

# 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

Supporting Knowledge Acquisition From Text Collection  
by Visualization of Exploration Process

大塚 直也\*<sup>1</sup>   岡本 香帆里\*<sup>2</sup>   松下 光範\*<sup>2</sup>  
Naoya Otsuka   Kahori Okamoto   Mitsunori Matsushita

\*<sup>1</sup>関西大学大学院総合情報学研究科   Graduate School of Informatics, Kansai University  
\*<sup>2</sup>関西大学総合情報学部   Kansai University, Faculty of Informatics

Our goal is to support users in finding knowledge from a large amount of text data. For such purpose, text-mining techniques are often used. On a text-mining task, users need to adaptively combine various technologies, such as natural language processing, data mining, and information visualization. In addition, users' information need is not generally clear at first. By iterating exploration, users gradually clarify their information need.

In this paper, we regard text mining as two concepts: (1) exploratory search task, and (2) process for creative knowledge work. Hence, we propose a design principle for building text-mining applications based on these concepts. The goals of our design principle are to (1) facilitate exploration actions, (2) support the externalization of findings and hypotheses, and (3) support reflection actions. In addition, we implement a prototype system that uses our design principle based on the existing TETDM system.

## 1. はじめに

ブログや SNS (Social Networking Service) などの普及に伴い、一般のユーザが様々な情報をテキストデータとして公開することが一般的となった。その結果として、我々が得られ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だけでなく、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ることで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 [Hearst 99]。テキストマイニングは、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、これらを適応的に用いることによって、テキスト集合から有益な情報や新たな知識を発見・獲得することができる [那須川 01]。それを目的として、情報抽出や文書要約、文書分類などのテキストマイニングの要諦技術について研究開発が行われており、既に多くの技術やそれらを活用したツールが成果として発表されている [Fan 06]。

砂山らは、そのような多種多様なテキスト分析ツールを組み合わせることでできる環境の構築を目指して、Total Environment for Text Data Mining プロジェクト (以下 TETDM プロジェクト) を提唱している [砂山 14]。TETDM プロジェクトは、テキスト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。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ものにはなっていない。そのため、現在の TETDM システムのインタフェースでは、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。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に資するために、TETDM システムのインタフェースの改良を試みる。

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である。このことから我々は、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。本研究では、これらの 2 つの観点から、テキストマイニングを支援するために必要な要件を整理する。加えて、これらに基づき、テキストマイニングのための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、(2) 探索プロセスにおける外在化、(3) 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。

本稿では、提案するデザイン指針に基づいて TETDM システムのインタフェースの再デザインを試みたので報告する。

## 2. テキスト集合からの知識獲得

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、それによって得られた知見を解釈したり、生成された仮説を検証したりしながら分析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する必要がある。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き、それを知識へと昇華させていく探索的な情報アクセスを必要とするタスクと捉えることができる。

これらを踏まえて、我々はテキストマイニングを、(1) 曖昧な情報要求に基づく探索的な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 [White 09] の枠組みに当てはめることができる。加えて、(2) のような活動

連絡先: 松下光範 関西大学総合情報学部 〒 569-1095 大阪府高槻市霊仙寺町 2-1-1 Tel:(072)690-2437 e-mail: mat@res.kutc.kansai-u.ac.jp

は、創造的活動 (Creativity Support) [中小路 04] と捉えることができる。

## 2.1 Exploratory Search

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検索行為を説明するためのモデルである [White 09]。

検索活動には Lookup, Learn, Investigate の 3 種類がある [Marchionini 06]。これらの中でも特に、検索を繰り返しながら徐々に目的の情報を発見していくような、Learn や Investigate の検索活動が Exploratory Search に該当する。

Taylor は、ユーザの情報要求の明確さの程度を (1) 直感的な要求、(2) 意識されている要求、(3) 形式化された要求、(4) 解決に至る要求、の 4 段階に分類している [Taylor 68]。(1) や (2) の段階のような、ユーザの情報要求が明確である場合、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、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、(3) や (4) の段階のような、ユーザの情報要求が曖昧な場合、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ていき、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、興味に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 には Exploratory Browsing と Focused Searching の 2 つのプロセスがある。Exploratory Browsing は、探索空間を広げる方向への検索である。このプロセスでは、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。そのため、ユーザは様々な情報を俯瞰することによって検索対象への理解を深めていく。Focused Searching は、探索空間を絞る方向への検索である。このプロセスでは、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。そのため、ユーザは検索結果について詳細に調べたり、Exploratory Browsing によって生成された仮説を検証したりする。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ていき、探索の目的を明確にしていく。

## 2.2 創造活動の支援

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 [中小路 04]。

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、最終的に何をどのように創出するのかを徐々に決定していく。つまり、創造活動のプロセスにおいて、ユーザは知識の結晶化と液状化 [Hori 04] 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。

## 3. デザイン指針

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した創造的活動である。本研究では、それを支援するシステムを構築するためのデザイン指針を (1) 探索プロセスの円滑化、(2) 探索プロセスにおける外在化、(3) 探索プロセスの振り返り支援の 3 つの観点から策定した。

### 3.1 探索プロセスの円滑化

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ながら、テキスト集合を様々な観点で分析する必要がある。そのような探索プロセスでは、ユーザは自身の興味や探索の目的に応じて、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繰り返して行く。

テキストマイニングはこのような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、ユーザの探索活動を妨げないような直感的な操作方法を提供することが望ましい。また、探索においてユーザが試行錯誤を円滑に行えるようにするために、ユーザがテキスト分析ツールを円滑に選択・切り替えできるようにする必要がある。加えて、このようなインタラクティブな情報アクセスを必要とするタスクの支援を目的としたシステムでは、テキスト集合やテキスト分析ツールなどの各要素を直接操作 (Direct Manipulation) [Shneiderman 93] できるようなインタフェースを提供することが望ましい。

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、分析ツールの選択や切り替えを容易にしたり、直感的な操作方法の提供したりすることはテキストマイニングの支援に寄与すると考えられる。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、データや創造的活動のプロセスを可視化することは創造的活動の支援においても必要とされている。

### 3.2 探索プロセスにおける外在化

ユーザは、探索によって得られた知見や生成された仮説を外在化し、それをきっかけとして次の探索を行っていく。すなわち、探索プロセスにおいて外在化された知見や仮説は、次の思考への“補助線”として用いられる。そのため、探索的な情報アクセスを支援するためには、そのプロセスにおいて得られた知見や生成された仮説を何らかの表現として外在化できる手段を提供することが望ましい。

テキストマイニングでは、様々なテキスト分析手法を適応的に組み合わせながら探索を進めていくため、探索内容が非常に複雑となる。そのようなタスクでは、前述した探索プロセスにおける知見や仮説の外在化のような、ユーザが意識的に“何か気づいた”と考えた場合に行われる直接的な外在化だけではなく、システム上のオブジェクトの移動やサイズの変更などの暗示的外在化についても、次の思考への“補助線”として用いられると考えられる。そのため、個々の分析によって得られた結果だけではなく、過去のユーザのシステムへの操作や用いた分析手法についても履歴として残しておき、探索の手がかりとして利用できるようにすることが必要である。

### 3.3 探索プロセスの振り返り支援

テキストマイニングでは、探索の繰り返し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て最終的に創出する知識を構築していく。つまり、ユーザは、自身が直面しているタスクにおいて得られた知見や生成された仮説を利用

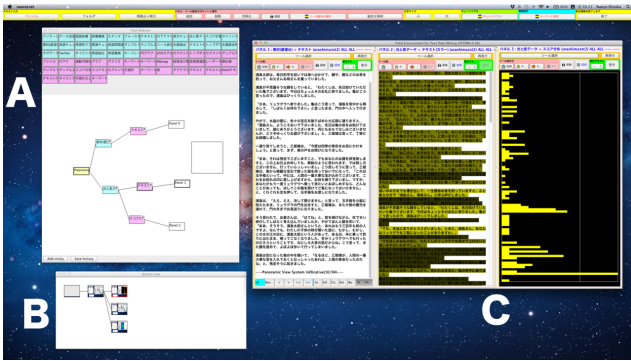


図 1: 提案インタフェースの画面構成

するために、改めてそれらを整理し、解釈するといった振り返り行為を行う。そのため、3.2 節で述べた外在化表現を基に、ユーザが自身の探索行為を振り返ることができるようなインタフェースを提供する必要がある。

#### 4. 提案インタフェースの実装

本節では、3. 節で示したデザイン指針に基づいて実装した提案インタフェースのプロトタイプについて述べる。提案インタフェースは TETDM システム\*1 をベースとし、Java を用いて実装した。

提案インタフェースは、ツール選択や探索によって得られた知見を外在化するためのワークスペースウィンドウ (図 1-A)、探索履歴を表示するための履歴ツリーウィンドウ (図 1-B)、可視化ツールによる処理結果を表示するための結果表示ウィンドウ (図 1-C) から構成されている。

ワークスペースウィンドウには、現在の探索の状況が表されている (図 2)。ユーザは、ワークスペース上のオブジェクトを操作することにより、利用するテキスト分析ツールを選択・変更したり、探索によって得た知見や仮説を外在化したりすることができる。履歴ツリーウィンドウには、探索履歴をノードとした履歴ツリーが表示されている (図 3)。これによりユーザは、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る。結果表示ウィンドウは、TETDM システムのものを踏襲しており、ユーザが選択したマイニングツールと可視化ツールによる処理結果が表示されている。図 1-C では、3 つの異なるツールによる結果が表示されている。

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が更新されるようになっている。

##### 4.1 グラフ表現による分析状況の視覚化

提案インタフェースでは、各分析ツールをノード、それらを処理の順に繋いだ線をエッジとするグラフ表現を採用した。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することができる。

TETDM システムに入力されたテキストファイルは、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処理結果が表示される。そのため、提案インタフェースでは、グラフのノードとして、(1) 前処理、(2) マイニングツール、(3) 可視化ツール、(4) 結果表示パネル、の 4 種類のノードを用意した。インタ

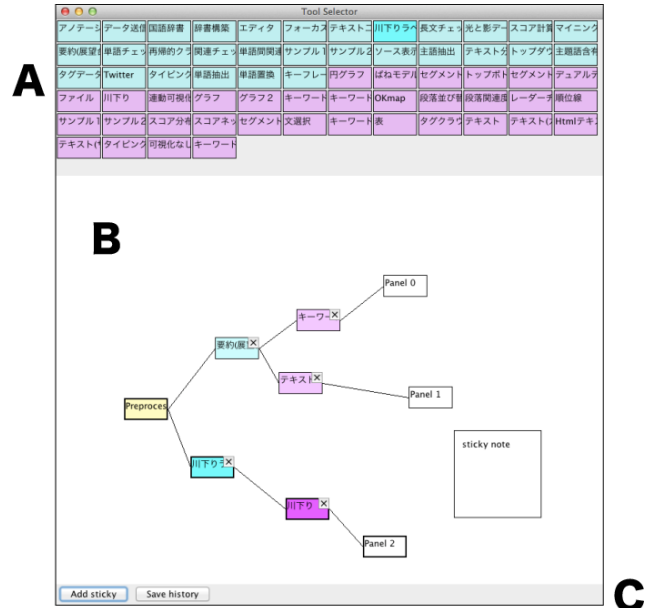


図 2: ワークスペースウィンドウ

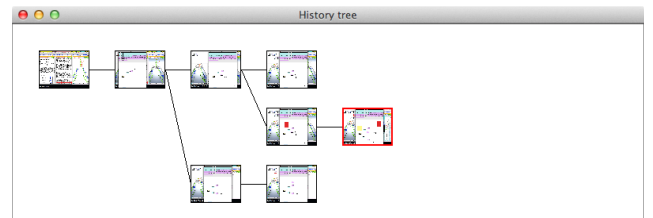


図 3: 履歴ツリーウィンドウ

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、処理の段階に応じて、各ノードをそれぞれ異なる色で彩色した。

##### 4.2 直接操作によるテキスト分析ツールの選択

提案インタフェースでは、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で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。各ノードは、マウスでドラッグ & 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せることができる。任意のノードを組み合わせ可能なノードに近づくことによって、エッジが自動で張られる。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られて利用されていることを表している。

##### 4.3 外在化のための付箋機能

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、それらの知見や仮説をテキスト情報を持つ付箋としてワークスペース上に外在化できるようにした。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操作によってワークスペース上の任意の位置に移動させたり、任意の色に変更することができる。これによりユーザは、外在化された知見や仮説を整理することができる。

\*1 ver. 0.56 (stable ver., 2014/3/31 公開)

#### 4.4 探索履歴の視覚化

提案インタフェースでは、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の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。

提案システムでは、探索プロセスの履歴を、システム起動時のシステムの状態をルート、ある時点でのシステムの状態をノードとするツリー構造として表現した (図 3)。履歴ツリーにおいて、現在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。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、以前の状態を振り返ることができる。

ユーザは、ワークスペースウィンドウ上の Save history ボタンを押すことにより、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ノードとして履歴ツリーに追加される。その際には、(1) 入力されているテキストデータ、(2) ワークスペースの状態、(3) 結果表示パネルの状態、が履歴として保存される。ユーザが履歴ツリーの任意のノードをクリックした際に、システムはこれらの情報を用いて以前の探索の状態を復元する。

#### 5. 議論

Fan らは、テキストマイニングに用いられる要素技術を整理し、それらを基にテキストマイニングのプロセスモデルを提案している [Fan 06]。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。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、このモデルでは、テキストマイニングの目的である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述べられていない。また、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知識獲得のプロセスは考慮していない。また、繰り返しの伴う情報アクセスであるという点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、単にそれら技術を採用するだけでは、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。

一方で、情報可視化に関する研究では、情報をどのように見せるのかという可視化の技術ではなく、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に着目して、プロセスモデルの提案やタスクの分類が行われている [Card 99]。テキストマイニングにおいても同様に、テキスト集合から知識を得る場合に人はどのようにそれ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れらがなされるべきであろう。

#### 6. おわりに

本研究の目的は、テキスト集合からの知識獲得の支援である。我々はテキストマイニングを、Exploratory Search に該

当するタスクであり、創造的活動であると見做している。

本稿では、これらの 2 つの観点から、それを支援するシステムのインタフェースを構築するためのデザイン指針を提案した。加えて、提案したデザイン指針に基づいたインタフェースのプロトタイプを実装した。

#### 謝辞

本研究の遂行にあたり、文部科学省科学研究費 (課題番号: 22300048、24650040) の助成を受けた。記して謝意を表す。

#### 参考文献

- [Card 99] Card, S. K., Mackinlay, J. D., and Shneiderman, B.: *Readings in information visualization: using vision to think*, Morgan Kaufmann (1999)
- [Fan 06] 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)
- [Hearst 99] Hearst, M. A.: Untangling text data mining, in *In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics*, pp. 3–10 (1999)
- [Hori 04] Hori, K., Nakakoji, K., Yamamoto, Y., and Ostwald, J.: Organic Perspectives of Knowledge Management: Knowledge Evolution through a Cycle of Knowledge Liquidization and Crystallization, *Journal of Universal Computer Science*, Vol. 10, No. 3, pp. 252–261 (2004)
- [Marchionini 06] Marchionini, G.: Exploratory Search: From Finding to Understanding, *Commun. ACM*, Vol. 49, No. 4, pp. 41–46 (2006)
- [中小路 04] 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)
- [那須川 01] 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)
- [Shneiderman 93] Shneiderman, B.: 1.1 direct manipulation: a step beyond programming languages, *Sparks of innovation in human-computer interaction*, Vol. 17, p. 1993 (1993)
- [砂山 14] 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)
- [Taylor 68] Taylor, R. S.: Question-negotiation and Information Seeking in Libraries, *College & Research Libraries*, Vol. 29, No. 3, pp. 178–194 (1968)
- [White 09] White, R. W. and Roth, R. A.: *Exploratory Search: Beyond the Query-Response Paradigm*, Morgan & Claypool Publishers (2009)