は、ドキュメント各々が独立し、各ドキュメント単位で内容が完結していることに起因する。複数ドキュメントを横断しての情報閲覧が困難であるために、ドキュメント数が増加するほど、全体を通してのトレーサビリティが低下していく。また、内容の一部に変更が生じた際も、影響が波及する部分をドキュメント毎に修正する必要がある、メンテナンスコストが大きい。これらは、ドキュメントの内容に不整合が生じる原因にもなっている。ドキュメントの内容の不備は、それらを参照して作成したソフトウェアにも引き継がれるため、結果的にソフトウェア自体の品質の低下を招く恐れがある。三つ目の問題は、多くのドキュメントが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るために、執筆と同時に書式設定を行う必要が生じることである。これは非本質的な作業であるため、生産性の低下に繋がっている。

一方、近年、マニュアルやヘルプファイルなどの技術文書において、XML ドキュメントのアーキテクチャ標準である DITA (Darwin Information Typing Architecture) が用いられている。DITA に基づくドキュメントは、再利用可能な情報の単位である「トピック」と、トピックの組み合わせにより文書構造を定義する「マップ」で構成される。この構造をソフトウェアドキュメントに適用すると、関連する情報を一括して参照することや、同じ記述を複数ドキュメントで再利用することが容易になる。しかし、従来のソフトウェアドキュメントはドキュメント単位での管理を想定して記述されているため、容易にトピック単位に分割することができない。DITA をソフトウェアドキュメントに適用するためには、どのようにトピックに分割するか、そして、どのようにトレーサビリティを実現するかを解決する必要がある。また、それらを踏まえ、ソフトウェアドキュメンテーションを行うための作業環境を作る必要がある。

2. 研究の目的

本研究は、DITA を適用することでソフトウェアドキュメントにおける問題を解決し、ドキュメントの作成・管理を効率化することを目的とする。ドキュメント間のトレーサビリティを確保することで、内容に変更が生じた際の修正漏れを防ぐことができ、結果として不整合が生じる可能性が低下する。ドキュメントの整合性は、それらを参照して作成されるソフトウェアの品質にも影響するため、

不整合の解消はソフトウェアの品質の向上に繋がる。

3. 研究の方法

本研究は以下の三つのサブテーマに分けて段階的に実施した。

(1) ソフトウェアドキュメントの構造化と粒度の検討

DITA ではトピックを単位としてドキュメントを管理することから、1 項目毎に内容が独立し、トピック単位に分割しやすいマニュアルやヘルプファイルといったドキュメントに主に用いられてきた。一方、一般的なウオーターフォール型開発におけるドキュメントはそれぞれ文脈を持っており、容易にトピックへ分割することはできない。分割するためには、一連のドキュメントに含まれる情報とそれらの関係を分析し、適切なトピックの粒度を決定する必要がある。そこで本論文では、ソフトウェアドキュメントの構造化を行い、適切なトピックの粒度について考察する。

(2) 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が必要となる。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。そこで、トピック同士を予め関連付けておき、その関連情報を元にトピックを検索できるような仕組みを構築する。

(3) 統合ドキュメンテーション環境の構築

DITA におけるトピックやマップなどのコンテンツは、一般的にコンテンツマネジメントシステム (CMS) で管理される。ソフトウェアドキュメンテーションを CMS 上で行うと、複数の開発者によるドキュメントを介した協調作業や、分散環境におけるドキュメントの共同編集が容易になる。また、DITA に基づくドキュメンテーションでは、特殊化やトピック・マップの執筆など、従来のソフトウェアドキュメンテーションとは異なる作業も発生する。さらに、(2) の意味検索手法をサポートする必要もある。以上の作業を補助するために、統合ソフトウェアドキュメンテーション環境を構築する。このドキュメンテーション環境は既存の CMS を拡張する形で構築する。

4. 研究成果

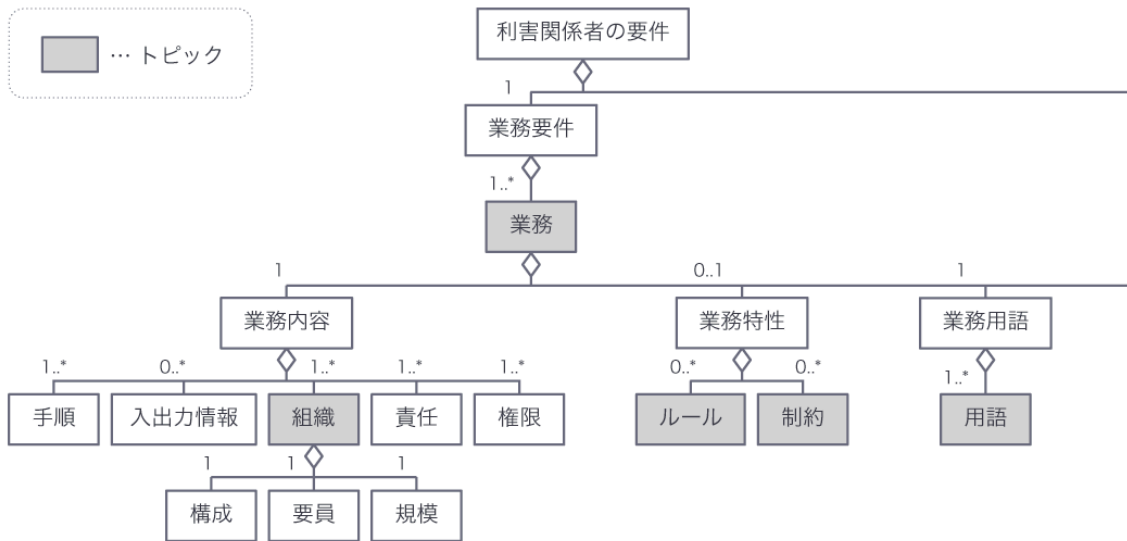


図1 「利害関係者の要件」のドキュメント構造化モデルの一部

(1) ソフトウェアドキュメントの構造化と粒度の検討

DITAでは、トピックの粒度が細かいほど再利用性が増す一方で、管理が煩雑になる。そこで本研究では、ソフトウェアドキュメントの構造化モデルを作成し、適切なトピックの粒度について検討を行った。

ドキュメントに含まれる情報とその関係性を可視化するために、構造化を行った研究事例があるが、詳細にモデルを定義するためには、ドメイン（ソフトウェアの対象となる業務領域）を限定しなければならない。モデルがドメインに依存してしまうと、他のドメインに適用する際にモデルの再構築が必要となる。そこで、本研究では、ドメインに依存せず、汎用的に適用できるモデルの作成を行うことにした。そこで、ドメインに依存しない標準的なソフトウェアドキュメントとして、共通フレーム 2007 の主ライフサイクルプロセスにおいて作成が推奨されているドキュメント群を用いるものとした。

トピックとなり得る最小単位の項目を洗い出すため、各ドキュメントの構成と記述される具体的な項目をモデルとして可視化する必要がある。共通フレーム 2007 では、各ドキュメントに記述すべき項目も定義されている。それらの各項目をノードとし、ドキュメント毎にツリー状の構造化モデルを作成した。また、ノード毎に親ノードに対して一つしか存在しないもの、複数存在するものを区別するため、多重度を記載した。これは、UML のクラス図における多重度と同様の記法を用いている。「利害関係者の要件」ドキュメントについて、作成した構造化モデルの一部を図1に示す。

ドキュメント構造化モデルにおいて、ソフトウェアドキュメントの文書構造には特徴が見られた。構造化モデルにおける葉ノード同士の関係性から、それらを「レコードパターン」「集合パターン」「サブセットパターン」

の3種類に分類した。これらのパターン別に情報の再利用性を検討し、適切なトピックの粒度を定義した。その際、ドキュメントの実例として、情報処理推進機構ソフトウェア・エンジニアリング・センターが公開している「超上流から攻める IT 化の事例集」を用いた。

定義した粒度を評価する目的で、「利害関係者の要件」「システム要件」の2つのドキュメントに、定義した粒度を適用した。その上で、トピック数や DITA で記述した場合の XML タグ数を試算し、作成した構造化モデルの葉ノード、すなわち最小単位をトピックとした場合との比較を行った。「利害関係者の要件」のトピック数は最小粒度の場合と比較して 41%、XML タグ数は 25%減少した。同様に、「システム要件」のトピック数は 80%、XML タグ数は 81%減少した。この結果は、ドキュメント内での再利用性を損なわない範囲でトピックの粒度を大きくすることが可能であり、結果として管理コストを抑制することが可能であることを示している。

(2) トピックの意味検索手法の構築

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。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。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などが提案されている。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を構築している。RDF でトピック間の関係を表現すると、

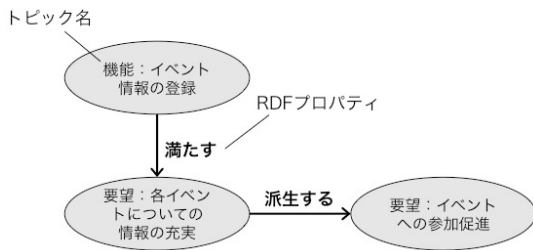


図2 トピック間の関係のモデル化

SPARQL クエリを用いて条件に該当するトピッ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。

RDF を用いる場合、ユーザがトピックを検索する際は、トピックの型とそれらを結ぶプロパティを検索の条件として指定する。しかし、使用されているトピックの型やプロパティが不明であったり、同じ意味を持つ型やプロパティが複数存在したりすると、検索漏れが生じる原因となる。そのため、利用できるトピックの型とプロパティの種類を標準化する必要がある。プロパティの標準化には、Ramesh らの定義した関連メタモデルを用いる。Ramesh らは、要求トレーサビリティを確保するために要求や設計事項などの関係を記述する関連メタモデルを定義している。このモデルでは、Requirements, System Objectives といった関連モデルを構成するエンティティや、それらのエンティティを繋ぐ depend-on, describe といったリンクタイプの種類が定義されており、ソフトウェア開発のトレーサビリティ管理に汎用的に適用できる。トピック間に同様のリンクタイプを設定した場合、要求トレーサビリティの観点からトピック間の関係に意味付けをすることが可能である。そのため、Ramesh らの研究において定義されているリンクタイプを、トピック間を結ぶプロパティとして利用する。

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2）。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った。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。

構築した意味検索手法と可視化手法について、評価実験を行った。実験は、PBL プロジェクトとして実施したイベント告知システムの開発において作成された要件定義書を用いて行った。このシステムは、大学で開催されるイベントを登録・Web 上で閲覧可能にすることを目的としたものであり、開発者8名・工期4ヶ月のプロジェクトである。

まず、要件定義書をトピック化し、意味的関連付けを行った。そして、意味検索あり（深さ3まで検索）と意味検索なしの場合で、各機能の関連トピックの検索を行い、検索されたトピック数の比較を行った。その結果、意味検索なしの場合には352トピック、意味検索ありの場合には114トピックとなり、意味検索を用いることで、検索結果が約3分の1に絞り込まれた。この結果は、意味検索の導入により、求める情報をよりの確に絞り込めるようになることを示している。

次に、グラフ表現の効果について実験を行った。グラフがドキュメント理解の補助になるかを確かめることを目的として、要求定義書の内容に関する質問に被験者が回答する実験とした。実験は、1. 要件定義書の閲覧、2. 質問に回答、3. グラフの提示、4. 再度質問に回答という手順で実施した。システム開発PBLでドキュメント作成・閲覧経験のある学生6人を被験者として8問の質問を用いた実験を行った。

いずれの被験者についても、1回目の質問よりも2回目の質問の正当が多く、1回目の平均正当人数が2.6人であったのに対して2回目は4.4人であった。特に、要求トレーサビリティに関する質問（4問）に限定すると、1回目が1.5人、2回目が4.5人であった。この結果より、グラフ表示が、要件定義書のみでは理解し難い要求トレーサビリティの理解に有効であることが明らかになった。

(3) 統合ドキュメンテーション環境の構築

本研究では、オープンソース CMS である Drupal を拡張することで、DITA ベース CMS を構築した。Drupal は PHP で記述されたオープンソースの CMS であり、モジュールを作成することで必要な機能を自由に追加することができる。Drupal を用いる理由は、1. Web CMS であり、Web ブラウザ上でコンテンツの作成・管理を行うことができる、2. オープンソースで、カスタマイズ可能である、3. Apache と PHP が動作する全てのプラットフォーム上で稼働する、4. DITA コンテンツを作成・管理するための Drupal モジュール（プラグイン）である “DITA integration for Drupal” が開発・公開されている、の4点である。本研究では Drupal を拡張し、統合ドキュメンテーション環境の開発を行った。以下、主要な機能について説明する。

①DITA トピック・マップのオーサリング

Drupal 上で、Task 型、Concept 型のトピックと、マップを作成することができる。図

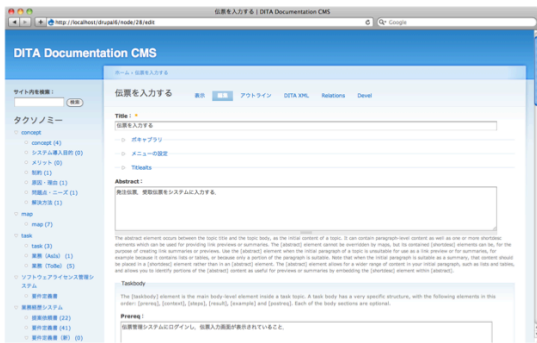


図 3 Task トピック作成画面

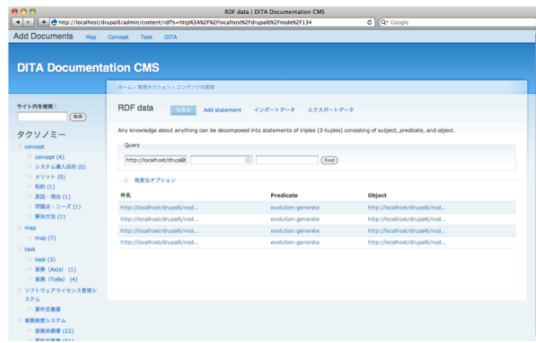


図 6 RDF データ検索画面



図 4 トピック閲覧画面

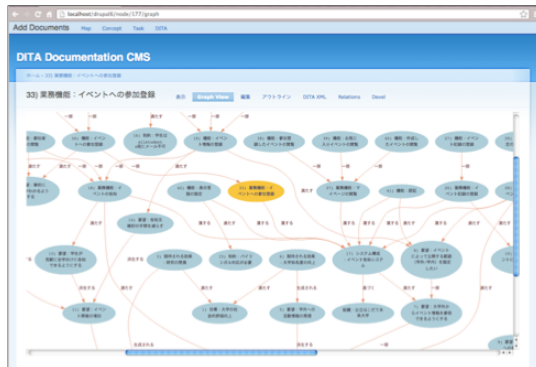


図 7 ラベル付き有向グラフ表示画面

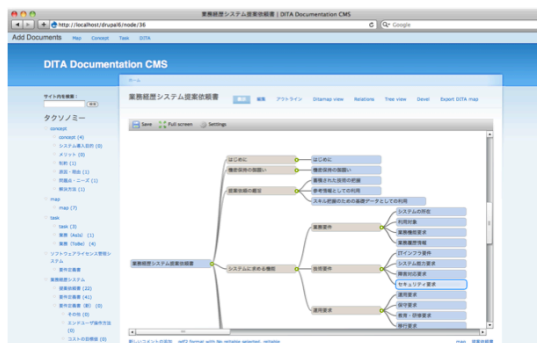


図 5 マップ作成・閲覧画面

3 に、Task トピックの作成画面を示す。“Prereq”, “Context” などトピックの構造に合わせて入力フォームが用意されており、フォームに従って項目を入力することでトピックを作成することができる。図 4 は、作成したトピックの閲覧画面である。これは Task トピックの例であり、〈step〉として記述した作業手順が番号付きリストの形で表示されている。

マップの作成・閲覧画面を図 5 に示す。この機能はマインドマップを作成するための Drupal モジュール “GraphMind” を拡張して作成されており、トピックの追加や並べ替えを GUI で行うことができる。図のグレーのノードは、マップ上で自由に作成できる見出しノードである。ブルーのノードは、それぞれトピックを表している。トピックは、ノードを右クリックすることにより表示される「Load Drupal Item」のメニューから呼び出して、マップに配置することができる。これらのノードをマウス操作で並べ替えるこ

とにより、出力するドキュメントの構造を定義する。

②ドキュメント出力

作成したマップを元に、PDF や XHTML 形式のドキュメントに変換することができる。マップのエクスポート画面で出力するドキュメント形式を選択し、エクスポートを行うことができる。エクスポートされたマップは変換待ちの状態となり、コマンドラインから Drupal を操作できるツール「Drush」を別途動作させることで、指定した形式のドキュメントに変換される。

③RDF によるトピックの意味検索

RDF によるトピックの関連付け・検索を実現するために、“RDF” という Drupal モジュールを用いた。このモジュールにより、RDF データの登録・検索・エクスポートが可能になる。RDF データの検索画面を図 6 に示す。主語・述語・目的語のいずれかまたは複数を選択し、「Query」のフォームに入力することで、指定した条件に合致する RDF データが一覧で表示される。

④ラベル付き有効グラフ表示

グラフ表示には、グラフ描画のためのオープンソースソフトウェアである Graphviz を用いた。Graphviz による RDF モデルの視覚化は、PHP で RDF データを扱うためのパッケージである RAP によってサポートされており、PHP アプリケーションから容易に利用することができる。本システムでは、モジュール “RDF” の機能を利用し、XML 形式でエクスポートした RDF データを Graphviz に渡すことで、グラフの描画を行っている。実

装したグラフ表示画面が図7である。このグラフは、各トピックの閲覧画面にある「Graph View」というメニューから閲覧することができる。その際、自身と直接関連するトピックをグラフ上で探しやすいように、閲覧の起点となったトピックの色を変えて表示するようにしている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[学会発表] (計 11 件)

①坂井 麻里恵, 大場 みち子, 伊藤 恵, 奥野 拓, 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, 情報処理学会第 76 回全国大会, 2014. 3. 13, 東京.

②坂井 麻里恵, 大場 みち子, 伊藤 恵, 奥野 拓, 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手法の提案, 日本ソフトウェア科学会第 30 回大会, 2013. 9. 13, 東京.

③金谷 祥平, 伊藤 恵, 大場 みち子, 奥野 拓, DITA によるソフトウェア関連文書とソースコードの統合管理環境の提案, 日本ソフトウェア科学会第 30 回大会, 2013. 9. 13, 東京.

④赤石 裕里花, 坂井 麻里恵, 奥野 拓, 伊藤 恵, 整合性維持に着目したソースコードとドキュメントの一元管理環境の提案, 日本ソフトウェア科学会第 30 回大会, 2013. 9. 13, 東京.

⑤坂井 麻里恵, 大場 みち子, 伊藤 恵, 奥野 拓, 統合ソフトウェアドキュメンテーション環境としての DITA ベース CMS の開発, 第 88 回情報処理学会デジタルドキュメント研究会, 2013. 1. 18, 東京.

⑥Marie Sakai and Taku Okuno, Building DITA-Based CMS as Integrated Documentation Environment, IEEE The 6th International Conference on Soft Computing and Intelligent Systems/The 13th International Symposium on Advanced Intelligent Systems (SCIS-ISIS 2012), November 20, 2012, Kobe, Japan.

⑦奥野 拓, 坂井 麻里恵, 伊藤 恵, 大場 みち子, DITA に基づくソフトウェアドキュメンテーション, 第 11 回情報科学技術フォーラム (FIT2012), 2012. 9. 5, 東京.

⑧坂井 麻里恵, 奥野 拓, DITA 適用に向けたソフトウェアドキュメントの構造化, 第 11 回複雑系マイクロシンポジウム, 2012. 3. 3, 札幌.

幌.

⑨奥野 拓, 坂井 麻里恵, 伊藤 恵, 大場 みち子, DITA に基づくソフトウェアドキュメンテーション環境の提案, 計測自動制御学会第 12 回システムインテグレーション部門講演会 (SI2011), 2011. 12. 23, 京都.

⑩坂井 麻里恵, 伊藤 恵, 奥野 拓, DITA 適用のためのソフトウェアドキュメントの構造化, 日本ソフトウェア科学会第 18 回ソフトウェア工学の基礎ワークショップ (FOSE 2011), 2011. 11. 24, 青森.

⑪坂井 麻里恵, 奥野 拓, DITA 適用のためのソフトウェアドキュメントの構造化, 情報処理北海道シンポジウム 2011, 2011. 10. 1, 北見.

6. 研究組織

(1) 研究代表者

奥野 拓 (OKUNO, Taku)

公立ほこだて未来大学・システム情報科学部・准教授

研究者番号 : 30360936

(2) 研究分担者

伊藤 恵 (ITO, Kei)

公立ほこだて未来大学・システム情報科学部・准教授

研究者番号 : 30303324

(3) 研究分担者

大場 みち子 (OBA, Michiko)

公立ほこだて未来大学・システム情報科学部・教授

研究者番号 : 30588223