

デジタルコンテンツの部分関連付けと 文脈情報の可視化に基づく論文執筆支援

棚瀬 達央[†] 大平 茂輝^{††} 長尾 確[†]

[†]名古屋大学 大学院情報科学研究科 ^{††}名古屋大学 情報基盤センター

1 はじめに

近年、研究活動において、文書の電子化やクラウド技術などの情報技術の発達により多様なコンテンツの記録・管理が容易になっている。具体的な行為としては、調査のための論文などの文献の管理、研究内容の整理のためのノートの作成、成果報告のための発表資料の作成などが例として挙げられる。一方で、論文を書き始めるまでに、研究者は長期にわたる研究活動を行っており、蓄積されたコンテンツの情報量は非常に膨大になると考えられる。そのため、それらのコンテンツの中から論文に必要な情報のみを探し出し整理することは一般に困難なタスクである。不慣れな学生にとっては論文に必要な情報を頭の中だけで整理することは特に困難であるため、整理できないまま論文作成を始めてしまうことで、後で大幅な修正を加えなければならないといったことがある。効率的な論文作成を行うためには、論文を書き始める前に、著者が、執筆に必要な情報を整理・把握している必要があると考えられる。そこで、本研究では、論文アウトラインに対して、それまでに著者が作成・閲覧したコンテンツを容易に関連付ける仕組みと、関連付けたコンテンツの文脈情報を提示する仕組みを提供する。ここで述べるコンテンツの文脈情報の提示とは、あるコンテンツに対し、コンテンツ作成時に参照・引用された論文やノートなどの、外部の関連するコンテンツを提示することを指す。本研究では、研究活動における文書作成時における参照・引用行為に着目し、文脈情報として論文執筆時に利用する。

2 文書作成に基づく部分引用関係の収集

本研究では、文書作成時に論文・研究ノート・発表資料・画像・映像といったあらゆるコンテンツを検索・引用可能とする仕組みを提供する。我々は、その仕組みを実現するエディタである TDEditor を開発した。図 1 に TDEditor を用いた文書作成支援環境の構成を示す。クラウド上には我々の研究室で開発・運用されてきた研究活動支援アプリケーションを用いて作成されたコンテンツが蓄積されており、TDEditor 上でそれらのコンテンツを検索・引用できる。これらのアプリケーションの中では、コンテンツに付与されるメタ情報であるアノテーションが収集され、コンテンツの検索に利用される。具体的なアノテーション行為としては、論文や研究ノートに対するマーキングやコメント

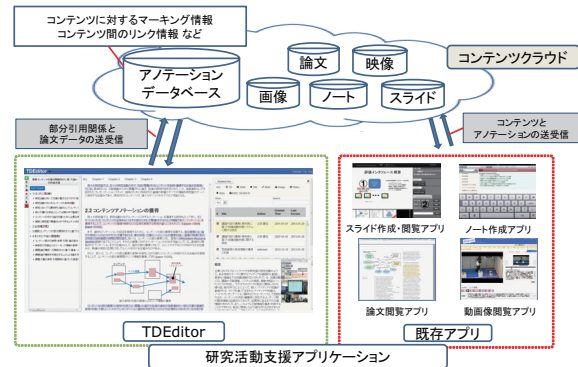


図 1: 文書作成支援環境

の付与、論文や映像などの引用行為が挙げられる。本研究では、特に引用行為に着目し、TDEditor 上で文書の作成における引用関係を収集する。TDEditor 上では、コンテンツ全体だけではなくその部分要素を引用する仕組みである部分引用を実現する。

2.1 コンテンツの部分引用

コンテンツの部分引用とは、既存のコンテンツの部分要素を、作成するコンテンツの一部として掲載する行為である。部分要素とは、論文や研究ノートなどのテキスト文書であれば、章や段落、文章といったレベルの要素を指し、画像や映像であれば特定の矩形範囲や時間区間を指す。我々は、コンテンツの部分要素に対し、固有の URI (Uniform Resource Identifier) を割り当てることにより、それらの要素の引用を実現する。

2.2 文書作成時のコンテンツの検索

図 2 にコンテンツの検索インタフェースを示す。本論文では、論文、ノート、プレゼンテーションスライド、画像、映像の検索を実現している。図 2(左)では、著者が研究活動も中で作成・閲覧したコンテンツのタイトルが、閲覧日時などのメタデータと共にリストで表示されている。研究活動の中でタグなどが付与されていた場合はその情報も表示され、キーワードを入力することでコンテンツを絞り込むこともできる。タイトルをマウスオーバーすると下にコンテンツの概要の文書や、サムネイル画像が表示され、タイトルをクリックするコンテンツの内容全体を表示する詳細ビューモードに切り替わる。図 2(右) は論文の場合である。その際に、エディタ上で図 1 に示した研究活動支援アプリケーションの中であらかじめマーキングされた部分をハイライトすることで、著者が効率的に目的の情報に辿り着けるように設計されている。

2.3 部分引用関係の収集

検索したコンテンツの部分要素を必要に応じて作成中の文章内に引用することができる。部分引用は、

Technical Document Creation Support Based on Deep Association and Visualization of Context Information among Digital Contents

[†] TANASE, Tatsuo (tanase@nagao.nuie.nagoya-u.ac.jp)

^{††} OHIRA, Shigeki (ohira@nagao.nuie.nagoya-u.ac.jp)

[†] NAGAO, Katashi (nagao@nuie.nagoya-u.ac.jp)

Graduate School of Information Science, Nagoya University

(†) Dept. of Information Engineering, School of Engineering, Nagoya University (††)

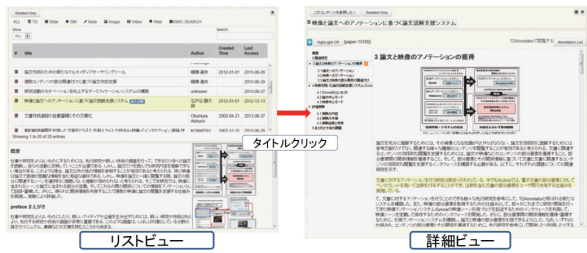


図 2: 検索インターフェース

テキストの場合は、コピーアンドペースト、画像や映像の場合は、ドラッグアンドドロップにより行うことができる。本エディタは HTML エディタとして動作し、コンテンツの引用元の部分要素へのポインタ情報を HTML のタグの属性情報に埋め込むことで部分引用関係をシステムの中で保持している。作成・公開された文書は、さらに他の文書の作成時に引用可能となる。本研究では、文書作成中の引用行為の中で引用関係を収集したが、我々の研究室ではミーティングにおける電子ボード、議事録、TODO ノートなどの作成時にこのようなコンテンツの引用関係を収集してきた [1][2]。これらの引用関係と文書作成時の引用関係を統合することで、引用関係に基づいたコンテンツの俯瞰的な検索が実現できる。

3 論文執筆支援

本章では、2章で説明したコンテンツの部分引用関係を論文執筆に利用する仕組みについて述べる。提案する仕組みは TDEditor の中で実現する。本研究で提案する論文執筆の手順を以下に示す。

1. 論文アウトライン (章・節・段落構成) の作成
2. アウトラインへのコンテンツの部分関連付け
3. アウトラインごとの内容の文章の執筆

論文アウトラインの作成は一般的に行われている行為であるが、章・節・項の構造に加えて段落構成まで記述する仕組みを TDEditor の上に提供する。段落構成まで記述することは、論文に書くべき情報を整理するために有効であると酒井の著書 [3] で述べられている。手順 2 のアウトラインへの部分関連付けは、アウトラインごとにどのような情報が必要であるか整理する行為をシステムで支援するものである。

3.1 論文アウトラインへのコンテンツの部分関連付け

アウトラインへのコンテンツ部分の関連付けは、TDEditor で検索したコンテンツの部分要素を選択してアウトラインに対してドラッグすることで行われる。関連付けが行われると、図 3 左に示すようにタイトルの右にアイコンが表示される。アイコンの横の数字は、アウトラインに関連付けられたコンテンツの数を表している。

3.2 文脈情報の提示

関連付けられたコンテンツは、アウトライン項目のタイトルをクリックすると検索インターフェースのコン

テンツリストの先頭に表示されるようになる。また図 3 右のようなグラフ表示により、直感的に各アウトライン項目に対してどのコンテンツの部分要素が関連付けられているか把握することが可能である。インターフェースとしては、グラフのノードをクリックするとその内容が図 3 右下の領域に表示されるようになっている。図 3 の場合は、アウトラインノードに対して論文の部分要素ノードとスライドの部分要素ノードを表すノードがエッジで結ばれている。部分要素ノードは、それぞれのコンテンツ全体のノードと結ばれている。さらに、論文 (全体) ノードからエッジで結ばれている論文部分ノードが存在している。これは 2 章で述べたコンテンツの部分引用関係に基づいて提示されている。この場合は、関連付けられた論文の引用元のノートと論文部分が表示されている。現在はコンテンツ全体の単位でアウトラインノードからの距離が 2 までのノードを表示しているが、部分引用が継続的に行われている場合は、さらに深く過去のコンテンツまで遡ることも可能となる。このように部分引用関係をコンテンツの文脈情報として提示することで、過去のコンテンツから論文に必要な情報を発見できる可能性がある。

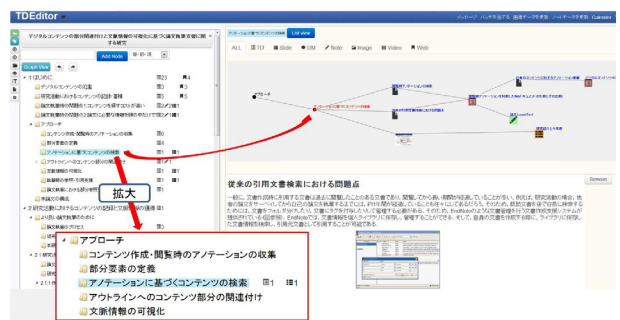


図 3: 関連付けられたコンテンツの可視化

4 おわりに

本研究では、文書作成時にコンテンツを検索する仕組み、部分引用に基づくコンテンツの文脈情報を収集する仕組み、コンテンツのアウトラインへの関連付けと文脈情報の提示に基づく論文執筆支援の仕組みについて述べた。今後の予定としては、文脈情報の提示に関する定量的な評価、大量の部分引用関係に基づいた高度なコンテンツの検索や推薦、部分引用関係の研究活動支援アプリケーション間での相互運用の実現を行う予定である。

参考文献

- [1] 石戸谷 顕太郎, 大平 茂輝, 長尾 確, 継続的ミーティング支援システム, 情報処理学会論文誌, Vol.53, No.8, pp.2044-2048, 2012
- [2] 土田 貴裕, 大平 茂輝, 長尾 確, ゼミコンテンツの再利用に基づく研究活動支援, 情報処理学会論文誌, Vol.51, No.6, pp.1357-1370, 2010
- [3] 酒井 聡樹, これから論文を書く若者のために, 共立出版, 2002