

文章作成過程における棄却テキスト断片の活用に関する一検討

生田 泰章^{†1} 才記 駿平^{†1} 西本 一志^{†2}

概要：従来、知識の再活用に関する研究が多数行われてきた。これらの研究における再活用の対象は、その有用性が認識されている知識であり、ほとんどの場合、論文や技術資料などとして形式的に記録されているものである。一方、そのような「有用な知識」を構築する過程で、いったん生成されたにもかかわらず、主題にそぐわないなどの理由で棄却される情報（たとえば論文執筆において、いったん執筆したものの最終的に削除されたパラグラフなど）も多く存在する。このような棄却情報も、主題が異なれば有用な知識として活用できる可能性があるが、従来このような棄却情報を蓄積し、活用する試みはなされてこなかった。本稿では、このような棄却情報の活用に向けた基礎的検討として、文章作成を対象とし、その過程において執筆者が削除したテキスト断片を収集・分析することで、棄却情報活用の可能性と、その適切な収集方法について検討する。

Are Discarded Textual Pieces in a Text Writing Process Useful?

HIROAKI IKUTA^{†1} SHUMPEI SAIKI^{†1} KAZUSHI NISHIMOTO^{†2}

Abstract: A lot of studies on reuse of knowledge have been performed so far. The ordinary studies have mainly attempted to reuse “useful knowledge” most of which are formally described as, for example, technical papers. On the other hand, there are many pieces of information that had been once generated but finally discarded because of some reasons such as that they were not consistent with the subject. There are possibilities that such discarded pieces of information would become useful in some different subject. However, to the best of our knowledge, there have been no attempts to collect and utilize the discarded pieces of information. Towards utilization of the discarded pieces of information, in this paper, we collect and analyze discarded textual pieces in a text writing process, and discuss possibilities of utilizing them as well as appropriate methods to collect them.

1. はじめに

今日まで、論文や技術資料等の文章化された知識が数多く創出されてきた。一般的に、論文や技術資料等の文章は、ある主題に基づいて論旨が展開されることにより、首尾一貫した内容で構成されることが好ましい。文章作成過程において、執筆者は、考えや思想の表出作業、表出内容の修正作業を再帰的に行い、首尾一貫性を高めた文章を作成する[1]。このような文章作成過程において、執筆者の表現したい主題と関連のない知識や情報は、いったん創出されても最終的には棄却されてしまい、通常は活用されることはない。

しかしながら、このような棄却された知識や情報が、元の主題とは関連のない別の新たな主題の下における知識の創出に有効に活用される場合がある。例えば、ポスト・イット®に使用される接着剤は、一般的な接着剤の開発過程において創り出された失敗作であり、本来棄却されるものであったが偶然保管されていた。そして、この接着剤は、付箋を接着面に付け外し自在とするための主要な構成要素

として、今日においても活用されている[2]。また、本稿第3 著者は、過去に実施したピアノ演奏の表現生成に関する研究[3]の中で、打鍵とそれによる発音のタイミングに微小な遅延がある場合、演奏者はそれを発音の遅延としてではなく、鍵盤の重さの増加として認知することを見出し、論文[3]の執筆過程でいったんその発見を文章化した。しかし、最終的な論文では、この知見に関する文章は論文[3]の主題と関連しないために棄却された。後年、本稿第3 著者は、この棄却された文章を読み返した際、当時進行中であったドラム演奏支援の研究への応用可能性を見出し、論文[4]として研究成果をまとめた。

しかしながら、このようなケースはきわめてまれである。上述のように、文書作成過程においては、主題に関連のない知識や情報はほとんどすべて単純に棄却されてきたため、それらが新たな知識の創出につながる機会が失われているおそれがある。そこで本稿においては、文章作成過程における執筆者の棄却情報（すなわち棄却テキスト断片）を収集し、その活用可能性を検討するとともに、棄却テキスト断片の適切な収集方法を考案する。

^{†1} 北陸先端科学技術大学院大学 知識科学研究科
School of Knowledge Science, Japan Advanced Institute of Science and Technology

^{†2} 北陸先端科学技術大学院大学 ライフスタイルデザイン研究センター
Research Center for Innovative Lifestyle Design, Japan Advanced Institute of Science and Technology

2. 関連研究

従来、組織等の知識を埋没させずに共有して活用可能とするために、各種知識の収集・蓄積が取り組まれてきた。知識ベースシステムの研究分野では、各種専門分野に関する専門家の知識を収集・蓄積し、再利用可能とすることが試みられている[5]。これは、有用な知識を再利用可能とする取り組みである。文献[6]においては、失敗事例を蓄積した失敗知識データベースが構築されている。失敗事例をそのまま利用しても、おそらく同じ失敗を繰り返すだけであるので、有用な知識とはならない。しかし、失敗事例を吟味し、失敗の原因を読み取り、反面教師として活かすことで、失敗事例は有用な知識となりうるし、そもそもこの失敗知識データベースは、そういう意図で構築されている。したがって、これらの研究はいずれも、収集対象となる知識は有用と判断されている知識である。これに対し、本研究では、有用ではないと判断された知識を収集し、これを別の主題下で活用することを試みる。

3. 棄却テキスト断片の収集方法

棄却情報を別の主題下で活用可能かどうかを検証するためには、まず実際に棄却情報を収集し、これを分析することが不可欠である。そこで本稿では、文章作成作業を対象とし、その作業過程における棄却情報（すなわち棄却テキスト断片）を収集するため、棄却テキスト断片収集用のテキストエディタ（以下、棄却テキスト断片収集エディタと呼ぶ）を、C#を用いて Windows 上に実装した。棄却テキスト断片収集エディタは、図 1 に示すような編集画面を有している。また、「コピー」、「切り取り」、「貼り付け」、「検索」、「置換」等の編集機能を実装し、執筆者はこれらの編集機能を用いながら文章を編集することができる。図 2 に、検索・置換機能画面を示す。

執筆者が文章を編集中に、文字、単語、文、段落等のテキスト断片を削除した場合、棄却テキスト断片収集エディタは、削除されたテキスト断片を棄却テキスト断片として収集する。具体的には、棄却テキスト断片は執筆者が以下の 3 つの操作を実行したときに収集される。

- (1) 削除キーの操作
- (2) 文字列が範囲選択された状態での文字入力
- (3) 置換機能の実行

ここで削除キーとは、「Delete キー」及び「Backspace キー」の両方を表す。また、上記の操作が行われた際には、棄却テキスト断片に加え、当該棄却テキスト断片の周辺情報を併せて収集する。周辺情報とは棄却テキスト断片の前後の情報であり、棄却テキスト断片とその前後それぞれにある区切り文字とで囲まれた文字列である。本稿における区切り文字は「.」「。」とした。例えば、ある文に含まれる単語が削除された場合、その文に含まれる、削除された単語を除く残りの文字列を周辺情報として収集する。

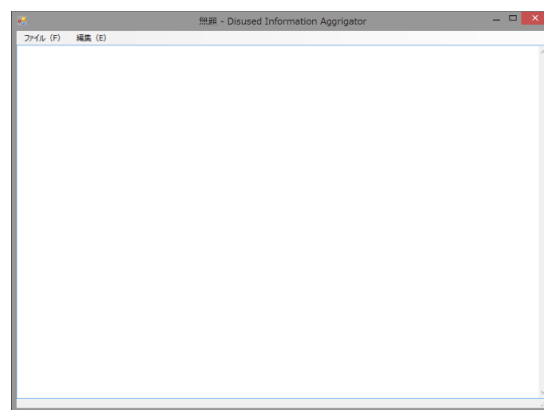


図 1. 棄却情報収集エディタの編集画面



図 2. 検索・置換機能画面

以下、棄却テキスト断片が収集されるべき各操作が執筆者により実行されたときにおける棄却テキスト断片収集エディタの処理手順について説明する。

3.1 削除キーが操作されたときにおける棄却テキスト断片収集エディタの処理

図 3 は、削除キーが執筆者に操作されたときにおける棄却テキスト断片収集エディタの処理手順を示すフローチャートである。まず、執筆者によって削除キーが押下されたか否かを判定する（ステップ S1）。削除キーが押下されていないと判定した場合（ステップ S1：NO）、執筆者から受け付けたその他の処理を実行し、削除キーが押下されるまで処理を待つ。削除キーが押下されたと判定した場合（ステップ S1：YES）、棄却テキスト断片を取得する（ステップ S2）。ステップ S2 において削除キーに応じた文字又は文字列を切り取り、一時的に記憶しておく。削除キーに応じた文字とは、「Delete キー」が削除キーであった場合は、カーレット直前の 1 文字又は範囲選択された文字列であり、「Backspace キー」が削除キーであった場合は、カーレット直後の文字又は範囲選択された文字列である。その後、周辺情報を取得する（ステップ S3）。すなわち、カーレットの直前及び直後にある区切り文字までの文字列をコピーし、一時的に記憶しておく。

次いで、削除動作の終了を示す操作があったか否かを判定する（ステップ S4）。削除動作の終了を示す操作とは、削除動作が継続して行われないことを示す操作であり、今回は削除キー以外のキーボード操作、及びマウスクリックの操作とした。削除動作の終了を示す操作がなかったと判定した場合（ステップ S4：NO）、ステップ S1 に処理を戻す。そして、再びステップ S2 の処理を行うに際して、先に

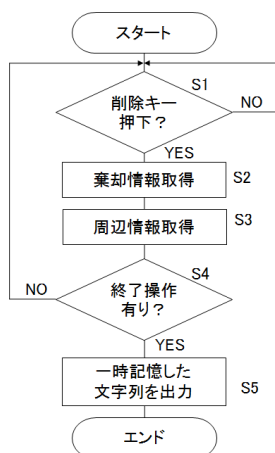


図 3. 棄却テキスト断片を収集するときの処理手順

一時記憶した文字列又は文字に追加するように対象の文字列又は文字の一時記憶を行う。また、再びステップ S3 の処理を行うに際して、周辺情報の更新を行う。削除動作の終了を示す操作があったと判定した場合 (ステップ S4: YES), ステップ S2 で取得された棄却テキスト断片及びステップ S3 で取得された周辺情報それぞれを表す文字列をファイルに出力する (ステップ S5)。ステップ S1～ステップ S4 の処理を行うことによって、複数回にわたって削除キーが押下された場合に、押下された回数分の文字列を一単位としてファイルに出力することができ、最後に押下されたときにおける周辺情報をファイルに出力することができる。

ステップ S5 の処理を行った後、棄却テキスト断片を収集する処理を終え、「日時」、「出力動作」、「文字列」、「周辺情報」を属性とする xml 形式にてファイル出力を行う。出力動作は、ファイルに出力するときに執筆者が行った動作を表し、本稿においては「削除」、「置換」に加え、「切り取り」及び「貼り付け」も含む。なお、「切り取り」及び「貼り付け」を出力動作として含む理由は後述する。

3.2 文字列が範囲選択された状態で文字入力が行われたときにおける棄却テキスト断片収集エディタの処理

文字列が範囲選択された状態で文字入力操作が執筆者によって行われた場合、範囲選択された文字列の表示位置に入力された文字列が表示される。すなわち、上述のような削除キーの操作によらず、範囲選択された文字列が削除される。そこで、棄却テキスト断片収集エディタは、文字列が範囲選択されるつど、その文字列を一時記憶しておく。その後、範囲選択された状態で文字入力操作が行われたことを検知した場合に、上述のステップ S6 同様、一時記憶された文字列をファイルに出力することによって棄却テキスト断片を収集する。なお、棄却テキスト断片収集エディタが検知する文字入力操作とは、執筆者による文字入力に係るキーボード操作の他、貼り付け機能の実行による文字入力も含む。

3.3 置換機能が実行されたときにおける棄却テキスト断片収集エディタの処理

文字列の置換機能が実行された場合、検索文字列 (図 2 における「検索する文字列」のテキストボックス内に格納された文字列) に代えて、置換文字列 (図 2 における「置換後の文字列」のテキストボックス内に格納された文字列) が表示される。すなわち、3.2 節同様、削除キーによらず、検索文字列が削除される。そこで、置換機能が執筆者によって実行された場合に、検索文字列をファイルに出力する。その際、「出力動作」の属性において置換機能が実行されたことを識別するように、検索文字列をファイルに出力する。

その他、「切り取り」及び「貼り付け」の機能を執筆者が行った場合にも、当該操作の対象となる文字列をファイルに出力するように実装した。執筆者によって切り取られた文字列は、執筆者が他のテキスト等に利用する意図があると考えられるため、現在の編集画面にて切り取られ、かつ当該編集画面にて貼り付けられていない文字列を棄却テキスト断片とすべく、このような実装を行った。

4. 実験

インタラクション 2016 のインタラクティブ発表用の日本語論文の一部を 4 人の被験者それぞれに棄却テキスト断片収集エディタを用いて作成してもらうことにより、棄却テキスト断片を収集した。表 1 に実験結果を示す。各被験者が作成した文章の文字数及び文の数と、当該文章の作成に際して収集された棄却テキスト断片の数とが示されている。なお、表 1 中の被験者 1 は、本稿の第 1 著者であり、本実験に際して第 4 章～第 6 章を除く部分を棄却テキスト断片収集エディタにより作成した。ただし、第 4 章執筆後の加筆修正箇所については、実験結果に含めていない。

本実験によって収集された棄却テキスト断片を精査したところ、被験者が以下の 3 つの編集作業を行うことにより棄却テキスト断片が生成されることが分かった。

- (1) 誤字の訂正
- (2) 表現の修正
- (3) 主題と異なる内容の削除

ただし、本実験では、棄却テキスト断片の大半は(1)及び(2)の編集作業により生成され、(3)の編集作業により生成された棄却テキスト断片の数は被験者 1 により生成された 1 つだけであった。具体的には、「2002 年にノーベル化学賞を受賞した田中耕一氏は、実験用の試料の配合に失敗したが、失敗作の試料を捨てずに実験に用いることでノーベル賞受賞につながる発見をすることができた。」という文である。この文は、本稿の第 1 章において、ある主題では棄却されるようなものが別の主題では有効となる場合があるということを示す例として記述されていたが、同じ主題において有効となる場合があることを示す内容であり、適した記述ではないと判断されて削除された。その後、当該文に代

表 1. 実験結果

	被験者 1	被験者 2	被験者 3	被験者 4
文字数	4726	2468	535	418
文の数	72	43	12	10
棄却テキスト断片数	551	124	16	99

えて、上述したポスト・イット@の例が記述されている。

5. 議論

上述の実験結果より、本稿で提案した棄却テキスト断片収集エディタを使用することによって、文章作成過程における棄却テキスト断片を収集することができることを確認した。また、本実験においては1つだけであったが、他の主題において活用されるかもしれない棄却テキスト断片を収集することができることを確認した。しかしながら、上述の実験結果から、棄却テキスト断片収集エディタによる棄却テキスト断片の収集について、少なくとも2つの課題があることを確認した。そこで、本章においては、各課題を解決するための方法について議論する。

まず、1つ目の課題は、収集された棄却テキスト断片がきわめて雑多であることである。特に、第4章で示した(1)及び(2)の編集作業によって生成された棄却テキスト断片は、文章作成過程における誤字又は文章表現に係る言葉等を表す情報であるため、異なる主題で活用される可能性は極めて低い。よって、(1)及び(2)の編集作業によって生成された棄却テキスト断片をノイズとして除去する必要がある。このようなノイズを除去する方法として、棄却テキスト断片と文章本文とを比較する方法が考えられる。例えば、棄却テキスト断片及び周辺情報からなる文字列と、当該周辺情報を含む文字列との編集距離が閾値以下であるか否かによりノイズであるか否かを判定する方法が考えられる。

2つ目の課題は、他の主題において活用されるかもしれない棄却テキスト断片の数が少ないことである。この課題は、そもそもそのような棄却テキスト断片が少ないということに加え、棄却テキスト断片収集エディタが、文章の作成過程における最終状態のみを表示可能とする、いわゆる清書用のメディアであることが一因であると考えられる。つまり、執筆者は、棄却テキスト断片収集エディタにて文章の作成を行う際には、当該文章の主題に関する内容をすでになんらかの別の手段によって整理した状態であることが考えられる。本実験においても、被験者4は、棄却テキスト断片収集エディタにて文章の作成を行う前に、文章構成を紙に整理していた。Nakakoji らが開発した ART#001[7]は、全体の文章を作成するに際して、部分的な文章を表すチャンクを線形的につなぎ合わせることができる GUI を備えている。すなわち、執筆者は ART#001 を用いることで、

チャンクを試行錯誤的に組み合わせることによって、文章の作成過程を視覚的に把握しながら自身が目的とする文章を作成することができる。棄却テキスト断片収集エディタにおいても、このような GUI を採用することによって、文章作成過程において生成されたが使用されなかったチャンクを棄却テキスト断片として収集することができる。これにより、2つ目の課題を克服できると考えられる。また、ART#001のように、チャンクを線形的につなぎ合わせて全体の文章を作成することによって、誤字又は文章表現に係る言葉等を表す情報を収集する必要がないため、上述の1つ目の課題も同時に克服できると考えられる。

6. まとめ

本稿においては、作成対象の文章における主題とは異なる情報を棄却テキスト断片として収集するために、棄却テキスト断片収集エディタを実装した。また、被験者に棄却テキスト断片収集エディタを使用してもらう実験を行うことによって、棄却テキスト断片収集エディタにより棄却テキスト断片が取得可能であることを確認すると共に、当該棄却テキスト断片収集エディタの課題を明確にすることができた。今後は、上述の ART#001 の GUI のように、文章作成過程を視覚的に把握可能で、かつ直感的に編集作業が行うことが可能な GUI を備えることで、上述の2つの課題を克服したシステムの開発を進める予定である。

謝辞 本研究は、JSPS 科研費 15K12093 の助成を受けたものです。本稿の執筆に当たり、実験に協力下さった被験者の方々に謝意を表します。

参考文献

- 1) 丸野俊一：対話の視点から捉えた書く時の推敲過程，人工知能学会誌，Vol. 23 No. 1, pp.293-302 (2008).
- 2) 宮永博史：セレンディピティと MOT，三菱 UFJ リサーチ&コンサルティング 季刊 政策・経営研究 2009 vol.3 pp. 50-60 (2009).
- 3) 大島千佳，西本一志，阿部明典：ピアノ演奏における離鍵速度の重要性と特性に関する考察，情報処理学会論文誌，Vol.47, NO.5, pp.1546-1557 (2006).
- 4) 池之上あかり，小倉加奈代，鶴木祐史，西本一志：微小遅延聴覚フィードバックを応用したドラム演奏フォーム改善支援システム，Vol.1, No.1, pp. 15-24 (2013).
- 5) 元田浩：明示的理解に魅せられて，人工知能学会誌 14(4), pp. 615-625 (1999).
- 6) 畑村洋太郎，中尾政之，飯野謙次：失敗知識データベース構築の試み，情報処理 44(7), 733-739 (2003).
- 7) Nakakoji, K., Yamamoto, Y., Reeves, B.N., Takada, S., Two-Dimensional Positioning as a Means for Reflection in Design, Proceedings of Design of Interactive Systems (DIS'2000), ACM Press, New York, NY, pp.145-154, 2000.