

# オントロジーを用いた製造現場の 不具合情報検索手法に関する研究

## A Study on Ontology-based Search Method for Failure Reports at Fabrication Shop

稗方和夫<sup>1</sup> 大和裕幸<sup>2</sup> 辻本翔<sup>2</sup>

Kazuo Hiekata<sup>1</sup>, Hiroyuki Yamato<sup>2</sup>, and Sho Tsujimoto<sup>2</sup>

<sup>1</sup> 東京大学大学院 工学系研究科

<sup>1</sup> Department of Systems Innovation, Graduate School of Engineering, The University of Tokyo

<sup>2</sup> 東京大学大学院 新領域創成科学研究科

<sup>2</sup> Graduate School of Frontier Sciences, The University of Tokyo

**要旨:** 設計品質向上に貢献する不具合情報を不具合文書群から検索する手法の開発を目的とする。係受け関係に注目して不具合の情報の要約をデータベースに蓄積する。このデータベースに対し、製品の部品構成オントロジーによる網羅的な不具合事例の検索、生産現場作業を細分化したオントロジーにより作業別に分類された不具合事例の検索を実行する。9604 件の不具合文書に適用したところ、文字列一致検索では発見出来ない不具合情報の検索が可能なることを示した。

## 1 緒言

製造現場では、設計品質低下によって製造工程における製品の不具合が後を絶たない<sup>[1]</sup>。このような不具合の情報は電子化されて不具合文書として蓄積されている。そこで、製造現場における不具合情報を設計時に提供することで設計品質の向上と不具合文書の有効利用を目指す。

ここでは、設計者が入力したキーワードに関連した不具合情報を含む不具合文書を検索することを想定する。単純な文字列一致による全文検索手法では、不具合文書の文脈や同義語を考慮した検索を実行することが出来ない。例えば、キーワードとして「取付け」と入力した場合、「取付け不良」という取付けに関連する不具合情報が検索されるが、同時に「取付け後の清掃が不十分」といった取付けに関連しない不具合情報も検索される。また、「取付け」の同義語である「設置」が含まれる不具合情報は検索することができない。

そこで、本研究では自然言語処理技術とオントロジー技術を用いることによって、不具合の箇所と現象との文脈上のつながりや同義語を考慮した不具合情報検索手法を提案する。なお、提案手法を造船会社設計部門の不具合文書に対して適用し有効性を検証した。

## 2 提案手法概要

### 2.1 対象とした不具合文書

一般に不具合文書には文書の ID、記述者の氏名、不具合の起こった日付、タイトルや不具合に関する記述等が記録される。本研究では、不具合に関する記述部分を検索対象とした。

### 2.2 提案手法の概要

本論文の提案手法の処理構成を Fig.1 に示す。本手法は、不具合情報抽出部、検索処理部から構成される。まず、係受け関係に注目した処理で不具合文書から不要な情報を除去し、不具合の情報をデータベースに蓄積する。このデータベースに対し、製品の部品構成オントロジーによる網羅的な不具合事例の検索、現場作業を細分化したオントロジーにより作業別に分類された不具合事例の検索を実行する。まず、係受け関係に注目した処理で不具合文書から不具合が発生した部品や現象について整理された不具合情報を抽出し、データベースに蓄積する。このデータベースに対して製品の部品構成オントロジーや製造現場作業を細分化したオントロジーによる網羅的な検索及び、オントロジーを同義語辞書として用いた同義語を考慮した検索を実行する。

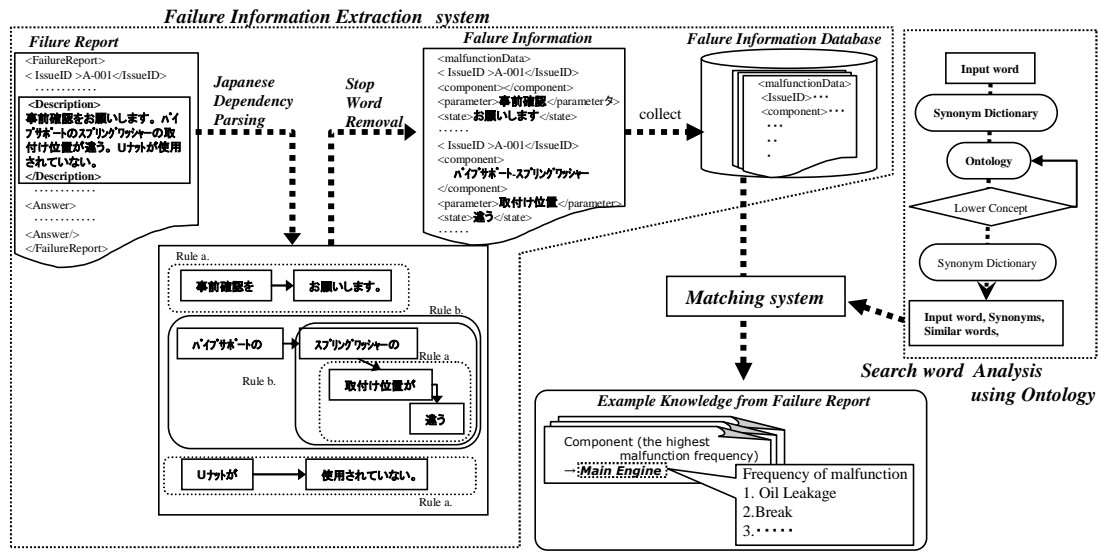


Fig.1 Proposed System Overview

## 2.3 不具合情報抽出部

### 2.3.1 不具合情報抽出ルール

係受け関係を利用することにより、固有表現抽出手法を利用した単語の頻度による抽出手法<sup>[2]</sup>では考慮できない不具合の箇所と現象との文脈上のつながりを考慮した不具合の情報を抽出することが出来る。係受け解析結果を考察し、不具合の情報は以下のパターンで記述される。

ア. (不具合箇所) において /で/ の (パラメータ) は /が/ を /に (状態量)

イ. (不具合箇所) において /で/ の (不具合事象) それぞれのパターンごとに以下に示すようなルールを利用して不具合情報を抽出する。

ルール a.: 名詞もしくは未知語と格助詞もしくは係助詞を含む文節かどうか

ルール b.: 「において /で/ の /に」 のような一部の助詞を含む文節かどうか

本研究では、これらのルールを利用して、不具合の情報を抽出する。

### 2.3.2 不具合情報抽出処理

Fig.2 は抽出処理のフローチャートである。フローチャートの制御には次の変数を用いる。n はノード番号で文末を 1 として文頭へ向かって順に振る。Node(n) は n 番目のノードを表し、Max は最大のノード番号を表す。Flag はルール a. もしくはルール b. を初めて満たした時に false に変わる変数である。Flag 変数は、フローの制御のみに使う。n=1、Flag=true として処理を開始する。ここで、Fig.3 のような係受

け関係の文があったとする。各ノードは文節を、矢印は係受け関係を表す。この場合、Max=10 である。Node(1) の時、ルール a.、ルール b. 共に合致しない。n は Max でなく、端点でも無いので、n=2 が代入されて次の文節の処理へ移行する。Node(3) の時、ルール a. を満たすので、その文節は parameter を表すと判定されるので Flag=false に変わり、次の文節へ処理が移る。Node(4) の時、その文節はルール a. を満たすので parameter と判定され、次の文節へ処理が移行する。Node(5) の時、この文節はルール b. を満たすので、component を表す文節であると判定される。前の段階で Flag=false が代入されているので、次は n が最大ノード番号かどうかの処理へ移る。Node(5) は Fig.3 の例では端点であるから、不具合情報の出力が行われる。出力は、ノード番号 n からノード番号 1 までをたどって、parameter・state・component・malfunction に該当する文節から助詞等の不要な品詞を除外してテーブルに出力する。このようにノード番号順に文節に処理を施していき、最終ノードに到達した時に、不具合情報抽出処理が終了し、不具合の発生した箇所とその現象で整理されたテーブルが完成する。Fig.3 から抽出される不具合情報は Table.1 である。

### 2.3.3 不具合情報データベース

以上の処理から抽出した不具合情報は、Table.2 のような不具合が発生した部品とその現象について整理されて蓄積される。記述者情報、不具合対応や対応者の情報をメタ情報として保持するために、不具合情報は文書 ID と関連づけされる。不具合情報の内で欠落している情報については、欠落したまま蓄積される。1 つの文書に複数の不具合情報が含まれる

場合もある。

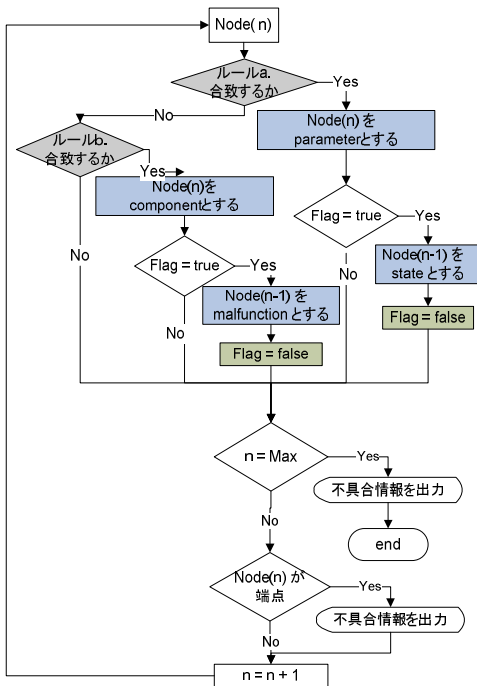


Fig.2 Failure Information extraction process

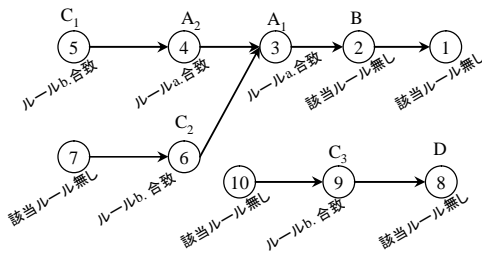


Fig.3 Result of Dependency Parsing

Table.1 Result of Failure Information

component	malfunction	
	parameter	state
⑤	③, ④	②
⑥	③	②
⑨	⑧	

Table.2 Example of Failure Information Database

ID	component	malfunction	
		parameter	state
A-001	制御パネル	設置不良	
A-002	主機	オイル漏洩	
A-002	--	溶接	不良
A-003	主機	温度	上昇
A-003	パイプサポートースプリングワッシャー	取り付け位置	違う
...	...	...	...

## 2.4 オントロジーによる検索処理部

検索語の同義語や近い概念の単語の推論を行う。そこで、本研究では情報の厳密性を定義し整理することができるオントロジー技術<sup>[3]</sup>に注目した。オントロジー技術により、あらゆる物が持つ概念とそれらの関係性を明示的に定義し、整理することができる。

### 2.4.1 構築したオントロジー

オントロジーとして component (不具合箇所) クラスの定義の一部を Fig.4 の上部に示す。クラスは複数の属性項目を有する。

#### operation クラス

<属性項目> component\_Name (名称属性)  
operation (構成属性)

#### component クラス

<属性項目> component\_Name (名称属性)  
sub\_component (構成属性)

#### synonym(同義語)クラス

<属性項目> synonym\_Name(同義語名称)

オントロジー辞書に登録する部品や同義語は、クラスのインスタンスとみなす。Fig.4 下部はその例であり、溶接は、ID「Op\_001」で「溶接」という component\_Name 属性、ID「Op\_101」という作業を is a 関係として持つ component クラスのインスタンスということを示している。また、「融接」は、ID「S\_001」という synonym クラスのインスタンスは、「溶接」を正規単語として、その同義語が「融接」であることを示している。構成部品を整理した component クラスの場合も operation クラスと同様に下位概念や同義語を持つ。

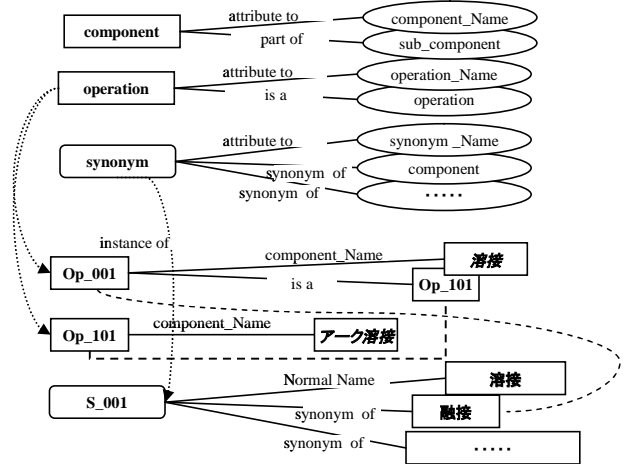


Fig.4 Example of Component Ontology

## 2.4.2 オントロジーと同義語辞書による処理

例えば、検索語として「溶接」と入力したケースを考える。検索語「溶接」は同義語「ID:S\_001」と結び付けられる。そうすると、ID: Op\_001 の正規表現は「溶接」であり、同義語として「融接」が推論される。この正規表現は ID:Op\_001 という作業の名前要素である。ID:Op\_001 は ID:Op\_101 と is a 関係にある。ID: Op\_101 についても同義語を参照する。以降、検索単語の下位概念全てに対して以上の処理を行う。以上から検索単語「溶接」を入力すると、「融接」「アーク溶接」も類似検索語として追加される。このようにして全ての下位概念とその同義語を検索語として追加する。component クラスの場合も operation クラスと同様に処理を行い、全ての下位概念とその類義語を検索語として追加する。

## 2.5 類似不具合事例検索部

Fig.1 の「Matching System」に該当するのが類似不具合事例検索部である。検索語に対するオントロジー処理の後に提示された全ての単語を用いて、不具合情報データベースとのマッチングを行う。各単語と完全に一致する単語を含む不具合情報を不具合情報データベースから検索する。その後、不具合情報にメタ情報として付加してある文書 ID を検索結果として列挙する。

## 3 評価実験

### 3.1 実験対象データ

造船会社設計部門が所有する不具合文書 9604 件を実験に用いた。その構成を Table.3 に記す。不具合に関する記述である Description 部分が対象である。

Table.3 Example of Failure Report

		記述例
Issue ID	不具合文書ID	No.001
ProductID	製品の種類	Ship
Status	不具合の対応状態	Closed
IssuedBy	記述者氏名	鈴木
Sent Date	日付	29-Apr-08
Title	タイトル	装完成検査
Description	不具合に関する記述	パイプサポートのスプリングワッシャの取付け位置間違い
Attached File	添付ファイル	無し

### 3.4 不具合情報抽出実験結果

本研究で提案した不具合情報抽出部がどの程度正

確に不具合情報を抽出できているか評価する実験を行った。まず、9604 件からランダムに選んだ 111 件の文書に対して本手法を適用し抽出された不具合情報と、同じ 111 件の文書からベテラン設計者が手動で抽出した不具合情報とを比較した。その結果は Table.4 のようになった。ベテラン設計者は 185 件の不具合情報を抽出した。一方、本提案手法では、711 件の単語を抽出し、その内 136 件は不具合情報を正確に抽出することができた。以上から、本手法では、文書に含まれる不具合情報の抽出漏れは少ないが、余分な情報をかなり抽出してしまうという特長があると言える。

Table.4 Result of Malfunction Information Extraction System for Shipbuilding Company's Data

	Extracted Key Word	NOT Extracted Key Word	Total
Correct Key Word	136・・・(W)	49・・・(X)	185
NOT Correct Key Word	575・・・(Y)		
Total	711		

### 3.3 文書検索実験結果

前節で述べた対象データを利用して、本研究で提案した手法の有効性を評価した。係受け解析器には CaboCha<sup>[4]</sup>を利用した。本実験では、検索語として「施工」「溶接」、「主機」という単語を入力した。これらの単語はオントロジー上では Fig.5 のように定義されている。「施工」「溶接」は現場作業を細分化したオントロジーに記述されており、「施工」は「溶接」や「塗装」を内包する上位概念で、「溶接」は具体的作業を表す下位概念である。「主機」は部品構造を示すオントロジーに記述されており、「シリンダ」や「クランクシャフト」といった部品から成り立っていることが描かれている。

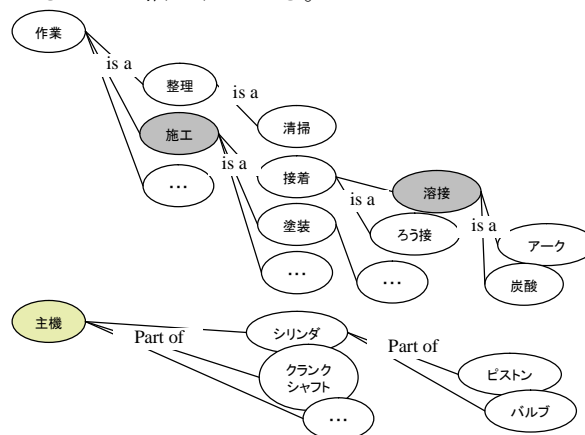


Fig.5 A Part of Ontology including the Term using in Experiment

### 3.3.1 検索語「施工」の実験結果

文字列の完全一致による全文検索の結果は Table.5 のようになった。( ) 内の数字は、収集した文書のうちでベテラン設計者が「施工」の不具合であると判断した文書数を表す。収集された文書の件数と比較すると、「施工」という抽象的な検索語の場合、提案手法の方が文字列完全一致による手法よりも多くの文書を収集することが出来た。抽出した文書のうちでベテラン設計者が期待した文書数の割合を精度と定義した。精度も抽出した文書数も提案手法の方が、文字列検索の結果よりも上回った。

Table.5 Precision Results in case of [施工]

	The Proposed Method	String Pattern Matching Method
The number of documents	1718 (1368)	258 (170)
precision	79.6 %	65.9 %

### 3.3.2 検索語「溶接」の実験結果

結果は Table.6 のようになった。( ) 内の数字は、収集した文書のうちでベテラン設計者が「溶接」の不具合であると判断した文書数を表す。検索結果は、提案手法の方が収集した文書が少なかったが、精度は提案手法の方が文字列検索の結果よりも上回った。

Table.6 Precision Results in case of [溶接]

	The Proposed Method	String Pattern Matching Method
The number of documents	106 (106)	436 (286)
precision	100.0 %	65.6 %

### 3.3.2 検索語「主機」の実験結果

結果は Table.7 のようになった。( ) 内の数字は、収集した文書のうちでベテラン設計者が「主機」の不具合であると判断した文書数を表す。収集された文書の件数と比較すると、提案手法の方が文字列完全一致による手法よりも多くの文書を収集することが出来た。抽出した文書のうちでベテラン設計者が期待した文書数の割合を精度と定義した。精度は提案手法の方が、文字列検索の結果よりも下回った。

Table.7 Precision Results in case of [主機]

	The Proposed Method	String Pattern Matching Method
The number of documents	1623 (1200)	58 (34)
precision	73.9 %	58.6 %

## 3.5 考察

不具合情報抽出手法の実験結果から、本手法では不具合文書から精度良く不具合情報を抽出することが出来た。一方で、余分な情報が 80%以上含まれていた。精度向上のために、抽出結果に含まれる余分な情報を除去する方策が必要である。次に前節で挙げた 3 つの文書群についてそれぞれ考察する。不具合情報が全く抽出されなかったグループについては、不具合に関する記述が不十分であったため、不具合情報を抽出出来なかった。このような文書に対しては、別途タイトル情報を利用すること有効である。その他のグループについては、不適切語句や表現を除去することにより、精度向上が望まれる。

次に文書検索実験について考察する。Table.5、Table.6 及び Table.7 の Precision の値を比較すると、提案手法の方が文字列検索よりも類似文書を正確に収集することが出来た。文字列検索では文脈を無視して検索語が含まれる文書を全て収集しているのに対して、本手法では不具合情報データベースの検索を行っている。このため、本手法の検索では、不具合文書に記述された不具合に関係が深い単語との検索に限られる。これにより、不具合と無関係に出現している検索語とマッチングすることを避けることができる。以上より、不具合情報データベースを利用すると、正確に類似文書を収集できると言える。

最後に、検索された文書数について考察する。「施工」で検索した場合、提案手法の方が多くの文書を収集することが出来た。Fig.5 のオントロジーにある通り、「施工」は「接着」や「塗装」以下の多数の下位概念を持つ単語である。したがって、オントロジーや同義語辞書から推論される単語数が多い。多数の単語で不具合情報データベースを検索することができたため、多くの不具合事例が収集された。この結果については、設計経験のある技術者から定性的に高い評価を得た。Table.7 についても同様に考察出来る。検索語として「主機」を入力すると、「シリンダ」や「バルブ」といった主機を構成する部品名でも検索が実行される。したがって、主機に関連する不具合は構成部品の不具合に至るまで網羅的に検索される。

一方、「溶接」による検索では、提案手法による収集件数は少なかった。Fig.5 のオントロジーを参照すると、「溶接」は下位概念を 2 つしか持たない概念である。したがって、オントロジーから推論される単語が少ない。このことは原因 1 つである。さらに、「溶接」を含む不具合情報を抽出に失敗している例が多かったことも原因として挙げられる。抽出に失敗した記述の例を以下に挙げる。

- ・ 下地処理後の溶接不良
- ・ 溶接部隙間あり

以上の記述からは不具合情報が抽出することができない。上記の例のように助詞が抜けている記述からは係り受け関係を解析できず、不具合情報を抽出することができないからである。

## 4. 関連研究

故障報告書の分析手法、自然言語処理による情報抽出及び、オントロジーを用いた情報検索における関連研究との比較を以下に示す。

### 4.1 故障報告書の分析手法

安藤ら<sup>[5]</sup>は、故障報告書分析の効率化を目的として、tf-idf法を用いて故障報告書から特徴語を抽出し、これを利用して故障報告書を整理する手法を開発した。さらに、各特徴語の同義語辞書としてオントロジーを利用している。本研究では、不具合が発生した箇所とその現象とのペアで文章の特徴付けを行っている。本研究においても同義語辞書としてオントロジーを利用しているが、その他に対象製品の部品構造や製造現場における作業もオントロジーとして整理し、情報検索の際に有効に利用している。

### 4.2 情報抽出

Blog や SNS に記述された意見文から評判情報を抽出する研究が盛んである。Hu ら<sup>[2]</sup>は高頻度の名詞句を連想ルールマイニングにより抽出、評価対象とし、その評価は評価対象と隣接する形容詞としている。本研究では、不具合情報の抽出を行っており、抽出対象は異なるが、抽出手法の本質的な部分において、単語の頻度ではなく、文節の係り受け関係を利用した抽出パターンにより、不具合が発生した箇所とその現象を同定している。この手法により、情報抽出における再現率の向上や単語の文脈上のつながりを反映した情報抽出が可能であると考えられる。

### 4.3 オントロジーを用いた情報検索

Trappey ら<sup>[6]</sup>は、特許文の自動分類を目指し、各特許文の内容を示すキーワードの生成手法を開発した。まず、tf-idf法を用いて各特許文から特徴語を抽出する。次に、特許に関する単語の意味的な関連性、階層構造を記述したオントロジーを利用して、抽出された特徴語と類似の意味を持つ単語を網羅的に収集する。本研究では、ユーザの入力した検索語と類似な単語を収集する際にオントロジーを利用している。本研究において利用したオントロジーは、意味的に

整理したオントロジーの他に、対象製品の部品構造を記述したオントロジーも構築している。これにより、対象製品の構造についても網羅的に単語の収集が可能となっている。

## 5. 結言

係り受け関係を利用して不具合情報を抜き出し、オントロジー技術を利用して、類似不具合情報を検索する手法を提案した。さらに、造船会社の不具合文書に対して適用し、網羅的に類似不具合事例を検索可能であることおよび、単純な文字列一致による検索では手法では発見することが出来ない不具合情報を検索可能なことを確認し、本手法の有効性を検証した。本手法によって、不具合文書から不具合情報を収集し、それらの情報を設計工程へ反映させ、設計品質の向上させる効果が期待される。

## 謝辞

本研究を行うにあたり、御助言、御協力いただいた株式会社川崎造船の野々口修次氏、松野二郎氏に深く感謝の意を表します。本研究は科研費若手研究 (B) 20760556 の助成を受けたものである。

## 参考文献

- [1] 田村泰彦, 飯塚悦功: 不具合に関する設計知識の運用に関する研究: ストレス-ストレングスマデルによる知識獲得, 日本品質管理学会, vol.31, No.1, pp168-186 (2001)
- [2] Mingqing Hu and Bing Liu: Mining opinion features in customer reviews, Proceedings of the 19th National Conference on Artificial Intelligence (AAAI), pp.755-760(2004)
- [3] 來村徳信, 溝口理一郎: オントロジー工学に基づく機能的知識体系化の枠組み, 人工知能学会論文誌, vol.17, No.2, pp.61-72 (2002).
- [4] CaboCha, <http://chasen.org/~taku/software/cabocha/>
- [5] 安藤英幸, 大和裕幸, 中野知彦, 増田宏: 故障報告書分析のためのコーディング手法に関する研究, 日本造船学会論文集, vol.195, pp53-61, (2004)
- [6] Amy J. C. Trappey, Charles V. Trappey, and Chun-Yi Wu: A Semantic Based Approach for Automatic Patent Document Summarization, Proceedings of the 15th ISPE International Conference on Concurrent Engineering, pp.485-494, (2008)