

Title	テキストマイニングに関する調査研究 [課題研究報告書]
Author(s)	藤井, 晃
Citation	
Issue Date	2017-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/14183
Rights	
Description	Supervisor:白井 清昭, 情報科学研究科, 修士

課題研究報告書

テキストマイニングに関する調査研究

北陸先端科学技術大学院大学
情報科学研究科

藤井 晃

2017年3月

課題研究報告書

テキストマイニングに関する調査研究

1210907 藤井 晃

指導教員 白井 清昭 准教授

審査委員主査 白井 清昭 准教授

審査委員 東条 敏 教授

審査委員 池田 心 准教授

北陸先端科学技術大学院大学
情報科学研究科

2017年2月

概要

近年、ビッグデータという言葉に代表される大量のデータの中から、有益な情報を取り出す技術に注目が集まっている。本課題研究では、ビッグデータの中でも特にテキストデータに着目する。大量のテキストデータを処理する方法として、自然言語処理の方法とデータマイニングの手法を合わせたテキストマイニングは有用な技術である。テキストマイニングとは、構造化されていないテキストから目的に応じて情報や知識を取り出す方法と技術の総称である。テキストデータは常に増加し、その中に有用な知識が埋もれていることを考えると、テキストマイニングは今後も発展が期待される技術である。そのため、過去の研究の概要や最新の動向をまとめることは、今後のテキストマイニング研究の方向性を決めるための貴重な資料となり、その意義は大きい。

本課題研究では、特にテキストマイニングの適用可能性に着目し、マイニングの対象となるテキストのトピックとメディアという観点から研究動向を分析することを目的とする。トピックとは、医療分野、経営分野など、テキストの分野やジャンルを指す。テキストマイニングの過去の文献をトピックによって分類することで、これまでどのような知識の獲得が試みられていたかを明らかにする。一方、メディアとは、ウェブ、アンケート調査、学術論文など、テキストが発行された媒体を指す。過去の文献をメディアによって分類することで、これまでどのようなタイプのテキストが解析の対象となっていたかを明らかにする。

Google Scholar を用いて、「テキストマイニング」をキーワードとして検索された上位 100 件の日本語文献、「text mining」をキーワードとして検索された上位 100 件の英語文献を調査の対象とする。また、トピックとして 22、メディアとして 19 のカテゴリを定義する。調査対象の文献に対し、該当するトピック及びメディアのカテゴリを人手で選び、トピックやメディアのカテゴリの分布や、その言語による違い、発行年による変化を調査する。

日本語文献では、出現頻度が最大のトピックは「医学・生命科学」の 10 件で、全体の 10% である。また、1 つの文献しか該当しないカテゴリが 8 件、2 つの文献しか該当しないカテゴリが 5 件存在する。特定のトピックへの偏りが見られず、多様なトピックのテキストを対象にテキストマイニングの研究が行われていることが分かる。メディアとしては「アンケート」のカテゴリが 21 件と最多で、「インターネットコミュニティ」17 件、「報告書」10 件と続く。これはトピックが多岐に渡っていることと関連があるだろう。アンケートやインターネットコミュニティには様々なトピックのテキストが存在すると考えられるからである。

一方、英語文献では、「医学・生命科学」のトピックに分類された文献が 32 件あり、医学のテキストに対するマイニングが盛んに研究されている傾向がうかがえる。

日本語でも「医学・生命科学」の文献は最多であったが、件数は10件であり、英語は日本語と比べて3.2倍の医学・生命科学分野の文献が存在する。また、新しい手法を提案する論文だけでなく、同分野のサーベイ論文や、医学・生命科学分野のデータベースを活用するのに特化したツールやシステムに関する論文も数多く存在する。メディアとしては、文献データベースを含む「論文」のカテゴリに属する文献が21件と最多であるが、それ以外に10件以上のカテゴリはなく、研究対象が論文に集中していることが分かる。これはトピックで文献を分類したとき「医学・生命科学」のカテゴリが最多となっていることと関連する。すなわち、英語文献では、医学・生命科学のデータベースに収録されている論文・アブストラクトをテキストマイニングの対象とした研究が多い。

調査対象とした文献の発行年が年によってどのように変化しているかを考察する。日本語文献については2004年の13件、英語文献については2005年と2008年の11件が最多となっており、これらの年にテキストマイニングの研究が盛んに行われていたことが分かる。また、その後現在に至るまで、テキストマイニングの論文は減少傾向にあることが分かる。次に、発行年によるばらつきの影響を軽減し、文献数の年毎の変化をより明確に調べるため、年毎に、その年の前後2年、合計5年間の文献数を調べた。その結果、英語文献については、2006年を中心とした5年間(2004年～2008年)で発行された論文数45件がピークとなり、その後論文数は減っている。一方、日本語文献については、2006年(40件)、2007年(35件)、2008年(41件)、2009年(41件)、2010年(40件)付近で文献数が最も多く、比較的長い期間にテキストマイニング研究のピークが存在したと言える。

当初の想定では、ビッグデータと呼ばれるようなインターネット上に膨大にあるテキストから有意義な知見を自動的に獲得することを目的とした研究が多く、経済、経営、社会調査などの分野でのテキストマイニングの活用を想定していた。実際は、全体の傾向では、「医学・生命科学」のデータベースを活用したテキストマイニングの研究が主流であることが分かった。ただし、英語文献は日本語文献に比べてはるかに多いので(調査時点におけるGoogle Scholarの検索で、日本語3,010件に対し、英語は2,390,000件)、対象とする英語文献の数を増やせば、様々なトピックやメディアを対象とした研究が見つかるものと思われる。

今後の課題としては、調査する文献の数を増やして同様の調査を行うこと、各論文の内容を精査して、テキストマイニングに使われている要素技術を調べ、どの要素技術がよく使われているかを整理することが挙げられる。また、現在利用可能なテキストマイニングのツールに関して、その特徴と活用事例を整理することも重要な課題である。

目次

第1章 はじめに	1
1.1 背景	1
1.2 目的	1
1.3 本課題研究報告書の構成	2
第2章 関連研究	3
2.1 テキストマイニングの定義	3
2.2 テキストマイニングのサーベイ論文	3
2.3 本課題研究の特色	4
第3章 テキストマイニングの研究動向の分析	5
3.1 調査方法	5
3.1.1 トピックの定義	6
3.1.2 メディアの定義	8
3.2 日本語文献の調査結果	10
3.3 英語文献の調査結果	16
3.4 考察	23
3.4.1 トピックに関する考察	23
3.4.2 メディアに関する考察	24
3.4.3 医学・生命科学を対象としたテキストマイニングに関する考察	25
3.4.4 年代による傾向の考察	26
3.4.5 被引用数の比較	28
第4章 おわりに	30
謝辞	31
参考文献	32

第1章 はじめに

1.1 背景

近年、ビッグデータという言葉に代表される大量のデータの中から、有益な情報を取り出す技術に注目が集まっている。情報を取得する対象となるビッグデータには様々なものがあるが、その中でもテキストデータは、インターネット上のブログ、SNS、マイクロブログ、オンラインフォーラムを通じた個人のコメントなど、様々な形で増加の一途をたどっている。そうしたテキストデータの中に、社会にとって有益な知識が埋もれている可能性は高い。大量のテキストを入手することは容易であるが、そこから有用な情報を得ることは簡単なことではない。

本課題研究では、ビッグデータの中でも特にテキストデータに着目する。大量のテキストデータを処理し、適切な情報を見出す方法に対するニーズは大きい。大量のテキストデータを処理する方法として、自然言語処理の方法とデータマイニングの手法を合わせたテキストマイニングは有用な技術である。テキストマイニングとは、構造化されていないテキストから目的に応じて情報や知識を取り出す方法と技術の総称である[201]。

先に述べたようにテキストデータは常に増加し、その中に有用な知識が埋もれていることを考えると、テキストマイニングは今後も発展が期待される技術である。そのため、過去の研究の概要や最新の動向をまとめることは、今後のテキストマイニング研究の方向性を決めるための貴重な資料となり、その意義は大きい。

1.2 目的

本課題研究の目的は、テキストマイニングの研究動向を調査することである。テキストマイニングに関する過去の研究を網羅的に調査し、大量のテキストデータから知識を発見する技術がどのような分野に応用されているか、応用される分野は言語・年代によってどのような違いがあるかを調査する。本課題研究では特にテキストマイニングの適用可能性に着目する。テキストデータには様々な種類がある。ウェブ上のテキストを始め、新聞記事、学術論文など、今後も増加し続けるテキストの種類は多い。過去の研究を概観し、テキストマイニングの技術はどのようなタイプのテキストに対して適用可能か、あるいは実際に適用されてきたかを知ることは重要である。また、大量のテキストからどのような知識を獲得するためにテキストマイニングの技術が使われてきたかを調査する。獲得できる知識はマイニングの対象とするテキストの分野(もしくはトピック)に大きく依存する。例えば、医学分野のテキストからは医学の知識が、経済関連のテキストからは株価や市場に関する知識が得られると期待できる。そのため、テキストマイニングにおいて、分析の対象とするテキストのトピックにはどのようなものが多いのかを調査する。これにより、テキストマイニングがどのような知識を得るために適用可能かを知ることができる。

1.3 本課題研究報告書の構成

本課題研究報告書の構成は以下の通りである。2章では、テキストマイニングの定義を述べる。また、過去のテキストマイニング研究のサーベイについて報告する。3章では、テキストマイニングの研究動向の分析を行う。テキストマイニングに関する日本語論文、英語論文を、分析対象とするテキストのトピックとメディアによって分類する。またトピック及びメディアの言語及び年による傾向の違いを分析する。最後に4章で本課題研究を総括する。

第2章 関連研究

本章では関連研究について述べる。まず、2.1節でテキストマイニングの定義を説明する。2.2節では、これまでに行われたテキストマイニングの調査研究について報告する。2.3節では本課題研究の特色について述べる。

2.1 テキストマイニングの定義

テキストマイニングとは、構造化されていない自然言語で書かれたテキストから目的に応じて情報や知識を掘り出す方法と技術の総称である[201]。テキストを単語や文節に分割する自然言語処理方法を介し、語句やモデリングしたパターンを集計し、データマイニングの手法で有用な情報を発見する。その基礎技術は、テキストマイニングを進めていくうえでの情報処理の流れに合わせて以下の3つに分類される[16]。

- ・ 概念抽出技術
自然言語で書かれた文書情報からその内容をあらわす概念を抽出する
- ・ マイニング技術
抽出された概念を統計的に分析する
- ・ 可視化技術
マイニング結果を人間が理解しやすい形に可視化（視覚化）し対話的な分析を実現する

2.2 テキストマイニングのサーベイ論文

これまでもテキストマイニングの研究を調査し、その研究動向を報告したサーベイ論文がいくつか存在する。これらのサーベイでは、テキストマイニングの要素技術を解説したり、テキストマイニングツールを用いた応用事例について紹介している。渡部は、テキストマイニングの基礎となる要素技術として文書から概念情報を抽出する技術について解説し、応用事例として特許検索・分析支援システム ACCENT の概要を紹介している [16]。林田と脇森は、テキストマイニングの適用分野、技術、製品開発、適用事例について紹介し、製品事例として「MiningPro21 文書マイニング・システム」と「TopicExplore」について触れ、それを適用した事例を紹介している [25]。大隅と保田は、テキストマイニングについて広範なサーベイを行った[33]。彼らはテキストマイニングの適用範囲として以下を報告している。

- ・ テキスト・カテゴリーゼーション
- ・ ドキュメント分類
- ・ ルール探索、ルール発見
- ・ 概念抽出、関係の発見
- ・ 情報の統合化、有機的統合化

- ・ 特定なトピックスの検出
- ・ テキストの分割
- ・ テキスト、文書の要約化と情報収集
- ・ 知識取得と理解
- ・ テキスト・ナビゲーション、視覚化のためのユーザー・インターフェース
- ・ Web への応用 (Web マイニング、テキスト学習、知的エージェント化)
- ・ 生物情報学への応用 (ゲノム解析、生物文献情報処理など)
- ・ ビジネスへの応用 (CRM、意見のマイニング)
- ・ 調査データの分析への応用 (自由回答、自由記述)
- ・ テキスト検索、全文検索、文書検索
- ・ 情報抽出

野村は、アンケート調査で自由回答欄に書かれたテキストを分析する手法としてテキストマイニングを紹介している [47]。また、テキストを解析するツールとして形態素解析システム茶筌¹を挙げている。奥村は、インターネット上のブログを対象としたテキストマイニング、すなわちブログマイニングに関する研究を紹介している [13]。

2.3 本課題研究の特色

本課題研究によるテキストマイニングの研究動向の分析は、これまでのサーベイとは異なり、トピックとメディアに着目している点に特色がある。トピックとは、ここではマイニングの対象とするテキストの分野やジャンルを指す。また、メディアとは、テキストがリリースされている媒体を指す。トピックとメディアの詳細は 3.1 節で述べる。テキストマイニングがどのようなトピックのテキストに適用されているか、どのようなメディアのテキストを分析しているかを分類し、整理する。トピックとメディアに注目して過去のテキストマイニングに関する論文を分類することで、これまでのテキストマイニングの研究が対象としてきたテキストの傾向を明らかにする。

¹ <http://chasen.naist.jp/hiki/ChaSen/>

第3章 テキストマイニングの研究動向の分析

3.1 調査方法

本課題研究では、テキストマイニングの研究の動向を調査するため Google Scholar (<https://scholar.google.co.jp/>) を利用して調査対象の文献を検索する。また、調査対象とする論文は、日本語で書かれた論文(以下、日本語文献と呼ぶ)と英語で書かれた論文(英語文献と呼ぶ)とする。日本語文献を検索するときは「テキストマイニング」を検索クエリとして検索する。英語文献を検索するときは「text mining」をクエリとして検索する。それぞれのクエリに対し、Google Scholar が表示する上位 100 件の日本語文献および英語文献を調査対象とする。Google Scholar の掲載順位は以下のように決められている。

Google Scholar では、Google ウェブ検索と同様、最も関連性の高い情報がページの上部に表示されます。Google のランキング技術では、それぞれの記事の全文、著者、記事が掲載された出版物、他の学術資料に引用された回数が考慮されます。²

2016 年 10 月 27 日時点で検索を実施した。検索された論文の数は、日本語文献が 3,010 件、英語文献は 2,390,000 件であった。日本語と英語のそれぞれについて上位 100 件の論文リストを得た。

次に、調査対象の文献を「トピック」と「メディア」の 2 つの観点から分類する。「トピック」による分類では、テキストマイニングに関する文献がどのようなトピックのテキストを対象に知識の獲得を試みているかで分類する。テキストには医学や教育など様々なトピックのものがあるが、知識獲得の対象とされやすいトピックは何か、あるいはテキストマイニングの対象とされるトピックに偏りがあるかを調査する。一方、「メディア」による分類では、どのような媒体(メディア)で公開されているテキストをテキストマイニングの対象としているかで文献を分類する。テキストは新聞、ウェブ、学術論文など様々な形式でリリースされているが、どのようなメディアのテキストがテキストマイニングの対象となりやすいかを分析する。

今回の調査では、トピックならびにメディアのカテゴリは以下の手順で定義した。最初に、各文献にキーワードを自由に与えて分類した。その後、似ているキーワードをまとめることカテゴリを定義した。最後に、カテゴリのセットを確定させた後、調査対象の各文献に対し、その文献が取り扱っているテキストのトピックおよびメディアを確定する作業を行った。

以下、本課題研究におけるトピックとメディアのカテゴリについて説明する。

²Google Scholar.(n.d.). Google Scholar について, 2017 年 1 月 30 日取得, from <http://www.google.co.jp/intl/ja/scholar/about.html>

3.1.1 トピックの定義

トピックとして以下に挙げる 22 個のカテゴリを定義した。

- 医学・生命科学
PubMed や MEDLINE のような医学・生命科学のデータベースや、医学分野の論文をテキストマイニング対象とした文献をこのカテゴリに分類する。
- 介護
介護専門職への調査の自由回答欄などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 化学
化学論文をテキストマイニング対象とした文献をこのカテゴリに分類する。
- 看護
看護研究抄録や看護に関する調査の自由回答欄などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 教育
教育学部の学生の報告書や、大学のゼミナールに関する質問紙調査をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 経営
マーケティングへの活用のための顧客アンケートや、組織革新の調査のための有価証券報告書をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 経済
経済動向や金融市場動向を把握するために Twitter や新聞記事などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 公衆衛生
インフルエンザの流行を分析するため SNS をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- コールセンター
コールセンターでの顧客とのやり取りの記録をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 災害
災害に関する新聞記事などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 社会調査
社会調査における自由回答データの解析のためのツールを紹介する文献や、都市農村交流のニーズを把握するために落書き帳をテキストマイニングの対象とした文献をこのカテゴリに分類する。

- 書誌学
文章の書き手の同定分析や計量書誌学に関連するテキストマイニングツールに関わる文献をこのカテゴリに分類する。
- 政治
議会の議事録をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 船舶
船舶の故障報告書などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- ソフトウェア
ソフトウェアのソースコードや設計書、ログ情報などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 体育
身体知覚のアンケートをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- テレビ CM
テレビ CM に関するアンケートをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 都市開発
都市イメージのアンケート調査の自由記述欄や、生活環境評価の自由記述欄などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 特許
特許に関するテキストをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- ナレッジマネジメント
ナレッジマネジメントのためのテキストマイニングツールに関する文献をこのカテゴリに分類する。
- 農業
農産物に対する消費者アンケートの自由記述欄などをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- ホットスポット検出
オンラインフォーラム内の発言を対象としてテキストマイニングを行う文献をこのカテゴリに分類する。
- テキストマイニングの手法・技術*
特定のトピックのテキストを対象としているのではなく、テキストマイニングの汎用的な手法や技術について論じている文献

- テキストマイニングのツール・システム*
テキストマイニングのためのツールやシステムについて論じている文献
- テキストマイニングのサーベイ*
テキストマイニングのサーベイ論文
- 対象外*
Google Scholar では書籍が検索されることがあるが、書籍は内容が多岐に渡るため、調査の対象外とした。また、学会の発表資料(スライド)がヒットした場合も今回の調査の対象外とした。

上記のカテゴリのうち、*のついている最後の4つのカテゴリは、トピックでの分類が不可能であるが、便宜上設けたカテゴリである。ただし、テキストマイニングの手法・技術、ツール・システム、サーベイに関する論文のうち、対象となるテキストのトピックが限定されているときは、その対象のトピックのカテゴリに分類する。例えば、医学に関するテキストマイニング技術のサーベイ論文は「医学・生命科学」のカテゴリに分類した。

3.1.2 メディアの定義

メディアとして19個のカテゴリを定義した。

- CMU テキストデータセット
カーネギーメロン大学(Carnegie Mellon University, CMU) の提供するテキストデータのセットで、様々なトピックのニュース記事などが含まれている。これをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- ウェブサイト・ニュース
ウェブサイトやウェブ上のニュース記事をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- アンケート
アンケートや調査における自由記述欄をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- インターネットコミュニティ
SNS、ブログ、マイクロブログ、オンラインフォーラム上の口コミなどをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- コーパス
ロイターコーパスやニュース記事のコーパスをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- コールセンターの記録
コールセンターの記録をテキストマイニングの対象とした文献をこのカテゴリ

りに分類する。

- 書籍
書籍、小説、日記、作文をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 新聞記事
新聞記事をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- ソフトウェア
ソフトウェアのソースコード、関連文書、設計書、ログ情報をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 電子メール
電子メールをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 特許文書
特許データベースや特許論文をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 報告書
各種報告書やレポートをテキストマイニングの対象とした文献をこのカテゴリに分類する
- 落書き帳
落書き帳をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 論文
学術論文や要旨 (Abstract)、それらを収めたデータベースをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 地方議会の審議項目
地方議会の審議項目をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- プレスリリース
企業のプレスリリースをテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 有価証券報告書
有価証券報告書をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 求人票
人材紹介の求人票をテキストマイニングの対象とした文献をこのカテゴリに分類する。
- 説明書・技術書

ソフトウェアなどの説明書・技術書をテキストマイニングの対象とした文献をこのカテゴリに分類する。

● 対象外*

書籍や学会の発表資料など。トピックと同様に、これらは調査対象外とした。

*のついている「対象外」のカテゴリは、メディアでの分類が不可能であるが、便宜上設けたカテゴリである。

3.2 日本語文献の調査結果

日本語文献の調査結果を以下の表にまとめる。この表では、参考文献番号、文献のタイトル、発行年、被引用数、トピックのカテゴリ、メディアのカテゴリの順に記載している。また、この表における文献の順序は、Google Scholar で表示された順序と同じである。

参考 文献	タイトル	発行 年	被引 用数	トピック	メディア
[1]	テキストマイニングによる評価表現の収集	2003	44	手法	インターネット コミュニティ
[2]	Excel で学ぶテキストマイニング入門	2002	55	対象外（書 籍）	対象外
[3]	blog ページの自動収集と監視に基づくテキストマイニング	2004	38	手法	インターネット コミュニティ
[4]	大規模テキストからの経験マイニング	2009	25	手法	インターネット コミュニティ
[5]	見えログ: 情報視覚化とテキストマイニングを用いたログ 情報	2000	22	ツール・シ ステム	ソフトウェア (ソースコー ド、関連文書、 設計書、ログ情 報)
[6]	テキスト型データの計量的分析	2004	76	ツール・シ ステム	対象外
[7]	単語の連想関係によるテキストマイニング	1999	11	ツール・シ ステム	新聞記事
[8]	テキストマイニングを活用したアンケートにおける自由回 答の分析と生活環境評価	2012	9	都市開発	アンケート

[9]	組織革新の認知的研究; 認知変化・知識の可視化と組織科学へのテキストマイニングの導入	2008	8	経営	有価証券報告書
[10]	テキストデータマイニングのための統合環境 TETDM の開発	2013	10	ツール・システム	対象外
[11]	SPSS によるテキストマイニング入門	2012	14	対象外 (書籍)	対象外
[12]	テキストマイニングと統計解析言語 R による特許情報の可視化	2009	8	特許	特許文書
[13]	ブログマイニング技術の最新動向	2008	11	サーベイ	インターネット コミュニティ
[14]	高齢者向け住宅改造の効果に関する介護専門職の評価指標と要介護反別特徴: テキストマイニングによる自由回答の分析	2004	6	介護	アンケート
[15]	テキストマイニングのための連想関係の可視化技術	1999	10	手法	対象外
[16]	テキストマイニングの技術と応用	2003	10	サーベイ	対象外
[17]	Total Environment for Text Data Mining テキストデータマイニングのための統合環境	2011	7	ツール・システム	対象外
[18]	看護大学生の精神科保護室に対する受け止めおよび視点の変化: テキストマイニングによる非構造型データの分析から	2007	9	看護	アンケート
[19]	テキストマイニングを用いた都市農村交流ニーズの把握ー岡山県吉永町ふるさと村の八塔寺山荘の落書き帳を対象としてー	2004	8	社会調査	落書き帳
[20]	新聞記事のテキストマイニングによる長期市場動向の分析	2013	6	経済	新聞記事
[21]	ソーシャルメディアを対象としたテキストマイニング	2013	7	手法	インターネット コミュニティ
[22]	統合物語生成システムの現状と特に物語内容生成メカニズム	2014	6	ツール・システム	対象外
[23]	テキストマイニングのための情報抽出	2000	5	コールセンター	コールセンター の記録
[24]	テキストマイニングを用いた米に関する消費者アンケートの解析	2006	6	農業	アンケート
[25]	テキストマイニング技術とその応用	2005	5	サーベイ	対象外
[26]	テキスト情報による金融市場変動の要因分析	2010	14	経済	報告書
[27]	経済テキスト情報を用いた長期的な市場動向推定	2011	7	経済	報告書

[28]	ソーシャルメディアへのテキストマイニングの適用に関する検討	2008	5	ツール・システム	インターネット コミュニティ
[29]	統合物語生成システム, 間テキスト性, テキストマイニング	2014	4	ツール・システム	対象外
[30]	テキストマイニングハンドブック	2010	6	対象外 (書籍)	対象外
[31]	教員養成初期段階の学生のティーチング・ポートフォリオのテキストマイニング分析: INTASC 観点 「コミュニケーション」 に関するリフレクションの記述から	2009	6	教育	報告書
[32]	医薬品使用の安全性に関するアンケートの解析 テキストマイニング手法の適用	2005	5	医学・生命科学	アンケート
[33]	テキスト型データのマイニング	2004	5	サーベイ	対象外
[34]	括弧表現に基づく Web テキストマイニングを用いた流行語への自動読み付与の提案	2009	4	手法	インターネット コミュニティ
[35]	インシデントレポート解析のための多重接続型階層的テキストマイニング手法の提案	2006	4	医学・生命科学	報告書
[36]	実務実習事前教育に向けたテキストマイニング手法の活用	2008	5	医学・生命科学	アンケート
[37]	言語情報を利用したテキストマイニング	2002	6	手法	対象外
[38]	ダンス学習の動機づけに関するテキストマイニング分析: 中学生の 「現代的なリズムのダンス」 の授業を事例として	2013	4	教育	アンケート
[39]	統合環境 TETDM を用いたマイニングツールの開発と利用の実践	2014	9	ツール・システム	対象外
[40]	Web マイニングを用いた因果ネットワークの自動構築手法の開発	2006	11	手法	インターネット コミュニティ
[41]	希死念慮者へのメッセージにみる, 自殺予防に対する医師の説明モデル - テキストマイニングによる分析	2009	6	医学・生命科学	アンケート
[42]	質的データ分析におけるグラウンデッドなテキストマイニング・アプローチの提案- がん告知の可否をめぐるフォーカスグループでの議論の分析から-	2011	4	医学・生命科学	アンケート
[43]	大学生のもつ 「死」 のイメージ: テキストマイニングによる分析	2003	6	社会調査	アンケート
[44]	調査における自由回答データの解析-InfoMiner による探索的テキスト型データ解析-	2000	25	社会調査 (システム)	アンケート

[45]	テキストマイニングの活用	2000	4	ツール・システム	対象外
[46]	テキストマイニングと強化学習を用いた電子メール自動分配	2004	5	手法	電子メール
[47]	テキストマイニング	2003	3	サーバイ	対象外
[48]	テキストマイニング技術を用いたソフトウェアの類似性分析	2010	3	ソフトウェア	ソフトウェア (ソースコード、関連文書、設計書、ログ情報)
[49]	これからの看護研究- 基礎と応用	2000	6	対象外 (書籍)	対象外
[50]	テキストマイニングによる農産物に対する消費者ニーズの把握	2010	3	農業	アンケート
[51]	産学連携のマッチング性分析におけるテキストマイニングの有効性	2009	3	特許	論文
[52]	顧客の声マネジメント: テキストマイニングで本音を「見る」	2007	4	対象外 (書籍)	対象外
[53]	半構造データマイニングにおけるパターン発見技法	2004	12	手法	対象外
[54]	形態素解析とチャンキングの組み合わせによる日本語テキスト中の未知語出現箇所同定	2003	7	手法	対象外
[55]	大規模テキストマイニングによる医学・生命科学分野の社会課題・技術トレンド抽出	2007	3	医学・生命科学 (ツール・システム)	新聞記事
[56]	テキストマイニングによる学部ゼミナールの魅力・不満の検討	2012	4	教育	アンケート
[57]	ネットワーク分析を用いたテキスト・マイニング: アスベスト問題の新聞報道について	2007	3	医学・生命科学	新聞記事
[58]	テキストマイニング・類似文書検索システムによる EBM に基づく診療ガイドライン作成支援の有効性	2004	3	医学・生命科学 (ツール・システム)	論文
[59]	Wikipedia マイニングによる大規模 Web オントロジの実現	2008	6	手法	インターネット コミュニティ

[60]	ユーザ興味を反映した情報推薦のための潜在的ディリクレ配分法を用いた協調フィルタリング	2014	3	手法	インターネット コミュニティ
[61]	ライフサイエンス分野におけるテキストマイニング技術適用の動向	2003	3	医学・生命 科学（サー ベイ）	論文
[62]	テキストマイニングによる知財ポートフォリオ分析	2008	4	特許	特許文書
[63]	自由記述データを用いたテキストマイニングによる都市のイメージ分析	2012	3	都市開発	アンケート
[64]	テキストマイニングを用いた故障報告書分析手法の研究	2002	4	船舶	報告書
[65]	意思決定支援のための主観情報マイニング	2006	5	ツール・シ ステム	対象外
[66]	半構造化テキストの分類のためのブースティングアルゴリズム	2004	22	手法	対象外
[67]	ランダムフォレスト法による文章の書き手の同定	2007	15	書誌学	書籍（小説、日 記、作文）
[68]	データマイニング技術を応用したソフトウェア構築・保守支援の研究動向	2010	7	ソフトウェ ア	ソフトウェア （ソースコー ド、関連文書、 設計書、ログ情 報）
[69]	褥そう対策の成果についての自由記入分析結果	2008	5	介護	アンケート
[70]	テキストデータマイニングのための統合環境: TETDM プロジェクト	2011	2	ツール・シ ステム	対象外
[71]	WWW とテキストマイニングの統合による問題構造可視化支援	1999	2	手法	インターネット コミュニティ
[72]	グラフ分析を利用した文書集合からの話題構造マイニング	2007	8	手法	新聞記事
[73]	データ・テキストマイニングの最新動向: 外れ値検出と評判分析を例に	2002	3	ツール・シ ステム	対象外
[74]	中国特許解析・テキストマイニングによる KW 分析	2014	2	特許	特許文書
[75]	大学生における「就職しないこと」イメージの構造と進路未決定: テキストマイニングを用いた検討	2009	3	教育	アンケート
[76]	Twitter テキストマイニングによる経済動向分析	2013	2	経済	インターネット コミュニティ
[77]	文脈マイニングモデルを用いた討議過程の可視化手法に関する研究	2011	5	都市開発	報告書

[78]	拡張固有表現タグ付きコーパスの構築	2008	8	対象外	対象外
[79]	学生の看護研究抄録におけるテキストマイニング法の検討	2004	3	看護	論文
[80]	文書クラスタリングによるトピック抽出および課題発見	2008	17	手法	新聞記事
[81]	テキストマイニングを用いた災害状況イメージネーション支援システムの構築に関する研究	2005	2	災害	アンケート
[82]	フェルデンクライス・メソッドによる身体知覚変容に関する一考察～身体感覚の言語化についてのテキストマイニング分析～	2012	2	体育学	アンケート
[83]	テレビ CM に対する視聴者反応の分析ー自由回答文のテキストマイニング	2005	3	テレビ CM	アンケート
[84]	大規模ウェブテキストからの片仮名用言の自動獲得	2009	5	手法	インターネット コミュニティ
[85]	金融テキストマイニング研究の紹介	2012	2	経済	インターネット コミュニティ
[86]	概念抽出型テキストマイニングによるアンケート分析手法の提案	2003	3	手法	アンケート
[87]	マーケティングミックスとテキストマイニングを用いた市場分析支援	2010	2	経営	アンケート
[88]	テキストマイニングによる薬学生実務実習レポートの分析	2010	2	教育	報告書
[89]	blog の自動収集と監視	2004	30	手法	インターネット コミュニティ
[90]	産後 1 ヶ月の母親のストレスの本質の探索: テキストマイニング分析によるストレス内容の結びつきから	2013	4	対象外	対象外
[91]	テキストマイニングと NLP ビジネス	2003	2	対象外 (発表資料)	対象外
[92]	感情表現の抽出手法に関する提案	2004	6	手法	インターネット コミュニティ
[93]	ナレッジマネジメントにおけるテキストマイニング	2006	2	ナレッジマネジメント	インターネット コミュニティ
[94]	テキストマイニングによる退院時サマリーからの時系列関係知識抽出	2007	3	医学・生命科学	報告書
[95]	テキストマイニングを用いた株価予測	2010	2	対象外	対象外
[96]	フランス地方議会の審議項目のテキストマイニング分析	2010	2	政治	地方議会の審議 項目
[97]	看護におけるテキストマイニングとその活用事例	2008	2	看護	対象外

[98]	言語統計分析による住宅建設五箇年計画及び答申の特性分析: 政策の立案と評価における非定型・大量情報の活用可能性	2004	8	都市開発	報告書
[99]	フィールドの魅力を掘り起こすフィールドマイニング	2008	8	対象外	対象外
[100]	語横断テキストマイニング	2010	2	手法	報告書

3.3 英語文献の調査結果

100 件の英語文献に対する調査結果を以下の表に示す。表のフォーマットは 3.2 節に示した日本語文献の調査結果の表と同じである。

参考文献	タイトル	発行年	被引用数	トピック	メディア
[101]	GENIA corpus—a semantically annotated corpus for bio-textmining	2003	775	医学・生命科学	論文
[102]	Text mining: The state of the art and the challenges	1999	535	サーベイ	対象外
[103]	Document warehousing and text mining: techniques for improving business operations, marketing, and sales	2001	359	対象外 (書籍)	対象外
[104]	Survey of text mining	2004	619	サーベイ	対象外
[105]	Fast and effective text mining using linear-time document clustering	1999	918	手法・技術	対象外
[106]	The text mining handbook: advanced approaches in analyzing unstructured data	2007	1484	対象外 (書籍)	対象外
[107]	survey of current work in biomedical text mining	2005	662	医学・生命科学 (サーベイ)	論文
[108]	Mining text data	2012	323	対象外 (書籍)	対象外
[109]	Discovering evolutionary theme patterns from text: an exploration of temporal text mining	2005	524	災害	ウェブニュース 論文
[110]	Maximizing text-mining performance	1999	290	手法・技術	電子メール
[111]	Text mining for biology and biomedicine	2006	277	対象外 (書籍)	対象外

[112]	A text-mining-based patent network: Analytical tool for high-technology trend	2004	385	特許	特許文書
[113]	A brief survey of text mining.	2005	517	サーベイ	対象外
[114]	Text mining application programming	2006	155	対象外（書籍）	対象外
[115]	A text-mining analysis of the human phenome	2006	365	医学・生命科学	論文
[116]	Tapping the power of text mining	2006	258	サーベイ	対象外
[117]	What is text mining	2003	166	サーベイ	対象外
[118]	Text mining: finding nuggets in mountains of textual data	1999	206	サーベイ	対象外
[119]	Text mining infrastructure in R	2008	463	ツール・システム	対象外
[120]	Text mining: predictive methods for analyzing unstructured information	2010	562	対象外（書籍）	対象外
[121]	Text mining techniques for patent analysis	2007	477	特許	特許文書
[122]	Text mining at the term level	1998	244	手法・技術	新聞記事
[123]	Practical text mining and statistical analysis for non-structured text data applications	2012	164	対象外（書籍）	対象外
[124]	Text mining and ontologies in biomedicine: making sense of raw text	2005	271	医学・生命科学	論文
[125]	Text mining and its potential applications in systems biology	2006	281	医学・生命科学	論文
[126]	Getting started in text mining	2008	207	医学・生命科学	論文
[127]	Mining knowledge from text using information extraction	2005	254	手法・技術	論文 求人票 製品説明
[128]	A cross-collection mixture model for comparative text mining	2004	226	手法・技術	新聞記事 インターネット コミュニティ
[129]	Visualizing association rules for text mining	1999	191	手法・技術	コーパス
[130]	Text mining: generating hypotheses from MEDLINE	2004	303	医学・生命科学	論文

[131]	MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling	1999	263	医学・生命科学	論文
[132]	Text mining with decision rules and decision trees	1998	144	手法・技術	対象外
[133]	[HTML] Text-mining and information-retrieval services for molecular biology	2005	183	医学・生命科学	論文
[134]	Text analysis and knowledge mining system	2001	245	経営	コールセンターの記録
[135]	Natural language processing and text mining	2007	155	対象外（書籍）	対象外
[136]	Frontiers of biomedical text mining: current progress	2007	242	医学・生命科学（サーベイ）	対象外
[137]	Integration of text-and data-mining using ontologies successfully selects disease gene candidates	2005	210	医学・生命科学	論文
[138]	Text Mining	2010	138	対象外（書籍）	対象外
[139]	Text mining with information extraction	2002	195	手法・技術	対象外
[140]	Facts from text—is text mining ready to deliver?	2005	149	医学・生命科学（サーベイ）	対象外
[141]	Forecasting intraday stock price trends with text mining techniques	2004	215	経済	プレスリリース
[142]	Text Mining Using Non-Negative Matrix Factorizations.	2004	184	手法・技術	対象外
[143]	BioCaster: detecting public health rumors with a Web-based text mining system	2008	153	公衆衛生	ウェブサイト・ニュース
[144]	Method and system for text mining using multidimensional subspaces	2003	240	特許	特許文書
[145]	PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites	2008	171	医学・生命科学（ツール・システム）	対象外
[146]	Using text mining and sentiment analysis for online forums hotspot detection and forecast	2010	208	ホットスポット検出	インターネットコミュニティ

[147]	Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup	2003	170	医学・生命科学 (ツール・システム)	対象外
[148]	Effective pattern discovery for text mining	2012	201	手法	対象外
[149]	Linking genes to literature: text mining, information extraction, and retrieval applications for biology	2008	184	医学・生命科学	論文
[150]	Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge	2008	166	医学・生命科学 (ツール・システム)	対象外
[151]	Deploying approaches for pattern refinement in text mining	2006	127	手法	対象外
[152]	Event extraction for systems biology by text mining the literature	2010	169	医学・生命科学	論文
[153]	PubMeth: a cancer methylation database combining text-mining and expert annotation	2008	141	医学・生命科学 (ツール・システム)	対象外
[154]	Text-mining approaches in molecular biology and biomedicine	2005	135	医学・生命科学 (サーベイ)	対象外
[155]	Hybrid text mining for finding abbreviations and their definitions	2001	139	手法	技術書 プレスリリース
[156]	Text mining: natural language techniques and text mining applications	1998	129	手法	コーパス
[157]	Discovering interesting usage patterns in text collections: integrating text mining with visualization	2007	140	ツール・システム	書籍
[158]	Improved feature selection approach TFIDF in text mining	2002	98	手法	CMU テキストデータセット
[159]	Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach	2011	188	経営	インターネット トコミュニティ
[160]	Text mining applied to patent mapping: a practical business case	2003	130	特許	特許文書

[161]	Social media competitive analysis and text mining: A case study in the pizza industry	2013	175	経営	インターネット トコミュニテ イ
[162]	Gender-preferential text mining of e-mail discourse	2002	134	手法	電子メール
[163]	Text mining for metabolic pathways, signaling cascades, and protein networks	2005	95	医学・生命科 学	論文
[164]	Mining text data: special features and patterns	2000	113	対象外（書 籍）	対象外
[165]	Sentiment analysis and opinion mining	2012	1708	対象外（書 籍）	対象外
[166]	U-Compare: share and compare text mining tools with UIMA	2009	88	医学・生命科 学（ツール・ システム）	対象外
[167]	Text and structural data mining of influenza mentions in web and social media	2010	132	公衆衛生	インターネット トコミュニテ イ
[168]	GIS: a biomedical text-mining system for gene information discovery	2004	73	医学・生命科 学（ツール・ システム）	対象外
[169]	Automatic keyword extraction from individual documents	2010	112	手法	ウェブニュー ス
[170]	Citation mining: Integrating text mining and bibliometrics for research user profiling	2001	116	書誌学	書籍
[171]	Text mining: Classification, clustering, and applications	2009	112	対象外（書 籍）	対象外
[172]	Anni 2.0: a multipurpose text-mining tool for the life sciences	2008	116	医学・生命科 学（ツール・ システム）	対象外
[173]	Text mining and visualization tools—impressions of emerging capabilities	2008	114	ツール・シス テム	対象外
[174]	Combining data and text mining techniques for analysing financial reports	2004	119	経済	報告書
[175]	Mine your own business: Market-structure surveillance through text mining	2012	162	経営	インターネット トコミュニテ イ

[176]	Knowledge Management: A Text Mining Approach.	1998	107	経営	新聞記事
[177]	More than words: Social networks' text mining for consumer brand sentiments	2013	134	経営	インターネット トコミュニテ イ
[178]	Text mining for causal relations.	2002	146	手法	新聞記事
[179]	Knowledge management systems: a text mining perspective	2001	122	対象外 (書 籍)	対象外
[180]	MeSHmap: a text mining tool for MEDLINE.	2001	87	医学・生命科 学 (ツール・ システム)	対象外
[181]	Text mining for product attribute extraction	2006	128	経営	ウェブサイト
[182]	Method for data and text mining and literature-based discovery	2005	178	手法	論文
[183]	Yale: Rapid prototyping for complex data mining tasks	2006	1106	ツール・シス テム	対象外
[184]	A text-mining system for knowledge discovery from biomedical documents	2004	107	医学・生命科 学 (ツール・ システム)	対象外
[185]	Overview and semantic issues of text mining	2007	113	サーベイ	対象外
[186]	Text mining	2000	63	サーベイ	対象外
[187]	Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism	2010	111	医学・生命科 学	論文
[188]	Partially labeled topic models for interpretable text mining	2011	113	手法	対象外
[189]	CREATING KNOWLEDGE REPOSITORIES FROM BIOMEDICAL REPORTS: THE MEDSYNDIKATE TEXT MINING SYSTEM	2001	116	医学・生命科 学 (ツール・ システム)	対象外
[190]	PubTator: a web-based text mining tool for assisting biocuration	2013	92	医学・生命科 学	論文
[191]	Translation of web queries using anchor text mining	2002	93	手法	ウェブサイト
[192]	OSCAR4: a flexible architecture for chemical text-mining	2011	102	化学	論文
[193]	Overview of Mondou web search engine using text mining and information visualizing technologies	2000	85	ツール・シス テム	対象外

[194]	Assisted Curation: Does Text Mining Really Help?	2008	93	医学・生命科学 (ツール・システム)	対象外
[195]	Text mining using database tomography and bibliometrics: A review	2001	118	書誌学	書籍
[196]	Biomedical text mining and its applications	2009	91	医学・生命科学 (サーベイ)	対象外
[197]	GTP (General Text Parser) software for text mining	2003	83	ツール・システム	対象外
[198]	Text mining-knowledge extraction from unstructured textual data	1998	74	対象外 (書籍)	対象外
[199]	Text mining for biology-the way forward: opinions from leading scientists	2008	98	医学・生命科学	論文
[200]	Exploring text mining from MEDLINE	2002	81	医学・生命科学	論文

3.4 考察

3.2 節、3.3 節の調査結果を元に、テキストマイニングがどのようなトピック・メディアのテキストに応用されてきたか、年によって論文数にどのような変化があるか、考察を行う。なお、調査対象の文献を Google Scholar のみを用いて選定しているため、対象文献にある程度の偏りが生じており、それに対する考察も正確性を欠いている可能性がある点は注意が必要である。例えば、3.1 節で紹介したように、Google Scholar では「それぞれの記事の全文、著者、記事が掲載された出版物、他の学術資料に引用された回数が考慮され」て掲載順位が決まるため、最近の論文は上位に現われにくい可能性がある。調査対象論文に偏りがないようにするために、複数の検索エンジンや論文データベースを用いて対象論文を選定する必要があるが、これは今後の課題としたい。

3.4.1 トピックに関する考察

トピックの個々のカテゴリについて、そのカテゴリに分類された日本語文献、英語文献の数を表 1 に示す。以下、テキストマイニングの文献に見られるトピックの傾向を日本語文献と英語文献に分けて考察する。

表 1 トピックの分布

トピック	日本語	英語
医学・生命科学	10	32
介護	2	0
化学	0	1
看護	2	0
教育	4	0
経営	2	7
経済	6	2
公衆衛生	0	2
コールセンター	1	0
災害	1	1
社会調査	3	0
書誌学	1	2
政治	1	0
船舶	1	0
ソフトウェア	2	0
体育	1	0
テレビCM	1	0
都市開発	4	0
特許	4	4
ナレジマネジメント	1	0
農業	2	0
ホットスポット検出	0	1
テキストマイニングの手法・技術	22	20
テキストマイニングのツール・システム	13	6
テキストマイニングのサーベイ	6	8
対象外（書籍、発表資料etc）	10	14

日本語文献では、特定のトピックへの偏りが見られず、多様なトピックのテキストを対象にテキストマイニングの研究が行われていることが分かる。出現頻度が最大のトピックは「医学・生命科学」の10件で、全体の10%である。また、1つの文献しか該当しないカテゴリが8件、2つの文献しか該当しないカテゴリが5件存在する。これは英語文献と比べると件数が多い。例えば、「災害」のカテゴリに該当するのは文献[81]のみである。この文献では、災害時の状況を想起するためにテキストマイニングを活用している。「政治」のカテゴリに該当するのは文献[96]のみである。文献[96]では、フランスの地方議会の審議項目(議案)の一覧に対し、テキストマイニングの技術を用いてその内容を分析し、最終的に日本国内の地方議会の議案との比較を試みている。2つの文献しか該当しないカテゴリの例は「ソフトウェア」である。文献[48]と[68]はともにソフトウェアを分析の対象としている。

英語文献では、「医学・生命科学」のカテゴリに分類された文献が32件あり、医学のテキストに対するテキストマイニングが盛んに研究されている傾向がうかがえる。日本語文献でも「医学・生命科学」は最多であったが、件数は10件であり、英語では日本語と比べて3.2倍の「医学・生命科学」分野の文献が存在している。また、「医学・生命科学」に関するテキストをマイニングする手法や技術を提案する論文だけでなく、同分野のサーベイ論文(文献[136][140][154][196])や、「医学・生命科学」分野のデータベースを活用するのに特化したツールやシステムに関する論文(文献[145][147][150][153][166][172][180][184][189][194])が数多く存在する。例えば、文献[145]のPolySearchは、科学論文や生命科学のデータベースから必要な情報を検索するためのツールである。PolySearchの典型的なクエリは以下のようなになる。

Given X, find all Y's' where X or Y can be diseases, tissues, cell compartments, gene/protein names, SNPs, mutations, drugs and metabolites.

文献[153]のPubMethは、テキストマイニングの技術を使って抽出した情報をもとに各種がんの遺伝子のメチル化頻度を調べることができるツールで、遺伝子をクエリとする検索とがんの種類をクエリとする検索が可能である。一方、文献が1つしか該当しないカテゴリは3つ、文献が2つしか該当しないカテゴリは3つで、日本語文献と比べて少ない。このことから、英語文献は日本語文献に比べて、扱うテキストのトピックに偏りが見られると言える。

3.4.2 メディアに関する考察

メディアの個々のカテゴリについて、そのカテゴリに分類された日本語文献、英語文献の数を表2に示す。以下、テキストマイニングの文献に見られるメディアの傾向を日本語文献と英語文献に分けて考察する。

表 2 メディアの分布

メディア	日本語	英語
CMUテキストデータセット	0	1
ウェブサイト・ニュース	0	5
アンケート	21	0
インターネットコミュニティ	17	7
コーパス	0	2
コールセンターの記録	1	1
書誌	1	3
新聞記事	6	4
ソフトウェア	3	0
電子メール	1	2
特許文書	3	4
報告書	10	1
落書き帳	1	0
論文	4	21
地方議会の審議項目	1	0
プレスリリース	0	1
有価証券報告書	1	0
求人票	0	1
説明書・技術書	0	2
対象外	30	49

日本語文献では、「アンケート」のカテゴリに属する文献が 21 件と最多で、「インターネットコミュニティ」17 件、「報告書」10 件と続いていることが分かる。これは日本語文献のトピックが多岐に渡っていることと関連があるだろう。アンケートやインターネットコミュニティ、報告書には様々なトピックのテキストが存在すると考えられるからである。

英語文献では、「論文」のカテゴリに属する文献が 21 件と最多であるが、それ以外に 10 件以上のカテゴリはなく、研究対象が「論文」に集中していることが分かる。これはトピックで文献を分類したとき「医学・生命科学」のカテゴリが最多となっていることと関連するだろう。すなわち、英語文献では、医学・生命科学のデータベースに収録されている論文・要旨(アブストラクト)をテキストマイニングの対象とした研究が多い。英語では、医学・生命科学のオンラインデータベース (MEDLINE / PubMed) が一般的に利用されており、それをいかに有効活用するかが大きなテーマとなっている。近年では、年間 50 万件の医学・生命科学の論文が新規に発行されている [61]。膨大に膨れ上がる論文情報から、テキストマイニングの技術を使い、効率的に必要な情報、知識を発見するための研究が求められていることが分かる。

3.4.3 医学・生命科学を対象としたテキストマイニングに関する考察

表 1 より、医学・生命科学の分野は、日本語、英語ともに最多のトピックであるが、

その研究内容は大きく異なっている。

日本語文献では、アンケートの自由記述欄や報告書のテキストを分析の対象として、何らかの評価や有意義な情報・知識を引き出すことを目的とするものが多い。具体的には、評価医薬品使用の安全性に関するアンケートを解析して安全対策を評価する試み[32]、病院内のインシデントレポートを解析し医療事故を防止する試み[35]、退院時サマリーから自動的に疾患分類をする試み[94]などがある。英語文献で数多く見られる PubMed / MEDLINE 形式のデータベースを活用した日本語文献は文献[58]のみである。この文献は、IBM が開発する MedTAKMI を利用して、診療ガイドラインの作成を支援することを目的としている。

英語論文では、前述のとおり、PubMed / MEDLINE に収められているデータ・情報を効率的に取り扱う研究が多く行われている。例えば、Kim らは、MEDLINE の要旨(アブストラクト)から、論文アブストラクトに専門用語情報等をタグ付けした GENIA Corpus というコーパスを作っている[101]。GENIA corpus version 3.0 は MEDLINE 上の 2,000 のアブストラクトからなり、400,000 語を収録する。

3.4.4 年代による傾向の考察

本項では、調査対象とした文献(日本語 100 件、英語 100 件)の発行年が年によってどのように変化しているかを考察する。

表 3 は、1998 年から 2014 年の各年において、その年に発行された日本語文献の数と英語文献の数を示している。また、図 1 は、表 3 の調査結果をグラフで表わしたものである。日本語文献については 2004 年の 13 件、英語文献については 2005 年と 2008 年の 11 件が最多となっており、これらの年にテキストマイニングの研究が盛んに行われていたことが分かる。また、その後現在に至るまで、テキストマイニングの論文は減少傾向にあることが分かる。

図 1 のグラフに示されているように、文献数は一年毎に増減を繰り返すこともある。これは、テキストマイニングが盛んな時期でも、発行年によって文献数にばらつきがあることを意味する。発行年によるばらつきの影響を軽減し、文献数の年毎の変化をより明確に調べるため、年毎に、その年の前後 2 年、合計 5 年間の文献数を調べた。

表 4 は 5 年間に発行された日本語と英語の論文数を示している。今回の調査対象の文献の発行年は 1998 年から 2014 年までであるが、ここでは前後を含めて 5 年間に発行された論文数を調査しているため、5 年分の論文数を集計できない 1998 年と 1999 年、および 2013 年と 2014 年は対象からはずし、2000 年から 2012 年までの論文数を調べた。また、図 2 は表 4 の結果をグラフで表わしたものである。これらから、英語文献については、2006 年を中心とした 5 年間(2004 年～2008 年)で発行された論文数 45 件がピークとなり、その後論文数は減っている。一方、日本語文献については、2006 年(40 件)、2007 年(35 件)、2008 年(41 件)、2009 年(41 件)、2010 年(40 件)付近で文献

数が最も多く、比較的長い期間にテキストマイニング研究のピークが存在したと言える。その後、現在に至るまで論文数は減少している。

2006年及びその前後2年間に発行された文献のトピックの分布を確認すると、英語文献では、45件の文献が発行されており、その内「医学・生命科学」をトピックとするものが21件で49%を占めている。また、「医学・生命科学」をトピックとする論文の総数は32件なので、全体の66%がこの5年間に発行されたことになる。日本語文献についても、「医学・生命科学」をトピックとする全10件の文献のうち7件(70%)がこの期間に発行されていることが分かる。

表 3 発行年と論文数

年	日本語	英語
1998	0	5
1999	3	6
2000	5	3
2001	0	8
2002	4	6
2003	8	6
2004	13	9
2005	4	11
2006	5	8
2007	7	6
2008	11	11
2009	8	3
2010	10	7
2011	5	3
2012	6	5
2013	6	3
2014	5	0

図 1 発行論文数の年による変化

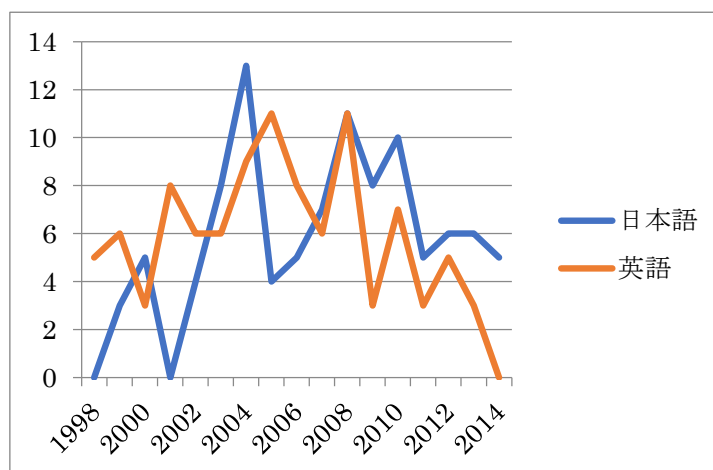
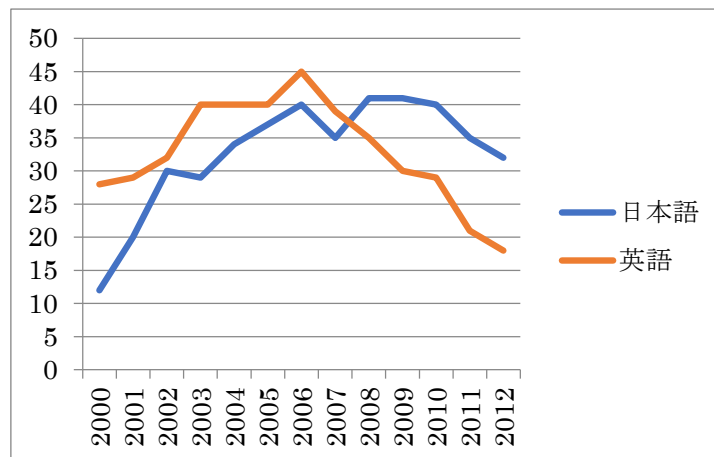


表 4 5年以内に発行された論文数

年	日本語	英語
2000	12	28
2001	20	29
2002	30	32
2003	29	40
2004	34	40
2005	37	40
2006	40	45
2007	35	39
2008	41	35
2009	41	30
2010	40	29
2011	35	21
2012	32	18

図 2 5年以内に発行された論文数の変化



3.4.5 被引用数の比較

表 5は、日本語文献100件と英語文献100件の被引用数の和を比較したものである。英語文献の被引用数は日本語文献の被引用数のおよそ30倍である。一般に、日本語で書かれた論文は日本の研究者しか引用しないのに対し、英語で書かれた論文は世界中の研究者が引用するため、被引用数もはるかに多い。テキストマイニングを対象とした論文に限ったときも、被引用数について同様の傾向が見られることが確認された。

表 5 調査対象の日本語・英語論文の被引用数の比較

	日本語	英語
被引用数	839	24,683

第4章 おわりに

本課題研究では、テキストのトピックとメディアに着目してテキストマイニングの研究動向の分析を行った。日本語文献では、特定のトピックへの偏りが見られず、多様なトピックのテキストを対象にテキストマイニングの研究が行われていること、メディアとしては「アンケート」「インターネットコミュニティ」「報告書」を対象としていることが多いことが分かった。一方、英語文献では、「医学・生命科学」のトピックに分類された文献が多く、メディアとしては文献データベースを含む「論文」を対象としているものが多いことが分かった。「医学・生命科学」の分野では、年間50万件の新規論文が投稿されていると言われており、日々新しく登録される論文やその中に含まれる遺伝子情報やタンパク質の情報から、新たな知見を効率よく発見する手法が求められているためと考えられる。

当初の想定では、ビッグデータと呼ばれるようなインターネット上に膨大にあるテキストから有意義な知見を自動的に獲得することを目的とした研究が多く、経済、経営、社会調査などの分野でのテキストマイニングの活用を想定していた。実際は、日本語文献では想定していた「インターネットコミュニティ」を対象とした研究も多かった一方、「アンケート」「報告書」といったウェブ以外のテキストを対象として多様なトピックを取り扱う研究が多かった。英語文献では、「医学・生命科学」のデータベースを活用したテキストマイニングの研究が主流であることが分かった。英語文献は日本語文献に比べてけた違いに多いので（調査時点における Google Scholar の検索で、日本語 3,010 件に対し、英語は 2,390,000 件）、対象とする英語文献の数を増やせば、様々なトピックやメディアを対象とした研究が存在する可能性は高いと思われる。

今後の課題としては、英語文献の調査量を増やして先行研究が対象としたテキストのトピックをより正確に調べることが挙げられる。また、各論文の内容を精査して、テキストマイニングに使われている要素技術を調べ、どの要素技術がよく使われているかを整理することも重要な課題である。テキストマイニングの要素技術についての包括的な知見が得られれば、新しい技術を開発したり、これまで使われていなかった技術を新たに適用したりするなど、テキストマイニング研究の更なる発展につながると考えられる。また、年代毎のトピックとメディアの件数の変化を調査することは、過去のテキストマイニング研究のトレンドを把握するのに役立つと考えられる。一方、既にテキストマイニングのツールはいくつか販売されており、それらを実際に活用した事例も多い。既存のテキストマイニングのツールの特徴とそれらの活用事例を整理することや、更には実際にツールを使って例えば経済動向の分析を試みることも今後取り組みたい課題である。

謝辞

本課題研究を進めるにあたり、多大なご支援、ご指導をいただいた白井清昭准教授に深く御礼申し上げます。中間審査の場において貴重なご意見をいただいた東条敏教授、池田心准教授に深く感謝いたします。副テーマ指導教官としてご指導いただいた鶴木祐史教授に心より御礼申し上げます。最後に、北陸先端科学技術大学院大学先端領域社会人教育院関係者各位に深く御礼申し上げます。

参考文献

- [1] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. (2003). テキストマイニングによる評価表現の収集. 情報処理学会研究報告自然言語処理 (NL), 2003(23 (2002-NL-154)), 77-84.
- [2] 林俊克. (2002). Excel で学ぶテキストマイニング入門. 株式会社 オーム社.
- [3] 奥村学, 南野朋之, 藤木稔明, 鈴木泰裕. (2004). blog ページの自動収集と監視に基づくテキストマイニング. 人工知能学会, セマンティックウェブとオントロジー研究会, SIG-SWO-A401-01.
- [4] 倉島健, 藤村考, 奥田英範. (2009). 大規模テキストからの経験マイニング. 電子情報通信学会論文誌 D, 92(3), 301-310.
- [5] 高田哲司, 小池英樹. (2000). 見えログ: 情報視覚化とテキストマイニングを用いたログ情報ブラウザ. 情報処理学会論文誌, 41(12), 3265-3275.
- [6] 樋口耕一. (2004). テキスト型データの計量的分析. 理論と方法, 19(1), 101-115.
- [7] 渡部勇, 三末和男. (1999). 単語の連想関係によるテキストマイニング. 情報処理学会研究報告情報学基礎 (FI), 1999(57 (1999-FI-055)), 57-64.
- [8] 小林祐司, 寺田充伸, & 佐藤誠治. (2012). テキストマイニングを活用したアンケートにおける自由回答の分析と生活環境評価. 日本建築学会計画系論文集, 77(671), 85-93.
- [9] 喜田昌樹. (2008). 組織革新の認知的研究: 認知変化・知識の可視化と組織科学へのテキストマイニングの導入. 認知科学, 15(4), 723-724.
- [10] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, & 梶並知記. (2013). テキストデータマイニングのための統合環境 TETDM の開発. 人工知能学会論文誌, 28(1), 1-12.
- [11] 内田治, 川嶋敦子, & 磯崎幸子. (2012). SPSS によるテキストマイニング入門. 株式会社 オーム社.
- [12] 安藤俊幸. (2009). テキストマイニングと統計解析言語 R による特許情報の可視化. 情報管理, 52(1), 20-31.
- [13] 奥村学. (2008). ブログマイニング技術の最新動向. 電子情報通信学会誌, 91(12), 1054-1059.
- [14] 鳩間亜紀子, 児玉桂子, 田村静子. (2004). 高齢者向け住宅改造の効果に関する介護専門職の評価指標と要介護反別特徴: テキストマイニングによる自由回答の分析. 社会福祉学, 45(2), 67-80.
- [15] 三末和男, 渡部勇. (1999). テキストマイニングのための連想関係の可視化技術. 情報処理学会研究報告デジタルドキュメント (DD), 1999(57 (1999-DD-019)), 65-72.
- [16] 渡部勇. (2003). テキストマイニングの技術と応用 (< 特集> 情報の分析・解析法).

情報の科学と技術, 53(1), 28-33.

- [17] 砂山渡, 高間康史, 西原陽子, 徳永秀和, 串間宗夫, 阿部秀尚, 梶並知記. (2013). テキストデータマイニングのための統合環境 TETDM の開発. 人工知能学会論文誌, 28(1), 1-12.
- [18] 入江拓, 小平朋江. (2007). 看護大学生の精神科保護室に対する受け止めおよび視点の変化: テキストマイニングによる非構造型データの分析から. 聖隷クリストファー大学看護学部紀要, 15, 1-10.
- [19] 石川修, 星野敏. (2004). テキストマイニングを用いた都市農村交流ニーズの把握—岡山県吉永町ふるさと村の八塔寺山荘の落書き帳を対象として—. 農村計画学会誌, 23, 181-186.
- [20] 藏本貴久, 和泉潔, 吉村忍, 石田智也, 中嶋啓浩, 松井藤五郎, 吉田稔, 中川裕志. (2013). 新聞記事のテキストマイニングによる長期市場動向の分析. 人工知能学会論文誌, 28(3), 291-296.
- [21] 奥村学. (2013). ソーシャルメディアを対象としたテキストマイニング. 電子情報通信学会 基礎・境界ソサイエティ Fundamentals Review, 6(4), 285-293.
- [22] 秋元泰介, 小方孝. (2014). 統合物語生成システムの現状と特に物語内容生成メカニズム (言語生成・心的影響, 第 4 回テキストマイニング・シンポジウム). 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, 113(429), 27-32.
- [23] 長野徹, 武田浩一, 那須川哲哉. (2000). テキストマイニングのための情報抽出. 情報処理学会研究報告情報学基礎 (FI), 2000(91 (2000-FI-060)), 31-38.
- [24] 磯島昭代. (2006). テキストマイニングを用いた米に関する消費者アンケートの解析. 農業情報研究, 15(1), 49-60.
- [25] 林田英雄, 脇森浩志. (2005). テキストマイニング技術とその応用. UNISYS TECHNOLOGY REVIEW, (84), 29-44.
- [26] 和泉潔, 後藤卓, 松井藤五郎. (2010). テキスト情報による金融市場変動の要因分析. 人工知能学会論文誌, 25(3), 383-387.
- [27] 和泉潔, 後藤卓, 松井藤五郎. (2011). 経済テキスト情報を用いた長期的な市場動向推定. 情報処理学会論文誌, 52(12), 3309-3315.
- [28] 大野邦夫, 渡辺篤史. (2008). ソーシャルメディアへのテキストマイニングの適用に関する検討. 情報処理学会研究報告電子化知的財産・社会基盤 (EIP), 2008(10 (2008-EIP-039)), 47-54.
- [29] 小方孝, 小野淳平. (2014). 統合物語生成システム, 間テクスト性, テキストマイニング (言語生成・心的影響, 第 4 回テキストマイニング・シンポジウム). 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, 113(429), 33-38.
- [30] 辻井潤一. (2010). テキストマイニングハンドブック. 東京電機大学出版局
- [31] 谷塚光典, 東原義訓. (2009). 教員養成初期段階の学生のティーチング・ポートフォ

- リオのテキストマイニング分析: INTASC 観点「コミュニケーション」に関するリフレクションの記述から. 日本教育工学会論文誌, 33(Suppl.), 153-156.
- [32] 木村昌臣, 古川裕之, 塚本均, 田崎久夫, 空閑正浩, 大倉典子, 土屋文人. (2005). 医薬品使用の安全性に関するアンケートの解析 テキストマイニング手法の適用. 人間工学, 41(5), 297-305.
- [33] 大隅昇, 保田明夫. (2004). テキスト型データのマイニング. 理論と方法, 19(2), 135-159.
- [34] 三宅純平, 竹内翔大, 川波弘道, 猿渡洋, 鹿野清宏. (2009). 括弧表現に基づく Web テキストマイニングを用いた流行語への自動読み付与の提案. 電子情報通信学会技術研究報告. SP, 音声, 108(422), 1-6.
- [35] 岡部貴博, 吉川大弘, 古橋武. (2006). インシデントレポート解析のための多重接続型階層的テキストマイニング手法の提案. In 日本知能情報ファジィ学会 ファジィシステム シンポジウム 講演論文集 第 22 回ファジィ システム シンポジウム (pp. 54-54). 日本知能情報ファジィ学会.
- [36] 濃沼政美, 小池勝也, 中村均. (2008). 実務実習事前教育に向けたテキストマイニング手法の活用. YAKUGAKU ZASSHI, 128(6), 925-931.
- [37] 工藤拓, 山本薫, 坪井祐太, 松本裕治. (2002). 言語情報を利用したテキストマイニング. 情報処理学会研究報告自然言語処理 (NL), 2002(20 (2001-NL-148)), 65-72.
- [38] 内山須美子, 松尾健太, 奥山美希. (2013). ダンス学習の動機づけに関するテキストマイニング分析: 中学生の「現代的なリズムのダンス」の授業を事例として. 白鴎大学教育学部論集, 7(1), 71-108.
- [39] 砂山渡, 高間康史, 西原陽子, 梶並知記, 串間宗夫, 徳永秀和. (2014). 統合環境 TETDM を用いたマイニングツールの開発と利用の実践. 人工知能学会論文誌, 29(1), 100-112.
- [40] 佐藤岳文, 堀田昌英. (2006). Web マイニングを用いた因果ネットワークの自動構築手法の開発. 社会技術研究論文集, 4, 66-74.
- [41] 川島大輔, 小山達也, 川野健治, 伊藤弘人. (2009). 希死念慮者へのメッセージにみる, 自殺予防に対する医師の説明モデル. パーソナリティ研究, 17(2), 121-132.
- [42] 稲葉光行, 抱井尚子. (2011). 質的データ分析におけるグラウンデッドなテキストマイニング・アプローチの提案—がん告知の可否をめぐるフォーカスグループでの議論の分析から—. 政策科学, 18(3), 255-276.
- [43] 藤井美和. (2003). 大学生のもつ「死」のイメージ: テキストマイニングによる分析. 関西学院大学社会学部紀要, 95, 145-155.
- [44] 大隅昇, Lebart, L. (2000). 調査における自由回答データの解析—InfoMiner による探索的テキスト型データ解析—. 統計数理, 48(2), 339-376.
- [45] 吉田稔, 中川裕志. (2010). テキストマイニングの活用 (< 特集> データマイニン

- グの活用). 情報の科学と技術, 60(6), 230-235.
- [46] 上田芳弘, 成田仁志, 加藤直孝, 林克明, 南保英孝, 木村春彦. (2004). テキストマイニングと強化学習を用いた電子メール自動分配. 電子情報通信学会論文誌 D, 87(10), 887-898.
- [47] 野村義明. (2003). テキストマイニング. ヘルスサイエンス・ヘルスケア, 3(1), 42-43.
- [48] 岸本康成, 坂本啓, 佐藤宏之, 小林透. (2010). テキストマイニング技術を用いたソフトウェアの類似性分析. 電子情報通信学会技術研究報告. SS, ソフトウェアサイエンス, 109(456), 79-84.
- [49] 松木光子, 小笠原知枝. (2000). これからの看護研究- 基礎と応用. 廣川 書店.
- [50] 磯島昭代. "テキストマイニングによる農産物に対する消費者ニーズの把握." フードシステム研究 16.4 (2010): 38-42.
- [51] 山本外茂男. (2009). 産学連携のマッチング性分析におけるテキストマイニングの有効性. 情報の科学と技術, 59(6), 291-297.
- [52] 三室克哉, 鈴木賢治, 神田晴彦. (2007). 顧客の声マネジメント: テキストマイニングで本音を「見る」. 株式会社 オーム社.
- [53] 浅井達哉, 有村博紀. (2004). 半構造データマイニングにおけるパターン発見技法. 電子情報通信学会論文誌 D, 87(2), 79-96.
- [54] 浅原正幸, 松本裕治. (2003). 形態素解析とチャンキングの組み合わせによる日本語テキスト中の未知語出現箇所同定. 情報処理学会研究報告自然言語処理 (NL), 2003(23 (2002-NL-154)), 47-54.
- [55] 内海和夫, 乾孝司, 村上浩司, 橋本泰一, 石川正道. (2007). 大規模テキストマイニングによる医療分野の社会課題・技術トレンド抽出. 研究・技術計画学会第 22 回年次学術大会, 684-687.
- [56] 伏木田稚子, 北村智, 山内祐平. (2012). テキストマイニングによる学部ゼミナールの魅力・不満の検討. 日本教育工学会論文誌, 36(Suppl.), 165-168.
- [57] 高橋由光, 宮木幸一, 新保卓郎, 中山健夫. (2007). ネットワーク分析を用いたテキスト・マイニング: アスベスト問題の新聞報道について. 医療情報学, 27(1), 83-89.
- [58] 佐藤康仁, 竹内広宜, 星佳芳, 浦本直彦, 佐藤敏彦, 稲岡則子, 武田浩一, 山口直人. (2004). テキストマイニング・類似文書検索システムによる EBM に基づく診療ガイドライン作成支援の有効性. 医療情報学, 24(2), 315-322.
- [59] 中山浩太郎, 原隆浩, 西尾章治郎. (2008). Wikipedia マイニングによる大規模 Web オントロジの実現. In 人工知能学会全国大会論文集 2008 年度人工知能学会全国大会 (第 22 回) 論文集 (pp. 96-96). 社団法人 人工知能学会.
- [60] 渡邊恵太, 加藤昇平. (2014). ユーザ興味を反映した情報推薦のための潜在的ディクレ配分法を用いた協調フィルタリング (抽出, 第 4 回テキストマイニング・シンポジウム). 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション

ン, 113(429), 15-20.

- [61] 浦本直彦, 松澤裕史, 猪口明博, 武田浩一. (2003). ライフサイエンス分野におけるテキストマイニング技術適用の動向. 情報処理学会研究報告データベースシステム (DBS), 2003(51 (2003-DBS-130)), 25-32.
- [62] 中居隆. (2008). テキストマイニングによる知財ポートフォリオ分析. 情報管理, 51(3), 194-206.
- [63] 森田哲夫, 入澤覚, 長塩彩夏, 野村和広, 塚田伸也, 大塚裕子, 杉田浩. (2012). 自由記述データを用いたテキストマイニングによる都市のイメージ分析. 土木学会論文集 D3 (土木計画学), 68(5), I_315-I_323.
- [64] 安藤英幸, 大和裕幸, 堀晃, 増田宏, 白山晋. (2002). テキストマイニングを用いた故障報告書分析手法の研究. 日本造船学会論文集, 2002(192), 475-483.
- [65] 佐々木千晴, 藤井敦, 石川徹也. (2006). 意思決定支援のための主観情報マイニング. 言語処理学会第 12 回年次大会発表論文集, 77-80.
- [66] 工藤拓, 松本裕治. (2004). 半構造化テキストの分類のためのブースティングアルゴリズム. 情報処理学会論文誌, 45(9), 2146-2156.
- [67] 金明哲, 村上征勝. (2007). ランダムフォレスト法による文章の書き手の同定. 統計数理, 55(2), 255-268.
- [68] 小林隆志, 林晋平. (2010). データマイニング技術を応用したソフトウェア構築・保守支援の研究動向. コンピュータ ソフトウェア, 27(3), 3_13-3_23.
- [69] 村上亜紀, 木村哲彦, 滝沢茂男, 牛澤賢二, 高田一, 森田能子. (2008). 褥そう対策の成果についての自由記入分析結果. In バイオフィリア リハビリテーション学会研究大会予稿集 バイオフィリア リハビリテーション学会第 12 回大会 (pp. 8-8). バイオフィリア リハビリテーション学会.
- [70] 砂山渡, 高間康史, ボレガラダヌシカ, 西原陽子, 徳永秀和, 串間宗夫, 松下光範. (2011). テキストデータマイニングのための統合環境: TETDM プロジェクト (解析手法, 第 1 回テキストマイニング・シンポジウム). 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, 111(119), 15-20.
- [71] 土橋喜. (1999). WWW とテキストマイニングの統合による問題構造可視化支援. 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, 99(447), 51-58.
- [72] 戸田浩之, 北川博之, 藤村考, 片岡良治, 奥雅博. (2007). グラフ分析を利用した文書集合からの話題構造マイニング. 電子情報通信学会論文誌 D, 90(2), 292-310.
- [73] 山西健司. (2002). データ・テキストマイニングの最新動向: 外れ値検出と評判分析を例に (< 特集> データ・テキストマイニング). 応用数理, 12(4), 341-356.
- [74] 安藤俊幸, 桐山勉. (2014). 中国特許解析・テキストマイニングによる KW 分析. In 情報プロフェッショナルシンポジウム予稿集 第 11 回情報プロフェッショナルシンポジウム (pp. 31-36). 国立研究開発法人 科学技術振興機構 一般社団法人

情報科学技術協会.

- [75] 杉本英晴. (2009). 大学生における「就職しないこと」イメージの構造と進路未決定: テキストマイニングを用いた検討.
- [76] 迫村光秋, 和泉潔. (2013). Twitter テキストマイニングによる経済動向分析. 第 9 回人工知能学会 ファイナンスにおける人工知能応用研究会 発表論文.
- [77] 難波雄二, 塚井誠人, 桑野将司. (2011). 文脈マイニングモデルを用いた討議過程の可視化手法に関する研究. 土木学会論文集 D3 (土木計画学), 67(5), 67_I_209-67_I_219.
- [78] 橋本泰一, 乾孝司, 村上浩司. (2008). 拡張固有表現タグ付きコーパスの構築. 情報処理学会研究報告自然言語処理 (NL), 2008(113 (2008-NL-188)), 113-120.
- [79] 佐川輝高, 岡田ルリ子, 青木光子. (2004). 学生の看護研究抄録におけるテキストマイニング法の検討. 看護と情報: 看護図書館協議会会誌, 11, 36-41.
- [80] 橋本泰一, 村上浩司, 乾孝司, 内海和夫, 石川正道. (2008). 文書クラスタリングによるトピック抽出および課題発見. 社会技術研究論文集, 5, 216-226.
- [81] 近藤伸也, 大山宗則, 目黒公郎. (2005). テキストマイニングを用いた災害状況イメージネーション支援システムの構築に関する研究. 生産研究, 57(4), 348-351.
- [82] 大山康彦. (2012). フェルデンクライス・メソッドによる身体知覚変容に関する一考察~ 身体感覚の言語化についてのテキストマイニング分析~. 茨城キリスト教大学紀要. II, 社会・自然科学, 46, 263-272.
- [83] 浅川雅美, 岡野雅雄. (2005). テレビ CM に対する視聴者 反応の分析—自由回答文のテキストマイニング. 文教 大学女子短期 大学部研究紀要, (48).
- [84] 鍛冶伸裕, 福島健一, 喜連川優. (2009). 大規模ウェブテキストからの片仮名用言の自動獲得. 電子情報通信学会論文誌 D, 92(3), 293-300.
- [85] 和泉潔, 松井藤五郎. (2012). 金融市場における最新情報技術: 8. 金融テキストマイニング研究の紹介. 情報処理, 53(9), 932-937.
- [86] 相川勇之, 伊藤山彦, 高山泰博, 鈴木克志, 今村誠. (2003). 概念抽出型テキストマイニングによるアンケート分析手法の提案. 情報処理学会研究報告情報学基礎 (FI), 2003(37 (2002-FI-070)), 1-6.
- [87] 竹内広宜, 杉山喜昭, 太田千景, 山口高平. (2010). マーケティングミックスとテキストマイニングを用いた市場分析支援. 第 24 回人工知能学会全国大会予稿集. 3B3-01.
- [88] 中村光浩, 寺町ひとみ, 足立哲夫, 土屋照雄. (2010). テキストマイニングによる薬学生実務実習レポートの分析. 医療薬学, 36(1), 25-30.
- [89] 南野朋之, 鈴木泰裕, 藤木稔明, 奥村学. (2004). blog の自動収集と監視. 人工知能学会論文誌, 19(6), 511-520.
- [90] 武田江里子, 小林康江, 加藤千晶. (2013). 産後 1 ヶ月の母親のストレスの本質の

- 探索: テキストマイニング分析によるストレス内容の結びつきから. 母性衛生, 54(1), 86-92.
- [91] 山西健司. (2003). テキストマイニングと NLP ビジネス. JEITA 自然言語処理技術に関するシンポジウム, 33-39.
- [92] 中山記男, 江口浩二, 神門典子. (2004). 感情表現の抽出手法に関する提案. 情報処理学会研究報告自然言語処理 (NL), 2004(108 (2004-NL-164)), 13-18.
- [93] 松井くにお, 渡部勇, 内野寛治. (2006). ナレッジマネジメントにおけるテキストマイニング. 情報処理, 47(8), 893-899.
- [94] 阿部秀尚, 平野章二, 津本周作. (2007). テキストマイニングによる退院時サマリーからの時系列関係知識抽出. 医療情報学, 27(1), 33-38.
- [95] 仲矢浩二. (2010). テキストマイニングを用いた株価予測. 高知工科大学 2009 年度学位論文.
- [96] 増田正. (2003). フランス地方議会の審議項目のテキストマイニング分析. 地域政策研究, 13(2), 3.
- [97] 鷺田万帆, 服部兼敏. (2008). 2. 看護におけるテキストマイニングとその活用事例. 看護研究, 41(3), 249-258.
- [98] 崔延敏, 浅見泰司. (2004). 言語統計分析による住宅建設五箇年計画及び答申の特性分析: 政策の立案と評価における非定型・大量情報の活用可能性. 日本建築学会計画系論文集, (579), 89-96.
- [99] 松村真宏. (2008). フィールドの魅力を掘り起こすフィールドマイニング. 信学誌, 93(3), 237-241.
- [100] 海野裕也, 那須川哲哉. (2010). 言語横断テキストマイニング. In 『The 24th Annual Conference of the Japanese Society for Artificial Intelligence』, 人工知能学会 (3A4-1) (pp. 1-4).
- [101] Kim, J. D., Ohta, T., Tateisi, Y., & Tsujii, J. I. (2003). GENIA corpus—a semantically annotated corpus for bio-textmining. *Bioinformatics*, 19(suppl 1), i180-i182.
- [102] Tan, A. H. (1999, April). Text mining: The state of the art and the challenges. In *Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases* (Vol. 8, pp. 65-70).
- [103] Sullivan, D. (2001). *Document warehousing and text mining: techniques for improving business operations, marketing, and sales*. John Wiley & Sons, Inc..
- [104] Berry, M. W. (2004). Survey of text mining. *Computing Reviews*, 45(9), 548.
- [105] Larsen, B., & Aone, C. (1999, August). Fast and effective text mining using linear-time document clustering. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 16-22).

ACM.

- [106] Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- [107] Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1), 57-71.
- [108] Aggarwal, C. C., & Zhai, C. (Eds.). (2012). *Mining text data*. Springer Science & Business Media.
- [109] Mei, Q., & Zhai, C. (2005, August). Discovering evolutionary theme patterns from text: an exploration of temporal text mining. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining* (pp. 198-207). ACM.
- [110] Weiss, S. M., Apte, C., Damerau, F. J., Johnson, D. E., Oles, F. J., Goetz, T., & Hampf, T. (1999). Maximizing text-mining performance. *IEEE Intelligent Systems and their applications*, 14(4), 63-69.
- [111] Ananiadou, S., & McNaught, J. (2006). *Text mining for biology and biomedicine* (pp. 1-12). London: Artech House.
- [112] Yoon, B., & Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *The Journal of High Technology Management Research*, 15(1), 37-50.
- [113] Hotho, A., Nürnberger, A., & Paaß, G. (2005, May). A brief survey of text mining. In *Ldv Forum* (Vol. 20, No. 1, pp. 19-62).
- [114] Konchady, M. (2006). *Text mining application programming*. Charles River Media, Inc..
- [115] Van Driel, M. A., Bruggeman, J., Vriend, G., Brunner, H. G., & Leunissen, J. A. (2006). A text-mining analysis of the human phenome. *European journal of human genetics*, 14(5), 535-542.
- [116] Fan, W., Wallace, L., Rich, S., & Zhang, Z. (2006). Tapping the power of text mining. *Communications of the ACM*, 49(9), 76-82.
- [117] Hearst, M. (2003). *What is text mining*. SIMS, UC Berkeley.
- [118] Dörre, J., Gerstl, P., & Seiffert, R. (1999, August). Text mining: finding nuggets in mountains of textual data. In *Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 398-401). ACM.
- [119] Meyer, D., Hornik, K., & Feinerer, I. (2008). Text mining infrastructure in R. *Journal of statistical software*, 25(5), 1-54.
- [120] Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent

- analysis. *Information Processing & Management*, 43(5), 1216-1247.
- [121] Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5), 1216-1247.
- [122] Feldman, R., Fresko, M., Kinar, Y., Lindell, Y., Liphstat, O., Rajman, M., ... & Zamir, O. (1998, September). Text mining at the term level. In *European Symposium on Principles of Data Mining and Knowledge Discovery* (pp. 65-73). Springer Berlin Heidelberg.
- [123] Miner, G., Elder IV, J., & Hill, T. (2012). *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press.
- [124] Spasic, I., Ananiadou, S., McNaught, J., & Kumar, A. (2005). Text mining and ontologies in biomedicine: making sense of raw text. *Briefings in bioinformatics*, 6(3), 239-251.
- [125] Ananiadou, S., Kell, D. B., & Tsujii, J. I. (2006). Text mining and its potential applications in systems biology. *Trends in biotechnology*, 24(12), 571-579.
- [126] Rzhetsky, A., Seringhaus, M., & Gerstein, M. B. (2009). Getting started in text mining: part two. *PLoS Comput Biol*, 5(7), e1000411.
- [127] Mooney, R. J., & Bunescu, R. (2005). Mining knowledge from text using information extraction. *ACM SIGKDD explorations newsletter*, 7(1), 3-10.
- [128] Zhai, C., Velivelli, A., & Yu, B. (2004, August). A cross-collection mixture model for comparative text mining. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 743-748). ACM.
- [129] Wong, P. C., Whitney, P., & Thomas, J. (1999). Visualizing association rules for text mining. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on* (pp. 120-123). IEEE.
- [130] Srinivasan, P. (2004). Text mining: generating hypotheses from MEDLINE. *Journal of the American Society for Information Science and Technology*, 55(5), 396-413.
- [131] Tanabe, L., Scherf, U., Smith, L. H., Lee, J. K., Hunter, L., & Weinstein, J. N. (1999). MedMiner: an Internet text-mining tool for biomedical information, with application to gene expression profiling. *Biotechniques*, 27(6), 1210-4.
- [132] Apte, C., Damerau, F., & Weiss, S. (1998). *Text mining with decision rules and decision trees*. IBM Thomas J. Watson Research Division.
- [133] Krallinger, M., & Valencia, A. (2005). Text-mining and information-retrieval services for molecular biology. *Genome biology*, 6(7), 224.
- [134] Nasukawa, T., & Nagano, T. (2001). *Text analysis and knowledge mining*

- system. *IBM systems journal*, 40(4), 967-984.
- [135] Kao, A., & Poteet, S. R. (Eds.). (2007). *Natural language processing and text mining*. Springer Science & Business Media.
- [136] Zweigenbaum, P., Demner-Fushman, D., Yu, H., & Cohen, K. B. (2007). *Frontiers of biomedical text mining: current progress*. *Briefings in bioinformatics*, 8(5), 358-375.
- [137] Tiffin, N., Kelso, J. F., Powell, A. R., Pan, H., Bajic, V. B., & Hide, W. A. (2005). *Integration of text-and data-mining using ontologies successfully selects disease gene candidates*. *Nucleic acids research*, 33(5), 1544-1552.
- [138] Berry, M. W., & Kogan, J. (Eds.). (2010). *Text mining: applications and theory*. John Wiley & Sons.
- [139] Nahm, U. Y., & Mooney, R. J. (2002, March). *Text mining with information extraction*. In *Proceedings of the AAAI 2002 Spring Symposium on Mining Answers from Texts and Knowledge Bases* (pp. 60-67).
- [140] Rebholz-Schuhmann, D., Kirsch, H., & Couto, F. (2005). *Facts from text—is text mining ready to deliver?*. *PLoS Biol*, 3(2), e65.
- [141] Mittermayer, M. A. (2004, January). *Forecasting intraday stock price trends with text mining techniques*. In *system sciences, 2004. proceedings of the 37th annual hawaii international conference on* (pp. 10-pp). IEEE.
- [142] Pauca, V. P., Shahnaz, F., Berry, M. W., & Plemmons, R. J. (2004, April). *Text mining using non-negative matrix factorizations*. In *Proceedings of the 2004 SIAM International Conference on Data Mining* (pp. 452-456). Society for Industrial and Applied Mathematics.
- [143] Collier, N., Doan, S., Kawazoe, A., Goodwin, R. M., Conway, M., Tateno, Y., ... & Shigematsu, M. (2008). *BioCaster: detecting public health rumors with a Web-based text mining system*. *Bioinformatics*, 24(24), 2940-2941.
- [144] Billheimer, D. D., Booker, A. J., Condliff, M. K., Greaves, M. T., Holt, F. B., Kao, A. S. W., ... & Wu, Y. J. (2003). *U.S. Patent No. 6,611,825*. Washington, DC: U.S. Patent and Trademark Office.
- [145] Cheng, D., Knox, C., Young, N., Stothard, P., Damaraju, S., & Wishart, D. S. (2008). *PolySearch: a web-based text mining system for extracting relationships between human diseases, genes, mutations, drugs and metabolites*. *Nucleic acids research*, 36(suppl 2), W399-W405.
- [146] Li, N., & Wu, D. D. (2010). *Using text mining and sentiment analysis for online forums hotspot detection and forecast*. *Decision support systems*, 48(2), 354-368.

- [147] Yeh, A. S., Hirschman, L., & Morgan, A. A. (2003). Evaluation of text data mining for database curation: lessons learned from the KDD Challenge Cup. *Bioinformatics*, 19(suppl 1), i331-i339.
- [148] Zhong, N., Li, Y., & Wu, S. T. (2012). Effective pattern discovery for text mining. *IEEE transactions on knowledge and data engineering*, 24(1), 30-44.
- [149] Krallinger, M., Valencia, A., & Hirschman, L. (2008). Linking genes to literature: text mining, information extraction, and retrieval applications for biology. *Genome biology*, 9(2), S8.
- [150] Krallinger, M., Morgan, A., Smith, L., Leitner, F., Tanabe, L., Wilbur, J., Hirschman, L., & Valencia, A. (2008). Evaluation of text-mining systems for biology: overview of the Second BioCreative community challenge. *Genome biology*, 9(2), S1.
- [151] Wu, S. T., Li, Y., & Xu, Y. (2006, December). Deploying approaches for pattern refinement in text mining. In *Data Mining, 2006. ICDM'06. Sixth International Conference on* (pp. 1157-1161). IEEE.
- [152] Ananiadou, S., Pyysalo, S., Tsujii, J. I., & Kell, D. B. (2010). Event extraction for systems biology by text mining the literature. *Trends in biotechnology*, 28(7), 381-390.
- [153] Ongenaert, M., Van Neste, L., De Meyer, T., Menschaert, G., Bekaert, S., & Van Criekinge, W. (2008). PubMeth: a cancer methylation database combining text-mining and expert annotation. *Nucleic acids research*, 36(suppl 1), D842-D846.
- [154] Krallinger, M., Erhardt, R. A. A., & Valencia, A. (2005). Text-mining approaches in molecular biology and biomedicine. *Drug discovery today*, 10(6), 439-445.
- [155] Park, Y., & Byrd, R. J. (2001, June). Hybrid text mining for finding abbreviations and their definitions. In *Proceedings of the 2001 conference on empirical methods in natural language processing* (pp. 126-133).
- [156] Rajman, M., & Besançon, R. (1998). Text mining: natural language techniques and text mining applications. In *Data mining and reverse engineering* (pp. 50-64). Springer US.
- [157] Don, A., Zheleva, E., Gregory, M., Tarkan, S., Auvil, L., Clement, T., Shneiderman, B., & Plaisant, C. (2007, November). Discovering interesting usage patterns in text collections: integrating text mining with visualization. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management* (pp. 213-222). ACM.

- [158] Jing, L. P., Huang, H. K., & Shi, H. B. (2002). Improved feature selection approach TFIDF in text mining. In *Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on* (Vol. 2, pp. 944-946). IEEE.
- [159] Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the “helpfulness” of online user reviews: A text mining approach. *Decision Support Systems*, 50(2), 511-521.
- [160] Fattori, M., Pedrazzi, G., & Turra, R. (2003). Text mining applied to patent mapping: a practical business case. *World Patent Information*, 25(4), 335-342.
- [161] He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472.
- [162] Corney, M., De Vel, O., Anderson, A., & Mohay, G. (2002). Gender-preferential text mining of e-mail discourse. In *Computer Security Applications Conference, 2002. Proceedings. 18th Annual* (pp. 282-289). IEEE.
- [163] Hoffmann, R., Krallinger, M., Andres, E., Tamames, J., Blaschke, C., & Valencia, A. (2005). Text mining for metabolic pathways, signaling cascades, and protein networks. *Sci STKE*, 283, e21.
- [164] Delgado, M., Martín-Bautista, M. J., Sánchez, D., & Vila, M. A. (2000). Mining text data: special features and patterns. In *Pattern Detection and Discovery* (pp. 140-153). Springer Berlin Heidelberg.
- [165] Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*, 5(1), 1-167.
- [166] Kano, Y., Baumgartner, W. A., McCrohon, L., Ananiadou, S., Cohen, K. B., Hunter, L., & Tsujii, J. I. (2009). U-Compare: share and compare text mining tools with UIMA. *Bioinformatics*, 25(15), 1997-1998.
- [167] Corley, C. D., Cook, D. J., Mikler, A. R., & Singh, K. P. (2010). Text and structural data mining of influenza mentions in web and social media. *International journal of environmental research and public health*, 7(2), 596-615.
- [168] Chiang, J. H., Yu, H. C., & Hsu, H. J. (2004). GIS: a biomedical text-mining system for gene information discovery. *Bioinformatics*, 20(1), 120-121.
- [169] Rose, S., Engel, D., Cramer, N., & Cowley, W. (2010). Automatic keyword extraction from individual documents. *Text Mining*, 1-20.
- [170] Kostoff, R. N., del Rio, J. A., Humenik, J. A., Garcia, E. O., & Ramirez, A. M. (2001). Citation mining: Integrating text mining and bibliometrics for research user profiling. *Journal of the American Society for Information Science and*

Technology, 52(13), 1148-1156.

- [171] Srivastava, A. N., & Sahami, M. (Eds.). (2009). Text mining: Classification, clustering, and applications. CRC Press.
- [172] Jelier, R., Schuemie, M. J., Veldhoven, A., Dorssers, L. C., Jenster, G., & Kors, J. A. (2008). Anni 2.0: a multipurpose text-mining tool for the life sciences. *Genome biology*, 9(6), R96.
- [173] Yang, Y., Akers, L., Klose, T., & Yang, C. B. (2008). Text mining and visualization tools—impressions of emerging capabilities. *World Patent Information*, 30(4), 280-293.
- [174] Kloptchenko, A., Eklund, T., Karlsson, J., Back, B., Vanharanta, H., & Visa, A. (2004). Combining data and text mining techniques for analysing financial reports. *Intelligent systems in accounting, finance and management*, 12(1), 29-41.
- [175] Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine your own business: Market-structure surveillance through text mining. *Marketing Science*, 31(3), 521-543.
- [176] Feldman, R., Fresko, M., Hirsh, H., Aumann, Y., Liphstat, O., Schler, Y., & Rajman, M. (1998, October). Knowledge Management: A Text Mining Approach. In *PAKM* (Vol. 98, p. 9).
- [177] Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. *Expert Systems with Applications*, 40(10), 4241-4251.
- [178] Girju, R., & Moldovan, D. I. (2002, May). Text mining for causal relations. In *FLAIRS Conference* (pp. 360-364).
- [179] Chen, H. (2001). Knowledge management systems: a text mining perspective. Knowledge Computing Corporation.
- [180] Srinivasan, P. (2001). MeSHmap: a text mining tool for MEDLINE. In *Proceedings of the AMIA Symposium* (p. 642). American Medical Informatics Association.
- [181] Ghani, R., Probst, K., Liu, Y., Krema, M., & Fano, A. (2006). Text mining for product attribute extraction. *ACM SIGKDD Explorations Newsletter*, 8(1), 41-48.
- [182] Kostoff, R. N. (2005). U.S. Patent No. 6,886,010. Washington, DC: U.S. Patent and Trademark Office.
- [183] Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006, August). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings*

- of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 935-940). ACM.
- [184] Uramoto, N., Matsuzawa, H., Nagano, T., Murakami, A., Takeuchi, H., & Takeda, K. (2004). A text-mining system for knowledge discovery from biomedical documents. *IBM Systems Journal*, 43(3), 516-533.
- [185] Stavrianou, A., Andritsos, P., & Nicoloyannis, N. (2007). Overview and semantic issues of text mining. *ACM Sigmod Record*, 36(3), 23-34.
- [186] Liddy, E. D. (2000). Text mining. *Bulletin of the American Society for Information Science and Technology*, 27(1), 13-14.
- [187] Tari, L., Anwar, S., Liang, S., Cai, J., & Baral, C. (2010). Discovering drug-drug interactions: a text-mining and reasoning approach based on properties of drug metabolism. *Bioinformatics*, 26(18), i547-i553.
- [188] Ramage, D., Manning, C. D., & Dumais, S. (2011, August). Partially labeled topic models for interpretable text mining. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 457-465). ACM.
- [189] ROMACKER, U. H. M., & SCHULZ, S. (2001). CREATING KNOWLEDGE REPOSITORIES FROM BIOMEDICAL REPORTS: THE MEDSYNDIKATE TEXT MINING SYSTEM. In *Pacific Symposium on Biocomputing 2002, Kauai, Hawaii, 3-7 January 2002* (p. 338). World Scientific.
- [190] Wei, C. H., Kao, H. Y., & Lu, Z. (2013). PubTator: a web-based text mining tool for assisting biocuration. *Nucleic acids research*, gkt441.
- [191] Lu, W. H., Chien, L. F., & Lee, H. J. (2002). Translation of web queries using anchor text mining. *ACM Transactions on Asian Language Information Processing (TALIP)*, 1(2), 159-172.
- [192] Jessop, D. M., Adams, S. E., Willighagen, E. L., Hawizy, L., & Murray-Rust, P. (2011). OSCAR4: a flexible architecture for chemical text-mining. *Journal of cheminformatics*, 3(1), 41.
- [193] Kawano, H. (2000). Overview of Mondou web search engine using text mining and information visualizing technologies. In *Digital Libraries: Research and Practice, 2000 Kyoto, International Conference on.* (pp. 234-244). IEEE.
- [194] Alex, B., Grover, C., Haddow, B., Kabadjor, M., Klein, E., Matthews, M., ROEBUCK, S., & Wang, X. (2008, January). Assisted Curation: Does Text Mining Really Help?. In *Pacific Symposium on Biocomputing (Vol. 13, pp. 556-567)*.
- [195] Kostoff, R. N., Toothman, D. R., Eberhart, H. J., & Humenik, J. A. (2001).

- Text mining using database tomography and bibliometrics: A review. *Technological Forecasting and Social Change*, 68(3), 223-253.
- [196] Rodriguez-Esteban, R. (2009). Biomedical text mining and its applications. *PLoS Comput Biol*, 5(12), e1000597.
- [197] Giles, J. T., Wo, L., & Berry, M. W. (2003). GTP (General Text Parser) software for text mining. *Statistical data mining and knowledge discovery*, 455-471.
- [198] Rajman, M., & Besançon, R. (1998). Text mining-knowledge extraction from unstructured textual data. In *Advances in data science and classification* (pp. 473-480). Springer Berlin Heidelberg.
- [199] Altman, R. B., Bergman, C. M., Blake, J., Blaschke, C., Cohen, A., Gannon, F., Grivell, L., Hahn, U., Hersh, W., Hirschman, L., Jensen, L. J., Krallinger, M., Mons, B., O'Donoghue, S. I., Peitsch, M.C., Rebholz-Schuhmann, D., Shatkay, H., & Valencia, A. (2008). Text mining for biology-the way forward: opinions from leading scientists. *Genome biology*, 9(2), S7.
- [200] Srinivasan, P., & Rindfleisch, T. (2002). Exploring text mining from MEDLINE. In *Proceedings of the AMIA Symposium* (p. 722). American Medical Informatics Association.
- [201] 石田基広, 金明哲.(2012) コーパスとテキストマイニング. 共立出版.