

探索過程の視覚化による
テキスト集合からの知識獲得支援
に関する研究

総合情報学研究科
知識情報学専攻

インタラクションデザインの理論と実践

13M7110
大塚 直也

要旨

本研究は

目 次

1	序論	1
1.1	本研究の背景	1
1.2	テキスト分析による新たな知識の獲得	2
1.3	本研究で取り組む課題	2
1.4	本論文の構成	4
2	関連研究	5
2.1	テキストマイニング	5
2.2	Total Environment for Text Data Mining プロジェクト	9
2.3	Exploratory Search	12
2.4	知識創発支援に関する研究	19
3	デザイン指針	23
3.1	探索プロセスの円滑化	23
3.2	探索プロセスにおける外在化	25
3.3	探索プロセスの振り返り支援	26
4	プロトタイプシステムの実装	28
4.1	画面構成	28
4.2	グラフ表現による分析状況の視覚化	29
4.3	マウス操作によるテキスト分析ツールの選択・変更	30
4.4	探索プロセスの振り返り支援のための履歴	31
4.5	探索プロセスにおける思考の外在化の支援	33
5	議論	35
5.1	テキストマイニングと情報可視化のモデル	36
5.2	TETDM システムの問題点	38
6	結論	41

1 序論

本章では、本研究の実施に至った背景を説明し、対象とする課題を明確にする。

1.1 本研究の背景

我々の身の回りには膨大な量のテキストデータが存在している。例えば、我々が日常的に利用しているニュース記事や E メールなどが挙げられる。また、BBS やブログ、SNS (Social Networking Service) などの CGM (Consumer Generated Media) の普及に伴い、一般のユーザーが様々な情報をテキストデータとして公開することも一般的となった。結果として、我々が得ることのできるテキストデータの量は爆発的に増加している。我々は、それらの大量のテキストデータを閲覧することによって、自身の興味や関心のある情報を得ることができる。加えて、単にそれらの情報を得るだけではなく、それらを分析することによって有益な情報を発見したり、それらを組み合わせて知識へと昇華させたりするための情報資源として活用することができる。

このような膨大なテキストデータを分析して有益な情報や新たな知識を発見・獲得するための技術としてテキストマイニングがある [27]。テキストマイニングとは、自然言語処理、データマイニング、情報可視化などの複合技術であり、これらを適応的に利用することによって、テキスト集合を対象とした有益な情報や新たな知識の発見・獲得を実現している [20]。それを目的として、情報抽出 (information extraction) やトピック追跡 (topic tracking)、文書要約 (summarization)、文書分類 (categorization)、文書クラスタリング (clustering)、概念結合 (concept linkage)、情報可視化 (information visualization)、質問応答 (question answering) などのテキストマイニングの基盤技術について研究開発が行われており、既に多くの技術が成果として発表されている [4]。

テキストマイニングの活用事例として、コールセンターに蓄積された顧客からの問い合わせデータの分析 [19] や生命科学に関する文献情報の分析 [39]、特許情報の分析 [1] がある。その他にも、医療や顧客関係管理 (CRM: Customer Relationship Management)、人的資源管理 (HRM: Human Resource management)、金融市場などの様々な領域において活用されている [5]。このようなことから、大量のテキストデータを分析することによって有益な情報や新たな知識を発見することの必要性は高まっている。それに伴い、テキストマイニングの利用場面は更に広がっていくと考えられる。

このように、テキストデータを利用して有益な情報や新たな知識を獲得することを目的として、テキストマイニングに関する技術が開発されており、様々な分析ツールやシステムが提案・提供されている。しかし、それらの分析ツールの多くは单一の目的のために作られているため、ユーザはそれらの分析ツールが想定していないような様々な視点からテキストデータを分析することができない。そのため、ユーザがより多角的な分析を行いたいと考えた場合、複数の分析ツールを同時並行的に用いる必要がある。また、それらの分析ツールを用いて円滑に分析を進めるためには、複数の分析ツールによる分析結果を比較するために、データフォーマットを統一したり、比較のためのインタフェースを新たに実装したりする必要がある。

これらのプロセスは、膨大なテキストデータを分析し、有益な情報や新たな知識を発見した

いと考えるユーザの創造的活動の妨げになる。

このような問題を解決するために、砂山らは、Total Environment for Text Data Mining プロジェクト（以下 TETDM プロジェクト）¹ を提唱している [34]。TETDM プロジェクトの目的は、多種多様なテキストマイニングツールを画一的に扱える統合環境を構築することによって、テキストデータを扱うユーザの創造的活動を支援することである。

1.2 テキスト分析による新たな知識の獲得

一般に、ユーザがテキスト集合から新たな知識を獲得したいと考えた場合、ユーザの情報要求は必ずしも明確にはなっていない。すなわち、探索開始時にユーザは、“最終的にどのような知識を得たいのか”という目的と“どのような分析を行えば新たな知識を獲得することができるのか”という明確なプランを持っていない。そのため、ユーザはテキスト集合を様々な観点から分析し、それを繰り返すことによって、徐々に自身の情報要求を精緻化して必要がある。加えて、ユーザはそのような分析プロセスにおいて生成された仮説を検証したり、得られた知見を解釈しながら分析を繰り返していく [6]。すなわち、テキストマイニングによる知識発見は、あらかじめ探索のゴールが明確に定まっているような one-shot の探索プロセスで構成されたタスクではなく、試行錯誤を繰り返しながら徐々に有益な情報を発見していく探索的な情報アクセスを必要とするプロセスで構成されたタスクと据えることができる。

那須川は、テキストマイニングとは知識の候補、すなわち“気づき”を与えてくれる技術である、としている [21]。テキストマイニングによって得られた“気づき”が実際に有益な知識となるかどうかを判断するためには、ユーザは様々な情報を考慮する必要がある。つまり、何らかの知見に繋がりそうな“気づき”を与えるのがテキストマイニングの役割であり、この“気づき”をどう活用するのか判断することはユーザの役割である。そのため、テキストマイニングとは、システムが自動的に知識を発見するような技術ではなく、知識発見を支援する技術であり、その支援を目的としたシステムでは、ユーザがインタラクティブに操作をしながら知識を獲得していくことが重要である。これは、テキストマイニングを用いて有益な情報や新たな知識を獲得したい場合、ユーザがその探索プロセスにおいて積極的に関与する必要があることを示唆している。

これらのことから、テキストマイニングは Exploratory Search に該当するタスクとみなすことができる。Exploratory Search とは、曖昧な情報要求に基づき、試行錯誤を繰り返しながら、徐々に目的の情報へと近づいていくといったような情報探索行為を説明するためのモデルである [40]。

1.3 本研究で取り組む課題

本研究の目的は、テキスト集合からの新たな知識の発見を支援することである。1.1 節で述べたように、テキストマイニングとは、様々な技術を適応的に組み合わせて用いることによって、テキスト集合から有益な情報の発見や新たな知識の獲得を支援するための技術である。しかし、“テキストマイニング”という語の定義は曖昧であるため、ユーザによりテキストマイニングへの要求が異なったり、単なる情報検索や自動分類がテキストマイニングであると捉えら

¹<http://tetdm.jp>

れていることもある [21]。本研究ではまず、テキストマイニングに関する研究を俯瞰しながら、本研究における“テキストマイニング”的定義を明確にする。加えて、テキストマイニングのタスクの特徴について整理する。

テキストマイニングによる知識獲得は、試行錯誤を必要とする探索的な情報アクセスを必要とするタスクである。本研究では、そのようなテキストマイニングを支援するためのアプリケーションシステムが満たすべき要件を 2 つの観点から整理する。我々は、テキストマイニングを(1) Exploratory Search に該当するタスクである、(2) 創造的活動である、と見做している。

Exploratory Search では、ユーザは、自身の興味や目的に応じて、試行錯誤を繰り返しながら探索を進めていく。そのようなプロセスを円滑に行えるように支援するためには、システムのインターフェースはユーザの試行錯誤を妨げるようなものであってはならない。また、ユーザが探索を繰り返していくにつれて、探索作業が複雑になるため、それを支援するためには、ユーザが今までどのような探索プロセスを経ているのかという探索のコンテキストを把握できるような仕組みも必要となる。テキストマイニングにおいては、ユーザは様々なテキスト分析ツールを適応的に組み合わせて用いながら、知識を獲得していく。そのため、分析の進行に沿って、ユーザがテキスト分析ツールを選択したり切り替えたりすることを円滑に行えるようにするだけでなく、どのような分析ツールによってどのような知見が得られたかを概観できるようにする必要である。

新たな知識の獲得というテキストマイニングの目的に着目すると、テキストマイニングは創造的活動であると見做すことができる。創造的活動とは、曖昧な目的や意図を基に、(1) 何を情報として創出するのか、(2) どのようにそれを表現するのか、を同時並行的に同定していく作業である [18]。創造的活動において、ユーザはそのプロセスによって得られた知見や思考を外在化していく。また、それらを整理したり、再構築しながら、徐々に最終的に何をどのように創出するのかを決定していく。つまり、創造活動のプロセスにおいて、知識の結晶化と液状化 [8] を繰り返しながら、最終的に創出する知識を構築していく。そのようなプロセスにおいて、得られた知識や思考をどのような表現として外在化できるかは、創造的活動の成果に大きく影響を及ぼす。そのため、創造活動の支援を目的としたシステムをデザインする際には、外在化された表現の表現方法やそれらとのインタラクションの手法について、熟考する必要がある。

本研究では、これらの観点から、テキストマイニングのためのアプリケーションシステムを構築する際のデザイン指針を提案する。提案するデザイン指針は、(1) 探索プロセスの円滑化、(3) 探索プロセスにおける外在化、(2) 探索プロセスの振り返り支援、の 3 点に着目して策定した。

TETDM プロジェクトは、複数のテキスト分析ツールを組み合わせて用いることのできる環境を構築することによりユーザの創造的活動を支援することを目的としている。しかし、TETDM システムは現在開発途中であるため、そのインターフェースは、テキスト分析ツールの切り替えを円滑に行えるインターフェースにはなっていない。TETDM システムを用いた分析では、複数のテキスト分析ツールを組み合わせて分析を行うため、入力されたテキストデータに対する処理の流れが煩雑となる。しかし、ユーザが全てを把握できる手段が用意されておらず、ユーザの意図しない操作や処理結果が引き起こされる可能性がある。そのゆえ、現在の TETDM システムのインターフェースでは、試行錯誤を伴う分析の妨げになることが懸念され、創造的活動を行う場面で使用するには十分ではない。また、TETDM プロジェクトの目的はユーザの

創造活動を支援することであるが、それに対する取り組みは十分でない。そこで本研究では、TETDM プロジェクトの目的に則り、ユーザの試行錯誤を伴う探索行為をより円滑に支援し、それによってユーザの創造的活動を支援するために、提案するデザイン指針を用いて TETDM システムのインターフェースの改良を試みる。

1.4 本論文の構成

本論文の構成について述べる。1 章では、本研究の背景と問題を整理し、本研究の目的について述べた。2 章では、テキストマイニングのタスクの特徴を整理するために、テキストマイニングに関する研究を整理する。加えて、TETDM プロジェクトの概要について説明する。テキストマイニングは Exploratory Search に該当するタスクであり、その目的は新たな知識の創出であるため、Exploratory Search と創造活動支援に関する研究についてそれぞれ述べる。3 章では、テキスト集合からの知識獲得を支援するためのシステムを構築する際に用いるデザイン指針を策定する。4 章では、提案したデザイン指針に基づき実装したプロトタイプシステムについて述べる。5 章では、実装したプロトタイプシステムを基に、TETDM システムの問題点について議論する。6 章では、本論文の結論を述べる。

2 関連研究

本研究の目的はテキスト集合からの新たな知識の発見・創出を支援することである。

本章ではまず、そのような目的を達成するための技術である、テキストマイニングについて述べる。しかしながら、“テキストマイニング”という語の定義は明確でない。そのため、テキストマイニングとその要素技術について関連研究を俯瞰しながら、本研究におけるテキストマイニングの定義を明確にする。また、テキストマイニングというタスクの特長を整理し、その支援のために必要な要素を整理する。

我々はテキストマイニングを、(1) ユーザの曖昧な情報要求に基づいた試行錯誤を必要とする探索的情報アクセスである、(2) 新たな知識の発見・創出を目的とした創造的情報創出 (Creative Knowledge Work) 活動である、と捉えている。(1) のような情報アクセスは Exploratory Search の枠組みに当てはめることができるため、本章では、Exploratory Search について説明する。加えて、(2) のような活動の支援を目的とする創造活動支援 (Creativity Support) に関する研究について説明する。

2.1 テキストマイニング

テキストマイニングとは大量のテキストデータから有益な、新たな知識を得るための技術である [27]。また、テキストマイニングは、自然言語処理やデータマイニング、情報可視化などの様々な技術の複合技術であり [20]、それらを適応的に組み合わせながら、分析を進めていくことが要求される。

多量のデータ集合からパターンの発見や有益な情報の抽出を行うという観点から考えると、テキストマイニングは、データマイニングや情報抽出 (Information Extraction)、情報検索 (Information Retrieval) と類似している [4]。Rajman は、テキストマイニングとデータマイニングの違いについて「データマイニングは構造化されたデータからの情報抽出であり、一方で、テキストマイニングは構造化されていないデータからの情報抽出である」と説明している [27]。つまり、それらの違いはテキストマイニングは構造化されていないデータから情報を抽出するという点である。テキストのような構造化されていないデータからその文法構造などの非明示的構造を抽出するためには、自然言語処理技術が必要不可欠である [4]。このことから、テキストマイニングは、自然言語処理技術を用いてデータマイニングをテキストを対象として拡張したものであると考えることができる。加えて、ユーザが分析結果の理解を容易に行えるように支援するために、情報可視化技術がよく用いられる [4]。

“テキストマイニング”という語の定義は曖昧であり、このように、テキストを対象としたデータマイニングや情報抽出など、単なる情報抽出やパターン抽出として扱われることも多い。しかし、那須川や Fan らは、テキストマイニングの本質的な役割について、Hearst の文献 [6] を引用しながら、「新たな知識の発見を目的とした技術」と説明している [22][4]。本研究では、Hearst の定義に則り、「テキスト集合を分析することによって個々のテキストを見るだけではわからないような新たな知識を発見することを目的とした技術」をテキストマイニングと呼ぶ。Hearst は、テキストマイニングとデータマイニングや情報検索 (Information Retrieval) の関係について、表 2.1 のように整理している。

Hearst は情報検索やデータマイニングについて、それらとテキストマイニングの差異について述べている [6]。Hearst は、情報検索について、「情報検索システムが、ユーザが要求した情報を含む文章を結果として返すという事実は、新たな情報の発見がなされたわけではない、ということを示している。なぜならば、情報検索システムによって得られた情報は、それを含む文書の著者にとって既知の情報であるはずであり、真に新たな情報が含まれているわけではないからである。」と述べている。また、データマイニングは膨大なデータセットの中から自動的に傾向やパターンを抽出するための技術であると説明している。情報検索やデータマイニングなどの技術はそれ単体では新たな知識の発見・創出をもたらさないが、ユーザは様々な観点から分析を行なながら人間の持つ認知能力を活用することで新たな知識を発見することができる。Hearst は、新たな知識の発見をもらたすか否かという観点から、テキストマイニングと情報検索及びデータマイニングは異なるものであると説明しており、新たな知識の発見こそがテキストマイニングの重要な側面であるとしている。このことから、テキストマイニングとは、新しい、今まで知られていなかつた情報を発見するための技術であり、「価値のない岩から価値のある鉱石を発掘すること」という「mining」という言葉本来の意味に合致することから、真のマイニング (real text data mining) である、と述べている。

現在までに、テキストマイニングを行うためのシステムは数多く提案されている。

例えば、Nasukawa らは、テキストマイニングのためのシステムとして TAKMI (Text Analysis and Knowledge MIning) を提案している [23]。那須川は、TAKMI を用いて、PC のヘルプセンターに蓄積された顧客からの問い合わせに関するデータに対して、テキストマイニングを適用した事例を紹介している [19]。テキストマイニングによって大量の問合せデータを容易に分析することが可能になったことにより、分析時間の短縮のみならず、今まで分析に利用していなかつたデータについても有効活用できた、製品の問題点の早期発見が可能になった、と述べている。加えて、分析結果を FAQ の作成に用いたり、分析担当者だけではなく、企業において実際の意思決定に携わる上級管理者がそのデータをもとに議論をするなど、分析結果の新たな利用方法も創出されたとしている。

テキスト集合から有益な情報を得たり新たな知識を発見する場合には、ユーザはそれらのテキストを様々な観点から分析する必要がある。そのため、ユーザは、様々なテキスト分析ツールを組み合わせて用いることが要求される。砂山は、世の中に散在する様々なテキストマイニング技術を集め、それらを統合的に扱える環境の構築を目指して TETDM プロジェクトを提唱している [35]。TETDM プロジェクトでは、ユーザが自身の目的や興味に応じて、様々なテキストマイニング技術を組み合わせながら分析を行える環境を構築することによって、テキス

表 2.1: Hearst によるデータマイニングとテキストデータマイニングの分類 [6]

	Finding Patterns	Finding Nuggets	
		Novel	Non-Novel
Non-textual data	standard data mining	?	database queries
Textual data	computational linguistics	real TDM	infomation retrieval

ト集合を扱うユーザの創造的活動を支援することを目的としている。TETDM プロジェクトの詳細については後述する。

Fan らは、主要な商用ベンダーによって開発されたテキストマイニングのためのシステムについて整理し、それらに用いられている技術について以下に 8 つにまとめている [4]。

- **情報抽出 (Information Extraction)**

情報抽出とは、テキスト内の重要語句の抽出やそれらの関係を識別することである。例えば、キーワード抽出や、重要文抽出、固有表現抽出などが挙げられる。これは、コンピュータによって構造化されていないテキストを処理するために必須の技術であり、テキストマイニングにおける他の技術の要諦技術でもある。

- **トピック追跡 (Topic Tracking)**

トピック追跡とは、自動でトピックを検出し、その時間的な変化を追跡するための技術である。ユーザがあるトピックについて理解したいと考えた場合、時間的な変化を踏まえて、その全体像を把握できることが望ましい。トピック追跡の役割はそれを支援することである。

- **テキスト要約 (Summarization)**

大量のテキストが存在する場合、その内容を把握するためにそれら全てに目を通すことは不可能である。そのため、ユーザの負担を減らすためや、あるテキストがユーザの要求に合うものか、それを読む価値があるのか、という判断を助けるために、テキスト要約技術が必要となる。

要約はその利用目的に応じて、(1) 原文を参照する前の段階で用いる指示的 (indicative) 要約と (2) 原文の代わりとして用いる報知的 (informative) 要約に分けられる [25]。

- **テキスト分類 (Categorization)**

テキスト分類は、人手より事前に与えられたカテゴリに基づいて分類を行う技術である [36]。テキスト分類は教師あり分類であり、そのため、ユーザは分類対象の文書集合に対する背景知識を要求される。

- **テキストクラスタリング (Clustering)**

テキストクラスタリングは、人手により事前にカテゴリを与えるのではなく、入力テキスト集合の内容に応じてクラスタを検出し、各文書を適切なクラスタに配置する技術である [11]。文書クラスタリングでは、結果として与えられたクラスタの意味はユーザの解釈に委ねられるため、必ずしも期待するクラスタが得られるとは限らない [9]。

- **コンセプト結合 (Concept Linkage)**

コンセプトリンクージとは、共有する概念を識別することにより関係する文書を結びつける技術である。これは、従来の検索方法では発見できなかつたであろう情報の発見を支援する。テキストマイニングにおいて、単語や文書、トピック同士の関係を把握するために、コンセプトリンクージは重要である。

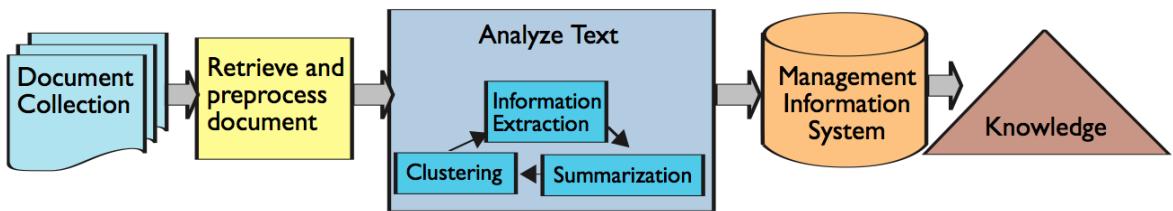


図 2.1: Generic process of model for text mining application (reprinted from [4])

- **情報可視化 (Information Visualization)**

情報可視化とは、インタラクティブな視覚表現を用いて、人がデータを理解することを支援するための技術である [2]。人にとって、単なるデータから情報や知識を得ることは容易ではない。情報可視化の役割は、データを人が認知しやすい視覚表現に変換し、データから十分な量の情報や知識を得ることを支援することである [3]。

- **質問応答 (Question Answering)**

質問応答とは、自然言語による質問に対して、解答を提示する技術である。

また、Gupta はこれを踏まえて、テキストマイニングにおける主要な操作を、特徴抽出 (Feature Extraction)、テキストベースのナビゲーション (Text-base navigation)、検索 (Search and Retrieval)、カテゴリ化 (教師有り分類) (Categorization (Supervised Classification))、クラスタリング (教師無し分類) (Clustering (Unsupervised Classification))、要約 (Summarization) の 6 つにまとめている [5]。

テキストマイニングを目的とするシステムを構築する際に、これらの技術を取り入れることが望ましい。

2.1.1 テキストマイニングにおける探索プロセス

Fan らはテキストマイニングアプリケーションのためのモデルの一般的なプロセスを示している (図 2.1) [4]。文書集合から始まり、テキストマイニングツールは特定の文書を検索し、それに対して前処理を行う。その後、時折様々な技術を繰り返し用いながら、目的の情報が得られるまでテキストを分析する。

上述したように、データマイニングや情報検索、情報抽出は新たな知識の獲得に直接的には寄与しない。しかしながら、ユーザはデータマイニングや情報検索、情報抽出によって得られた結果から、予想外の傾向やパターンを発見することがある。これは、新たな知識を発見するために重要であり、そのような探索には人間の認知能力が必要であることを示唆している。コンピュータはテキストデータの持つ語義曖昧性や文脈依存性を踏まえて容易に扱うことはできないが、人はテキストを見たときに暗黙のうちに言語的パターンを適用し、それを識別できる認知能力を持っている [5]。一方で、人は大量のテキストデータを高速に処理することはできないが、コンピュータはそのような処理能力を備えている。このように、新たな知識を発見するためには、多量のデータを扱えるというコンピュータの能力と、傾向やパターンの微妙な差異を識別できるという人間の能力を組み合わせることが必要である。那須川は、テキストマイニングという技術の本質的な役割は人手による分析処理の自動化ではない、と指摘している [21]。テ

キストマイニングは知識の候補となる“気づき”を与えてくれる技術である。そのため、テキストマイニングでは様々な観点から試行錯誤を繰り返すことが重要であり、単に結果を出力するだけではなく、インタラクティブに内容を確認することが重要な役割を果たす。

Hearst は、仮説検証のために未知の情報を発見するという観点からすれば、テキストマイニングは探索的データ分析 (Exploratory Data Analysis) [38] に該当するタスクである、としている [6]。コンピュータを用いた探索的データ分析を行う場合のプロセスは以下の通りである。

- (1) ユーザは、どのような情報を得たいのかという曖昧な考えをもとに、コンピュータにクエリを与える。
- (2) そのクエリをもとに、コンピュータは結果を得る。
- (3) 得られた結果に基づき、ユーザはデータに対する理解を深めたり、新たなクエリを構成したりする。

探索的データ分析では、ユーザは、目的の情報や知識が得られるまで、問題解決や意思決定に関わる有益な情報を収集していく [13]。加えて、コンピュータで探索的データ分析を効率的に行えるようにするために、ユーザの探索行動と探索行動の振り返りを支援する必要がある [15]。データ分析において、探索行動と振り返り行動の目的は異なる。探索行動は、試行錯誤のプロセスの繰り返しから得られた有益な分析結果を外在化することを目的として行われる。振り返り行動は、ユーザが直面しているタスクにおいて外在化された分析結果を利用するため、それを整理・解釈することを目的として行われる。これらの 2 つの行動は互いに独立しているのではなく、深く関連しており、ユーザは一般的にこれらの行為を適宜状況に応じて切り替えることによって、探索的データ分析を行う。

2.2 Total Environment for Text Data Mining プロジェクト

本節では、TETDM プロジェクトの概要について述べる。

これまで、テキストマイニング及びその要素技術について、多くの研究成果が発表されている。しかし、これらの研究により提案されたツールは論文用の試験的なものが多く、実際に世の中で使われている技術はごく一部に限られている。これらのツールは各研究者によって独自に構築されていることが多いため、個別に導入しなくてはならず、異なるツール間のデータ受け渡しを行う際には手作業でデータフォーマットを整えたり、複数のツールの処理結果を比較するために新たに比較用のインターフェースを実装したりするといった手間が必要であった。そのため、テキストデータを用いて情報を多角的に分析したい場合や従来のマイニングツールと新たに開発したマイニングツールの特徴を比較したい場合に、複数のツールを用いることが容易ではなかった。

このような問題を解決するために、砂山らは TETDM プロジェクトを提唱している [35]。TETDM プロジェクトでは世の中に分散しているテキストデータマイニングツールを画一的に扱える環境の構築を目指し、それに向けた統合環境 (図 2.2 参照) の構築を進めている (以後、本稿では、TETDM プロジェクトにより構築された統合環境を TETDM システムと呼ぶ)¹。

¹TETDM システムの最新版は TETDM プロジェクトのホームページ (<http://tetdm.jp/>) 上で公開されている (平成 27 年 2 月 16 日現在の最新版は ver. 0.62 である。)。

TETDM システムを用いることで、ユーザは複数のテキスト分析ツールを用いたり、それらを柔軟に組み合わせたりすることができる。

前述したように、TETDM の目的は同一環境上で複数のテキストマイニングツールを扱えるようにすることである。そのため、同一のインターフェース上で複数のツールを組み合わせたり、横に並べて表示することができる仕様になっている。ユーザは、自らの興味や目的に応じてツールを選択してそれらを自由に組み合わせたり、処理結果を比較することによって、多角的なテキスト分析を行うことができる。また、この統合環境上では複数のマイニングツール間のインタラクションを想定しており、あるツールの出力結果を他のツールの入力として用いたり、出力結果に対しての操作が並列に表示されている他のツールの処理結果にも即座に反映されたりするようになっている。

TETDM システムに入力されたテキストファイルは、TETDM システムによって共通に行われる前処理(例えば、形態素解析など)の後、各ツールによって処理されるようになっている。TETDM システムでは、各ツールは統合環境内のモジュールとして実装されてる。モジュールとは、テキストマイニングを行う機能ごとにまとめられた要素であり、これを TETDM システムに追加していくことで、取り扱えるテキストマイニング手法を拡張できるようになっている。TETDM システムではテキストマイニングの要諦である二つの技術、すなわちマイニング技術と可視化技術が統合環境内のモジュールとして実装されており、それらをユーザが切り替えることで処理内容と処理結果の表示手法を変更できる。

現状の TETDM システムのインターフェースは、横に並べられた複数のツールパネルから構成されている。その一例を図 2.2 に示す。図 2.2 では、3 つの異なるツールが同時に表示されている。それぞれのツールパネルの上部にある「ツール選択」ボタンをクリックすることにより、使用可能なマイニングツールと可視化ツールの一覧が表示され、それを通じてそれらを選択することができる(図 2.3)。TETDM システムでは、ユーザが自身の分析目的や興味に応じてマイニングツールと可視化ツールをそれぞれ 1 つずつ選択することにより、様々なテキストマイニング技術を利用することができます。

2.2.1 TETDM システムで利用可能なテキスト分析手法

既にいくつかのテキストマイニングツールが TETDM システムにモジュールとして実装されており、利用可能になっている。現在、利用可能なマイニングツールと可視化ツールの一覧を表 2.2、2.3 に示す。加えて、それらを組み合わせて用いることにより実現できるテキスト分析手法について、例をいくつか示す。

テキストの主題関連度を可視化するシステムとして「ひなたシステム」がある [24]。このシステムは、ユーザによって与えられた主題を表す名詞の集合と各文の関連度を計算し、各文の背景色を関連度が低いものほど黒く、高いものほど黄色く彩色する。主題と関連する箇所や全体に対する位置や割合を確認できるようにすることで、ユーザの文章理解の支援を試みている。TETDM システムでは「光と影データ」(表 2.2-1 参照)というマイニングツールと「分布」(表 2.3-1 参照)、「テキスト(カラー)」(表 2.3-3 参照)の 2 つの可視化ツールを組み合わせることで、この「ひなたシステム」を実現できる。図 2.2 の中央パネルが、「ひなたシステム」によって出力された可視化の例である。

また、テキストの主題に基づく一貫性評価を行えるシステムとして「川下りシステム」が考

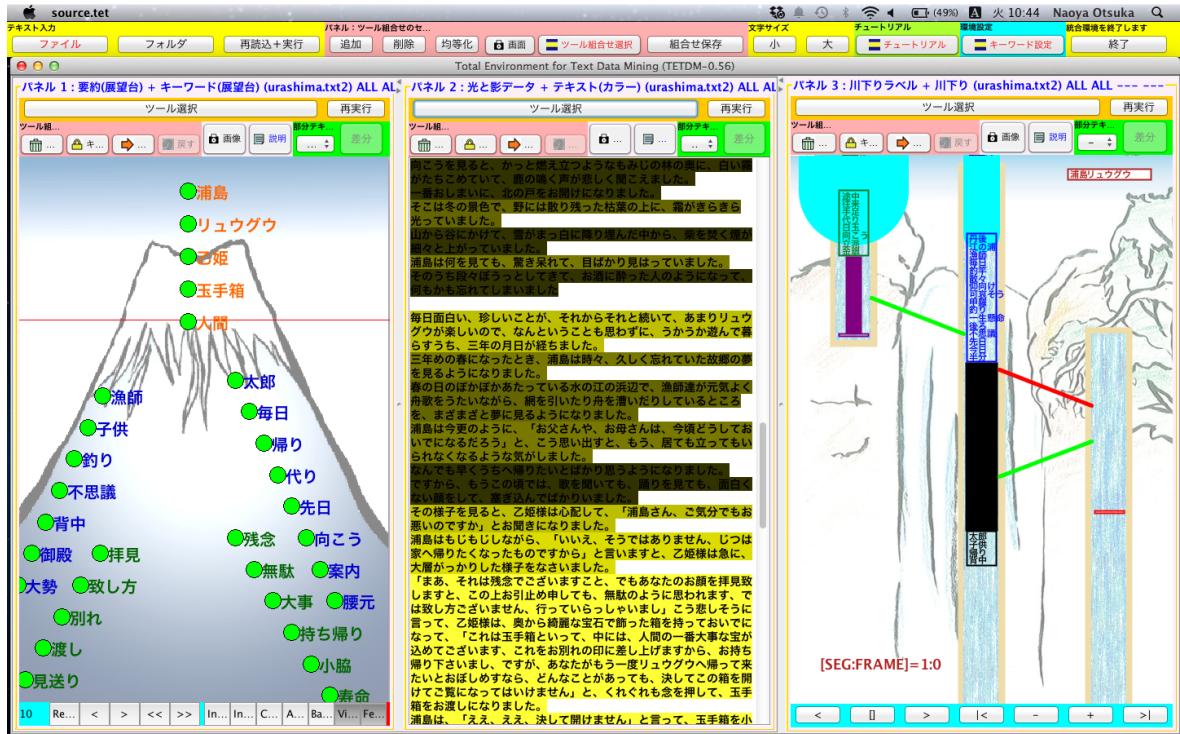


図 2.2: 複数の可視化結果を表示した時の TETDM の例

案されている [33]。川下りシステムでは、テキストとその主題を表す単語の集合を入力とし、テキスト中で使われている各単語に主題との関連を表すラベルを与え、それに基づいた単語分布を可視化することによりテキストの主題に基づく一貫性の評価を行える。TETDM システムでは、マイニングモジュール「川下りラベル」(表 2.2-2 参照) と可視化モジュール「川下り」(表 2.3-4 参照) によって実装されている。また、入力に用いられる主題を表す単語の集合は、統合環境上に実装された展望台システム [32] の出力結果を用いている。図 2.2 の右パネルが、川下りシステムによって出力された可視化の例である。

このように、TETDM システムでは、同一のマイニングツールからの出力に対し、ユーザの目的に応じて異なった可視化手法を適用したり、あるツールの出力結果を別のツールの入力として用いることが可能である。既存ないし将来の研究成果によるテキストマイニングツールを同一環境上に集め、画一的に扱える環境を構築することにより、複数ツールを用いた分析を行う際に余計な作業をする必要がなくなり、分析作業に集中できることが期待される。また、多くの研究成果の実用化の際や複数のマイニング手法の比較の際に利用が見込まれる。

TETDM システムは開発途中であるため、不十分な点も多い。そのため、様々な観点から統合環境の実現に向けた取り組みが行われている。本稿ではそれらの中から、ユーザの支援に関する取り組みについて説明する。

2.2.2 TETDM システムを用いたテキストマイニングの支援への取り組み

前述の通り、TETDM システムは開発途中であり、現状のインターフェースは、テキストマイニングのような探索的な情報アクセスを必要とするタスクを行うには十分ではない。我々は、以前の研究において、その問題を解決するためのアプローチとして、テキストマイニングを行



図 2.3: リスト表示によるツール選択画面

うユーザの試行錯誤の円滑化を目的とし、以下の 2 つの観点から TETDM のインターフェースの再デザインを試みた [26, 10]。

- (1) 直感的な操作によって、テキスト分析ツールの組合せ・選択を可能にすること。
- (2) ユーザが自身が行なっている現在の分析状況の理解を容易にすること。

このインターフェースでは、ユーザが自身の分析状況の把握を容易に行えるようにするために、各ツールをノード、ツール間の処理の流れをリンクとしたグラフ表現を採用している（図 2.4）。ユーザは、そのグラフのノードをマウスで直接操作することによって、使用するツールを切り替えたり、組み合わせを変更したりできる。本研究においてもこれを踏襲する。

中垣内らは、TETDM システムを用いたことがないようなテキストマイニングの初心者ユーザを対象とし、テキスト分析ツールの選択を支援するためのインターフェースを提案している [16]。中垣内らのインターフェースでは、(1) 直感的なツールを選択、と (2) 選択したツールの説明の表示を行えるようにしている（図 2.5）。

中垣内らは、テキストマイニングの初心者ユーザが TETDM システムを用いて、それを行うことができるようなスキルを身につけることを支援するために、TETDM システムにその使い方を学習するためのチュートリアルを実装している [17]。これは、TETDM システムの使い方に関する Q & A を提供するものであり、ユーザはこれに取り組むことによって、TETDM システムを用いてテキストマイニングを行うために必要な基本的なスキルを身につけることができる（図 2.5）。

2.3 Exploratory Search

我々の日常生活において、検索は、様々な状況下で、様々な目的のために行われる。Marchionini はそのような検索活動を Lookup、Learn、Investigate の 3 種類に分類している [12]（図 2.7）。Lookup は、事実検索 (Fact Retrieval) や質問応答 (Question Answering) のような、ある

表 2.2: 主な利用可能マイニングツール

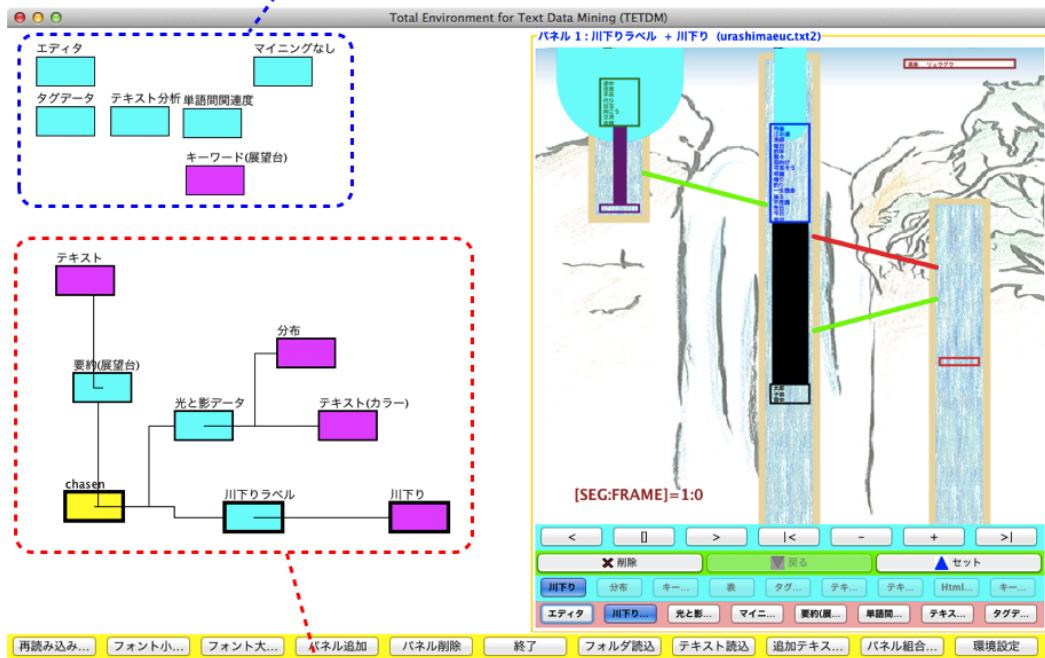
ツール名	内容
1 光と影データ	文章中の各文の主題関連度を評価
2 川下りラベル	テキストの一貫性評価
3 要約 (展望台)	特徴的な単語の抽出、重要文抽出
4 主語抽出	文書内の主語を抽出
5 単語チェック	文章作成時のチェック支援
6 長文チェック	文書中の長文の提示
7 テキスト分析	他モジュールの出力結果の集約
8 Twitter	Twitter の検索結果を入力テキストとする
9 単語間関連度	各単語の単語間関連度を出力
10 タグデータ	単語の出現頻度によるタグクラウド用データを出力
11 アノテーション	アノテーションを付与すべき文の抽出

特定の事柄について知るための検索活動である。Lookup では、ユーザが生成したクエリと検索対象の文書とのマッチングによって、検索を実現している(図 2.8)。これは、データベース管理システムや Web 検索エンジンの開発に伴い発展してきた、最も基本的な検索活動である。Learn は、知識獲得 (Knowledge acquisition) や ある情報や概念についての理解や解釈 (Comprehension/Interpretation)、それらの比較 (Comparison) のような、新たな事柄について学習するための検索活動である。探索過程で得られた情報の意味や考えの理解・解釈、情報や概念の比較といった、単に情報を獲得するのではなく、人が情報を知識として蓄えるような知的活動を含んでいる。Investigate は、分析 (Analysis) や 評価 (Evaluation)、発見 (Discovery) のような、ある事柄について、その周辺の知識を含めて理解を深めるといった調査のための検索である。これらの検索活動は、独立して行われるわけではなく、互いに関連して行われる。

表 2.3: 主な利用可能可視化ツール

ツール名	内容
1 分布	テキスト中の各文の評価値を横向きの棒グラフで表示
2 テキスト	プレーンテキストを表示
3 テキスト (カラー)	テキストの各文や単語に背景色をつけて表示
4 川下り	川下りラベル用
5 キーワード (展望台)	キーワードをその重要度に応じて富士山の画像の上に配置
6 HTML テキスト	html 形式でテキストを表示
7 キーワードマップ	キーワードを 2 次元平面上に配置
8 タグクラウド	単語の出現頻度によるタグクラウド

利用可能なモジュール



利用中のモジュール

図 2.4: 直接操作によるツール切り替えインターフェース [26]

例えば、Learn や Investigate の検索活動を行う場合においても、ある事柄に関する事実を知るために局所的に Lookup 型検索を行うことがある。

Marchionini は、Learn と Investigate を Exploratory Search として特に強調している [12]。このような、曖昧な情報要求に基づく情報探索を説明するモデルとして Exploratory Search がある。[12, 40] Exploratory Search とは、曖昧な情報要求に基づく情報探索を説明するためのモデルである [12, 40]。Exploratory Search は、制限がなく終わりが決められておらず、永続的であり、多面的な情報探索の問題背景と、場当たり的であり、反復を必要とする、多戦略的な情報探索のプロセスを説明するために使われる。

Exploratory Search を行うユーザには以下の特長がある。

- 検索を行っている対象や領域についてよく知らない
- 検索の目的を達成する方法について確信を持っていない
- 検索の目的が曖昧であり、それについて確信を持っていない

そのため、ユーザは探索を繰り返すことによって、検索対象の情報空間について徐々に理解を深めながら、自身の情報要求と目的の達成方法を明確にしていく。

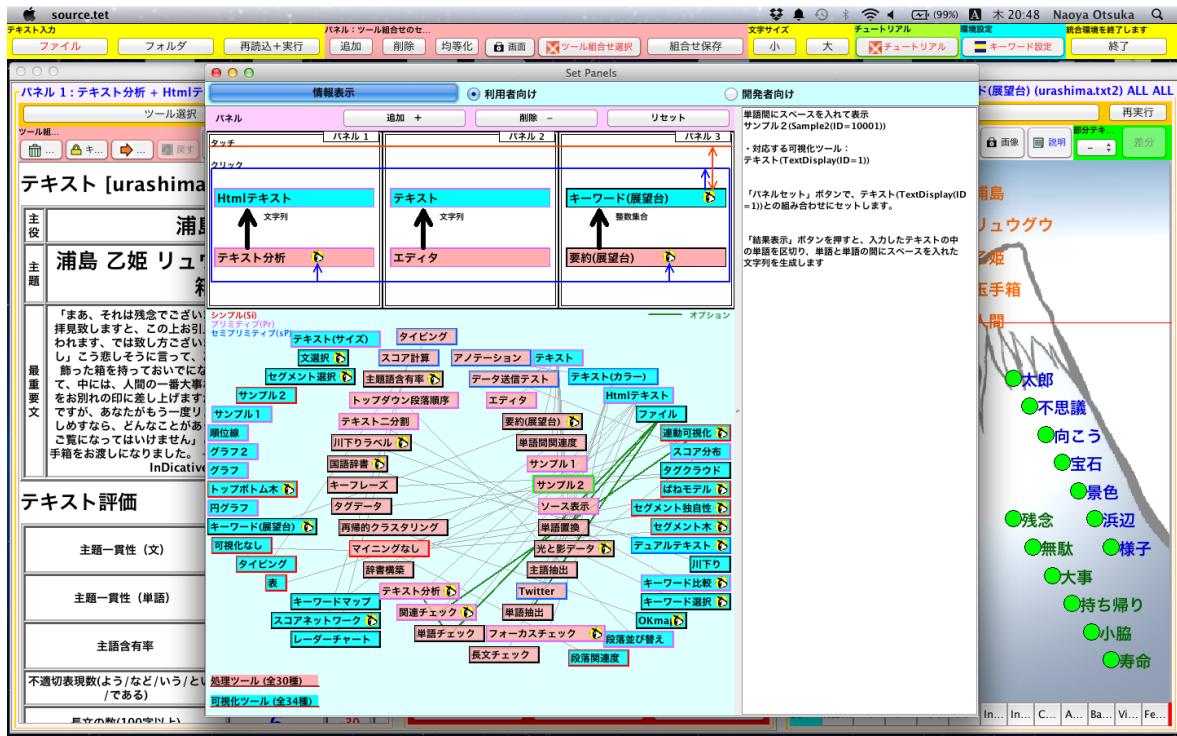


図 2.5: グラフ表示によるツール選択画面

2.3.1 情報要求と探索行動の関係

Taylor は、ユーザの情報要求の明確さの程度を以下のように 4 つの段階に分類している [37]

- 直感的な要求 (visceral need)

この段階では、ユーザは現状に対する漠然とした不満を持っているが、何が問題でそれをどのように解決できるのかがわかつていない。そのため、ユーザは自身の要求を具体的な言語表現として顕在化させることができない。

- 意識されている要求 (conscious need)

この段階では、visceral need と比べると明確に問題を認識しているが、それを曖昧な表現や整理されていない表現でしか顕在化できない。

- 形式化された要求 (formalized need)

この段階では、ユーザは明確に問題を整理し、それを具体的な言語表現として顕在化できるが、どうすれば解決できるかについてはわかつていない。

- 解決に至る要求 (compromised need)

この段階では、ユーザは明確かつ具体的に問題を顕在化させることができ、それだけでなくそれを解決する上でどのような情報源を利用すればよいかも理解している。

ユーザの情報要求が formalized need や compromised need の段階である場合、すなわち、明確である場合、ユーザは自身の情報要求をクエリとして表現することができる。このような、



図 2.6: TETDM システムのチュートリアル

ユーザの情報要求が明確であり、ユーザがそれをクエリとして生成可能な場合、検索エンジンなどの Lookup 型検索を利用することにより、ユーザは必要な情報を発見することができる。しかし、ユーザの情報要求が visceral need や conscious need の段階である場合、すなわち、曖昧である場合、ユーザは Lookup 型検索では自身の要求を満たすことができない。そのような場合、ユーザは自身の曖昧な情報要求を基に、ある事柄やその周辺領域についての知識を獲得しながら検索を繰り返し行い、自身の情報要求を明確にしていく必要がある。また、検索を繰り返すことによって、ユーザの要求、興味が変化する場合もある。それによって、ユーザは徐々に目的の情報へと近づいていく。

Exploratory Search は、多くの点で、特に、探索の初期段階において、探索的データ分析 [38] と類似している。Exploratory Data Analysis では、ユーザは尤もらしい “story” が出てくるまで、多くの観点でデータを探索していく。Exploratory Search においても、このように、ユーザは仮説生成と仮説検証を行なっていくことによって、問題背景を明らかにしていく。

2.3.2 Exploratory Search におけるユーザの振る舞い

Exploratory Search におけるユーザの振る舞いには、Exploratory Browsing と Focused Searching の 2 つの主要なプロセスがある。Exploratory Browsing は、探索空間を広げる方向への検索である。Exploratory Browsing のプロセスによって、ユーザは検索対象への理解を深めていく。また、仮説を生成したり、どのような情報が利用可能なのか理解するために、ユーザは複数の文書を閲覧する。Focused Searching は、探索空間を絞る方向への検索である。Focused Searching のプロセスでは、ユーザは明確な目的とそれを達成するための手がかりを持っている。それらによって、ユーザは検索結果を調査したり、Exploratory Browsing のプロ

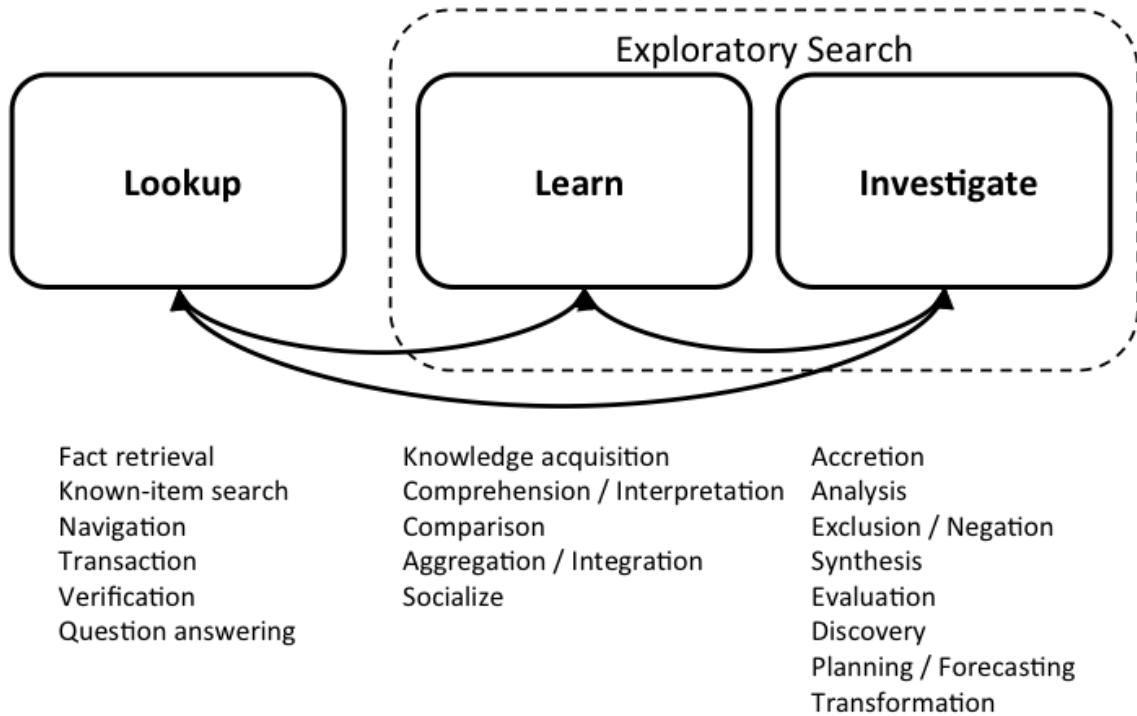


図 2.7: Three kinds of search activities; lookup, learn and investigate (reprinted from [12]).

セスによって生成された仮説を検証したりする。このように、Focused Searching のプロセスでは、ユーザは自身の目標を達成するために、関連する文書を抽出する。Exploratory Browsing の段階では、ユーザは一般的に、曖昧で不明確な情報要求しか持っていない。従って、ユーザは Exploratory Browsing と Focused Searching によって探索空間の発散と収束を繰り返しながら探索していくことにより、探索プロセスの中でユーザは自身の情報要求を明確にしていく。

Exploratory Search における重要な点は探索空間の拡大と縮小である。従来の検索活動では、探索空間の領域は一点に集中される(図 2.9 左)。つまり、反復検索 (iterative search)において、検索者は探索空間を縮小していく方向の検索により、目的の情報にたどり着く。一方で、Exploratory Search では、探索空間の領域は様々な点に集中される(図 2.9 右)。つまり、ユーザは探索空間を拡大しながら、様々な情報を探索していくことが重要である。しかし、Exploratory Search では、ユーザの情報要求は探索の開始時に必ずしも明確になっていない。そのため、探索の開始点が不適切であれば、適切な情報を得ることはできない。Exploratory Browsing は、そのどこを探すべきなのかそのものを探すための Search である。つまり、ユーザは、Focused Searching を行うための空間を探すために Exploratory Browsing を行う。

Exploratory Search では、ユーザは、このようなプロセスを繰り返しながら、新たな気付きや知見、興味を得たり、それに基づいて仮説を生成したりする。このように、探索空間の拡大と縮小を繰り返すことによって、探索プロセスの中で自身の要求を明確にしながら、目的の情

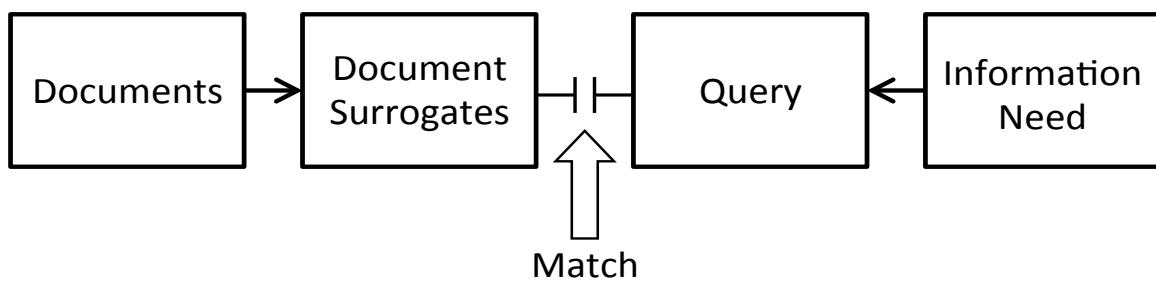


図 2.8: Information-seeking model of lookup search [12] (reprinted from [40]).

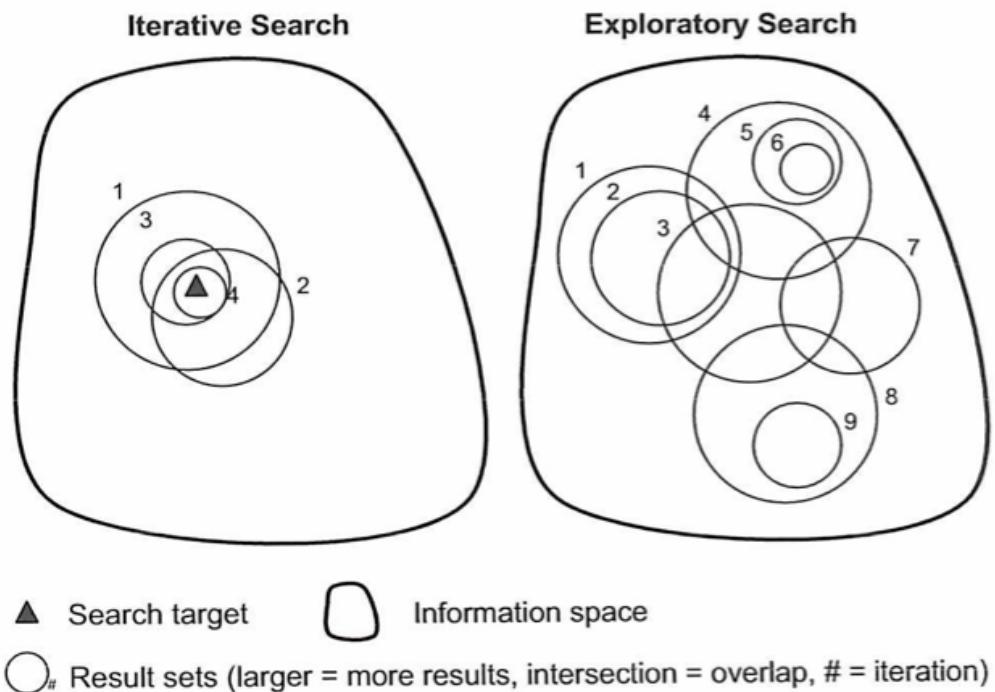


図 2.9: Comparison of iterative search and exploratory search (reprinted from [40]).

報を得ていく。

2.3.3 Exploratory Search のシステムの必要要件

White は Exploratory Search を支援するためのシステムの特徴を整理し、その必要要件として以下を挙げている [40]。

(1) クエリ入力とクエリ洗練の支援

システムは、ユーザがクエリを容易に定式化したり、変更・調整できるようにしたりすることによって、検索結果の表示を素早く調整できるように支援する必要がある。

(2) 検索結果のフィルタリング機能

システムは、ユーザがカテゴリやファセット、文書のメタデータを選択することによって、探し結果をフィルタリングできるようにする必要がある。

(3) 検索のコンテキストの利用

システムは、そのユーザやシステムの状況、現在の検索タスクについての利用可能な情報を活用すべきである。

(4) 洞察や意思決定を支援するための可視化

システムは、仮説生成や動向の把握を支援するために、探索対象の情報や検索結果を視覚表現として提示する必要がある。

(5) 学習と理解の支援

システムは、ユーザの現在の知識・技能レベルに応じて、ユーザが知識と技能の両方を獲得するのを支援する必要がある。

(6) 協調作業の促進

システムは、タスク分割と知識共有の支援において、ユーザ間の同期及び非同期のコラボレーションを促進する必要がある。

人は各々、異なる観点、経験、専門性を持っているため、複数人で協調的に情報を探すことは、1人だけで情報を探すよりも効果的な場合もある。

(7) 探索履歴や、ワークスペース、進捗アップデートの提供

システムは、ユーザが探索プロセスにおける有益な情報の断片を素早く振り返り、保存し、操作することを可能にする必要がある。また、目的の情報への進捗の更新を提供する必要がある。

(8) タスク管理の支援

システムは、複数セッションや複数ユーザが関わる Exploratory Search のシナリオを支援するために、ユーザが検索タスクを保存、取得、共有できるようにする必要がある。

Exploratory Search のような目的が不明確であるという検索特有の性質は、これらの特長を踏まえてユーザを支援するための新たなシステムデザインの必要性を示唆している。そのため、Exploratory Search に該当するタスクを支援するシステムを構築する際にはこれらの要件を考慮すべきである。

2.4 知識創発支援に関する研究

前述の通り、テキストマイニングという技術の本質は、新たな知識の発見・創出である。そのような新たな知識の発見・創出を目的として、創造的活動支援 (Creativity Support) や 創造的情報創出 (Creative Knowledge Work) と呼ばれる研究が行われている [7, 18]。

創造的活動とは、人々が行う、音楽や美術などの芸術活動、文学的な創作活動、人工物の設計と製造などの光学的な活動、理論創出などの科学的な活動などのことである。堀は、「創造的活動支援の研究とは、人々の創造的な活動のプロセスを変化させ、今までよりも更に創造的に仕事をすることが可能となるような支援の方法を考え、実際に主としてコンピュータを使つ

た支援システムを構築する研究である」と述べている [7]。また、創造とは「新しくてかつ意味のあるものをつくること」としている。堀は、通常の思考と創造的な思考の違いについて以下のように説明している。

- 通常の思考

ある定まった複数の軸から成る空間の中で変数の値をいろいろと変化させる

- 創造的な思考

異なる軸から構成される別の思考空間へと思考空間そのものが変化する

中小路らは、創造的情報創出とは「もやもやとしたゴールや意図を基に、具体的に何をどう創出するかという具体的なプランがない状態から開始し、漸次的に情報の断片を構築していくながら、何を情報として創出するかと、要求されるフォームとしてどう表現するかとを、同次並行的に同定していく作業を指す、としている [18]。また、山本らは、(1) ユーザは曖昧なゴールを基に知識創発のプロセス始め、(2) ユーザは徐々に情報の断片を得ながら情報を創出するための計画とその表現方法を同時並行的に考え出すという作業である、としている [41]。

このように、創造的活動において、創出される情報や知識はあらかじめそのフォームが明確に定まっているようなものではない。加えて、どのように目的の情報や知識を創出するために、どのような手順を取ればよいかも明確ではない。

このような点において、創造的活動のプロセスは Exploratory Search のプロセスと類似している。

Shneiderman は、創造活動に含まれる要素として、(1) Collect、(2) Relate、(3) Create、(4) Donate の 4 種類の活動があるとしている [30]。

また、創造的活動を支援するためには以下の 8 種類のタスクが必要である [31]。

- (1) **Searching:** Web やその他の電子的リソースの検索・閲覧すること
- (2) **Visualization:** 関係性を理解・発見するためにデータやプロセスを可視化すること
- (3) **Consulting:** 知的・精神的支援のために仲間やメンターに相談すること
- (4) **Thinking:** 新たなアイデアの組み合わせを作るために自由連想によって思考すること
- (5) **Exploring:** 解決方法を探索すること
- (6) **Composing:** 成果物を段階的に構成すること
- (7) **Reviewing:** 振り返りを支援するためにセッションの履歴を確認・再生すること
- (8) **Disseminating:** 認知度を高めるため、検索可能なリソースを追加するために結果を広めること

図 2.10 に、4 種類の活動と 8 種類のタスクの関連性を示す。ユーザは、これらのタスクを繰り返し適用することによって、創造的活動に含まれる 4 つの活動を達成することができる。

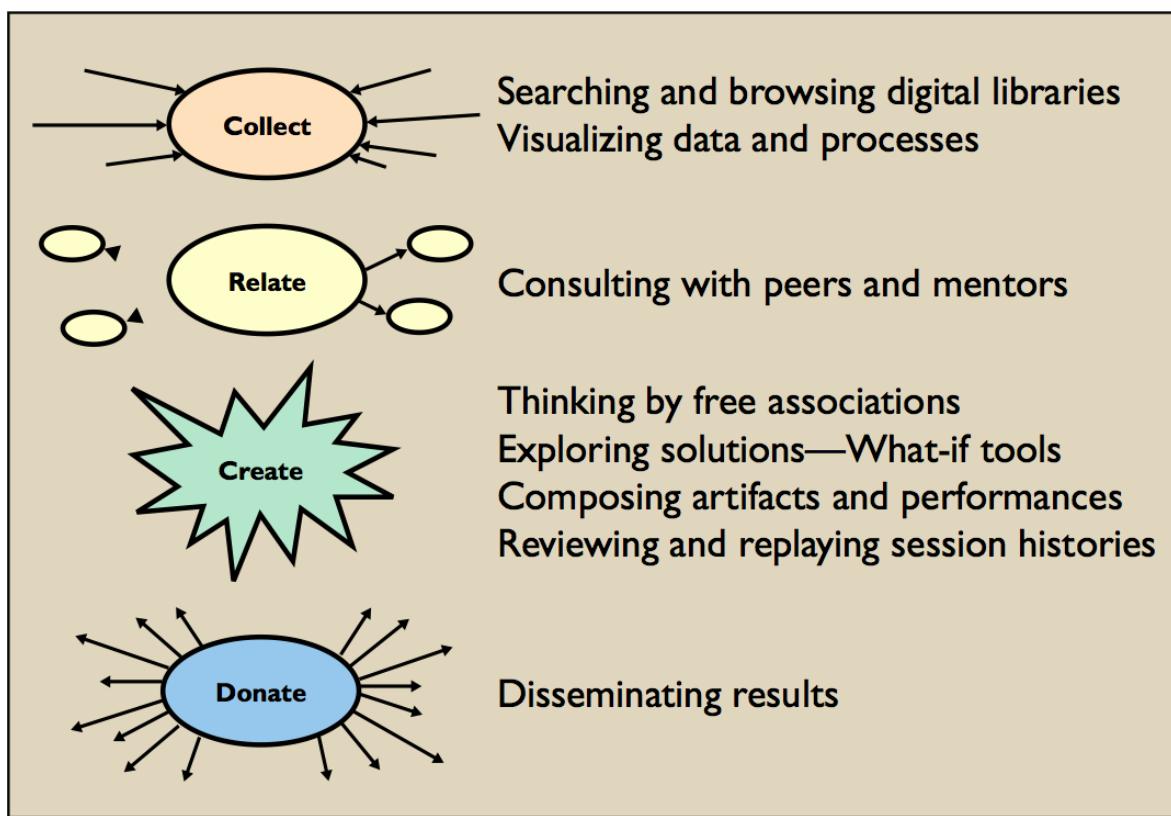


図 2.10: 創造活動における 4 種類の活動と 8 種類のタスクの主要な関係 (reprinted from [31])

2.4.1 創造活動における外在化の役割

中小路らは、創造的思考における外在化の役割とその重要性に着目し、アプリケーションシステムを創造的情報創出を行う際の外在化の手段とみなして様々なシステムをデザインしている [18, 41]。それに伴い、創造的情報創出のためのシステムをインターラクションデザインを行う際の外在化の要件を以下のようにまとめている [18]。

- 曖昧さを表現できる表現系
- “解”と“問題”とを表せる表現系
- 作りかけの“部分”と、出来上がりつつある“全体”とを同時に概観できる表現系
- これらを、“直感的に”操作できる操作系

創造活動のプロセスにおいて、思考を曖昧な表現として外在化できるようにすることで、ユーザは様々な視点で柔軟にそれを解釈でき、創造的な思考を促すことができる。これらの要件は、それぞれの項目が個別に満たされればよいものではなく、相互に依存しており、創造的情報創出のためのアプリケーションシステムのインターラクションデザインを行う際のチェックリストとしてみなされるべきものだとしている。

中小路らは、このような創造的情報創出のための外在化を踏まえたインターラクションデザインを行い、アプリケーションシステムを開発するための方法論として、ナレッジインターラクショ

ンデザイン (KID: Knowledge Interaction Design) を提案している [18]。これは、情報を創出する際の人間の思考と行為のモデルを構築し、それを沿ったインタラクションのモデルを作成する、というものである。思考と行為のモデルとは、たとえば、「並べて比較する」、「つながりを認識する」、「フォーカスを移動する」といった、情報を創出する際の人間の思考と行為のプロセスを領域横断的に認識できる粒度でまとめて表現したものである。ナレッジインタラクションデザインの枠組みでは、このような思考と行為のモデルに基づいて、インタラクションモデルを構築する。「インタラクションモデル」とは、

- タスクで対象とするオブジェクトを表現する上で必要となる視点の種類と表現の仕方
- ユーザがそれらの表現されたオブジェクトに対して行える操作とその意味解釈
- 視点間での表現間の連携

を記述するものである。

創造活動のプロセスにおいて、なにか思い付いたことがあったときに、それをどのように外在化表現として表現できるのかは、創造活動によって最終的に創出される情報の質に大きく影響を及ぼす [7]。そのため、創造活動支援のためのシステムをデザインする際には、その対象に応じて、外在化表現を表出す手段、表現方法、操作方法について、考慮する必要がある。

3 デザイン指針

本章では、本研究において提案するインターフェースをデザインするために我々が定めたデザイン指針について述べる。

本研究では、(1) 探索プロセスの円滑化、(2) 探索プロセスにおける思考の外在化、(3) 探索プロセスの振り返り支援、の 3 つに着目してデザイン指針を策定した。

テキストマイニングは Exploratory Search に該当するタスクであり、新たな知識を創出する創造的活動である。本章では、2 章で述べた、Exploratory Search と創造的活動のそれぞれの支援に必要な要件を参照しながら、テキストマイニングを支援するための要件について整理する。

3.1 探索プロセスの円滑化

テキストマイニングにおけるユーザの主な探索プロセスは、(1) 分析対象のテキスト集合を入力する、(2) テキスト分析手法の選択、(3) 得られた結果の解釈、である。加えて、ユーザは試行錯誤を行いながらこれらのプロセスを繰り返していく。テキストマイニングのためのアプリケーションシステムでは、これらの一連のプロセスを円滑に行えることが望ましい。そのため、ユーザの探索活動を妨げないような直感的な操作手法を提供し、探索におけるユーザの試行錯誤を支援することが必要である。

ユーザはテキスト集合を分析して何らかの結果を得ようとする場合、(1) テキスト集合からそれにどのような処理を施しどのようなデータ構造へ変換するのか、(2) 構造化されたデータをどのような視覚的表現と対応させるのか、を決定する必要がある。そのため、テキスト分析のためのアプリケーションシステムでは、ユーザがそれらの手法を円滑に決定できるよう支援することが必要である。

- 直感的な操作手法の提供

テキストマイニングにおいて、ユーザは、2.1.1 節で示したようなプロセスを遂行するために、分析対象のテキスト集合や分析ツールを選択・変更したり、各分析ツール内の変数を調整するなどの操作を行う。そのため、それらの操作を円滑に遂行できるようにするためには、その操作方法はユーザが直感的に理解できるものでなければならない。

ユーザは、自身の操作によって変更された処理結果を解釈して、更に分析を繰り返していく。このようなインタラクティブに情報へアクセスする必要のあるタスクを扱うシステムでは、現行の TETDM システムが提供しているようなメニュー形式のツール選択インターフェース (図 2.3 参照) ではなく、各要素 (テキスト集合やテキスト分析ツール) を直接操作 (Direct Manipulation) [28] により操作するようなインターフェースを提供することが望ましい。

- ユーザの試行錯誤の支援

テキストマイニングでは、様々なテキスト分析ツールを適応的に組み合わせて用いることによって、テキスト集合を様々な観点から分析する必要がある。そのような分析プロ

セスにおいて、ユーザは自身の興味や分析の目的に応じて、テキスト分析ツールを選択・変更したり、得られた結果を基に分析を更に繰り返していく。

しかし、2.3.1 節で述べたように、ユーザがテキスト集合を分析する場合、ユーザは、何をどのように分析し、どのような結果を得たいか、という具体的な問題や目的を持っているわけではない。すなわち、ユーザがテキスト分析においては、ユーザの情報要求は必ずしも明確ではない。このような場合、ユーザの情報要求は直感的な要求や意識されている要求の段階であり、ユーザは、何か問題であり、どのようにすれば解決できるについて明確に認識していない。そのため、ユーザが自身の要求に沿ったテキスト選択ツールを選択することは困難である。

そのような場合、ユーザは、様々なテキスト分析手法を用いながら試行錯誤を繰り返していく。それによって、ユーザは分析対象のテキスト集合に対する理解を深めて、徐々に自身の情報要求を明確にしながら、最終的に何を知識として創出するのかという探索の目的と、どのような方法を用いてそれを達成するのかという解決方法を明確にしていく。

こうした分析プロセスを効率良く行えるように支援するためには、そのためのアプリケーションシステムのインターフェースは、2.3.2 節で挙げたようなユーザの試行錯誤のための振る舞いを妨げるようなものであってはならない。そのため、ユーザがテキスト分析手法を円滑に選択・切替できるように、ユーザの試行錯誤の支援する必要がある。

- 次に実行可能な探索プロセスの把握

テキストマイニングでは、ユーザは自身の要求に応じて探索を繰り返していく。このとき、ユーザは、次に遷移可能な探索プロセス集合の中から自身が最適であると考えるものを選択する、という行為を繰り返す。加えて、ユーザは、次に行う探索プロセスを決定するとき、現在行っている探索プロセスを考慮する。つまり、ユーザは、自分が今までにどのような探索を経て、どのような知見や仮説を得て現在の状態に辿り着いたのかという、探索のコンテキストを考慮して次の探索プロセスを決定する。そのため、2.3.3 節で述べたように、ユーザは自身の探索のコンテキストを常に把握し、現在行っている探索に引き続いて、どのような探索が行えるのかを把握できることが望ましい。また、それを支援するために、システムは、探索のコンテキストと次に実行可能な探索プロセスをユーザへ明確に提示する必要がある。

Exploratory Search を支援するためには、ユーザがクエリを容易に変更・調整できるようにする必要がある。そのため、直感的な操作手法の提供や、ユーザが分析ツールが容易に変更・切り替えできるようにすることは、テキストマイニングの支援に繋がると考えられる。加えて、検索のコンテキストを利用することや、検索対象の情報や検索結果の可視化を用いることが必要である。また、創造的活動の支援においても、データや創造的活動のプロセスを可視化することは必要とされている。

本研究における提案インターフェースでは、これらの要件に答えるために、各々の分析ツールなどのテキストマイニングにおけるユーザの操作対象を視覚表現として表現し、それを直接操作することによって、テキスト分析ツールを選択・切り替えを行えるようにする。

3.2 探索プロセスにおける外在化

ユーザは、探索行為によって得られた知見や生成された仮説を外在化し、その外在化された表現をきっかけとして次の探索行為を行っていく。すなわち、探索プロセスにおいて外在化された知識や仮説は、次の思考への“補助線”として用いられる。

松下は、ユーザが情報を探索する際に行う外在化行為について、直接的外在化 (direct externalization) と暗示的外在化 (allusive externalization) を区別している [14]。直接的外在化とは、探索プロセスにおいてユーザが得た知見や生成した仮説などを外在化することである。つまり、ユーザによって意識的に行われる外在化行為である。暗示的外在化とは、オブジェクトの位置変更やサイズ変更などの、ユーザによって無意識的に行われる外在化である。ユーザは、直接的外在化と同様に暗示的外在化も過去の探索行為を振り返る際の手がかりとして用いる。

ユーザは探索プロセスにおいて、試行錯誤を繰り返しながら知見を得たり、仮説を生成していく。また、それらのプロセスにおいて、ユーザは、意図的に知見や仮説を外在化していく直接的外在化は、このような、ユーザが何らかの知見を得たり、仮説を生成するという、意識的に“何かに気づいた”考えた場合にのみ行われる。そのため、ユーザが探索プロセスにおいて得た知見は有益であったが、ユーザ自身がその重要性に気づいていない場合、直接的外在化は行われず、後にユーザがその重要性に気がついた場合においてそれを手がかりとして探索行為を振り返ることはできない。このような場合、ユーザは、自身の記憶と暗示的外在化を結びつけ、それを手がかりとして探索行為を振り返る。

テキストマイニングのような分析内容が複雑になるタスクにおいては、直接的外在化と暗示的外在化のどちらも探索プロセスの振り返りとして用いられる。そのため、直接的外在化と暗示的外在化にそれぞれに対して適した表現方法を提供することが必要となる。

- 探索プロセスによって得られた知見の外在化

テキスト集合を分析することによって、知見を得たりや仮説生成したユーザは、次の探索への手がかりとするために、それらを外在化する。そのため、コメントの付与などのそれを外在化できるような表現手法が必要となる。また、ユーザが、コメントなどの外在化されたオブジェクトを自由に移動させられるようにするなど、それらの外在化表現を整理できるようにする必要がある。

- 探索プロセスの視覚化

ユーザは、以前の探索プロセスを振り返る際に、上述したような直接的外在化のみではなく、システムの操作や視覚的情報、自身の記憶を頼りに振り返りを行うことも考えられる。そのため、現在の探索の状態や探索プロセスの履歴を保存しておき、それらを視覚的に概観できることが望ましい。

2.4.1 節で述べた創造活動支援のための外在化の要件によると、

- 曖昧さを表現できる表現系
- “解”と“問題”とを表せる表現系
- 作りかけの“部分”と、出来上がりつつある“全体”とを同時に概観できる表現系

- これらを、“直感的に”操作できる操作系

が必要であるとされている [18]。

直接的外在化は、ユーザが探索プロセスにおいて何らかの知見を得たときに行われる。そのため、直接的外在化の手段としては、アノテーションやコメントの付与などを明示的に表せる手法が必要である。また、それらの外在化したオブジェクトをユーザが移動させられるようにして、暗示的外在化を引き起こすことができる。これにより、“解”や“問題”を表現できる。また、現在の探索の状態や探索プロセスの履歴を視覚的に概観できるようにすることにより、探索における“部分”や“全体”を表現できる。

本研究において提案するインターフェースでは、ユーザが探索中に得た知見をテキスト情報として外在化できるようにする。

3.3 探索プロセスの振り返り支援

テキストマイニングは、一度の分析のみでその目的が満たされるような one-shot の探索プロセスではなく、試行錯誤を行いながら何度も探索を繰り返していくという、複数の探索プロセスから構成されるタスクである。そのため、ユーザは自身の興味や関心に基づいて分析を行い、それによって問題解決や意思決定のための知見を得たり、仮説を生成したりといった、プロセスを繰り返す。ユーザは、このようなプロセスの繰り返しによる一連の探索プロセスを経て蓄積された知見や仮説と、他の様々な情報を統合的に判断し、最終的に創出する知識を構築していく。つまり、ユーザは自分が直面しているタスクにおいて得られた知見や生成した仮説を利用するため、改めてそれを整理し、解釈するといった振り返り行為を行う。そのため、テキストマイニングによるユーザの知識創出を支援するためには、ユーザが以前の探索プロセスを振り返るという行為を支援する必要がある。そのためには、ユーザが過去の探索プロセスの履歴を保存しそれを参照できるようにし、ユーザが以前の探索プロセスを把握できるようにすることが必要である。

- 探索プロセスの履歴の保存

ユーザは探索プロセスで得た知見や仮説を確認したり整理したりするために、過去の探索プロセスを振り返ることがある。そのため、ユーザが得た知見や生成した仮説を探索プロセスの履歴として保存できるようにする必要がある。

2.1 節で概観したように、テキストマイニングでは、複数のテキスト分析ツールを用いて、多面的な探索を行うことが求められる。加えて、ユーザは、自身の興味や関心に応じて、テキスト分析ツールを試行錯誤的に組み合わせながら、探索を進めていく。そのため、ユーザは探索プロセスによって得られた知見や仮説だけではなく、それらを得るために用いた分析手法や可視化手法に対する振り返りを行うことも考えられる。そのため、探索プロセスにおける各分析手法による結果のみではなく、ユーザがどのように分析したのかという手法についても、探索プロセスの履歴として保存しておく必要がある。

- テキストマイニングにおけるユーザの思考の単位

テキストマイニングでは、ユーザは様々なテキスト分析手法を試行錯誤に用いて探索を繰り返しながら、知見を得たり仮説を生成したりする。ユーザの振り返り行為は、探索プロセスにおいて得られた知見や生成された仮説をきっかけとして行われると考えられる。すなわち、テキストマイニングにおいて、ユーザが以前の探索への振り返りを行うのは、自身が探索プロセスにおいて“何かに気づいた”ときである。そのため、ユーザは、知見を得たときや仮説を生成したときを思考の区切りとして認識していると考えられる。これは、テキストマイニングにおけるユーザの思考の単位は、ユーザが気付きを得た瞬間を区切りとして表されることを示唆している。

ユーザが探索プロセスの履歴を振り返る際には、ユーザは探索プロセスの履歴へ自身の思考したプロセス単位でアクセスできることが望ましく、そのためには、ユーザが知見を得たり、仮説を生成したと判断したときに、履歴を保存が必要である。

● 振り返りによる知見の整理

テキストマイニングでは、ユーザは自身の目的や興味に基づいて試行錯誤を行いながら探索を繰り返す。そのため、その探索プロセスは、必ずしも一本のパスから構成されるものではなく、様々な分岐が存在する複数のパスから構成されるものであると考えられる。そのような一連の探索プロセスで得た知見を用いてそれを知識へ昇華させるためには、重複した探索プロセスや不要な探索プロセスを削除することによって、得られた知見を整理し、再解釈する必要がある。

それを行えるよう支援するためには、ユーザは過去の探索プロセス全体を俯瞰できることが望ましい。

Exploratory Search と創造活動の支援のための要件について、どちらにおいても、システムやユーザ、タスクの状況などの検索のコンテキストを利用することや、有益な知見を得た際の探索プロセスの振り返りの支援が必要であるということが示唆されている。また、ユーザは振り返り行為によって、探索プロセスで得られた知見を整理していく。振り返り行為を適切に行えるように支援することによって、Reviewing や Composing を促すことでユーザの“創造的な思考（2.4 節）”を導き、その帰結として知識を段階的に構成していくことができるようになる。これらは、創造的活動の支援において重要な要素である。

本研究では、これらの要件を満たすために、探索プロセスの履歴をツリー構造で表現することによって、ユーザが今までの一連の探索プロセスを把握できるようにする。

4 プロトタイプシステムの実装

本章では、3章で示したデザイン指針に基づいて実装した提案インターフェースのプロトタイプについて述べる。プロトタイプシステムは既存の TETDM システム¹をベースとし、Java を用いて実装した。

4.1 画面構成

プロトタイプシステムのスクリーンショットを図 4.1 に示す。プロトタイプシステムは、(1) 使用するツールを選択するためのツール選択ウィンドウ(図 4.1-A)、(2) 探索の履歴を表示するための履歴ツリーウィンドウ(図 4.1-B)、(3) 可視化ツールの結果を表示するための結果表示ウィンドウ(図 4.1-C)、から構成されている。なお、今回実装したデザイン指針に基づく提案インターフェースは(1)及び(2)であり、(3)は既存の TETDM システムから踏襲している。以下に、提案インターフェースにおけるそれぞれの画面の役割について述べる。

- ツール選択ウィンドウ

ツール選択ウィンドウは、システムにおいて使用可能なツールのリスト(図 4.2-A)とツールの利用状況を表すグラフが表示されているワークスペース(図 4.2-B)、後述する機能を使用するための操作ボタン部(図 4.2-C)から構成される。

ツールリストには、使用可能なマイニングツールと可視化ツールが表示されている。ツールリスト内のツールは、マウスでクリックすることによって、ワークスペースに追加することができる。

ワークスペースには、現在用いられているツールをノード、それらのツール間の処理の流れをエッジとするグラフが表示されている(図 4.3-a)。ユーザは、このグラフのノードを操作することによって、分析手法の選択・変更を行うことができる。また、ワークスペース上のマイニングツールノードと可視化ツールノードにはそれを削除するためのボタンが用意されており、ユーザは分析で不要になったツールをワークスペースから取り除くことができる。加えて、後述する付箋機能によって、ユーザが探索プロセスによって得た知見をテキスト情報として外在化できる。ユーザは、ワークスペースを用いて、グラフのノードを操作したり、得られた知見を外在化していくことにより、テキスト集合の探索を行うことができる。

- 履歴ツリーウィンドウ

履歴ツリーウィンドウには、探索プロセスの履歴をノードとした履歴ツリーが表示されている(図 4.4)。ユーザは、履歴ツリーを見ることによって、過去の探索履歴を概観したり振り返ったりすることができます。

- 結果表示ウィンドウ

結果表示ウィンドウには、ユーザが選択したマイニングツールと可視化ツールの組合せによる処理結果が表示されている。結果表示ウィンドウは、既存の TETDM システムの

¹version 0.56 (stable version, 2014/3/31 公開)

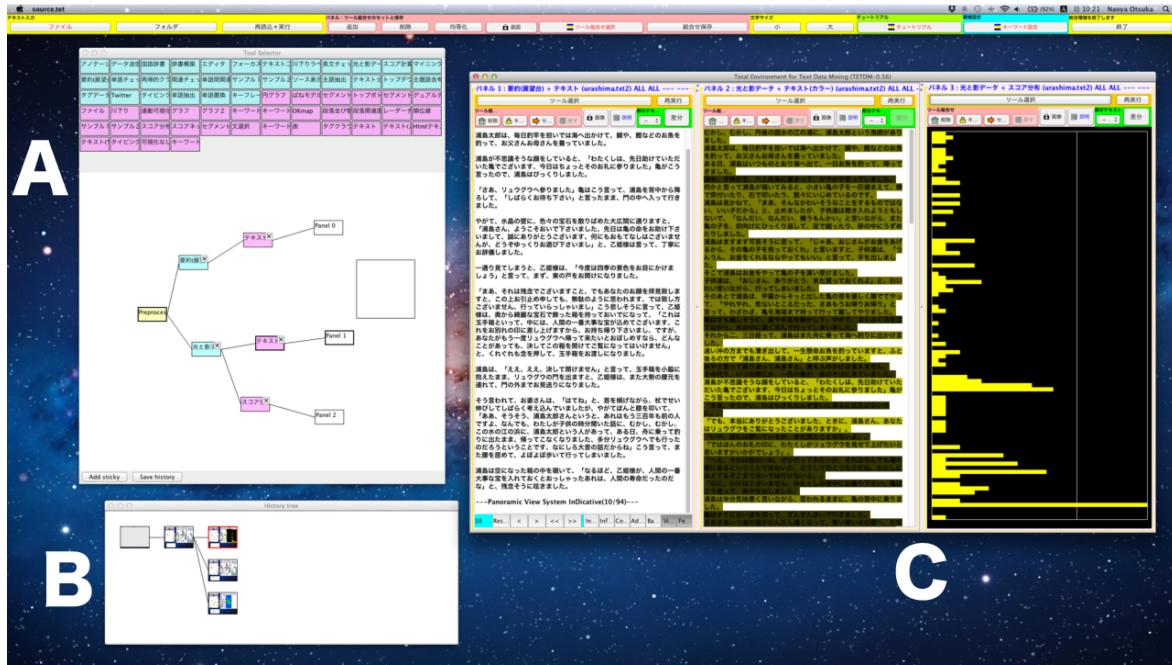


図 4.1: プロトタイプシステムの画面構成

ものを踏襲している。プロトタイプシステムでは、ツール選択ウィンドウ及び履歴ツリー ウィンドウ内での操作に応じて、TETDM システムのマイニングツールと可視化ツールの組合せが変更され、表示結果ウィンドウに表示される結果が変更されるようになって いる。

4.2 グラフ表現による分析状況の視覚化

提案インターフェースでは、現在の分析状況を TETDM システムにおける各ツールをノード、それらを処理の順に繋いだ線をエッジとするグラフとして表現した。グラフ表現により視覚化することによって、各ツールの利用状況を容易に把握できるようにした(図 4.3 参照)。

TETDM システムでは、マイニングツールと可視化ツールの 2 種類のツールが用意されており、ユーザはそれらを一つづつ選択して組み合わせることによって、様々なテキスト分析手法を用いることができる。TETDM システムに入力されたテキストファイルは、システムによつて共通に行われる形態素解析などの前処理の後、マイニングツール、可視化ツールによって処理され、選択されているパネルにその結果が表示される。提案インターフェースでは、TETDM システムによって用意されているマイニングツールと可視化ツールに加えて、システムによる形態素解析などの前処理と選択されたツールによる結果を表示するパネルを表すノードを用意した。これにより、図 4.3 のように、入力されたテキストから結果の表示に至るまでの処理の流れを把握することが可能となる。提案インターフェースにおけるノードの種類とその詳細を表 4.1 に示す。

提案インターフェース上に多数のツールが表示されている場合、それぞれのツールの種類を判別することが困難になる。そのため、TETDM システムにおける処理の段階に応じて、それぞ

A

Tool Selector														
アノテーション	データ送信	国語辞書	辞書構築	エディタ	フォーカス	テキスト二	川下りラベ	長文チェック	光と影データ	スコア計算	マイニング			
要約(展望)	単語チェック	再帰的クラ	関連チェック	単語間関連	サンプル1	サンプル2	ソース表示	主語抽出	テキスト分	トップダウ	主題語含有			
タグデータ	Twitter	タイピング	単語抽出	単語置換	キーフレ	円グラフ	ばねモデル	セグメント	トップボト	セグメント	デュアルテ			
ファイル	川下り	運動可視化	グラフ	グラフ2	キーワード	キーワード	OKmap	段落並び替	段落関連度	レーダーチャ	順位線			
サンプル1	サンプル2	スコア分布	スコアネッ	セグメント	文選択	キーワード表	タグクラウ	テキスト	テキスト(Htmlテキ				
テキスト(タイピング	可視化なし	キーワード											

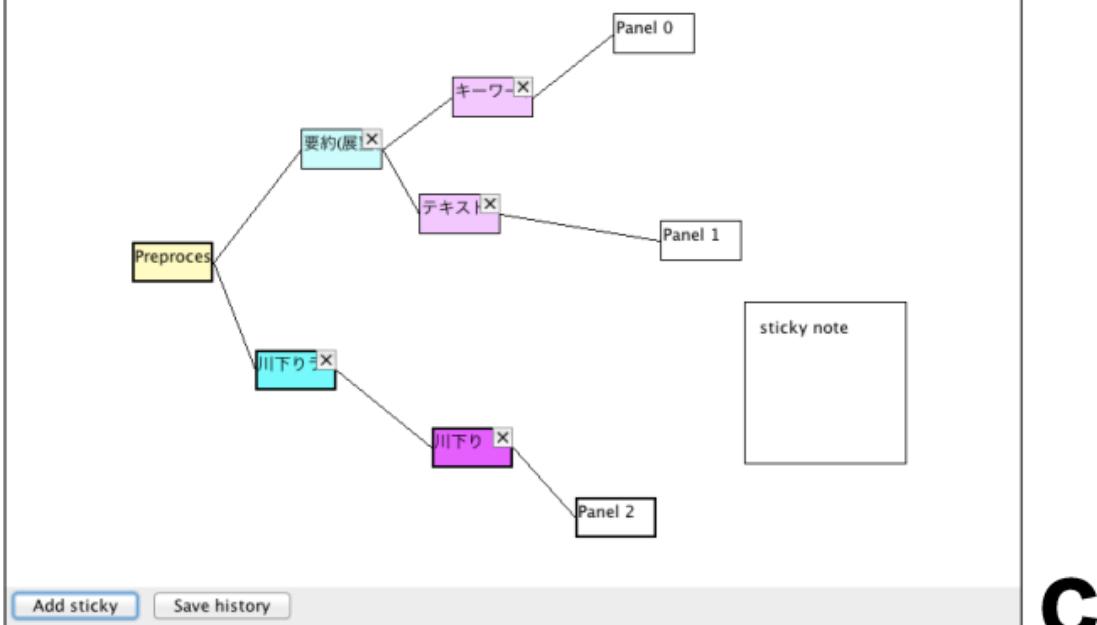
B**C**

図 4.2: ツール選択ウィンドウ

れ異なる色を採用した。各ノードはそれぞれ通常状態に加え、“選択”と“ハイライト”の合計3種類の状態を持っている。選択状態のノードはその輪郭が太く強調表示され、ハイライト状態のノードは色が強調される。

4.3 マウス操作によるテキスト分析ツールの選択・変更

提案インターフェースでは、ツールを選択する方法として、ツールを表すノードをマウスで直接操作する手法を採用した。

ユーザは、前述した4種類のノードをマウスにより直接操作することによって、使用するツールとそれらの組み合わせを選択・変更することができる。エッジで結ばれているノード同士は、それらが示すツール同士が組み合わせて利用されていることを表している。各ノードはマウスによってドラッグ & ドロップすることにより、ワークスペース上の任意の位置に移動さ

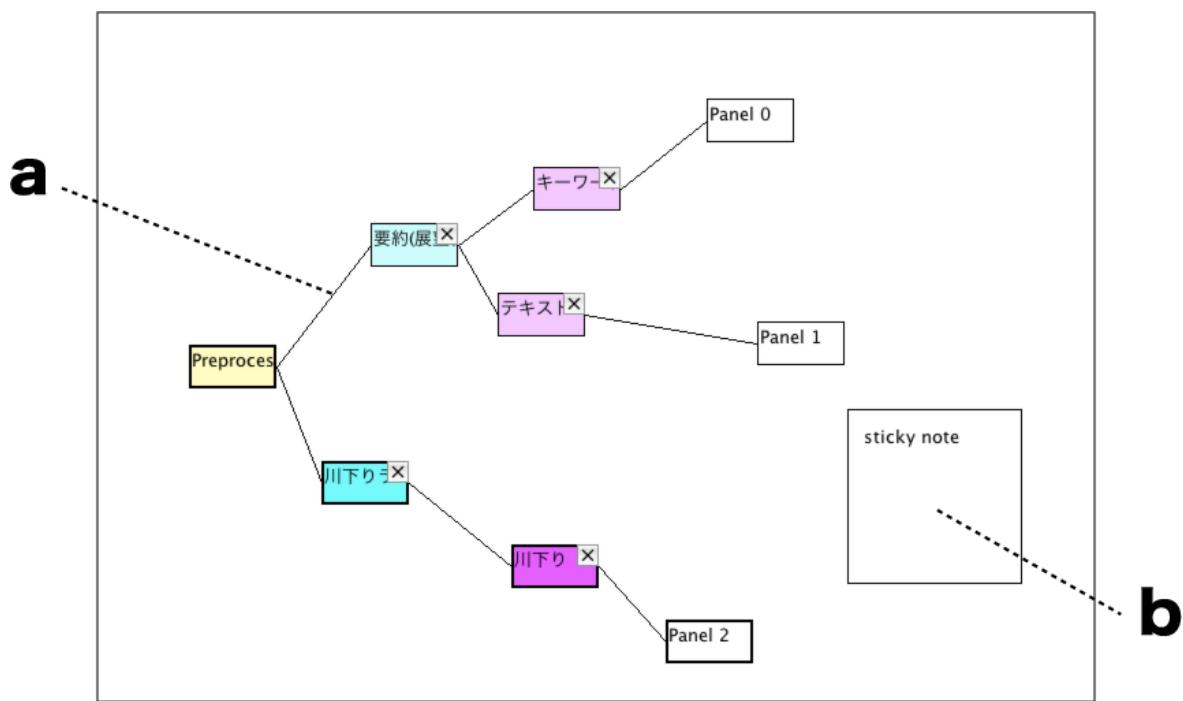


図 4.3: 探索と知見の外在化のためのワークスペース

せることができる。また、あるノードをそれと接続可能なノードに近づけることでエッジが自動で張られ、組合せ状態になる（図 4.5）。接続可能なノードの組合せを表 4.2 に示す。マイニングツールノードと可視化ツールノードは既存の TETDM システムで定められている組み合わせ可能なツールに基づいている。

4.4 探索プロセスの振り返り支援のための履歴

提案システムでは、ユーザが探索における任意の時点でのシステムの状態を履歴として保存し、必要に応じて以前の状態を復元することができる機能を実装した。これにより、ユーザは自身の探索プロセスを振り返ることができる。

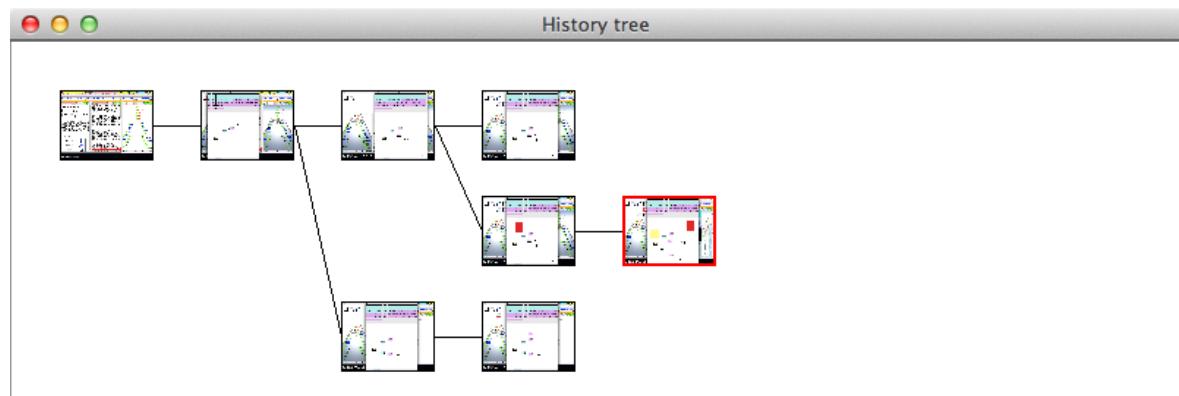


図 4.4: 履歴ツリーウィンドウ

表 4.1: 提案インターフェースにおけるノードの種類

ノードの種類	役割	色
前処理	TETDM システムにおける前処理を表す	Yellow
マイニングツール	TETDM システムにおけるマイニングツールを表す	Cyan
可視化ツール	TETDM システムにおける可視化ツールを表す	Magenta
結果表示パネル	TETDM システムにおける結果表示パネルを表す	White

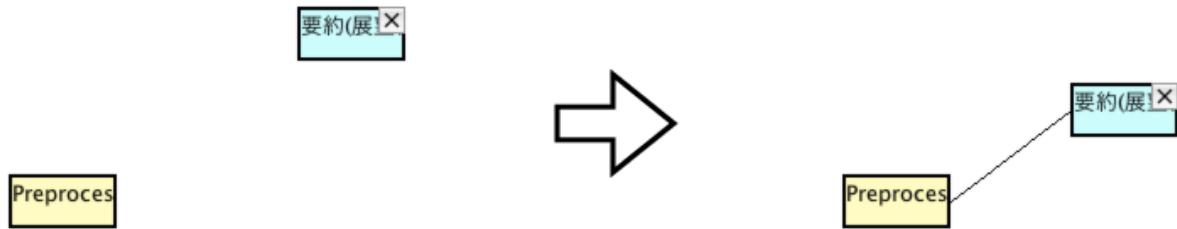


図 4.5: ノード同士を近づけることによるツール組み合わせの選択・変更

提案システムでは、探索プロセスの履歴は、システム起動時の状態をルート、ある時点でのシステムの状態をノードとする履歴ツリーとして表される（図 4.4）。履歴ツリーにおいて、現在の探索の状態を表すノードは枠線が赤色で強調表示される。各ノードには、履歴が保存された時点でのシステムのスクリーンショットがサムネイルとして表示される。また、それらをマウスオーバーすることにより、サムネイルがポップアップとして拡大表示され、それぞれの時点でのシステムの状態や探索の状態を確認することができる。

ユーザは、ツール選択ウィンドウ上の操作ボタン部にある“Save history”ボタン（図 4.2 下部）を押すことにより、現在のシステムの状態を履歴として保存することができる。保存された履歴は、現在の状態を表すノードの子要素として履歴ツリーに追加される。なお、プロトタイプシステムにおいて、履歴として保存される情報は、(1) 入力されているテキストデータ、(2) ツール選択ウィンドウ上のワークスペースの状態、(3) 結果表示パネルの状態、(4) システムのスクリーンショットの 4 種類である。

表 4.2: 接続可能なノードの組合せ

ノードの種類	接続可能なノード
前処理ノード	全てのマイニングツールノード
マイニングツールノード	前処理ノード及び、TETDM システムで定められている組み合わせ可能な可視化ツールノード
可視化ツールノード	TETDM システムで定められている組み合わせ可能なマイニングツールノード及び、結果表示パネルノード
結果表示パネルノード	全てのマイニングツールノード

ユーザは、履歴ツリーの任意のノードをクリックすることによって、以前の探索の状態を復元することができる。その際には、履歴として保存された情報を基に、(1) 入力されているテキストデータ、(2) ツール選択ウィンドウ上のワークスペースの状態、(3) 結果表示パネルの状態、が復元される。

これにより、ユーザは以前の探索プロセスを振り返ることができる。

4.5 探索プロセスにおける思考の外在化の支援

ユーザは探索プロセスを経て得た知見や生成した仮説を、振り返り行為によって整理・再構築しながら、最終的に創出する知識を構築していく。そのため、ユーザが探索プロセスを経て得た知見や生成した仮説を、ツール選択ウィンドウ上のワークスペースへ外在化し、それらを整理できるようにした。提案システムでは、ツール選択ウィンドウ上の操作ボタン部(図 4.2-C)にある“Add sticky”ボタンを押すことにより、任意のテキストを付箋としてツール選択ウィンドウ上のワークスペースに貼付できる機能を実装した(図 4.3-b)。これにより、ユーザは探索プロセスを経て得た知見や生成した仮説をテキスト情報として外在化することができる。なお、今回実装したプロトタイプシステムでは、外在化できる情報の種類はテキスト情報のみとした。

ワークスペース上の付箋は、ツールを表すノードと同様に、マウスでドラッグ & ドロップすることによってワークスペース上の任意の位置に移動させることができる。また、付箋をクリックすることによって、付箋編集画面を開くことができる(図 4.6)。付箋編集画面では、付箋に付与されているテキスト情報を編集したり、付箋の色を変更することができる。また、不要になった付箋をワークスペース上から削除することができる。これにより、ユーザは付箋として外在化した知見や仮説を整理することができる。



図 4.6: 付箋編集画面

5 議論

本研究では、テキストマイニングは Exploratory Search に該当するタスクであり、新たな知識の獲得を目的とする創造的活動であると見做して、それを支援するためのインターフェースを構築するためのデザイン指針を提案した。また、それに基づいて TETDM システムのインターフェースを改良し提案した。

2 章で述べたように、テキストマイニングは、ユーザの曖昧な情報要求に基づき試行錯誤を繰り返しながら目的となる情報を探索していくという Exploratory Search に該当するタスクであり、その目的は、ユーザの新たな知識の発見・獲得を支援することである。そのため、ユーザは様々なテキスト分析手法を適宜組み合わせて用いて試行錯誤を繰り返しながら、徐々に目的の情報を得ていく。

2.1 節で挙げたように、Fan らは、テキストマイニングに用いられる技術を 8 つに整理している [4]。それを踏まえて、Gupta は、テキストマイニングにおける主要な操作を 6 つにまとめている [5]。しかし、これらは、テキストマイニングのためのシステムに必要な要素技術を整理しているだけに過ぎず、テキストマイニングの目的である新たな知識の獲得にそれぞれの技術がどのように寄与するのかについて述べられていない。また、Fan らは、そのような要素技術を用いてテキストマイニングを行う際のプロセスのモデルを提案している(図 2.1) [4]。このモデルによると、ユーザは、まずテキスト集合に対して検索や前処理を行い、次に様々な技術を繰り返し用いながら目的の情報が得られるまで分析を繰り返す。そして、得られた結果はデータベースに蓄積され知識として活用される。しかし、このモデルは、知識獲得を目的とするユーザがそれらの技術を用いてシステムとどのようなインタラクションを行いながら知識を獲得していくのかというプロセスは考慮していない。また、分析の繰り返しという、試行錯誤を伴う情報アクセスの“形式”には言及しているものの、そのプロセスにおいてユーザがどのように目的の情報を得ていくのか、何に伴い試行錯誤を繰り返していくのか、といったユーザの“思考プロセス”については言及されていない。

すなわち、テキストマイニングに関する既存の取り組みは、データや技術などのシステム志向で行われており、ユーザがどのように情報にアクセスしていくのか、どのように知識を得ていくのか、といったユーザの振る舞いの観点で行われていない。そのため、テキストマイニングのためのアプリケーションシステムをデザインする際に、単にそれら技術やプロセスを採用するだけでは、必ずしもテキストマイニングの支援を行うことができるとは限らない。

TETDM プロジェクトの目的は、世の中に分散しているテキストマイニングツールを画一的に扱える環境を構築することである。しかし、この取り組みは、テキストマイニングに関する技術を同一のシステム上に集めるというものであり、それによって探索的な情報アクセスを支援できるのかはわからない。そのため、TETDM システムによって支援できるのか検証する必要がある。

テキストマイニングでは目的とするタスクが様々であるため、全てのタスクにに関して有用性を確認することは困難である。テキストマイニングに必要なタスクを整理して、それらを行うことができるのかどうかについて検討する。

テキストマイニングにおいて、テキストの分析結果は情報可視化技術によって行われることが多い。ユーザが自身の興味に応じて情報をまとめあげていくプロセスとその目的に着目する

と、テキストマイニングはテキスト情報を対象とした情報可視化であると見做すことができる。そのため、情報可視化のプロセスを参考にして、ユーザがテキストに円滑にアクセスできるのかについて考察する。

本章では、TETDM システムをベースとした提案したインターフェースを用いて、それらのタスクを行うことができるのか考察する。それによって、TETDM システムを使って実現できるタスクと実現できないタスクを整理する。加えて、TETDM システムの問題点について議論する。それによって、TETDM システムのテキストマイニングのためのプラットフォームとしての問題点を整理する。

5.1 テキストマイニングと情報可視化のモデル

テキストマイニングの目的は、テキスト集合から知識を得ることである。このことに着目すると、テキストマイニングのプロセスは、情報可視化のプロセスと類似している。そこで、本節では、情報可視化のモデルと情報可視化を支援するために必要なタスクについて述べる。

Card は、情報可視化の参照モデルを提案している [2]。このモデルは、(1) Data Transformation、(2) Visual Mapping、(3) View Transformation の 3 つのプロセスから構成される (図 5.1 参照)。

- **Data Transformation**

Raw Data から可視化に必要となるデータを取り出し Data Table へ変換する

- **Visual Mapping**

Data Table から、数値情報と利用可能な視覚的要素 (e.g. 空間的位置、軸情報、記号) の組み合った構造である Visual Structure へ変換する

- **View Transformation**

表示範囲の調整や視点の変更などの可視化におけるパラメータを変化させることによって、Visual Structure からユーザが認識しやすい表示 (View) へ変換する

これは、データから、視覚的情報を活用することによって有益な知見を得る際のプロセスをモデル化したものである。また、Card のモデルでは、データを有益な知見に変換するといったデータ思考のプロセスだけではなく、各々のプロセスにおいて、ユーザがそのプロセスにどのように関与するのかという、人とシステムのインタラクションに着目している。このように、情報可視化に関する研究においては、その目的が、データを人が認知しやすい視覚表現に変換することによってデータの理解を支援することである、ということから、ユーザの振る舞いを考慮したモデルが提案されている。

Shneiderman は、人が視覚的手がかりを用いて情報へアクセスする際のプロセスについて、“Overview first, zoom and filter, then details-on-demand” と説明している [29]。人は視覚化された情報へアクセスする際に、初めにその全体を見渡し (overview first)、次にその全体から絞り込み (zoom and filter)、更に、要求に応じて詳細を眺める (details-on-demand)。これは、視覚化された情報を探すときに、どのような振る舞いをするのかということを捉えたモデルである。

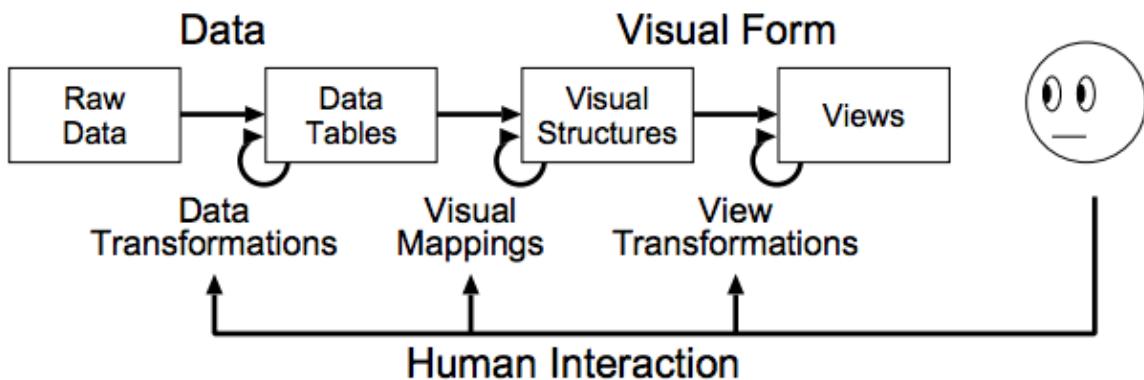


図 5.1: 情報可視化の参照モデル [2]

また、Shneiderman は、情報可視化のためのタスクを以下の 7 種類に分類している。

(1) **Overview** (見渡す)

コレクション全体の概観を把握する

(2) **Zoom** (拡大する)

興味のあるアイテムを深く見る

(3) **Filter** (絞り込む)

興味のないアイテムを除外する

(4) **Details-on-demand** (要求に応じて詳細を見る)

アイテムやグループを選択し、必要に応じて詳細を取得する

(5) **Relate** (関連付ける)

アイテム同士の関連を眺める

(6) **History** (履歴を保持する)

振り返り行為を支援するために行動履歴を保持する

(7) **Extract** (抽出する)

サブコレクションやクエリパラメータの抽出を可能にする

ユーザが視覚的情報へアクセスする際には、これらのタスクはそれぞれ独立して行われるのではなく、互いに関連して行われる。

このように、情報可視化の研究では、情報をどのように見せるのかという可視化の技術という観点ではなく、がどのように視覚情報へアクセスするのかという人の振る舞いの観点に着目して、そのプロセスモデルの提案やタスクの分類を行っている。

テキストマイニングにおいても、要素技術を整理したり、それらの技術を同一環境上で扱えるようにするだけではなく、人がテキスト集合から知識を得る場合に人はどのようにテキスト情報へアクセスするのか、という人の振る舞いの観点に着目すべきである。

表 5.1: TETDM システムにおける情報可視化の 7 タスク

情報可視化のタスク	TETDM システムの機能
Overview	文書要約、重要文の一覧表示 etc.
Zoom	部分テキスト、フォーカス運動
Filter	部分テキスト
Details-on-demand	フォーカス運動
Relate	単語間関連度、クラスタリング etc.
History	なし (提案インターフェースにより可能)
Extract	キーワード抽出、重要文抽出 etc.

次節では、これらの 7 種類のタスクをテキストマイニングのプロセスに適用し、各々のタスクが TETDM システムを用いて達成できるのかについて述べる。

5.2 TETDM システムの問題点

情報可視化の支援のために必要なタスクを参考に、TETDM システム及び 4 章で実装したプロトタイプシステムを用いて、それらのタスクを達成することができるのかについて考察する。

TETDM システムでは、上述した 7 種類のタスクに関して、“History” のタスク以外は、それぞれ独立した状態であれば達成することができる (表 5.1)。TETDM システムでは、様々な分析ツールを同一環境上で扱うことができる。それらの分析ツールはシステム内のモジュールとして扱われており、それを追加することによって、用いることのできる分析手法を拡張することができる。また、TETDM システムでは、同時に用いているツール間で、共通のデータ (フォーカスデータ) を参照して動作させる仕組み (フォーカス運動) や入力されたテキストデータから選択した部分のみを処理できる機能 (部分テキスト) が実装されている。そのため、上述した 8 種類のタスクに適したツールをそれぞれ実装することにより、それらのタスクを達成できる。しかし、これは、単に複数の技術を同一環境で扱えるようにするだけであり、それらを協調させて用いるような分析を行うことはできない。これは、TETDM プロジェクトの目的であるテキストを扱うユーザの創造的活動を支援することに寄与しない。

また、“History” のタスクは、他のタスクにおけるユーザの操作履歴を保存しておく必要があるため、それ単体では成立し得ないが、4 章で我々が実装したインターフェースに採用したようなアプローチによって達成することができる。

Zoom、Filter、Extract のタスクでは、ユーザは、データ集合の中からあるデータを抽出や絞り込みを行ったりするなど、データを加工した上で、それを他の分析に用いることも考えられる。TETDM システムでは、これらのようなユーザの操作の介入によって、分析対象のデータを変形し、それを基に更に分析を進めるようなタスクを円滑に行うことができない。

これらを踏まえて、TETDM システムの問題点を以下のように、(1) 入力テキストの取り扱いに関する問題、(2) 分析処理の粒度に関する問題、(3) 操作対象の把握可能性に関する問題、の 3 つに整理した。

- 入力テキストの取り扱いに関する問題

TETDM システムでは、複数のテキストを分析対象とする場合、予めそれらの全てをシステムに入力しておく必要がある。そのため、ユーザは分析途中で、それまで分析対象としていなかったような新たな新たなテキストを分析対象として追加することが困難である。

また、入力されたテキストから、あるテキストだけを取り出して、それに対して処理を行うような機能が用意されていない。それらのテキストを、単一のテキストと見做すか、複数のテキストと見做すかは、各ツールの実装に依存する。

- 分析処理の粒度に関する問題

TETDM システムにおける入力テキストに対する処理の単位は、(1) システムによって行われる前処理、(2) マイニングツールによって行われるマイニング処理 (3) 可視化ツールによって行われる可視化処理、の 3 つである。また、ユーザが任意で選択できるのは (2) と (3) である。これを、Card の参照モデルに当てはめて考えると、(1) や (2) は Data Transformation、(3) を選択することは Visual Mapping に相当すると考えることができる。また、View Transformation は、TETDM システムにおける可視化ツール内での操作であると捉えることができる。

Card の参照モデルによると、ユーザは、Data Transformation、Visual Mapping、View Transformation を繰り返しながら、自身が見たい観点に基づき可視化結果を調整していく。しかし、TETDM システムでは、そのような処理の繰り返しによるプロセスを円滑に行うことができない。たとえば、Zoom や Filter のタスクは、分析対象のテキスト集合から一部に着目をして分析を続けるために行われる。そのような処理は、(1) の処理が行われる前段階として行われるべきである。そのため、マイニングツールや可視化ツールの選択と同様に、ユーザがテキストデータから可視化結果までのプロセスを意識して Zoom や Filter のようなタスクを遂行できることは望ましいが、TETDM システムでは、そのような仕組みは提供されていない。

また、現在 TETDM システムに実装されているツールには、単一のツールで複数の処理を行うマイニングツールや、本来マイニングツールで行うべきのような可視化とは関係のない処理が行われているものもある。そのため、複数のツールを柔軟に組み合わせて用いることができない。

- 操作対象の把握可能性の問題

そのため、システムに入力されたテキストに対して、何らかの分析処理を行い、その結果を可視化する、といった基本的な分析プロセスは、新たなモジュールを作成しそれを用いることで達成できる。これは、Card の参照モデルにおける、Raw Data から View までのプロセスである。しかし、ユーザの操作が必要な場合、それを円滑に行うことが困難であったり、TETDM システムの制約により不可能であったりする。

Card のモデルにおいて、ユーザは、Data Transformation、Visual Mapping、View Transformation の各プロセスでシステムとインタラクションを行う。これを、効率的に行えるようにするために、直接操作 [28] による手法がよく用いられる。これは、操作対象の

データやクエリなど視覚的要素として表現し、それをマウスなどを用いて直接操作できるようにすることにより、データやクエリの変更を直感的なインタラクションを通して行えるようにするものである。

データに対する操作を行おうと考えた場合、直接操作による手法のような視覚的手がかりが提示されていないため、操作対象の把握が困難である。これは、Zoom や Filter、Extract のタスクを妨げる要因である。

TETDM システムでは、様々なテキスト処理を行えるように、分析ツールを追加できる仕組みや、分析ツール間を連動させ協調的に動作させる仕組み、入力されたテキストデータから選択した部分のみを分析するための機能が実装されている。これによって、情報可視化のための 7 つのタスクを独立に達成することができる。しかし、前節で述べたように、ユーザが情報へアクセスする際には、それらのタスクは独立して行われるのではなく、互いに関連して行われる。そのため、TETDM システムでは、そのような複数のタスクが関連する複雑なタスクを円滑に遂行することができない。これは、TETDM システムにおける入力テキストの扱いや、処理の粒度、複雑な操作方法などの TETDM システムのテキストマイニングのためのプラットフォームとしての制約に起因する問題である。これらは、新たな分析ツールの実装や本研究によるアプローチなどの外部的な働きかけによって解決されるものではない。

本研究では、テキスト集合からの知識獲得を支援するために、(1) 探索プロセスの円滑化、(2) 探索プロセスにおける思考の外在化、(3) 探索プロセスの振り返り支援、の 3 つに着目してデザイン指針を策定し、それに基づいて TETDM システムのインターフェースを改良した。改良したインターフェースでは、分析ツールを表すオブジェクトを直感的に操作できるインターフェースや、探索中に得た知見を外在化できる手段を提供したことにより、ユーザの探索行為と振り返り行為を円滑に行えるようにした。

6 結論

本研究の目的はテキスト集合からの新たな知識の獲得を支援することである。テキスト集合から新たな知識を発見するためには、ユーザは様々な手法を適応的に組合せ試行錯誤を繰り返しながら探索を進めていく必要がある。それを支援するシステムを構築するために、本稿では、(1) 探索の円滑化、(2) 探索の振り返り支援に着目してデザイン指針を作成した。加えて、それに基づいて実装したプロトタイプシステムを実装した。以下に本論文の内容を纏める。

1章では、まず研究背景として、テキストデータの増加に伴い、それらを単に閲覧するだけではなく、有益な情報や新たな知識を獲得するための情報資源として活用することの必要性、重要性が高まっていることについて述べた。また、それに伴いテキストマイニングの利用場面は今後益々広がっていくと考えられる。テキストマイニングは、テキスト集合から新たな知識を獲得することを目的としており、試行錯誤を必要とする探索的な情報アクセスである。そのため、それを支援するためには、ユーザが円滑に探索できるようにしたり、知識の結晶化と液状化のプロセスを支援するなど、様々なアプローチが必要となるため、それらを本研究で取り組む課題とした。

2章では、まず、関連研究として、テキストマイニングとその目的と特徴について述べた。加えて、テキストマイニングのための統合環境の構築を目指した TETDM プロジェクトについて述べた。TETDM プロジェクトの目的は、多種多様なテキストマイニングツールを画一的に扱える統合環境を構築することによって、テキストデータを扱うユーザの創造的活動を支援することである。次に、曖昧な情報要求に基づいた情報探索のモデルである Exploratory Search について説明した。テキストマイニングにおいて、ユーザは様々な分析手法を適応的に用いて、試行錯誤を繰り返しながら、徐々に目的の情報・知識を得ていく。そのため我々は、テキストマイニングが Exploratory Search に該当するタスクであると考えている。最後に、新たな知識の発見・創出を創出すること、つまり、創造活動の支援 (Creativity Support) に関する研究について説明した。特に、知識創出のプロセスにおける外在化の重要性について述べた。これは、テキストマイニングは、テキスト集合から新たな知識を発見・創出することを目的としているからである。

3章では、テキスト分析による知識獲得のためのアプリケーションシステムを構築するため策定したデザイン指針について述べた。本研究では、デザイン指針を策定するにあたって、(1) 探索プロセスの円滑化、(2) 探索プロセスの振り返り支援、(3) 探索プロセスにおける外在化、の 3 点に着目した。

4章では、3章で述べたデザイン指針に基づいて実装したプロトタイプシステムについて説明した。本研究では、既存の TETDM システムをベースとしてプロトタイプシステムを実装した。プロトタイプシステムには、既存の TETDM システムで採用されているインターフェースに加え、分析ツールの選択・切り替えを支援するためのインターフェースとユーザの探索履歴の振り返りを支援するためのインターフェース実装した。ツール選択を支援するためにのインターフェースには、現在用いられているテキスト分析ツールをノード、それらのツール間の処理の流れをエッジとするグラフ表現を用いた。これにより、ユーザはツールを表すノードをマウスで直接操作することによって、直感的かつ円滑なツール選択・切替を行える。ユーザの探索履歴の振り返りを支援するためのインターフェースには、システムを起動してからのユーザの探索

履歴がツリー構造として表示されている。これにより、ユーザは過去の探索の状況を振り返ることができる。

5章では、TETDMシステムの問題点について議論した。ユーザのインタラクションの観点から考えると、テキストマイニングのプロセスは情報可視化のプロセスと類似している。そのため、情報可視化のモデルに習い、TETDMシステム及び提案したインターフェースによって、情報可視化を支援するために必要なタスクを達成することができるのか、という観点から考察を行い、TETDMシステムの問題点を洗い出した。

6章では、本論文の要点をまとめた。

謝辞

本研究を纏めるにあたり、多大な御指導、御鞭撻を賜わりました、関西大学総合情報学部総合情報学科の松下光範教授に深甚の謝意を表します。

東京大学大学院総合文化研究科言語情報科学専攻の加藤恒昭教授には、インタラクティブ情報アクセスと可視化マイニング研究会にて学士の頃よりお世話になりました。心から感謝致します。

広島市立大学大学院情報科学研究科システム工学専攻の砂山渡准教授、首都大学東京大学院システムデザイン研究科システムデザイン専攻の高間康史教授、立命館大学情報理工学部メディア情報学科の西原陽子准教授及び、TETDM プロジェクトのコアメンバの方々には大変お世話になりました。私の研究は TETDM プロジェクトなしでは成立しないものです。心より感謝致します。

最後に、大学生活を支えて頂いた家族、友人、Emacs に心より感謝致します。

参考文献

- [1] 安藤俊幸: 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, Vol. 52, No. 1, pp. 20–31 (2009).
- [2] Card, S. K., Mackinlay, J. D. and Shneiderman, B.: *Readings in information visualization: using vision to think*, Morgan Kaufmann (1999).
- [3] Chen, M., Ebert, D., Hagen, H., Laramee, R. S., Van Liere, R., Ma, K.-L., Ribarsky, W., Scheuermann, G. and Silver, D.: Data, information, and knowledge in visualization, *Computer Graphics and Applications, IEEE*, Vol. 29, No. 1, pp. 12–19 (2009).
- [4] Fan, W., Wallace, L., Rich, S. and Zhang, Z.: Tapping the power of text mining, *Communications of the ACM*, Vol. 49, No. 9, pp. 76–82 (2006).
- [5] Gupta, V. and Lehal, G. S.: A survey of text mining techniques and applications, *Journal of emerging technologies in web intelligence*, Vol. 1, No. 1, pp. 60–76 (2009).
- [6] Hearst, M. A.: Untangling text data mining, *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 3–10 (1999).
- [7] 堀浩一: 創造活動支援の理論と応用, 人工知能学会オーム社 (2007).
- [8] Hori, K., Nakakoji, K., Yamamoto, Y. and Ostwald, J.: Organic Perspectives of Knowledge Management: Knowledge Evolution through a Cycle of Knowledge Liquidization and Crystallization, Vol. 10, No. 3, pp. 252–261 (2004).
- [9] 市村由美, 長谷川隆明, 渡部勇, 佐藤光弘: テキストマイニング : 事例紹介, 人工知能学会誌, Vol. 16, No. 2, pp. 192–200 (2001).
- [10] 井須弘恵, 大塚直也, 松下光範: 探索的情報アクセスの支援に向けた TETDM インターフェースの改良に関する基礎検討, 第 6 回 インタラクティブ情報アクセスと可視化マイニング研究会 予稿集 (2014).
- [11] 川谷隆彦: 多文書間の共通性分析に基づく文書クラスタリング, 情報処理学会論文誌, Vol. 47, No. 6, pp. 1903–1917 (2006).
- [12] Marchionini, G.: Exploratory Search: From Finding to Understanding, *Commun. ACM*, Vol. 49, No. 4, pp. 41–46 (2006).
- [13] Matsushita, M.: Supporting Exploratory Data Analysis by Preserving Contexts, *Proc. 9th International Conference on Knowledge-Based Intelligent Information and Engineering Systems*, pp. 540–546 (2005).
- [14] 松下光範: InTREND:ユーザの探索行為と振り返り行為に着目したデータ分析支援システム, 情報処理学会論文誌, Vol. 49, No. 7, pp. 2456–2467 (2008).

- [15] Matsushita, M. and Shirai, Y.: Supporting Exploration and Reflection in Exploratory Data Analysis, *Proc. 4th International Workshop on Chance Discovery*, pp. 3–8 (2005).
- [16] 中垣内李菜, 川本佳代, 砂山渡: 統合環境 TETDM を用いたテキストマイニング初心者のためのツール選択支援, 第 27 回人工知能学会全国大会 (2013).
- [17] 中垣内李菜, 川本佳代, 砂山渡: 統合環境 TETDM を用いたテキストマイニング初心者のスキル獲得支援, 第 28 回人工知能学会全国大会 (2014).
- [18] 中小路久美代, 山本恭裕: 創造的情報創出のためのナレッジインタラクションデザイン, 人工知能学会論文誌, Vol. 19, No. 2, pp. 154–165 (2004).
- [19] 那須川哲哉: コールセンターにおけるテキストマイニング, 人工知能学会誌, Vol. 16, No. 2, pp. 219–225 (2001).
- [20] 那須川哲哉, 河野浩之, 有村博紀: テキストマイニング基盤技術, 人工知能学会誌, Vol. 16, No. 2, pp. 201–211 (2001).
- [21] 那須川哲哉: テキストマイニングの普及に向けて : 研究を実用化につなぐ課題への取組み, 人工知能学会誌, Vol. 24, No. 2, pp. 275–282 (2009).
- [22] 那須川哲哉: テキストマイニングの可能性 : 有用性と研究の発展性, 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, Vol. 112, No. 196, pp. 19–24 (2012).
- [23] Nasukawa, T. and Nagano, T.: Text analysis and knowledge mining system, *IBM systems journal*, Vol. 40, No. 4, pp. 967–984 (2001).
- [24] 西原陽子, 佐藤圭太, 砂山渡: 光と影を用いたテキストのテーマ関連度の可視化, 人工知能学会論文誌, Vol. 24, No. 6, pp. 479–487 (2009).
- [25] 奥村学: テキスト自動要約, 情報処理, Vol. 45, No. 6, pp. 574–579 (2004).
- [26] Otsuka, N. and Mitsunori, M.: Graphical Interface that Supports Users' Trial-and-Error Process of Text Mining, *Proceedings of the 27th Annual Conference of the Japanese Society for Artificial Intelligence (JSAT 2013)* (2013).
- [27] Rajman, M. and Besançon, R.: Text Mining: Natural Language techniques and Text Mining applications, In *Proceedings of the 7 th IFIP Working Conference on Database Semantics (DS-7)*. Chapam, Hall, pp. 7–10 (1997).
- [28] Shneiderman, B.: 1.1 direct manipulation: a step beyond programming languages, *Sparks of innovation in human-computer interaction*, Vol. 17, p. 1993 (1993).
- [29] Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations, *Proceedings of the 1996 IEEE Symposium on Visual Languages*, VL '96, Washington, DC, USA, IEEE Computer Society, pp. 336– (1996).

- [30] Shneiderman, B.: Creating Creativity: User Interfaces for Supporting Innovation, *ACM Trans. Comput.-Hum. Interact.*, Vol. 7, No. 1, pp. 114–138 (2000).
- [31] Shneiderman, B.: Creativity Support Tools, *Commun. ACM*, Vol. 45, No. 10, pp. 116–120 (2002).
- [32] 砂山渡, 谷内田正彦: 文章の特徴を表すキーワードを発見して重要文を抽出する展望台システム, 電子情報通信学会論文誌, Vol. 84-D1, No. 2, pp. 146–154 (2001).
- [33] 砂山渡, 谷川信弘: テキストの主題に基づく一貫性評価と結論抽出への応用, 知能と情報 : 日本知能情報ファジィ学会誌, Vol. 23, No. 5, pp. 727–738 (2011).
- [34] 砂山渡, 高間康史, 西原陽子, 梶並知記, 串間宗夫, 徳永秀和: 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, Vol. 29, No. 1, pp. 100–112 (2014).
- [35] 砂山渡: Total Environment for Text Data Mining, 2010 (第24回人工知能学会全国大会).
- [36] 鈴木誠: カテゴリ間の単語頻度の差分を用いたテキストの自動分類, 日本経営工学会論文誌, Vol. 59, No. 4, pp. 355–363 (2008).
- [37] Taylor, R. S.: Question-negotiation and Information Seeking in Libraries, *College & Research Libraries*, Vol. 29, No. 3, pp. 178–194 (1968).
- [38] Tukey, J. W.: *Exploratory Data Analysis.*, Addison-Wesley Publishing Company (1977).
- [39] Uramoto, N., Matsuzawa, H., Nagano, T., Murakami, A., Takeuchi, H. and Takeda, K.: A text-mining system for knowledge discovery from biomedical documents, *IBM Systems Journal*, Vol. 43, No. 3, pp. 516–533 (2004).
- [40] White, R. W. and Roth, R. A.: *Exploratory Search: Beyond the Query-Response Paradigm*, Morgan & Claypool Publishers (2009).
- [41] Yamamoto, Y. and Nakakoji, K.: Interaction Design of Tools for Fostering Creativity in the Early Stages of Information Design, *Int. J. Hum.-Comput. Stud.*, Vol. 63, No. 4-5, pp. 513–535 (2005).