

複数コンテンツの部分関連付けに基づく論文作成支援

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。具体的には、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。研究者は、最終的に論文として、それらの作成したコンテンツに含まれる内容を整理して、一つのまとまった文書にする必要がある。

しかし、論文を書き始めるまでに、研究者は長期的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。不慣れた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初めてしまうことで、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。効率的な論文執筆を行うためには、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、章構成などの大まかな論文構成に加え、詳細な論文構成を作成する仕組みを提供し、さらに、それらの詳細な論文構成に対して、それまでに著者が作成・閲覧した複数のコンテンツを容易に関連付ける仕組みを提案する。この仕組みを実現するために、我々の研究室の既存のシステムを活用し、それらの多様なコンテンツの容易な参照・引用を可能とするエディタを開発した。

2 論文作成支援環境の構成

本研究で提案する論文作成支援の仕組みの全体の構成を図1に示す。

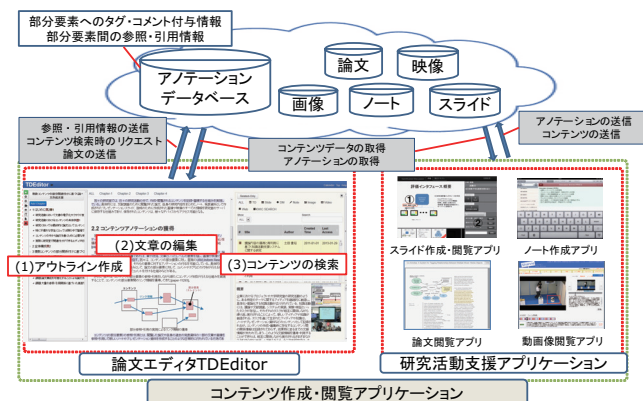


図1: 論文作成支援環境の構成

2.1 コンテンツアノテーションの獲得

コンテンツの容易な検索と参照・引用関係の収集を実現するために、本研究では、研究活動の中で、コンテンツのアノテーションを獲得する。アノテーションとは、コンテンツの全体あるいはその部分に対して関連付けられるメタ情報であり、アノテーションを獲得することで、コンテンツの管理・検索などの応用を実現する研究が盛んに行われている [1][2]。

適切なアノテーションの記述を実現するために、コンテンツの部分要素を定義する。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、それらの要素に対するアノテーションの付与を可能にしている。

具体的に関与されたアノテーションを行う仕組みとして、論文の部分要素に対して、コメントやタグなどの付与が行える仕組みや、映像の特定の区間に対してコメントを付ける仕組みなどがある。

さらに、我々は、コンテンツの部分要素を参照・引用しながら新たにコンテンツ作成が行える仕組みを実現することでコンテンツの部分要素間のリンク情報をアノテーションとして獲得してきた [3]。具体的には、専用アプリケーションを通して、発表資料やノートなどコンテンツが作成される際に、作成過程において自然に行われる論文の文章や画像といった他のコンテンツの部分要素の参照・引用行為を記録する。

本研究では、このような研究活動におけるコンテンツの部分要素に対するコメントの付与や部分参照・引用行為をアノテーションとして、データベースに記録し、論文作成において活用する。

2.2 論文エディタ (TDEditor)

研究活動の中で記録されたアノテーションを活用したコンテンツ部分の検索・参照・引用が行える論文エディタ: TDEditorを開発した。TDEditorは、図1のように主に、アウトライン作成機能、文章編集機能、コンテンツ部分の検索機能の3つで構成されている。アウトライン作成機能

アウトライン作成機能は一般的なテキストエディタにも標準的な機能として備えられているものと同様に、文書の構成を木構造で表現するものである。3章で詳細に述べるが、TDEditorでは、アウトラインに対してコンテンツを関連付けることが可能である。

文章編集機能

TDEditorでは通常の文章編集機能に加え、エディタの中でコンテンツを検索・参照しながら文書を執筆することができ、さらに、必要に応じて、作成中の文章内にコンテンツ部分を引用することができる。引用は、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。また、TDEditor上で行われたコンテンツの部分要素の参照・引用行為も前節で述べたアプリケーションと同様に、データベースに記録される。コンテンツ検索機能

図2にコンテンツ検索機能の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、映像の部分要素の検索を実現している。図2(左)では、著者が研究活動の中で作成・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

Technical Documents Creation Support Based on Deep Association among Multiple 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

([†])

Information Technology Center, Nagoya University (^{††})

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、アノテーションが付与されているコンテンツ部分がハイライトされるようになっており、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。

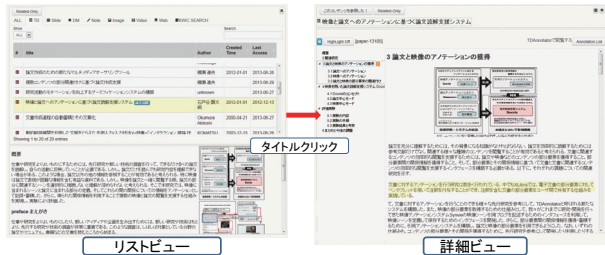


図2: 検索インタフェース

次に、これらの機能を組み合わせることによって行う、新しい論文作成の手法について説明する。

3 複数コンテンツの部分関連付けに基づく論文作成支援

3.1 論文作成手順

本研究で提案する論文作成の手順を以下に示す。

1. 大まかな論文構成(章・節など)を決める
2. コンテンツ部分を手順1で決定した論文構成ごとに関連付ける
3. 章ごとに詳細な構成を決める
4. 手順3で決定した論文構成に対してコンテンツ部分に関連付ける
5. 論文構成の各部分の文書を執筆する

手順1のように論文構成を決めることは一般的に行われており、手順2についても、論文の著者は、作成した章構成を決める際に、構成ごとにどのような情報が必要であるか頭の中で整理していると考えられるため自然な行為と思われる。

しかし、ここで決めることはあくまで大まかな論文構成であるため、書く内容を大まかに整理することはできても、実際に文章を書き始めてみると、思うようにならないといったことが(特に経験の少ない学生などには)生じやすい。

そこで、手順3として、論文構成ごとに詳細な構成を決める。詳細な論文構成の例を図3(左)に示す。図のように詳細な意味単位で構成を文書化して可視化することで、書くべき内容を細分化して整理することができると思われる。

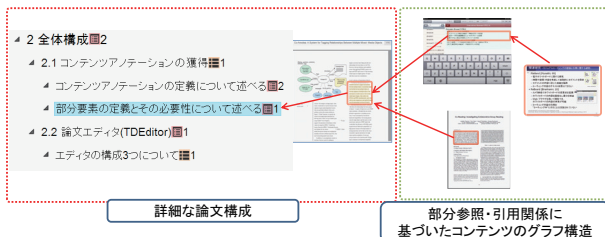


図3: 詳細な論文構成(左)と参照・引用関係に基づいたグラフ構造(右)

そして、手順4で、コンテンツの部分要素を、より細分化された論文構成に関係付けて整理・可視化することで、その構成を書くために必要な情報が十分に存在するかどうかを手順5の論文の本文を書き始める前に著者は把握することができると考えられる。

3.2 論文構成に対する複数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、コンテンツの部分要素を選択してコンテンツの部分要素をアウトラインに対してドラッグするか、アウトラインの内容を記述する編集エリアにコンテンツをコピーアンドペーストすることで行われる。関係付けが行われるとコンテンツに対して図3(左)のようにアウトラインの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全体の数を表している。

このように詳細にアウトラインに対してコンテンツ部分に関連付けておくことで、検索インタフェースで、容易にアクセス可能となり、論文構成を書くために必要な情報の有無を確認することができる。また、関係付いた文書や図表の量を把握することで、論文構成ごとの情報量のバランスなども、文章を書く前に気付くことができると考えられる。

しかし、論文を作成し始めるまでのコンテンツの量は膨大であるため、単純なキーワードなどによる検索だけでは、論文作成に必要な情報を含むコンテンツをもなく探し出すことは難しい。そこで、TDEditorでは、手順1または手順3で少なくとも1つコンテンツがアウトラインに関連付けられていれば、2.2節で説明した参照・引用関係に基づいた新たなコンテンツを著者に提示・推薦する。図3にその概念図を示す。これにより提示されるコンテンツ部分は、アウトラインに直接関連付けられたコンテンツの作成過程で参照・引用されたコンテンツであるので、著者が論文を作成する上で必要な情報を含んでいる可能性が高い。今後は、図のようなグラフ表示などの、より直感的に関連付けられたコンテンツの量が把握できる提示方法を実装する予定である。

4 まとめと今後の課題

論文アウトラインに対してコンテンツ部分の関連付けを行うことで、論文の文章の作成を行う前に著者に論文に必要な情報を整理・把握させる仕組みを実現した。今後の課題としては、論文構成へのコンテンツの部分関係付けに基づく著者への影響の評価、継続的な運用による大量のコンテンツの部分の参照・引用情報を収集及びそれらの統計的分析に基づいたより高度な検索の実現、アウトラインに関連付けられたコンテンツを検索・可視化することによる論文作成の効率や論文の質への影響について定量的な評価を行う予定である。

参考文献

- [1] 石戸谷 顕太郎, 山本 圭介, 大平 茂輝, 長尾 確: 映像と論文へのアノテーションに基づく論文読解支援システム, 映像情報メディア学会誌, Vol.66, No.11, pp.J461-J470, 2012.
- [2] Jane Hunter and Ronald Schroeter: Co-annotate: A system for tagging relationships between multiple mixed-media objects, IEEE MultiMedia, Vol.15, No.3, pp. 42-53, 2008.
- [3] 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.