

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月29日現在

機関番号：17102

研究種目：基盤研究（C）

研究期間：2010～2012

課題番号：22500220

研究課題名（和文）

医療情報テキストを対象にしたノントピカルカテゴリに対する自動分類手法の基礎的研究

研究課題名（英文）

Non-topical Text Classification for Healthcare Information

研究代表者

石田 栄美（ISHITA EMI）

九州大学・附属図書館・准教授

研究者番号：50364815

研究成果の概要（和文）：ノントピカル(非主題)カテゴリに対する分類の可能性を検討するため、まず、分類のための十分な学習用データが得られる、非主題カテゴリである「学术论文」を対象に学术论文と非論文に分類する実験を行った。その結果、テキストの構造や構成要素が分類の手掛かりになることが明らかになり、これは医療情報に対しても適用できる可能性がある。医療情報に関しては、文献調査等から有効な非主題カテゴリを検討し、ウェブ上の情報の提供実態を調査した。

研究成果の概要（英文）：The goal of this research project is to explore the potential for healthcare information classification based on non-topical categories. At first, this research focused on the detection of academic articles as one of the non-topical categories. The detection rules were based on the structures and elements of scientific papers. In addition, I conducted a survey of healthcare information on the Web for setting up appropriate non-topical categories for them.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,500,000	450,000	1,950,000
2011年度	900,000	270,000	1,170,000
2012年度	900,000	270,000	1,170,000
年度			
年度			
総計	3,300,000	990,000	4,290,000

研究分野：図書館情報学

科研費の分科・細目：情報学フロンティア，図書館情報学・人文社会情報学

キーワード：ノントピカルカテゴリ，医療情報，テキスト自動分類，学术论文自動判定，価値観推定

### 1. 研究開始当初の背景

現在、さまざまな種類の情報がオンライン上で利用可能となっている。利用者がある主題に関する情報を見つけない場合、既存のサーチエンジンをを用いることで、その主題に関する情報を得ることは可能となっている。しかしながら、ほとんどのサーチエンジンは、利用者が入力した検索語とページに含まれる

語の照合や主題的関連を重視しており、どのような読者を想定して書かれたページであるのか、書かれている情報を信頼することができるのか、客観的、あるいは主観的に書かれたページなのかというような側面を分析してはいない。たとえば、ぜんそくに関する2つのウェブページがあった場合を想定する。どちらも主題カテゴリとしては「ぜんそく」

となる。ただし、一つは医療機関や専門家がぜんそくの治療法について書いたものであり、もう一方はぜんそくの治療体験について書かれたものであったとする。この場合、あるぜんそく患者は、信憑性が低かったとしても、気持ちを共有するために、同じ患者の体験を読みたいと思っているかもしれないし、治療方法を検討している者は、どのような治療法が存在するのか、より専門性の高い情報を得たいと思っているかもしれない。このように、利用者によって、同じ主題でも異なる観点から情報を求めている場合が考えられる。現在のサーチエンジンでは、例にあげたような情報の側面まで示すことができない。テキストを非主題カテゴリに分類することができれば、これらの側面を分析することが可能となり、利用者は、より多面的な角度から情報にアクセスすることが可能となる。よって、本研究では、主題ではなく、ノントピカルなカテゴリ、すなわち、非主題カテゴリに対する分類手法の検討を行う。

主題カテゴリに対する自動分類研究は1990年代から盛んに行われるようになり、分類の手がかりとなる特徴素(主に単語)の抽出・選択方法や、Support Vector Machine (SVM)やk近傍法(kNN)に代表される優れた分類手法などが提案されてきた。パラメータの違いやテキストの違いによって多少の変化はあるが、分類手法に関しては基本的な分類手法は既に確立しているといえる。近年になり、非主題カテゴリを対象にした分類が注目され始め、たとえば、テキストのスタイルやジャンルの分類、意見分析、読者レベルの分析などに関する研究が行われている。

主題カテゴリの分類には主にテキスト中に出現する単語、およびその出現回数が手掛かりになるのに対し、非主題カテゴリは、それらに加え、特定の語、語彙、テキストのデザインなど主題を表現する単語以外の語や形態的属性も、分類の手がかりとして有効である可能性がある。本研究では、まず、非主題カテゴリに有効な分類の手がかりは何かを分析する。

## 2. 研究の目的

本課題では、テキストを非主題カテゴリへの分類するための基礎的な研究をすることが目的である。主題は、医療情報とする。医療情報は、ウェブ上で探索している利用者が多いこと、医療情報は明確な目的を持って探索している人が多いと考えられるため、より多くの側面から情報を提示するとは意義があるといえる。医療情報を対象にした分類を行うためには、医療情報テキストに対して非主題カテゴリが付与された実験用集合が必要であるが、どのような非主題カテゴリが適切であるか、それらが実際に提供されている

医療情報に対して有効かが明らかになっていない。そこで、まず、どのような非主題カテゴリを設定するのが適切であるのかを、既往研究や実際に提供されている医療情報を調査する。

それと並行して、非主題カテゴリへの分類の可能性の検討、単語、形態的属性などテキストのどのような特徴が非主題カテゴリの分類に対して有効かを検討する。これには、非主題カテゴリの一つの例として、テキストタイプの一つである「学術論文」を対象にし、ウェブ上のPDFファイルから学術論文を自動判定する手法を検討した。分類手法の検討のためには、大量の実験用集合が必要となるが、これを得るための手法がすでに確立しているために本研究の対象とした。

## 3. 研究の方法

### (1) 学術論文の自動判定手法

「学術論文」という非カテゴリを対象にし、ウェブ上のPDFファイルの中から学術論文を自動的に判定する手法の開発を行った。自動分類は、テキスト中に出現する語の出現情報を分類器に学習させ、学習した分類器を用いて分類対象テキストを分類する。これまでの成果から、出現語を特徴素とする「出現語アプローチ」よりも、判定ルールを用いる「ルールベースアプローチ」の方が、処理時間が少ないにもかかわらず、ほぼ同程度の性能で学術論文の自動識別できることから、本研究では判定ルールを用いるアプローチを採る。

具体的には、以下の手順で、判定ルールの構築、および実験を行った。①実際に出版されている学術論文の構成と要素を明らかにし、実態調査から判定ルールを構築する、②判定ルールを評価するための実験用PDFファイル集合を構築する、③判定ルールの有効性を検証する。

#### ①学術論文の構成と要素の実態調査

実際に出版されている学術論文の構造・構成要素を調査することにより、関連文献から得られた論文の構造や構成要素が反映されているかを調査した。

調査対象は、分野や学術雑誌の格付けに依存しない広範な学術論文集合を得るため、「CiNii 本文収録刊行物ディレクトリ」、Web of Science、Journal Citation Reports、Arts & Humanities Citation Indexに掲載されたタイトルから日本語、英語それぞれの雑誌タイトルを無作為に抽出した。次に、標本に含まれた雑誌タイトルに対してPDFファイルが提供されているかを、データベース、検索エンジンなどを用いて手作業で確認した。PDFファイルが入手できる場合は、2010年第1号までの中から最新号を選択し、当該号の雑誌の中で2番目に掲載された論文を調査対象としてダウンロードした。本調査では、

ウェブ上で PDF ファイルが入手できるもののみを対象としたが、できるだけ多くの学術論文を収集するため、有料のものも含めた。PDF ファイルに含まれる内容が、学術的でない場合は、調査対象から除外した。その結果、日本語集合では 490 ファイル、英語集合では 682 ファイルの計 1,172 ファイルを取得した。これらについて、論文名、著者名、所属、抄録、キーワード、引用文献、掲載雑誌名、ISSN、DOI の構成要素の有無、構造を表す見出しの有無、縦書か横書かを確認した。さらに、論文が見出しを持っている論文を対象に、明示的に構造が表現されている場合には構造に関するラベルを付与した。ラベルは、IMRAD を構成する「序論 (Introduction)」「方法 (Method)」「結果(Result)」「結論

(Discussion)」に加え、「文献レビュー」と「結語」を追加した 6 ラベルとした。

最後に、実態調査の結果をもとに、学術論文にある程度共通する構造、構成要素を明らかにし、これらを学術論文の自動判定ルールとして構築した。

## ②実験用 PDF ファイル集合の構築

英語と日本語のそれぞれ 20,000 件の PDF ファイル集合を、以下のような手順で構築した。まず、検索エンジンの API (Yahoo! Search BOSS) を用いて、PDF ファイルの URL を収集した。2010 年 7 月に、英語ファイルの URL を収集した。検索語には、WordNet 3.0 の名詞句辞書に収録された 117,797 語句を用いた。1 検索式あたりの検索結果のうち、上位 500 件までの URL を収集した。検索結果が 500 件に満たないものは、全検索結果の URL を収集した。日本語 PDF ファイルの収集には、日本語 WordNet と IPAdic の両方に登録されている名詞 27,384 語を検索語として用い、2010 年 12 月にファイルタイプを「PDF」に限定し、言語を「日本語」に指定し、同様の API を用いて URL を収集した。日本語の場合は、検索結果の上位 1,000 件までを取得した。収集した URL 集合には重複が含まれていたため、重複除去を行った。

次に、それぞれに収集した URL 集合から無作為に 3 万件の URL を抽出し、英語ファイルは 2010 年 8 月に、日本語ファイルは 2011 年 1 月に、URL に基づきダウンロードした。日本語 PDF ファイル集合に関しては、URL のサブドメインから中国語で書かれていると推定されるファイルは削除した。ダウンロードしたファイル集合に対し、Apache PDFBox 1.2.1 を用いて、テキスト抽出を行った。一部分でもテキスト抽出ができたものは、英語ファイルでは 27,848 件、日本語ファイルでは 27,158 件であった。さらに、これらのファイル群からそれぞれ 20,000 件を無作為に抽出し、英語と日本語の PDF ファイル実験用集合とした。これらに対し、5 名

の判定者が、学術論文であるか否かの確認を行った。学術論文の判定規準としては、1)論文の形態をとっている、2)タイトル、著者名、所属機関が明記されている、3)引用、参考文献がある、4)1 論文が 1 ファイルで構成されている、5)2 ページ以上である、を用いた。

## ③判定ルールの検証

構築した判定ルールがテキスト中に出現すれば 1、していなければ 0 という表現に各ファイルを変換し、これらを判定器に入力することにより判定実験を行った。判定器には、Weka で実装されている、SVM、アダブースト、ランダムフォレスト、ナイーブベイズ、決定木(C4.5)を用いた。評価尺度は、精度、再現率、F 値を用いた。

(2) 医療情報を対象にした非主題カテゴリへの自動分類のための基礎的調査

## ①ウェブ上の医療情報の実態

まず、ウェブ上で提供されている医療情報に対して内容の分析など行っている文献を中心に、どのような医療情報が提供されているかを調査した。また、ウェブ上で提供されている医療情報の実態を把握するため、サーチエンジンを用いて、「ぜんそく 原因」「ぜんそく 経験」の検索語を用いて、検索結果の URL を収集した。複数の語を用いたのは「ぜんそく」など一語で検索すると一通りの検索結果しか得られないためである。収集した URL に対し、どのような非主題カテゴリが有効であるかを検討するために、暫定的に以下の非主題カテゴリ「対象とする読者(一般人～専門家)」「信憑性」「読者レベル」「書かれた観点(主観的～客観的)」について、3 人の判定者により、1 から 5 段階で数字を付与した。それぞれの段階については、判定者の間で合議を行い、調整した。それぞれのページに対し、キーワードを付与した。

## 4. 研究成果

(1) 学術論文の自動判定手法

### ①学術論文の構成要素と構造

学術論文の構成要素と構造を調査した結果、論題名、著者名、所属、掲載誌名という学術論文を識別するための基本的な要素については、日英ともほぼすべての学術論文で明示されていた。抄録、引用文献に関しても、いずれも高い割合で明示されており、日英に大きな違いはなかった。キーワードに関しては、抄録や引用文献に比べては低いが、日本語で 75.5%、英語で 63.2%の論文に含まれていた。見出しについては、日本語集合のうち見出しがないものが 0.8%に対し、英語集合は 7.3%であり、英語集合の方が見出しのない論文の割合が高かった。ISSN は日本語集合では含まれていなかったが、英語集合では 4 割以上の論文に含まれていた。DOI は、日

本語集合にはほとんど含まれていなかった。

見出しを基準とした構造に関する調査は、レビュー論文を除き、見出しがあった 1,087 論文に対して行った。調査の結果、35 パターンの構造が検出された。序論以外の明示的な構造を持たない論文が最も多く、380 件と全体の 35% を占めていた。次に、IMRD、IMRDC 構造が多かった。完全な IMRAD 形式ではないが、結語が付加されているもの、序論を持たないものなどを派生的な形式も含めると、ウェブ上に存在する学術論文 PDF ファイルは 40% 程度が IMRAD 形式を採っていた。これらの結果から、学術論文に共通する構成要素は存在し、全体としてみれば IMRAD 形式に近い構造が多く採られていたことがわかった。

### ②判定ルールの構築

論文の構造と構成要素を反映した新しい判定ルールを構築した。判定ルールは、ファイルの属性、論文の基本構造、構成要素、論文中の表現、非論文中の表現という基本的なカテゴリから構成されている。ファイルの属性には、「ファイルの構造」「ファイルのドメイン」「URL に含まれる語」が特徴素として含まれている。さらに、ファイルの構造には、「ファイルサイズ」「ページ数」「レイアウト」、暗号化の有無がある。ファイルサイズとページ数は、数値をそのまま判定ルールの特徴素として用いた。論文の構造は、先に示した IMRAD 形式に従った。構成要素は、抄録、キーワード、引用文献、謝辞、図表、附録とした。判定ルールに適用する場合は、これらが見出し語として出現するか否かを基準としている。論文中に用いられる表現、非論文中で用いられる表現は、論文の特有の表現として本文中で用いられる可能性が高い語を選定した。

### ③判定結果

20,000 件の英語集合と日本語集合を用いて、判定ルールの有効性を実験により検証した。その結果、英語ファイル集合の関しては 8 割の確率、日本語ファイルに関しては 5 割程度の確率で学術論文を判定することができた。実験で用いた日本語集合は、論文の割合が極端に低かったため、日本語集合を英語集合の論文／非論文と同様の割合にした集合を新たに作成し再実験を行った。その結果、約 7 割の確率で論文判定が可能であり、実験用集合の論文の割合は大きく影響するが、言語の違いは性能に大きく影響しないことがわかった。

以上の結果から、学術論文の特徴を示す構造や要素を見つけることができれば、言語に関わらず、少ない判定ルールでも学術論文を自動的に判定することができる可能性を示すことができた。機械学習で一般的に行われている判定は、テキスト中の語を用いて統計

的分析を行い、それらの特徴からテキストを自動的に判別している場合が多いが、本研究で行ったアプローチのように構造や構成要素、特徴的に用いられる語を用いた判定ルールも有効性があることがわかった。本アプローチは、学術論文だけではなく、特定の形式に則ったかたちで記述されているテキストの判別にも適用できる可能性があるため、医療情報テキストを対象にした非主題カテゴリへの分類にも可能性を見いだすことができた。

### (2) ウェブ上の医療情報の実態

「ぜんそく 原因」「ぜんそく 経験」を検索語として、サーチエンジン (Google) を用いて検索した結果、それぞれ 953 件、844 の URL を取得した。これらについて、判定者がそれぞれ見たところ、閲覧できなかったページ、動画や画像、PDF ファイルなどを除き、調査対象となったページは、それぞれ 813 件、908 件であった。これら検索結果の上位のページには、個人の体験談を記したブログが多かったが、Yahoo や OKWave などの Q&A サイトも多かった。文献調査からは、多くの医療機関や保健所がホームページを開設し、情報提供を行っている例があったが、固有の病名でサーチエンジンを検索したこともあり、検索結果の上位には医療機関のページはあまり見られなかった。

判定者による結果を見てみると、対象とする読者は、1721 件中、一般人を読者対象にしたページが 1,659 件(96.4%)であり、圧倒的に多かった。専門家を対象にしたものは 60 件(3.5%)あり、研究機関の成果報告のページ、CiNii の書誌情報、J-GLOBAL の文献情報のページなどであった。書かれた観点は、1721 件のうち、最も客観的に書かれている(レベル 5)と判断されたページは 285 件(16.6%)、比較的客観的に書かれている(レベル 4)と判断されたページは 102 件(5.9%)、反対に、主観的に書かれている(レベル 1)ページは 460 件(26.7%)であった。レベル 2 は、422 件(24.5%)であった。客観的に書かれているページは、医療機関や医師などが作成していると判断できるページが多く、主観的に書かれているページには、患者の体験談などが多く含まれていた。

読者レベルについては、全体の 84.6% が中間的な知識レベル(1 から 5 段階評価の 3)となった。また、これらのほとんどは、対象とする読者が一般人と評価されたものと相関している。そのため、このカテゴリは対象とする読者で代替することが可能といえる。信憑性については、内容の真偽の判断が困難であり、信憑性を被リンク数など外部的属性で判断する手法を用いるほうが適切である。

(3) まとめ

非主題カテゴリへの分類の一例として、学術論文を対象にし、判定ルールを構築し、実際のデータを用いて実験を行った。これらの成果から、その非主題カテゴリの特徴を表現する固有の構造や構成要素を導くことができれば、それらが分類に有効である可能性がわかった。これらは、分類器のための学習用データが十分に得られない場合でも有効である。また、ウェブ上で提供される医療情報の実態調査からは、個人的な経験を記述したブログや Q&A サイトなどが多いことがわかった。非主題カテゴリとしては、読者レベルや書かれた観点などが候補となることがわかった。

5. 主な発表論文等

[雑誌論文] (計5件)

- ① 小泉公乃, 國本千裕, 石田栄美, 杉内真理恵, 羽生笑子, 楊雯欣, 上田修一. “日本において電子書籍はどのように論じられてきたか: 雑誌記事と新聞記事の内容分析による論点の整理.” *Library and Information Science*. 査読有, No. 68, p.1-22, 2012
- ② 宮田洋輔, 安形輝, 池内淳, 石田栄美, 上田修一. “深層ウェブの実態とその要因: 機関リポジトリに登録された文献を用いた調査.” *日本図書館情報学会誌*. 査読有, Vol. 58, No. 2, p. 97-109, 2012
- ③ 石田栄美. “九州大学大学院ライブラリーサイエンス専攻の概要と附属図書館との連携” *名古屋大学附属図書館研究年報*, 査読有, 第10号, p.1-11, 2012.03
- ④ An-Shou Cheng, Kenneth R. Fleischmann, Ping Wang, Emi Ishita, and Douglas W. Oard. “The Role of Innovation and Wealth in the Net Neutrality Debate: A Content Analysis of Human Values in Congressional and FCC Hearings.” *Journal of the American Society for Information Science and Technology*. 査読有, Vol.63, No.7, p.1360-1373, 2012.03 DOI: 10.1002/asi.22646
- ⑤ 石田栄美. “テキストの自動分類をめぐる現況” *現代の図書館*, 査読有, Vol.48, No.4, p.211-216(2010.12)

[学会発表] (計18件)

- ① Emi Ishita, Yukiko Watanabe, Naoya Mitani, Miki Horiuchi, Yuiko Higa, Takako Oda, and Ai Yamaguchi. “Citation Analysis of the Availability of Conference Proceedings Cited in Doctoral Dissertations” The 5th

international conference on Asia-pacific Library and Information Education & Practice (A-LIEP 2013), 査読有, Khon Kaen City, Thailand, 2013.07

- ② 三谷直也, 堀内美希, 比嘉由似子, 山口藍, 小田孝子, 石田栄美, 渡邊由紀子. “博士論文が引用している会議録の利用可能率” 第60回日本図書館情報学会研究大会発表要綱, 査読無, p.69-72, 2012.11.17-18, 於九州大学
- ③ 南友紀子, 倉田敬子, 宮田洋輔, 石田栄美. “「読み」におけるデジタルメディアの位置づけ” 第60回日本図書館情報学会研究大会発表要綱, 査読無, p.77-80, 2012.11.17-18, 於九州大学
- ④ Emi Ishita, Yukiko Watanabe, Ai Yamaguchi, Takako Oda, Yuiko Higa, Miki Horiuchi and Naoya Mitani "The Ratio of Conference Papers in Citations of Engineering Dissertations at Kyushu University" *Proceedings of the 14th international conference on Asia-pacific digital libraries (ICADL'12)*, 査読有, p.59-62, Taipei, Taiwan, 2012.11
- ⑤ 石田栄美, 安形輝, 宮田洋輔, 池内淳, 上田修一. “学術論文に特化した検索エンジンの構築と評価” 2012年度日本図書館情報学会春季研究集会発表要綱, 査読無, p.69-72, 2012-05-12, 於三重大学
- ⑥ 宮田洋輔, 石田栄美, 池内淳, 安形輝, 上田修一. “学術論文の構成要素と構造” 2012年度日本図書館情報学会春季研究集会発表要綱, 査読無, p.65-68, 2012-05-12, 於三重大学
- ⑦ Kenneth R. Fleischmann, Doug Oard, An-Shou Cheng, Jordan Boyd-Graber, Thomas Clay Templeton, Emi Ishita, Jes A. Koepfler, and William A. Wallace. "Content Analysis for Values Elicitation" *CHI 2012 Workshop on Methods for Accounting for Values in Human-Centered Computing*, 査読有, Austin, USA, 2012.05.
- ⑧ Emi Ishita and Yukiko Watanabe. "The Department of Library Science at Kyushu University and Our Collaboration with Librarians" *The proceedings of 4th Workshop of the Asia Library and Information Research Group, Session 3B: Open Access & Scholarly Communication*, 査読有, pp.145, Taipei, Taiwan, 2012.03.13.
- ⑨ 山口藍, 井川友利子, 小田孝子, 吉田素

- 文, 石田栄美, 渡邊由紀子. “大学図書館の利用と学習成果に関する研究 —先行研究に基づく考察—” 平成 23 年度西日本図書館学会秋季研究発表会, 査読無, 2011.12.03, 於鹿児島国際大学
- ⑩ 小泉公乃, 國本千裕, 楊雯欣, 杉内真理恵, 羽生笑子, 石田栄美, 上田修一. “日本において電子書籍はどのように論じられてきたか” 第 59 回日本図書館情報学会研究大会発表要綱, 査読無, p.93-96, 2011.11.12-13, 於日本大学文理学部
- ⑪ Kensuke Baba, Toshie Tanaka, Emi Ishita, Masao Mori, Eisuke Ito, and Sachio Hirokawa. "Evaluation of link system between repository and researcher database" Proceedings of the 13th international conference on Asia-pacific digital libraries (ICADL'11), 査読有, pp.381-382, Beijing, China, 2011.10
- ⑫ Emi Ishita, Teru Agata, Atsushi Ikeuchi, Yosuke Miyata, Shuichi Ueda. “Detecting academic papers on the web.” Proceedings of Joint 11th annual international ACM/IEEE joint conference on Digital Libraries (JCDL'11), 査読有, pp.413-414, Ottawa, Ontario, Canada, 2011.06
- ⑬ 石田栄美, 安形輝, 宮田洋輔, 池内淳, 上田修一. “大規模日本語 PDF ファイル集合からの学術論文の自動判定” 2011 年度日本図書館情報学会春季研究集会発表要綱, 査読無, p.71-74, 2011.05, 於東京学芸大学
- ⑭ Emi Ishita, Douglas W. Oard, Kenneth R. Fleischmann, An-Shou Cheng and Thomas Clay Templeton. "Investigating Multi-Label Classification for Human Values" Proceedings of the American Society for Information Science and Technology(ASIST2010), 査読有, Vol.47, No.1, pp.1-4, Pittsburgh, USA, 2010.10
- ⑮ 安形輝, 池内淳, 石田栄美, 宮田洋輔, 上田修一. "学術情報に特化した検索エンジンの開発: 機械学習による英語論文の自動判定" 第 58 回日本図書館情報学会研究大会発表要綱, 査読無, p.81-84, 2010.10, 於藤大学
- ⑯ Emi Ishita, Teru Agata, Atsushi Ikeuchi, Nozue Michiko, Yosuke Miyata and Shuichi Ueda. "A search engine for Japanese academic papers" Proceedings of Joint JCDL/ICADL International Digital Libraries Conference (JCDL10), 査読有, p.379, Gold Coast, Australia, 2010.06
- ⑰ Emi Ishita, Shinji Mine, Chihiro Kunimoto, Junko Shiozaki, Keiko Kurata and Shuichi Ueda. "Analyzing viewing patterns while reading picture books," Proceedings of Joint JCDL/ICADL International Digital Libraries Conference(JCDL10), 査読有, p.381, Gold Coast, Australia, 2010.06
- ⑱ 宮田洋輔, 安形輝, 池内淳, 石田栄美, 上田修一. "学術論文 PDF の自動判定: 学習用集合が判定性能に与える影響". 2010 年日本図書館情報学会春季研究集会発表要綱. 査読無, p. 71-74, 2010.05, 於同志社大学
- 〔図書〕(計 2 件)
- ① 大谷康晴, 安形輝, 石田栄美, 中島玲子著. 『情報検索演習』JLA 図書館情報学テキストシリーズ II 6, 担執筆: UNIT20-23, pp.106-124, 日本図書館協会. 2011
- ② 上田修一, 倉田敬子編著, 安形輝ほか著『図書館情報学』, 分担執筆: pp.166-171, 180-182, 勁草書房. 2013
6. 研究組織
- (1)研究代表者  
石田 栄美 (ISHITA EMI)  
九州大学・附属図書館研究開発室・准教授  
研究者番号: 50364815
- (2)研究分担者  
なし
- (3)連携研究者  
なし