

情報の関係性に着目した文書作成支援システム

Document writing support system focused on relationships between information pieces

重田 桂誓*¹ 松村 敦*¹ 宇陀 則彦*¹
Keisei SHIGETA Atsushi MATSUMURA Norihiko UDA

*¹筑波大学大学院 図書館情報メディア研究科

Graduate School of Library, Information and Media Studies, University of Tsukuba

We constructed a document writing support system focused on relationships between pieces of information. Using our system, scattered information pieces are collected at first. Next, relationships between the information pieces are defined on the two-dimensional space, by trial and error. Finally the pieces of information are converted into a complete document automatically. To evaluate effectiveness of our system, we conducted an experiment comparing with MS Word. As a result, our system induced trial and error to decide the order of sentences, and quality of the documents written by our system and by MS Word are equivalent.

1. はじめに

WWWの発展により、大量の情報が入手可能になった。同時に、1つの事柄に関する情報が複数のWebページに分散するようになった。そのため、ある事柄を深く理解するためには、何度も検索を繰り返して複数の文書を読まなければならない、多大な労力を要する。

その解決策として有効だと考えられるのが、情報のまとめである。分散している情報を1つにまとめたものがあれば、それを見た者は少ない手間でも面的かつ詳細な情報を入手できる。

情報をまとめることに着目した研究として、情報編纂[松下 09]がある。情報編纂では数値データ中心の情報を、自動でグラフにまとめることができる。しかし、より多様な表現ができるテキスト中心の情報を、1つの文書にまとめる研究は進んでいない。その原因の1つは、語の多義性や表記の揺れといった自然言語の曖昧性をコンピュータで自動処理するのが困難なことにある。それを解決するためのアプローチとして、システムの支援により人手で自然言語の曖昧性を解釈することが考えられる。実際に、デザインプロセスとしての文章作成支援[柴田 03]ではユーザが手動で文章を作成する作業をシステムにより支援する枠組みを提案し、その有効性を示している。ただし、文章の素材となる断片情報を組織化する際に、情報間の多様な関係を扱うことができないことが課題となっている。そのため、多種多様な情報をユーザの意図に沿ってまとめることには適していないと考えられる。本研究では、ユーザによる、テキスト中心の情報をまとめた1つの文書(まとめ文書)の作成を支援するシステムの構築を目的とする。

2. 本研究のアプローチ

まとめ文書の作成においては、まず文書の素材となる断片情報を複数生成し、次にそれらをユーザが考える文脈に沿って1つの文書にまとめあげる。したがって、複数の断片情報間の関係を解釈して、試行錯誤しながら文脈を構築することが必要となる。そこで本研究では、情報の関係性に着目し、複数の情報の関係を試行錯誤しながら、文脈を構築できるシステムを

指す。

本研究におけるまとめ文書の作成プロセスを図1に示す。まず、生成プロセスで文書の素材となる断片情報を作成する。対象とする断片情報の粒度は文章、文、単語などいくつか考えられるが、それらを「情報オブジェクト」という単位で一括して管理する。ユーザは自由に文章や文、単語を入力して情報オブジェクトを生成することができる。

次に、生成した複数の情報オブジェクトに対して、文脈を考えながら編集操作を行う。具体的には、情報オブジェクトの移動による文章の並び替えや、情報オブジェクト間の関係付けを行う。また、生成した情報オブジェクトが保持するテキストを修正したり、生成プロセスに戻って新たな情報オブジェクトを作成したりする。このプロセスを通して、徐々に文脈が構築される。

文書変換プロセスは、ユーザが情報オブジェクトに対して操作を加えるたびにシステムにより自動で実行される。変換結果は即座に編集画面へとフィードバックされるので、文書全体の構成や流れを常に確認しながら編集できる。

最後に、テキストエディタを用いてユーザが自由に文書を修正する。これは、自動での文書変換では十分に反映できないユーザの意図を、手動での修正により補うことをねらいとしている。

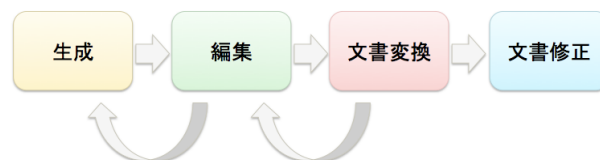


図1: まとめ文書の作成プロセス

3. システム

本システムは、情報オブジェクト間の関係の定義とそれに応じた文書変換を行うことを特徴とする文書作成支援システムである。システム構成を図2に示す。

本システムは、「情報オブジェクト生成部」、「編集支援部」、「文書変換部」、「文書修正部」という4つのモジュールで構成

連絡先: 重田 桂誓, 筑波大学大学院 図書館情報メディア研究科, 〒305-8550 茨城県つくば市春日 1-2, shigeta@slis.tsukuba.ac.jp

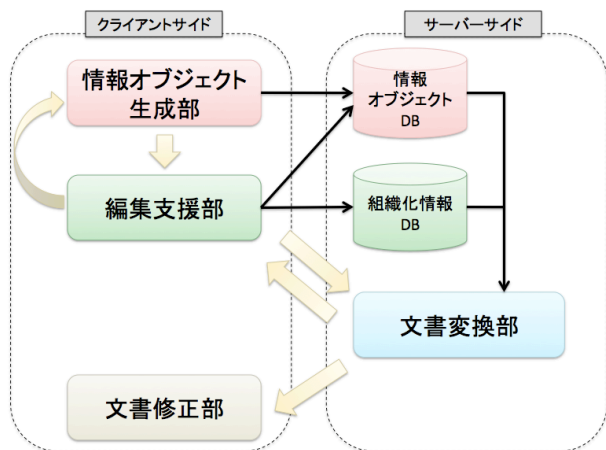


図 2: システム構成

されている。これら4つのモジュールは図1の各プロセスに対応している。

3.1 情報オブジェクト生成部

情報オブジェクト生成部では、テキストエディタによる手動での文章入力とRSSの読み込みにより、文書の素材となる情報オブジェクトを生成することができる。特に、RSSを読み込む機能を使えば一度に複数の情報オブジェクトを生成することができる。さらに、[tumblr](http://www.tumblr.com/)のような外部のウェブコンテンツをクリッピングするサービスのRSSを活用すれば、情報抽出を素早く簡単に行うことができる。生成した情報オブジェクトは情報オブジェクトデータベースに保存された後、編集支援部の二次元空間に表示される。

3.2 編集支援部

編集支援部では、二次元空間上で情報オブジェクトの移動や情報オブジェクト間の関係の定義、情報オブジェクトが保持するテキストの修正といった操作を行うことができる。

ドラッグ&ドロップにより、ユーザは二次元空間の自由な位置に情報オブジェクトを移動することができる。

情報オブジェクト間の関係の定義は、類似、従属、対立の3種類がある。複数の情報オブジェクトを重ねると、それらを囲む枠線が表示され、類似関係を定義することができる。これにより、似たようなことを述べている文章や関連のある文章をグルーピングでき、一度にまとめて移動することができる。従属関係は、2つの情報オブジェクトを指定して、それらを結ぶ一方向の矢印を引くことにより定義できる。対立関係は、2つの情報オブジェクトを指定して、それらを結ぶ双方向の矢印を引くことにより定義できる。

情報オブジェクトが保持するテキストは、エディタを用いて修正できる。これにより、編集過程で得られた洞察を文章に反映することができる。

図3に本システムを用いた編集状態の例を示す。画面の右側に、文書変換部の結果が表示されるので、文書全体の構成や流れを常に確認することができる。なお、編集操作が行われる度に、情報オブジェクトの位置と保持するテキストは情報オブジェクトデータベースへ保存され、オブジェクト間の関係は組織化情報データベースに保存される。

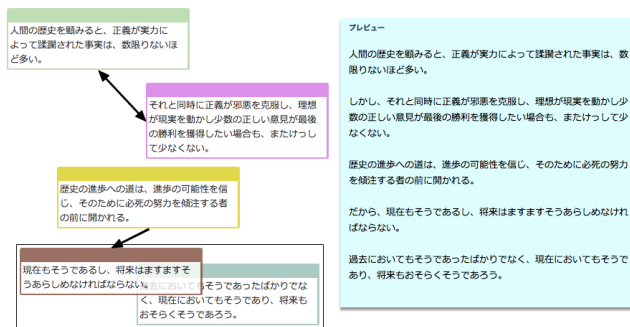


図 3: 編集状態の例

3.3 文書変換部

文書変換部では、システムが保持する文書変換ルールに則って、編集状態を文書に変換する。具体的には、情報オブジェクトが保持するテキストを上から順に結合すると同時に、定義された関係を接続詞に変換する。従属関係は「だから」、対立関係は「しかし」という接続詞に変換される。ただし、ここで付与される接続詞は、編集支援部においてユーザが自由に変更することができる。なお、類似関係は接続詞に変換されない。

3.4 文書修正部

文書修正部では、自動での文書変換では十分に反映できなかったユーザの意図を、ユーザ自身がテキストエディタで自由に修正することができる。

4. 評価実験

4.1 概要

本システムの編集支援部と文書変換部の有効性を評価するため、一般的なワードプロセッサである「MS Word」との比較実験を行った。対象は学生8名である。課題は公務員試験問題集[資格試験研究会 07]を参考に2題用意した。これは回答が一意に定まる課題を用いることで、完成した文章の質を点数化して比較するためである。具体的には、1つの文章を文単位で区切って文頭の接続詞を削除し、順不同に並べたものを被験者に提示する。それらを意味の通る文章に並び替え接続詞を補充するものである。これにより、文の並び替えと接続詞付与という2種類の操作から、文章作成における試行錯誤のしやすさを評価する。

実験では、まず、課題に慣れてもらうために事前テストを行った。事前テストは2つの問題をシステムを使わず紙で解くものである。次に、評価実験として課題1、2をWordと本システムを使って解いてもらった。このとき、表1に示すように課題とシステムの組み合わせがそれぞれ異なる4つのタスクを用意した。そして、表2に示すように、カウンターバランスを考慮して、被験者をA、B、C、Dの4グループに分け、各グループごとにタスクの組み合わせや、タスクを実行する順序を変化させた。これは、課題の難易度や被験者の能力による差が実験結果に及ぼす影響を吸収するためである。事前テスト及び課題は全て制限時間なしで行い、被験者の自己申告で終了とした。最後に、全部で10項目から成るアンケートに答えてもらい、実験終了とした。なお、取得したデータはデスクトップの動画キャプチャ、システムの操作ログ、課題に対する回答、文章完成までの時間、アンケートの5種類である。

表 1: 4つのタスク

	課題 1	課題 2
Word	①	②
本システム	③	④

表 2: 評価実験

グループ A	① → ④
グループ B	④ → ①
グループ C	② → ③
グループ D	③ → ②

表 3: 1分あたりの文の入れ替え回数及び接続詞の付与回数の平均 (回/分)

	文の入れ替え	接続詞の付与
Word	0.8	0.6
本システム	2.0	0.4

4.2 実験結果

表 3 に 1 分あたりの文の入れ替え回数と接続詞付与の回数の平均を示す。文の入れ替え回数は本システムの方が Word よりも多かったが、接続詞の付与回数は Word の方が多かった。

表 4 に文章完成までの時間の平均を示す。Word と本システムの間には大きな差は見られなかった。表 5 に完成した文章の点数の平均を示す。こちらも Word と本システムの間には大きな差は見られなかった。

図 4 に、本システムと Word を比較してそれぞれの優れている点を 8 項目の中から選択するアンケート (複数選択可) の結果を示す。

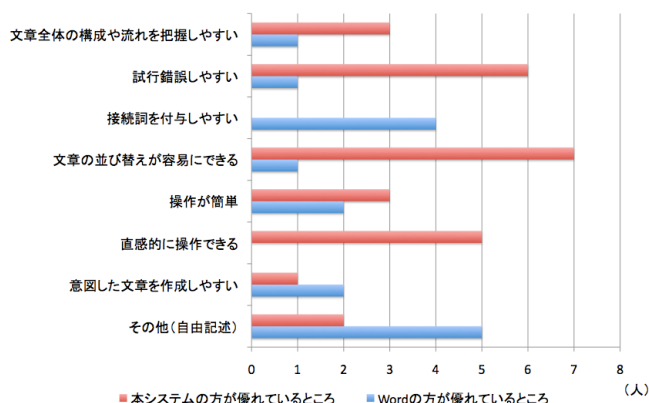


図 4: 本システムと Word の優れている点 (複数回答可)

なお、その他の項目では以下のような意見が得られた。

- 本システムの方が優れているところ
 - － グループ機能 (1名)
 - － オブジェクトの色分け (1名)
- Word の方が優れているところ

表 4: 文章完成までの時間の平均

	課題 1	課題 2	平均
Word	8 分 39 秒	8 分 25 秒	8 分 32 秒
本システム	10 分 11 秒	7 分 18 秒	8 分 45 秒
Word - 本システム	- 1 分 32 秒	1 分 7 秒	- 13 秒

表 5: 完成した文章の平均点

	課題 1	課題 2	平均
Word	80.0 点	76.3 点	78.1 点
本システム	77.5 点	81.3 点	79.4 点
本システム - Word	- 2.5 点	5.0 点	1.3 点

- － メモが取れる、コメントを付けることができる (2名)
- － フォントを変えることができる (1名)
- － 文章校正機能 (1名)
- － 特に無し (1名)

5. 考察

文の入れ替え回数は本システムの方が Word に比べて 2.5 倍多かった (表 3)。同じ時間でより多く文を入れ替えたということは、それだけ多く文の並びのパターンを考えたということであり、試行錯誤が促進されたといえる。実際に、アンケート結果 (図 4) から、本システムの優れているところとして 8 人中 7 人の被験者が「文章の並び替えが容易にできる」、6 人の被験者が「試行錯誤しやすい」という項目を挙げている。これは、Word で文を入れ替える際のコピー&ペーストに対して、本システムでのドラッグ&ドロップは操作の手間が少ないからだと考えられる。

接続詞の付与回数は Word の方が本システムに比べて 1.5 倍多く (表 3)。アンケート結果 (図 4) から「接続詞を付与しやすい」という項目は Word の方が評価が高い。これは、Word ではキーボードを使って素早く自由な接続詞を付与できるのに対して、本システムでは矢印を引いた上で意図にあうように矢印の種類や接続詞を変えるといった手間が生じるためだと考えられる。このような手間によって、本システムは接続詞付与に関する試行錯誤を妨げてしまったといえる。

文章完成までの時間と完成した文章の質を Word と本システムで比べたところ、共に大きな差は見られなかった (表 4,5) デスクトップの動画キャプチャを分析したところ、本システムを使った被験者の多くは絶えず何らかの操作を行っていた。その中には、文の入れ替えがない、つまり変換後の文章に影響を及ぼさないような情報オブジェクトの移動が多く含まれていた (表 6)。本システムは情報オブジェクトを移動する操作の自由度が高いため、十分に考えることなく、試行錯誤とは無関係な操作を行ったことを示唆している。一方 Word の場合は、本システムに比べて何も操作が行われない時間が長かった。文を入れ替える操作のしきいが本システムに比べて高いため、被験者は入れ替え後の文章を熟考してから、文を入れ替えたと考えられる。操作の自由度を高めることで、試行錯誤が促進されて文章の質が上がるという仮説を持っていたが、今回の実験ではそれを裏付ける結果は得られなかった。それどころか、本システムの情報オブジェクトを移動する際の自由度の高さが、

表 6: 情報オブジェクトの移動回数の割合

被験者	全移動	入れ替えがない移動	
		回数	割合 (%)
a	41	35	85.4
b	53	30	56.6
c	28	15	53.6
d	31	14	45.2
e	12	4	33.3
f	41	22	53.7
g	76	56	73.4
h	35	24	68.6
平均	39.6	25	58.7

熟考を妨げると同時に試行錯誤とは無関係と思われる操作を多く生むことになった。

6. おわりに

本研究では、テキスト中心の情報を1つの文書にまとめるプロセスを支援するシステムを構築した。特に情報の関係性に着目し、複数のテキストの接続関係を試行錯誤しながら、文脈を構築できるシステムを目指した。

MS Word との比較実験の結果、本システムの方が文の並び替えに関する試行錯誤のしやすさにおいて優れていることがわかった。一方で、接続詞の付与の試行錯誤はやりにくいこともわかった。また、完成した文章の質や、文章完成までの時間に大差はなかった。今後の課題は、文章の質を高めるような試行錯誤ができるシステムの設計と、効率的な熟考と試行錯誤のバランスを明らかにすることである。また、今回は比較対象に Word を用いたが、アウトラインプロセッサやアイデアプロセッサのような文章作成時の試行錯誤に焦点をあてたシステムとの比較実験も今後行いたい。

参考文献

- [松下 09] 松下光範, 加藤恒昭, 情報編纂研究促進のための試み, 人工知能学会論文誌, 2009, Vol. 24, No. 2, p. 272-283.
- [柴田 03] 柴田博仁, 堀浩一, デザインプロセスとしての文章作成を支援する枠組み, 情報処理学会論文誌, 2003, Vol. 44, No. 3, p. 1000-1012.
- [資格試験研究会 07] 資格試験研究会編, テキスト&ワーク [教養試験編] 一般知能 文章理解・資料解釈, 実務教育出版, 2007, p. 26-31, (公務員合格講座, 10) .