

情報検索タスクのモデル化を目的としたタスクオントロジー構築手法の提案

- 情報システム開発における設計知識管理への適用

Proposal of a Method to Build Task Ontology for Modeling of Search Tasks – A Case Study on Design Knowledge Management in Information System Development

中村覚^{*1} 大和裕幸^{*1} 稗方和夫^{*1} 岡田伊策^{*1*2} 齋藤稔^{*2} 笈田佳彰^{*1*2}

Satoru NAKAMURA Hiroyuki YAMATO Kazuo HIEKATA Isaac OKADA Minoru SAITO Yoshiaki OIDA

^{*1} 東京大学
THE UNIVERSITY OF TOKYO

^{*2} 富士通株式会社
FUJITSU LIMITED

The design document which contains design information of system development is useful resource for understanding the influence of system repairs from specification changes and for reference of technical know-how from past projects. As the characteristic of reuse of these documents, it requires not only to search documents based on information which is described directly in itself, but also to follow the relationship between design information. In this paper, by restricting an objective of knowledge processing to modeling of the search tasks, a method to build task ontology systematically is proposed. This method defines the requirement of a hearing process for experts in target domain and builds the ontology semi-automatically by use of existing design documents. The effectiveness of proposed method is verified by 2 case studies about search tasks in information systems companies.

1. はじめに

情報システムの開発や導入を行う企業では、プロジェクト毎に設計情報を記述した設計書を作成し、顧客に納品することが一般的である。これらの設計書はシステム導入後においても利用価値が高く、導入後のシステム改修における修正項目の把握や、新規プロジェクトにおける過去プロジェクトの設計情報の再利用等に用いられる。しかし、設計書の数が膨大となった際、全文検索では目的とする設計書を十分に絞り込むことが困難となる場合がある。この時、ファイル名やキーワード等のメタデータを用いた検索と組み合わせることで、検索対象の絞り込みを行うのが一般的である。さらに、情報システムに関する設計書の探索では、ファイルに直接記述された情報だけでなく、組織や業務固有の知識に基づき、関連する設計情報を辿ることで検索を行うケースが多い。この例として、システムの仕様変更に基づく影響分析を行う際の検索タスクを図 1 に示す。

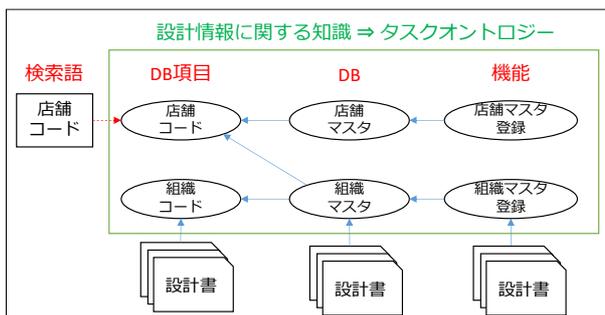


図 1 設計情報の関係に基づく検索タスク例

DB 項目 (ex. 店舗コード) の桁数の拡張等の仕様変更により、修正が必要となる設計書を検索するタスクである。DB 項目に関する設計書の検索については、仕様変更が生じた DB 項目を直接検索語として利用できるが、それに関連する DB や機能に関する設計書の検索には、「DB 項目と DB」「DB と機能」のよう

な設計情報の関係に基づいた検索が必要となる。組織における在籍年数や業務経験が豊富な熟練者は、これらの情報の関係性を暗黙的に理解し、目的とする設計書を検索することができる。一方、業務知識や設計経験が不足する新人や若手にとっては、目的とする設計書を検索することが困難である。

この課題に対し、本研究では上述したような設計情報の関係に基づく検索タスクを対象とし、この検索タスクをモデル化するタスクオントロジーの構築手法を提案する。さらに、本手法を情報システム企業 2 社における検索タスクに適用し、提案手法および構築したタスクオントロジーの有用性について考察する。

2. 関連研究

組織やタスクに関する知識を体系化し、計算機による知識処理を可能とする技術としてオントロジーがある。古崎ら[古崎 02]はオントロジーベースシステム開発を行う手順を「オントロジーの構築」「モデルの構築」「システム開発」の 3 点にまとめている。「オントロジーの構築」では、「対象世界の理解」「対象世界を構成する概念の分析・組織化」の 2 つの過程に分かれ、自己の内省や各種文献の参照、専門家との議論などを通じてオントロジーを設計する。「モデルの構築」では、オントロジーで定義した概念(クラス)から個々のインスタンスを作成し、接続情報や属性値など、そのインスタンスに特有の情報を与える。オントロジーベースのシステムの有用性を示す研究[関谷 99][研谷 07]は数多く存在するが、スキーマ定義における有識者との議論や、インスタンスの関係記述を行うモデル構築には多大なコストを要する[海谷 12]ため、オントロジーベースシステムが広く導入されているとは言い難い。これに対し、オントロジーの記述支援ツールやオントロジーの自動構築手法[白川 11]の提案もなされている。

本研究では、オントロジーに基づく知識処理の目的を、先述したような設計情報の関係に基づく検索タスクのモデル化に限定する。これにより、検索時に参照する設計情報(インスタンス)の関係記述を要件とし、簡易なスキーマ定義を利用したタスクオントロジーの構築手順の体系化を行う。具体的にはヒアリング項目の定型化による不確実性の低減と、既存の設計書を利用したオントロジーの半自動的な構築によるコスト削減を可能とする手法を提案する。

連絡先: 中村覚, 東京大学大学院新領域創成科学研究科, 〒277-8563 千葉県柏市柏の葉 5-1-5, Tel: 04-7136-4626, Fax: 04-7136-4626, e-mail: nakamura@is.k.u-tokyo.ac.jp

3. 提案手法

本研究で提案するタスクオントロジーの構築体系を図 2 に示す。本手法は「適用ドメインへのヒアリング」「スキーマの定義」「モデルの構築」から成り、以下それぞれについて説明する。

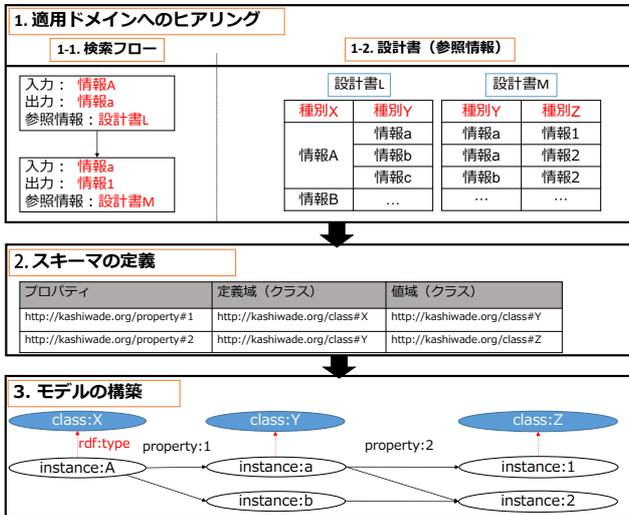


図 2 提案するタスクオントロジーの構築体系

3.1 適用ドメインへのヒアリング

ここでは、適用ドメインへのヒアリングを通じ、モデル化の対象である検索フローの把握と、検索フロー内で参照する情報の関係を記述した設計書の取得を目的とする。

検索フローの把握では、図 2 左上部に示すように、検索時に参照する情報の把握を行う。具体的には入力、出力、および参照情報の明確化を行う。ここでの入出力情報は、実際の検索時に用いる具体的な情報(「店舗コード」や「店舗マスタ」等のインスタンス)とする。このような具体的な情報を用いる目的は、ヒアリングにおける要求項目を明確化し、議論の発散等の不確実性を低減するためである。

参照情報とは、図 2 右上部に示すように、入出力情報の関係を記述した設計書を示す。先述したように、情報システム開発を行う企業では、設計情報を文書化することが一般的である。このような設計書から、設計情報の関係を形式的に記述した文書を取得する。この設計書は、ヒアリングを通じて得られたインスタンス例に基づく検索フローの汎化に必要な情報の取得に加え、後述するモデル構築プロセスの半自動化に利用する。

3.2 スキーマの定義

3.1 で得られた検索フローおよび参照情報として得た設計書に基づき、オントロジー構築に必要なスキーマの定義を行う。本手法では、定義する全てのクラスを「rdfs:Class」のサブクラス、プロパティを「rdf:Property」のインスタンスとして扱う。具体的には設計情報の区分を示すクラス、異なる情報区分の関係を記述するプロパティを定義する。

3.3 モデルの構築

3.2 で定義したスキーマに基づき、設計情報(インスタンス)の関係を記述する。具体的には、3.1 で得た設計書を RDF モデルに変換する作業である。各設計情報に一意の URI を割り当て、異なる情報区分の設計情報を定義済みのプロパティを用いて関連付ける。これにより、検索タスクにおける参照情報の関係について、図 2 下部に示すようなモデルとして記述する。

4. 適用事例 1: システム仕様変更時の影響調査

情報システム開発工程においては、画面や機能の変更、DB 項目の桁数の変更などの仕様変更が数多く生じ、その仕様変更によって影響を受ける設計書を修正する必要がある。本適用事例では、この仕様変更の一例として、DB 項目の桁数変更により、影響を受ける設計書を検索するプロセスを対象とする。例えば、店舗情報を管理する情報システムにおいて、管理対象とする店舗数の増加により、店舗情報を管理する DB「店舗マスタ」が持つ DB 項目「店舗コード」が 4 桁から 8 桁に増加した場合、この「店舗マスタ」に処理を加える「店舗マスタ登録」等の機能、およびこの機能に関する設計書に記載された情報を修正する必要がある。本適用事例では、このような検索タスクのモデル化を目的とする。図 3 にオントロジーの構築手順を示す。以下、図 3 に準じて説明する。

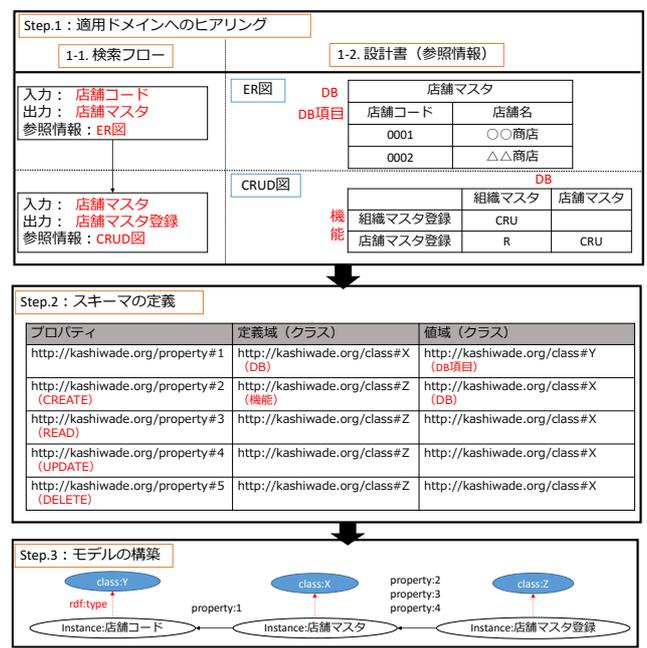


図 3 事例 1 のタスクオントロジー構築手順

4.1 適用ドメインへのヒアリング

適用組織へのヒアリングを通じ、図 3 上部に示す検索フローを得た。本対象プロセスは大きく 2 つのタスクからなる。1 つは「ER 図」を参照し、具体的なインスタンスを用いた検索フローの例として、「店舗コード」を入力とし、「店舗マスタ」を出力するプロセスである。なお「ER 図」とは、DB とその DB が持つ項目の関係を記述した設計書である。2 つ目のタスクは「CRUD 図」を参照し、「店舗マスタ」を入力とし、「店舗マスタ登録」を出力するプロセスである。「CRUD 図」とは、機能と DB の関係、具体的には DB に対する「登録 (CREATE)」、「参照 (READ)」、「更新 (UPDATE)」、「削除 (DELETE)」を行う機能の関係を記述した設計書である。

4.2 スキーマの定義

ヒアリング結果に基づき、クラスとプロパティに関するスキーマを定義する。参照情報として取得した「ER 図」「CRUD 図」から計 3 つのクラスを定義し、それぞれ URI として「class:X」「class:Y」「class:Z」を与えた。なお接頭名詞空間「class」は、本適用事例において独自に定義したものである。各クラスはそれぞれ「DB」「DB 項目」および「機能」を示す。

プロパティについては計 5 つを定義し、それぞれ URI として「property:1」「property:2」「property:3」「property:4」「property:5」を与えた。さらに、それぞれのプロパティの定義域と値域を与える。「property:1」は DB と DB 項目の関係を表すためのプロパティであり、定義域として「class:X」、値域として「class:Y」を与えた。また、「property:2」「property:3」「property:4」「property:5」は機能と DB の関係を表すためのプロパティであり、定義域として「class:Z」、値域として「class:X」を与えた。

4.3 モデルの構築

次に「ER 図」「CRUD 図」から「DB 項目」「DB」「機能」の設計情報の関係を記述したモデルを構築する。「ER 図」「CRUD 図」は MS Excel で作成された設計書であったため、これらの設計書内に記述されたインスタンス群を機械的に抽出し、インスタンス間の関係を定義済みのプロパティを用いて関連づけた。例えば、「CRUD 図」から抽出した機能「店舗マスタ登録」と DB「店舗マスタ」は「CREATE」「READ」「UPDATE」の関係にあるため、プロパティ「property:2」「property:3」「property:4」を用いて関連づけを行った。

これらのプロセスを経て、図 3 下部に示すタスクオントロジーを構築した。このタスクオントロジーを利用することにより、機能、DB、DB 項目の関係のモデル化を実現した。さらに、タスクオントロジーにおけるインスタンスを検索対象とする設計書と関連づけることにより、設計情報の関係に基づく設計書の検索が可能となる。

5. 適用事例 2: ERP パッケージの導入支援

パッケージ導入ベンダー企業は、顧客の要求に基づくシステム要件に対し、パッケージが提供する標準機能(以下、部品と呼ぶ)群を選択し、それらを組み合わせることでシステム設計を行う。例えば「債務管理」という「業務」は、「小切手印刷」や「債務伝票計上」という「機能」によって構築され、各機能はパッケージが提供する「小切手再印刷」や「未転記伝票計上」等の「部品」群の組み合わせによって構成される。しかし、パッケージが提供する部品、およびその組み合わせによって構築される機能や業務内容は多岐に渡る。したがって、パッケージシステムの導入においては、過去のプロジェクトにおける設計履歴等を探索、参照して設計を行う。本適用事例では、このような設計履歴を検索するタスクのモデル化を目的とし、図 4 に示す手順でタスクオントロジーを構築する。以下、これに準じて説明する。

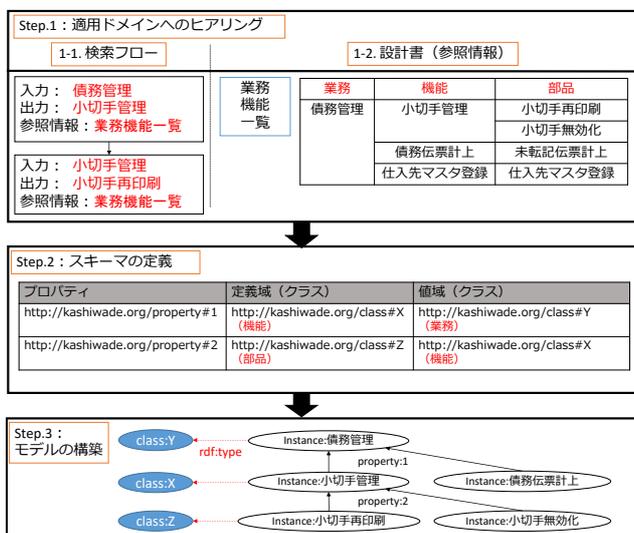


図 4 事例 2 のタスクオントロジー構築手順

5.1 適用ドメインへのヒアリング

適用組織へのヒアリングを通じ、図 4 上部に示す検索フローを得た。本対象プロセスは大きく 2 つのタスクからなる。1 つは「業務機能一覧」という設計書を参照情報とし、具体的なインスタンス例として、「債務管理」を入力し、「小切手管理」を出力するプロセスである。また 2 つ目のタスクは、先のプロセスと同様に「業務機能一覧」を参照し、「小切手管理」を入力、「小切手印刷」を出力とするプロセスである。なお「業務機能一覧」とは、プロジェクトにおいて構築した業務内容を構成する機能、およびその機能の構築に必要な部品の関係を記述した設計書である。

5.2 スキーマの定義

ヒアリング結果に基づき、クラスとプロパティに関するスキーマを定義する。参照情報として取得した「業務機能一覧」から計 3 つのクラスを定義し、それぞれ URI として「class:X」「class:Y」「class:Z」を与えた。各クラスが具体的に示すものは、それぞれ「機能」「業務」および「部品」である。

プロパティについては計 2 つを定義し、それぞれ URI として「property:1」「property:2」を与えた。さらに、それぞれのプロパティの定義域と値域を与える。「property:1」は機能と業務の関係を記述するためのプロパティであり、定義域として「class:X」、値域として「class:Y」を与えた。また、「property:2」は部品と機能の関係を記述するためのプロパティであり、定義域として「class:Z」、値域として「class:X」を与えた。

5.3 モデルの構築

次に「業務機能一覧」から「業務」「機能」「部品」間の関係を記述したモデルを構築する。「業務機能一覧」は MS Excel で作成された設計書であったため、これらの設計書内に記述されたインスタンス群を機械的に抽出し、インスタンス間の関係を定義済みのプロパティを用いて関連づける。例えば、業務「債務管理」と機能「小切手管理」「債務伝票計上」をプロパティ「property:1」を用いて関連づけた。

上述した手順を経て、図 4 下部に例を示すタスクオントロジーを構築した。このタスクオントロジーを利用することで、システム要件に対し、過去プロジェクトで構築した機能、およびその機能の構築に用いた部品の一覧を機械的に抽出することが可能となる。

6. 考察

本研究では、提案手法を 2 つの異なる組織における検索タスクに対して適用した。具体的には適用事例 1 における設計情報の依存関係、および適用事例 2 における設計情報の階層関係を考慮した検索タスクのモデル化を実現した。

以下では、構築したタスクオントロジーの有用性について、適用事例 1 におけるケーススタディ[稗方 14]を例に考察する。

6.1 ケーススタディ

ここでは、DB 項目「店舗コード」の仕様変更によって影響を受ける設計書を検索するタスクを対象とする。適用事例 1 において構築したタスクオントロジーを用い、検索対象とする 1153 ファイルの設計書と関連づけた。具体的には、タスクオントロジー内で定義された機能に関する設計書に対し、機能のインスタンスをメタデータとして付与した。なおこのタスクオントロジーと設計書の関連づけについては、筆者らが開発したデジタル資産管理システム「KASHIWADÉ」[中村 15]を利用した。

有識者が正解データとして選定した 346 ファイルに対し、全文検索と構築したタスクオントロジーを用いた検索を適合率と再現率、F 値で評価した結果を表 1 に示す。

表 1 ケーススタディの結果

	全文検索	タスクオントロジーを用いた検索	改良タスクオントロジーを用いた検索
検索結果数	8	660	731
正解数	8	285	329
適合率	100	43.2	45
再現率	2.3	82.4	95.1
F 値	4.5	56.7	61.1

DB 項目「店舗コード」という文字列を設計書に直接含む文書を検索する全文検索では、適合率は 100 を示すが、再現率は 2.3 となり、F 値は 4.5 となった。一方、タスクオントロジーを用いた検索では、「店舗コード」に関連する DB、および機能の関係を辿ることにより、適合率は 43.2 と減少したが、再現率が 82.4 に向上し、結果として F 値は 56.7 となった。本事例のような仕様変更によって影響を受ける設計書の検索タスクにおいては、検索結果の網羅性が重要であり、本タスクオントロジーの有用性を示す結果となった。

またタスクオントロジーを用いた検索について、適合率、再現率の改善に向けた考察を行う。適合率については、DB と機能の関係を CRUD 図から抽出したため、DB 項目と機能の関係を正確にとらえていないことが適合率の低下につながった。具体的には、「店舗コード」を含む DB として「店舗マスタ」「組織マスタ」を ER 図から抽出し、これらの DB に作用する機能として「店舗マスタ登録」や「組織マスタ登録」を CRUD 図から抽出した。しかし、機能「組織マスタ登録」は「店舗コード」には直接作用せず、DB「店舗マスタ」に含まれる他の DB 項目に作用する機能であった。このように DB 項目と機能の正確な関係を「ER 図」「CRUD 図」からモデル化することができず、検索クエリとして与えた DB 項目に直接作用しない機能に関する設計書が抽出され、低い適合率を示す結果となった。

再現率については、「主キー」と「外部キー」の関係にある DB 項目が設計書内で異なる文字列として表現されている例が存在した。具体的には、主キー「店舗コード」の外部キーが設計書内で「取扱部門コード」として表現されており、後者の DB 項目に関連する設計情報の抽出漏れが生じる原因となった。

6.2 タスクオントロジーの改良

上述したケーススタディの結果から、適合率の低下の原因となった「DB 項目」と「機能」の関係、および検索結果の抜け漏れの原因となった「主キー」と「外部キー」の関係など、有識者へのヒアリング時には得られなかった関係性の存在が確認された。これらの関係性について、再度提案手法を適用することにより、タスクオントロジーの改良を行った。具体的には後者のケースにおいて、「主キー」と「外部キー」の関係を検索するフローを新たに考える。これらの関係は「DB 項目定義書」と呼ばれる設計書に記述されていた。この設計書を参照情報として利用し、主キーと外部キーの関係を設計書から抽出し、構築済みのタスクオントロジーに追加した。この改良オントロジーを用いた検索結果を表 1 右端に示す。この改良により、再現率が 95.1 に向上し、F 値も 61.1 に向上した。

6.3 タスクオントロジーの構築プロセスに関する考察

上述した改良プロセスに見るように、タスクオントロジーの品質の向上には有識者へのヒアリングに加え、実利用に基づく評価

が重要となる。これは有識者が検索時に参照する情報を漏れなく提示することは困難なためである。このため、有識者へのヒアリングと、構築したオントロジーの実利用に基づく評価を繰り返すことにより、タスクオントロジーの改良を行う必要がある。このようなサイクルの効率化には、提案したタスクオントロジーの構築手順の体系化によるコスト削減が有用である。また、設計情報の関係性の追加や更新を繰り返すに当たり、関係性の拡張が容易な RDF の利用が有益だと考える。

7. 結論

本研究では知識処理の対象を設計情報の関連を考慮した検索タスクに限定することにより、検索タスクのモデル化を行うタスクオントロジーの構築手法を提案した。具体的には、対象ドメインへのヒアリング項目を定型化し、設計書の再利用による半自動的なモデル構築に基づくオントロジー構築手順の体系化を行った。本手法を情報システム企業 2 社における検索タスクに適用し、依存関係や階層関係にある設計情報のモデル化を実現した。またケーススタディを通じ、タスクオントロジーを用いた検索が全文検索に比べて高い再現率を示し、本手法に基づいて構築したタスクオントロジーの有用性を示した。さらにオントロジーの改良プロセスの観点から、オントロジーの構築手順の体系化の利点について考察した。

参考文献

- [古崎 02] 古崎晃司, 來村徳信, 佐野年伸, 本松慎一郎, 石川誠一, 溝口理一郎: オントロジー構築・利用環境「法造」の開発と利用—実規模プラントのオントロジーを例として, 人工知能学会論文誌, Vol.17, No. 4, pp. 407-419 (2002).
- [海谷 12] 海谷治彦, 清水悠太郎, 安井浩貴, 海尻賢二, 林晋平, 佐伯元司: 要求獲得のためのオントロジーを Web マイニングにより拡充する手法の提案と評価, 情報処理学会論文誌, Vol.53, No.2, pp.495- 509 (2012).
- [関谷 99] 関谷貴之, 吉岡真治, 富山哲男: オントロジーを用いた統合的設計支援環境の実現, 人工知能学会論文誌, Vol.14, No.6, pp.1051-1060 (1999).
- [研谷 07] 研谷紀夫, 馬場章: 文化資源オントロジーの構築とその活用(<特集>第 15 回(2007 年度)年次大会(研究報告会 & 総会)), 情報知識学会誌, Vol.17, No.2, pp.129-134 (2007).
- [白川 11] 白川真澄, 中山浩太郎, 荒牧英治, 原隆浩, 西尾章治郎: Wikipedia と Web の情報を組み合わせたオントロジー構築の試み(Web 情報オントロジー, <特集>データ工学論文), 電子情報通信学会論文誌. D, 情報・システム, Vol.94, No.3, pp.525- 539 (2011).
- [稗方 14] 稗方和夫, 大和裕幸, 深田直人, 中村覚, 岡田伊策, 齋藤稔, 笈田佳彰, 渡辺郁雄, 松本滋: システムの仕様変更調査における設計情報を用いた影響分析システムの開発, 日本機械学会第 24 回設計工学・システム部門講演会. pp.1305 (2014).
- [中村 15] 中村覚, 大和裕幸, 稗方和夫, 岡田伊策, 齋藤稔, 笈田佳彰: ドメイン知識の記述支援と活用のためのデジタル資産管理システム基盤「KASHIWADE」の開発 - 情報システム開発における設計書・設計知識への適用, 2015 年度人工知能学会全国大会, 2M5-4 (2015).