

# 論文の作成と再利用に基づく研究活動支援システム

A Research Activity Support System Based on Creation and Reuse of Technical Papers

林 亮介\*1      土田 貴裕\*1      大平 茂輝\*2      長尾 確\*3  
 Ryosuke Hayashi      Takahiro Tsuchida      Shigeki Ohira      Katashi Nagao

\*1名古屋大学大学院 情報科学研究科  
 Graduate School of Information Science, Nagoya University

\*2名古屋大学 エコトピア科学研究所  
 EcoTopia Science Institute, Nagoya University

\*3名古屋大学 情報メディア教育センター  
 Center for Information Media Studies, Nagoya University

In this research, we propose a method to promote more efficient research activities including creation of technical documents. Our method deals with tight interaction between creation (or authoring) and reuse (or quotation) of online technical papers. One of key technologies of our research is annotation. Marking and commenting during reading papers are typical examples of annotation. Our system also considers quoting some parts of papers as a kind of annotation. Researchers' creative activities are supported by retrieving and reusing annotation data.

## 1. はじめに

本研究では、研究活動をより効果的に促進させるシステムの構築を目的としている。その第一歩として、本論文では、研究活動における論文の閲覧・作成プロセスの重要性と関連性について述べ、閲覧・作成プロセス間を連携することにより可能になる新たな研究活動支援手法を提案する。

研究活動において、研究の新規性を示すことは重要である。研究の新規性を示すとは、自己と他者の研究を比較し、優位性を示すことである。そのためには、自己と他者の研究の現状を明らかにする必要がある。また、現状を明らかにすることは研究の方向性を明確にし、より研究に新規性を持たせることが可能になるだろう。そのような背景から、研究活動において研究の現状と方向性の明確化は重要な意味を持つ。

一般に、研究者は多くの時間を関連研究の調査（以後サーベイと呼ぶ）に費やす。基本的にサーベイの対象となる媒体は論文である。研究者は関連する論文を日々閲覧することで、自分の研究の現状を理解していく。そのため、論文の閲覧プロセスは研究の現状と方向性の明確化において重要な意味を持つ。

また、論文作成のプロセスも研究の現状と方向性の明確化において重要な意味を持つ。柴田ら [1] は、文章作成のプロセスにおいては、書き手は最初から書く内容が明確なのではなく、むしろ書いたことがトリガーとなり新たな思考を展開し、次に書く内容を考えると述べている。つまり、論文作成のプロセスも、単に今までの思考をまとめて文書化するだけでなく、研究に関する思考を深めるプロセスであると言える。また、論文には自分の研究の新規性を示すために他者の論文を適切に引用する。そのため、他者の論文を適切に引用することは、研究の現状について再考して、頭の中で曖昧であったものを文章として明確化することであると考えられる。

また図 1 のように、論文を作成する際に過去に閲覧した論文を適切に引用したり、作成された論文を再利用するために閲覧することから、本来論文の閲覧・作成というプロセス間には互いに関連性が存在すると考えられるため、互いに連携して支援することが望ましい。しかし、従来研究では、論文の閲覧を支援するシステムや論文の作成を支援するシステムなど、研究

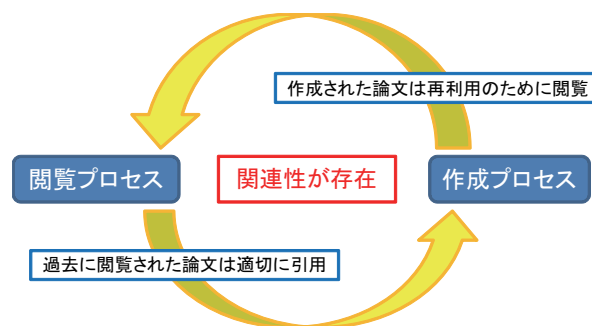


図 1: 論文閲覧・作成プロセスの連携

活動の個々のプロセスに特化したシステムが多い。

そこで本論文では、研究活動の現状と方向性の明確化という観点から、論文の閲覧・作成プロセスを関連付け、連携させることで研究活動を支援する仕組みを提案する。

## 2. 論文の閲覧・作成プロセスの連携による研究活動支援

本研究では、論文の閲覧・作成プロセスを連携させるアプローチとして、各プロセスにおけるユーザのアノテーションに着目する。アノテーションとは、コンテンツに対してメタデータを関連付け、従来困難であった高度な検索や要約などを実現する手法である [2]。

まず、論文の閲覧プロセスにおいて取得されるユーザの自発的な行為であるマーキングやコメントといったアノテーションを利用することで論文の作成プロセスにおける論文の引用支援を試みる。本研究では、これらのアノテーションを閲覧時アノテーションと呼ぶ。

さらに、論文の作成プロセスにおいて論文を引用した際に取得されるメタデータを利用し、論文の閲覧プロセスにおいてユーザの関連研究のサーベイ支援を行う。本研究では、このメタデータを引用アノテーションと呼ぶ。

本節では、これらのアノテーションを共有・利用することで、従来不可能であった支援を実現する手法を提案する。支援手法の詳細については以下で述べる。

連絡先: 氏名:林亮介, 所属:名古屋大学大学院情報科学研究科,  
 住所:名古屋市千種区不老町, 電話番号:052-789-5878, E  
 メール:hayashi@nagao.nuie.nagoya-u.ac.jp

## 2.1 閲覧時アノテーションを用いた論文の引用支援

論文の作成には、EndNote<sup>\*1</sup>のような文献管理を行う論文作成支援システムを利用している研究者が多い。EndNoteでは、文献情報を個人ライブラリに保存し、管理することができる。そして、論文作成の際に、ライブラリに保存した文献情報を検索し、引用文献として利用することができる。しかし、筆者らはEndNoteにおける論文の引用を支援する機能には、以下のような問題点があると考えている。

- 論文内のどの部分を引用したかったのか忘れる
- どのような意図で引用したかったのか忘れる

これらの問題点を解決するために、本研究では論文の閲覧プロセスにおける閲覧時アノテーションを利用する。先にも述べたように、閲覧時アノテーションとは論文の部分箇所に対するマーキングとコメントを指す。前者の問題点に関しては、論文の閲覧時にマーキングにより引用したい部分箇所を記録し、論文の作成時にマーキング箇所を容易に検索可能にすることで解決されることが考えられる。

閲覧時アノテーションには文脈や個人の背景知識といった様々なコンテキストが含まれる [3]。そのため、すべてのアノテーションを一様に取得すると、応用に利用することは困難であると考えられる。そこで本研究では、閲覧時アノテーションを文章の理解を目的としたアノテーションと論文の引用を目的としたアノテーションに分類している。これらの目的を取得するため、あらかじめユーザに閲覧時アノテーションに様々な属性を付与してもらう（属性例は表1に示す）。論文の部分箇所と属性を関連付け、アノテーションした意図を記録可能にすることで、後者の問題点の解決を試みる。また、副次的な効果として、意識的に属性を付与してもらうことでユーザの文章理解が深まると考えられる。

表 1: 閲覧時アノテーションの属性例

文章理解	辞書引き	まとめ	など
論文引用	論説根拠	論説対比	など

文章の理解を目的としたアノテーションは共有することで他者の論文閲覧支援に利用する。例えば、他者が英語の単語の意味を英和辞書で調べた場合などが挙げられる。辞書情報を共有することで、次に閲覧したユーザにその情報は提示可能になる。つまり、再度辞書で調べる必要がなくなるため、ユーザの閲覧にかかるコストを軽減できると考えられる。また、辞書情報を書き換えることが可能である。例えば、多義語の場合、提示された辞書情報が文脈に適していないことがある。その場合、辞書情報を書き換えることで、正しい情報に修正することが可能である。さらに、そのような修正が蓄積することで、提示される辞書情報に客観性や正確性を持たせることが可能になる。

さらに、論文の引用を目的としたアノテーションは論文作成支援に利用する。閲覧時アノテーションに属性を付与しておくことで、様々な目的に応じた引用が可能である。例えば、論説の根拠となる論文を引用したい場合、キーワード+マーキング属性として「論説根拠」を手掛かりとして検索することが可能になる。また、引用した際に引用元論文内の被参照箇所、引用

先論文内の参照箇所、引用目的という情報が取得される。被参照箇所はマーキング範囲、参照箇所は引用先論文の段落要素、引用目的は被参照箇所に含まれる閲覧時アノテーションの属性である。本研究では、これらの情報を引用アノテーションと呼ぶ。

## 2.2 引用アノテーションを用いた論文のサーベイ支援

論文のサーベイには、CiteSeer<sup>\*2</sup>等のシステムを利用している研究者が多い。CiteSeerでは、共引用の関係を利用して論文間の類似度を算出している。共引用の関係とは、同一の論文によって引用された論文同士の関係を指し、この関係にある論文同士には類似性があるとされている。この手法は、どのような共引用であっても、全て同一の類似度を示すことが前提となっている。しかし、論説の根拠を示す引用や論説の対比を示す引用等、引用には様々な意図があると考えられるため、この仮定には問題がある。そのような背景から、引用分析による引用意図の自動判別に関する研究も行われている [4]。しかし、筆者らはこのような手法には、以下のような問題点があると考えている。

- 引用意図の自動判別精度が完全ではない
- 論文のどの部分を引用しているのか明確でない

これらの問題点を解決するために本研究では、論文作成プロセスにおけるユーザの引用アノテーションを共有・利用する。前者の問題点に関しては、本手法では引用意図をアノテーションの属性として入力で入力しているため、精度向上が望める。さらに、引用アノテーションには引用元論文内の被参照箇所、引用先論文内の参照箇所といった論文の部分箇所に関する情報が含まれているため、後者の問題点も解決できると考えられる。

## 3. 研究活動支援システム

本節では、論文の閲覧・作成支援を行うためのシステムである研究活動支援システムについて述べる。本システムはブラウザ上で動作する Web アプリケーションである。Web アプリケーションとして構築することで、不特定多数のユーザからアノテーションを収集し、共有することが可能になる。

対象とする論文は、我々の研究室において作成・共有されている XML 形式の論文とする。本システムには、大きく分けて論文閲覧支援と論文作成支援の二つの機能が存在する。

### 3.1 論文閲覧支援

論文閲覧時におけるインタフェース画面例を図2に示す。ユーザに論文を表示する前処理として、論文に含まれる文字列を茶筌<sup>\*3</sup>を用いて形態素解析を行い、形態素に関する情報をタグとして分割して埋め込んでいる。前処理を行うことで、マーキング時における文字列の選択範囲の誤り防止等が可能になる。さらに、マーキング範囲を XPath を用いて指し示すことで、閲覧時アノテーションの再現が可能になる。本システムにおける主な論文閲覧支援機能は以下の4つである。

- アノテーション機能  
本システムは二通りのアノテーションを可能にしている。一つは、マーキングアノテーションである。ドラッグ&ドロップによるマウス操作を用いてパラグラフ内の任意の

\*1 EndNote - <http://www.endnote.com/>

\*2 CiteSeer - <http://citeseer.ist.psu.edu/>

\*3 茶筌 - <http://chasen.naist.jp/hiki/ChaSen/>

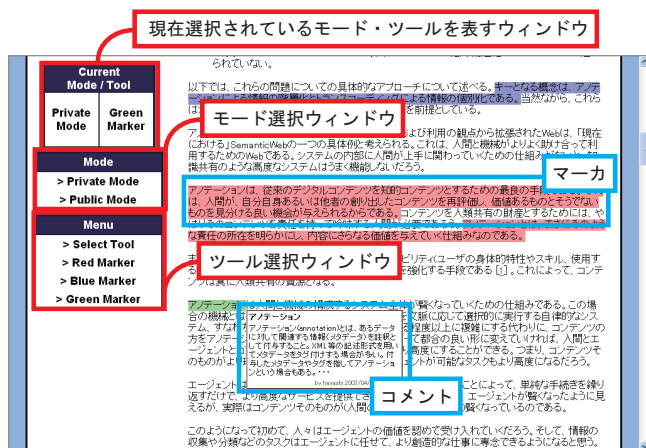


図 2: 論文閲覧支援インタフェース

文字列をハイライトさせることが可能である。あらかじめ、マーカの色に対して「論説根拠」や「辞書引き」等の属性を付与することが可能である。もう一つはコメントアノテーションである。コメントアノテーションはマーカ箇所に対して、メモ書きのようなコメントを付与する行為である。コメントにはタイトルと本文を記述することが可能である。さらに、コメントに対して「アイデア」や「まとめ」等の属性を付与することが可能である。書かれたコメントはマウスポインタをマーキングされた文字列上に置くことでポップアップ表示され、属性は背景色によって表される。

- 辞書引き機能  
 あらかじめ、マーキングの色に「辞書引き」属性を付与しておくことで、マーカ範囲の文字列に対して自動的に辞書引きを行う機能である。研究室内の専門用語辞書と Wikipedia<sup>\*1</sup> を利用する。さらに、専門用語辞書に存在しなかった項目はユーザが書き込むことで自動的に専門用語辞書に登録される。
- モード切替機能  
 ユーザ自身の閲覧時アノテーションのみを表示するプライベートモードと他者の閲覧時アノテーションを表示するパブリックモードの切替機能である。適切にモードを切り替えることで効率的な閲覧が可能になる。前者は、既読文書を効率的に閲覧するためのトリガーに、後者は未読文書を閲覧する際の文章理解の支援になると考えられる。
- サーベイ支援機能  
 引用アノテーションを用いることで、論文閲覧時に論文のサーベイ支援が可能になる。例えば、ある分野において重要とされる論文を閲覧したとする。その際、引用アノテーションを用いることで「この論文のこの箇所を引用して書かれた論文がある」という情報が論文のメタデータとしてユーザに提示される。さらに、引用目的が付与されているため、引用した論文がどのような目的で引用したのかということが参照可能である。また、通常引用元に向かってサーベイを行うが、引用アノテーションを用いることで、引用先へ向かう関連研究のサーベイを行うことが可能になる。

\*1 Wikipedia - <http://ja.wikipedia.org/>

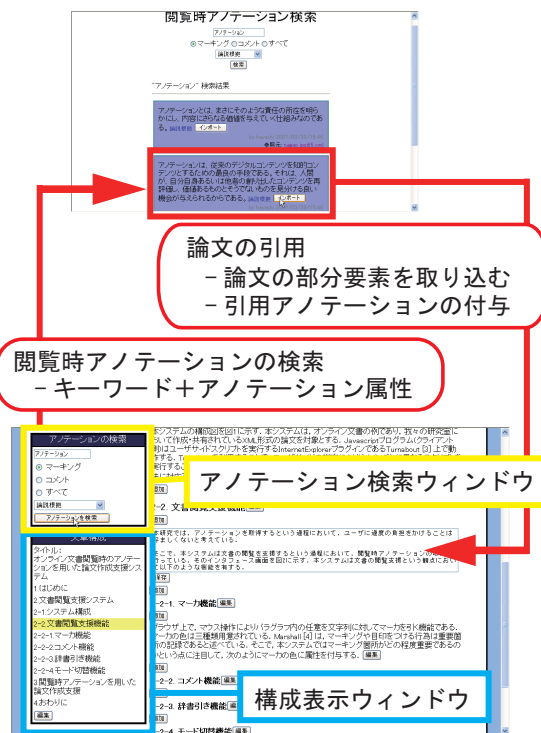


図 3: 論文作成支援インタフェース

### 3.2 論文作成支援

論文作成時におけるインタフェース画面例を図 3 に示す。本システムにおいて作成される論文は XML 形式の構造化文書である。

本システムの論文作成支援機能を、基本的な文章作成プロセスのモデルである Hunter[5] のモデルに基づいて述べる。Hunter のモデルでは、文章作成のプロセスを、情報を集め、書く内容やアイデアを思いつく「創作」(generating) プロセス、必要な内容を選択し、それを線形、または階層的にまとめあげる「組織化」(organizing) プロセス、計画に従って実際に文章を書く「作文」(composing) プロセス、必要に応じて挿入、削除、順番の変更、置き換え、計画の変更を行う「推敲」(revising) プロセス、という互いに関連し合った 4 つの基本プロセスで捉えている。以下に各プロセスにおける論文作成支援手法について述べる。

1. 創作 (generating)  
 論文における創作プロセスとは、日々の研究活動である。研究活動には、論文のサーベイやゼミ発表などが考えられる。そのため、論文閲覧支援は創作プロセスの支援に含まれる。さらに、我々の研究室において利用されている会議情報を収集する DiscussionMining システム [6] や知識活動を支援する DRIP システム [7] を用いて、アイデア等の情報を収集する。
2. 組織化 (organizing)  
 組織化プロセスでは、創作プロセスにおいて収集された断片的な情報を組織化していく。一般的に、組織化プロセスを支援する手法としては、木構造表現や二次元空間を用いる手法が考えられる。木構造表現はトップダウン、二次元空間はボトムアップの文章作成に適している。本システムの対象は論文であるため、初期段階においてあ

る程度構成を決定できると考えられる。つまり、トップダウンのプロセスであるため、木構造表現を用いた手法を適用する。さらに、DiscussionMining システムや DRIP システムと連携することで、書く内容やアイデアの容易な検索・挿入を行い、各構成要素に盛り込まれるべき内容を決定していくことが可能である。

### 3. 作文 (composing)

作文プロセスでは、組織化プロセスにおいて決定された内容の文書化を行う。閲覧時アノテーションを用いた論文の引用支援は作文プロセスにおける支援にあたる。さらに、専門用語辞書を用いて用語説明の引用を支援する。一般に論文に専門用語を記述する場合、用語の説明を記述する。閲覧時に辞書引き機能で利用・拡張した研究室内の専門用語辞書を用いることで容易な用語説明の引用が可能である。辞書を利用することでユーザが自分の言葉で記述するより、説明に客観性と正確性を持たせることができると考えられる。

### 4. 推敲 (revising)

論文閲覧支援におけるアノテーション機能を用いることで、筆者単独で考え、修正するのではなく、共著者等にコメントをもらうことが可能になる。そのため、多様な視点から草稿を見直すことが可能となり、論文のクオリティが向上すると考えられる。

## 4. まとめと今後の課題

本論文では、研究活動における論文の閲覧・作成プロセスの重要性と関連性について述べ、連携させることで可能になる研究活動支援手法を提案した。連携させる具体的な手法として、閲覧時アノテーションと引用アノテーションを利用することを述べた。さらに、各アノテーションを用いた支援手法の有効性について述べ、そのための研究活動支援システムの構築を行った。そして、研究活動支援システムには、論文閲覧支援と論文作成支援に関する機能があり、各機能の詳細について述べた。今後の課題として以下の点が挙げられる。

### 4.1 システムの評価

今後の課題としては、まず、本論文で提案した研究活動支援システムの評価が挙げられる。評価方法に関して、まず長期間ユーザにシステムを利用してもらう必要があると考えられる。その後、アンケート形式によるシステムのユーザビリティの評価や、システムを利用した人と利用しなかった人との比較評価などを行う必要がある。

### 4.2 未引用の関連論文の推薦

未引用の関連論文の推薦手法の検討が考えられる。本論文では、引用する論文は過去に閲覧したことがある論文に限定して引用支援した。しかし、論文を作成している際に、関連研究をサーベイするような状況も考えられる。そこで、引用すべき論文を推薦する仕組みを検討する必要があると考えられる。

## 参考文献

- [1] 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, Vol.44, No.3, pp.1000-1012, 2003.
- [2] Nagao K., Digital Content Annotation and Transcoding, Artech House Publishers, 2003.

- [3] Schilit B. N., Golovchinsky G., Price M., Beyond Paper: Supporting Active Reading with Freeform Digital Ink Annotations, Proceedings of CHI '98, 1998.
- [4] 難波英嗣, 神門典子, 奥村学, 論文間の参照情報を考慮した関連論文の組織化, Vol.42, No.11, 2640-2649, 情報処理学会論文誌, 2001.
- [5] Hunter W. J., Begoray J., A framework for the activities involved in the writing process, The Writing Notebook, 1990.
- [6] Nagao K., Kaji K., Yamamoto D., Tomobe H.: Discussion Mining: Annotation-Based Knowledge Discovery from Real World Activities, Proceeding of the Fifth Pacific-Rim Conference on Multimedia, 2004.
- [7] 土田貴裕, 友部博教, 大平茂輝, 長尾確, 議事録に基づく知識活動サイクルの活性化, 人工知能学会全国大会 (第 20 回), 2006.