

## 論文

## 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

## Support System based on Annotations of Documents and Video Scenes for Reading Technical Documents

石戸谷 顕太郎<sup>†</sup>, 山本 圭介<sup>††</sup>, 正会員 大平 茂輝<sup>†††</sup>, 長尾 確<sup>†</sup>Kentaro Ishitoya<sup>†</sup>, Keisuke Yamamoto<sup>††</sup>, Shigeki Ohira<sup>†††</sup> and Katashi Nagao<sup>†</sup>

**Abstract** Reading as many technical documents as possible is important to improve our research. When we read these documents, we can understand their contents more easily by referring to related resources such as images, audio clips, and videos. Videos contain a variety of helpful information, and facilitate our understanding of technical documents. We propose a method to define video scenes and document elements, and to annotate them with additional information such as relationships. Based on these annotations and relationships, we developed a support system that uses videos and helps readers understand technical documents. We performed some experiments to confirm whether the system was usable.

**キーワード**: アノテーション, 論文読解支援, 部分引用, 映像シーン

## 1. ま え が き

仕事や研究をよりよいものにしたたり、新しいアイデアや企画を生み出すためには、新しい研究や技術はもとより、先行する研究や技術の調査が非常に重要である。このような調査は、しばしば対象としている分野の論文やマニュアル、書籍などの文書を読むところから始まる。

ある文書が理解しやすいかどうかを考える場合、文書全体の構成、文章の論理性、文体や図・表が充実しているかなどが問われる。なかでも、著者の想定した読者が、実際の読者と合致しているかが特に重要な要素であろう。例えば、学習を始めて間もない初学者が専門分野の論文を十分に理解するためには、その文書が書かれた背景を認識することや基礎知識の蓄積が不可欠になる。その論文の文章や、挙げられている参考文献だけを読んでも、その内容を十分に理解するのは困難であろう。そして、特に論文などでは初学者が内容を理解するために必要な基礎的な事柄が詳述

されることは少ない。初学者でなくとも、システムの動作や実験風景などについて、文書に含まれるテキストや図だけでは状況を十分に理解することが困難な場合がある。

このような場合、参考文献に挙げられている文書だけでなく、元の文書の内容に関連する様々なコンテンツを参考にすることが理解の補助となる。例えば、システムのデモビデオや、文書に書かれている内容について会議や学会などで発表された時の記録映像、あるいはその発表資料や開発したシステムの詳細な設計図、関連分野の基礎的な内容に関する書籍など、元の文書そのものには含まれていないが、内容の理解の補助となるコンテンツである。

しかし、関連するコンテンツを寄せ集めただけでは、効果的な読解支援は期待できない。論文に対する理解を促進する目的で、1時間を超えるような映像が提示されたとしても、論文の一部を理解するためにそのような映像を閲覧するのは時間的な負荷が高い。そのため、ある文書の部分を理解するために必要な映像の部分を切り出して提示することが必要であると考えられる。

例えば、大学の研究室では、毎年、新たに学生が配属されてくる。一般に、このような学生に研究室の教授や先輩が執筆した論文を読ませて、理解させるのは困難である。その理由のひとつとして、研究室で行われてきた研究の背景知識が乏しく、その論文を読んだだけでは、その内容を十分に理解できないことが挙げられる。このような場合、論文の部分に対して、その部分の理解を補助するような映像が関連付いていれば読解の補助になるであろう。

説明のため、より具体的に、著者が所属する研究室に配属された学部4年生に本論文を読ませることを考える。本

2012年4月4日受付, 2012年7月27日再受付, 2012年10月10日採録

<sup>†</sup>名古屋大学 大学院情報科学研究科

(〒464-8603, 愛知県名古屋市中区不老町, TEL 052-789-5878, e-mail:ishitoya@nagoya-u.jp)

<sup>††</sup>株式会社デンソー IT ソリューションズ

(〒450-8626, 愛知県名古屋市中村区名駅南 1-27-2 11F)

<sup>†††</sup>名古屋大学 情報基盤センター

(愛知県名古屋市中区不老町, TEL 052-789-5878)

<sup>†</sup>Graduate School of Information Science, Nagoya University

(Furo-cho, Nagoya-shi, Aichi, 464-8603 Japan)

<sup>††</sup>DENSO IT SOLUTIONS

(11F 1-27-2 Meieki Minami, Nakamura-ku, Nagoya-shi, Aichi, 450-8626 Japan)

<sup>†††</sup>Information Technology Center, Nagoya University

(Furo-cho, Nagoya-shi, Aichi, 464-8603 Japan)

論文に、後述の Synvie (映像アノテーションシステム) や Docvie (論文読解支援システム) のインタフェースの動作を説明した映像の部分や、関連研究それぞれの論文の一部が適切に関連付いていれば、内容の理解がより容易になると考えられる。学生が、Synvie を拡張して実現した映像シーンに対するアノテーション画面の映像を閲覧し、その映像シーンと関連付けられた、Synvie について記述された論文<sup>1)</sup>の引用インタフェースに論文の部分や、Synvie の先行研究である iVAS\* や Co-Annotea<sup>2)</sup>、Annotea\*\* のインタフェースについて説明した論文の部分や映像の部分の参考にして読み進めていけば、本論文への理解が深まるであろう。

このような状況は、大学の研究室だけでなく一般的に起こりうる。本研究では、文書の読解、中でも論文の効率的な読解を支援するために、論文と映像の部分要素の定義とその関係に基づいて関連コンテンツを閲覧できるシステムについて提案する。

本研究では、まず、論文の部分 (例えば、段落、文、図、表など) に対するアノテーションを獲得する仕組み、映像の部分 (例えば、シーン、ショットなど) に対するアノテーションを獲得する仕組み、そして、それぞれのシステムによって蓄積された部分要素間の関係情報を獲得する仕組みについて、順に述べる。次に、それらのメタ情報に基づいて関連コンテンツを閲覧できるシステムについて説明する。そして、本システムを用いて被験者実験を行った結果、論文を効率的に読解できることを確認したので詳しく考察する。

## 2. 関連研究

論文を十分に理解するためには、その背景になる知識がなければならない。論文を効率的に読解するためには参考文献だけでなく、関連する様々な種類のコンテンツを閲覧することが有効であると考えられる。文書に関連するコンテンツの効率的な閲覧を支援するためには、論文や映像などのコンテンツの部分要素を獲得すること、部分要素間の関係情報を獲得すること、そして、部分要素とその関係情報に基づいて文書と文書に関連するコンテンツの効率的な閲覧を支援するインタフェースを構築する必要がある。以下に、それぞれの課題についての関連研究を示す。

文書に対するアノテーションを行う研究は数多く行われている<sup>3)~5)</sup>。中でも XLibris では、電子文書の部分要素に対してペンタブレットを用いて注釈を付与することができ、注釈を含む文書の部分要素をユーザ間で共有する仕組みを実現している<sup>6)</sup>。

映像の部分要素を獲得する手法に関係して、Sawhney らの映像の内部構造を表現するための仕組みである Hypervideo<sup>7)</sup> を挙げる。Hypervideo では、映像を複数の映像シーンに分割するメタ情報を付与し、テキストなどの他のコンテンツと関連付けて扱えるようにすることで、映像を取り扱いやしくするための仕組みである。また、長尾らは、映像シーンに対するメタ情報を Web 上のコミュニティ活動から獲得する、オンラインアノテーションを提案している<sup>8)</sup>。オンラインアノテーションの実装である山本らの Synvie<sup>1)</sup> では、映像シーンを引用したブログ記事を書けるようにすることで、映像シーンの定義とそれに対するテキストアノテーションを獲得し、映像シーンの検索に応用している。

また、本研究では取り扱わなかったが、映像だけでなく学会発表や講義で利用するスライドやミーティングの音声など、様々なマルチメディアコンテンツの部分要素に対するアノテーションを利用することで、論文読解を支援できると考えられる。例えば、Kam らの Livenotes<sup>9)</sup> や Davis らの Notepals<sup>10)</sup> では、講義中のスライドの一部に対して講義の受講者がアノテーションを付与できる。

これらのマルチメディアコンテンツ間の関係を獲得する手法に関して、Hunter らの Co-Annotea<sup>2)</sup> がある。Co-Annotea はマルチメディアコンテンツの部分要素に対するアノテーションと、部分要素間の関係を付与できるシステムである。付与したアノテーションに基づいて、複数のマルチメディアコンテンツを読み込んでアノテーションを確認しながら閲覧することができる。

コンテンツの部分要素を組合せて新たなコンテンツを作成することに注目した研究について、Chambel らの HTIMEL<sup>11)</sup> をあげる。HTIMEL では、映像シーンとテキストの部分要素、画像を組合せたハイパーメディアコンテンツをユーザが自由に作成できる環境を実現している。HTIMEL は、コンテンツを組合せてハイパーメディアを作成することができるが、そのコンテンツの閲覧者はハイパーメディアの作成者が意図した範囲のコンテンツしか閲覧することができない。

本研究では、コンテンツの部分要素とその関係を利用して、論文の読解を支援するためのインタフェースを実現する。論文読解を支援するためには、他のユーザが付与したアノテーションを手がかりに読解できるようにすることが有効であろう。テキストによるアノテーションだけでなく、映像や他のマルチメディアコンテンツの部分要素が互に関連付けていれば、閲覧している部分要素と関連付いているコンテンツを辿って閲覧することができる。先述した Co-Annotea においても、部分要素に関連付けられた他のコンテンツの部分要素を、単一のインタフェース上で閲覧できる仕組みを提供している。Co-Annotea のユーザはコンテンツの部分要素間が関連付いている限り、アノテーションを辿って行くことで、多種多様なコンテンツを閲覧することができるが、論文読解などの特定の用途に特化してはお

\* 本文の参考文献ではないので脚注に論文への参照を入れる。D Yamamoto and K Nagao. IVAS: Web-based Video Annotation System and its Applications. In 3rd International Semantic Web Conference (ISWC2004), pp. 7-11, 2004.

\*\* 同上, J. Kahan, M.-R. Koivunen, E. Prud'Hommeaux, and R.R. Swick. Annotea: An Open RDF Infrastructure for Shared Web Annotations. Computer Networks, Vol. 39, No. 5, pp. 589-608, 2002.

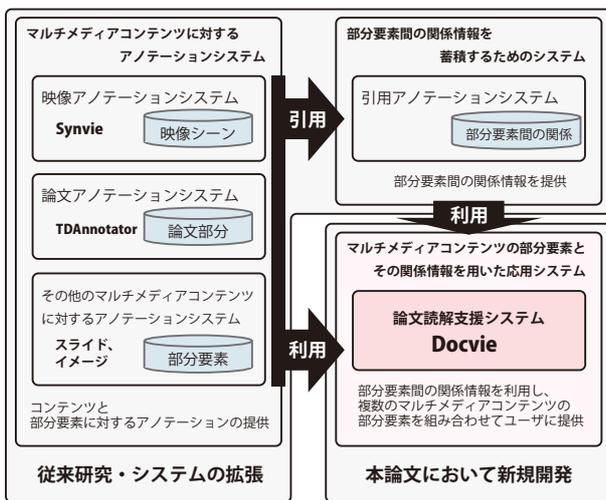


図1 コンテンツの部分要素間の関係を利用した論文読解支援の仕組み

Overview of reading-support system based on relationships among elements of multimedia contents.

らず、またインタフェースに対しての定量的な検証も行われていない。

そこで本研究では、まず我々の研究室で研究開発が行われてきた、コンテンツにアノテーションを行うためのシステムを適宜拡張することで、ユーザが日常的な研究活動で閲覧する論文や映像の部分要素に対するアノテーションと、部分要素間の関係を記録する。そして、獲得したアノテーションに基づいて、ユーザが注目しているコンテンツの種類（論文または映像）に応じた、関連するコンテンツの部分要素を効率的に閲覧できるインタフェースを実現する。さらに、実現したインタフェースの論文読解に対する有効性を検証した。

なお、ユーザのアノテーション行為そのものが学習や読解の支援につながるとする Active Reading という手法がある<sup>12)</sup>が、本研究では、次章で説明するアノテーションの仕組みをコンテンツの部分要素に対するアノテーションを獲得するためにのみ用い、アノテーションを付与する過程における学習・読解支援については別の機会に扱うことにする。

### 3. 論文と映像のアノテーションの獲得

図1に示すように、論文に関連するコンテンツの効率的な閲覧を実現するためには、論文と映像の部分要素、および部分要素間の関係情報を獲得する必要がある。本研究では、論文の部分要素を取得するためのシステムとして、文書に対するアノテーションを行うことのできる様々な先行研究<sup>3)~6)</sup>を参考にして、TDAannotatorと呼ばれる新たなシステムを構築した。また、映像の部分要素を取得するための仕組みとして、我々がこれまでに研究・開発を行ってきた映像アノテーションシステム Synvie の映像シーン引用ブログを記述するためのインタフェースを利用して、映像シーンを定義して保存するためのインタフェースを開発

した。さらに、部分要素間の関係情報を獲得・蓄積するために、引用アノテーションシステムを構築し、論文と映像の部分要素を引用できるようにした。なお、いずれの仕組みも、コンテンツの部分要素とその関係を獲得するために、先行研究を参考にして開発したり利用したりすることで実現しており、特に新規性はない。また、将来的には、関連付けるコンテンツの部分要素は映像に限らずスライドやイメージなども利用できるようにする。

先述したように、ユーザはこれらのアノテーションの仕組みを用いて、同じプロジェクトで研究している他のユーザと協調して、部分要素に対するアノテーションを行う。そして、日常的な研究活動の中で、矩形アノテーションを行える仕組みによって映像や論文へのアノテーションを蓄積していく。それぞれのアノテーションシステムは、別々のサーバで稼働しているが、ユーザ情報の管理を集中的に行うユーザ管理サーバを用いて、シングルサインオンを実現している。これによって、ユーザは認証を一度行うだけで全システムを利用することができ、それぞれのシステムが記録するアノテーションと関連付けされるユーザ情報は全システム共通である。また、それぞれのシステムは、他のシステムと連携するために、自身が管理しているコンテンツの情報を提供するAPIを実装している。

次に、映像と論文のそれぞれについて部分要素に対するアノテーションを作成する仕組み、両者の関係を獲得する仕組みについて詳述する。

#### 3.1 論文へのアノテーション

論文に対するアノテーションを作成・閲覧するために、TDAannotatorを開発した。TDAannotatorの論文読解インタフェースを図2に示す。この画面では論文の各ページが画像に変換され、アノテーションが付与された論文の部分が、状態ごとに色分けされた矩形で表示される。この矩形は、ユーザが論文の一部分に対してアノテーションを付与するために定義・作成する論文の部分要素である。以後この矩形領域を論文部分と呼ぶ。赤い矩形で表される論文部分は（図2中の論文中段右側の細線の矩形）は、アノテーションが付与されていることを示す。また、青い矩形で表される論文部分（同、左上の太線の矩形）は、3.3節で述べる関連付けの仕組みを用いて、映像シーンが論文部分に関連付けられていることを示しており、これは5章で述べる実験で利用する。

TDAannotatorでアノテーションを行うためには、まず、画面に表示される論文上でマウスドラッグ操作を行って矩形範囲を選択（続けて複数の矩形範囲を選択することで複数の矩形範囲に対するアノテーションも可能）し、選択した矩形範囲をクリックすると、その矩形に対するアノテーションを入力するダイアログが表示される。このダイアログで矩形内容（本文、数式、グラフ、表、図、タイトル）、文章タイプ（翻訳、書き起し、コメント）、アノテーションの本文を入力する。

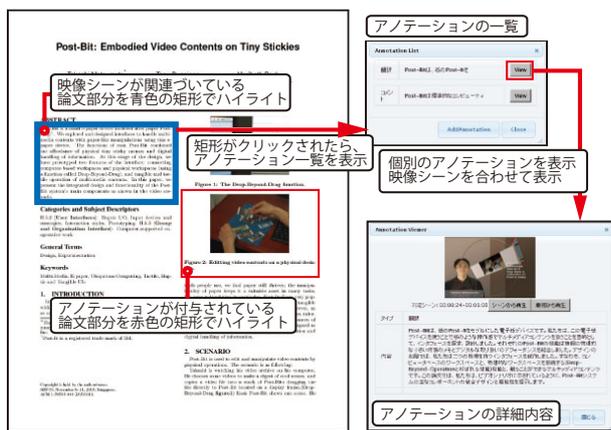


図 2 TDAnnotator の画面  
Interface of TDAnnotator.

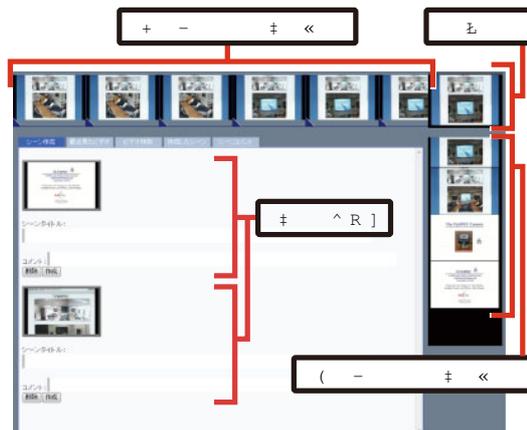


図 3 映像シーンを引用するインタフェース  
User interface for video scene quotation.

なお、すでに定義された論文部分の内側の部分に対してアノテーションを行う場合にも新たに、アノテーションを付与する場合と同様に、インタフェース上でドラッグ操作を行うことで新たな論文部分を定義できる。このような他の論文部分に含まれた論文部分を選択する場合には、1回目のクリックが論文部分の選択となり、選択された論文部分がハイライトされ、2回目のクリックでダイアログが表示される。

アノテーションの登録を行うと、データベースには上述の矩形範囲（論文のページ番号、位置、サイズ）の集合、矩形内容、文章タイプ、アノテーションの本文、論分部分を含む論文の ID、アノテーション作成者の全システム共通の ID が記録される。

また、他のユーザが定義した論文部分やユーザ自身が以前に定義した論文部分に対してアノテーションを付与することもできる。論文上に配置された論文部分をクリックするとアノテーション一覧が表示され、「Add Annotation」ボタンを押すことで既存の論文部分に対してアノテーションを付与することができる。この場合、データベースにはすでに定義された矩形範囲の ID とアノテーションが関連付けられて保存される。

付与されたアノテーションは付与したユーザのみが編集・削除を行える。ただし、矩形範囲については、指し先のないアノテーションを防ぐために、他者がアノテーションが行っていたり、後述の仕組みで関連付けが行われている場合は、削除することができない。

論文上に表示される矩形を選択することで、付与されているアノテーションの内容を閲覧できる。また、論文に対するアノテーションはユーザ間で共有され、あるユーザが付与したアノテーションが後続のユーザの読解時の手がかりとなると考えられる。

### 3.2 映像へのアノテーション

図 3 に映像シーン引用を行うためのユーザインタフェースを示す。引用を行うためには、まず引用したいシーンを含む映像コンテンツを、ユーザの閲覧履歴やキーワード検

索を用いて読み込む。画面は、右上に配置された映像を再生するためのプレイヤー、そして、プレイヤーの水平方向・垂直方向に一定時間ごとのサムネイル画像を並べたシークバー（以下、サムネイルシークバーと呼ぶ）によって構成される。映像の再生を開始すると映像に同期して左右、上下にそれぞれサムネイル画像が移動する。また、サムネイル画像をドラッグすることで、プレイヤーをシークできる。水平サムネイルシークバーは、1秒間隔でサムネイル画像が表示され、引用する映像シーンを決定するために利用する。垂直サムネイルシークバーは映像を素早くシークするために利用する。

シークして引用したいシーンを発見したら、水平サムネイルシークバー上で開始フレームと終了フレームを選択することで引用するシーン区間を設定できる。引用するシーン区間を決定すると、映像シーンがシーン作成欄に追加されタイトル入力欄が表示され、タイトルを設定するとシーンの作成が完了する。これにより、映像シーンの開始・終了時間、タイトルを獲得できる。

すでに定義されたシーンはサムネイルシークバー下部に表示され、自分で定義したシーンと同じように、他のユーザが定義したシーンに対して新たなアノテーションを付与することができる。

定義したシーンは、TDAnnotator と同様に指し先のないアノテーションを防ぐため、他人がアノテーションを付与していたり、後述の仕組みで関連付けされている場合には削除できない。

データベースには、シーンの開始時間、終了時間、タイトル、シーンを含む映像の ID、全システム共通のアノテーション作成者の ID が映像シーンの情報として記録される。

### 3.3 論文と映像の部分要素の関連付け

本研究では、論文・映像の部分要素間の関係情報を獲得するための仕組みとして、引用アノテーションシステム<sup>13)</sup>を利用した。このシステムは、ユーザがスライドや、そのスライドを利用して発表した会議において他者から得られた発言（映像とテキストによって表される）を引用しながらメ

モを書き、メモの内容を振り返ったり俯瞰したりしながら自身の活動をよりよくしていくためシステムである。ユーザは、クライアントアプリケーションを自身のPCにインストールして利用する。本研究では、引用アノテーションシステムで引用できるコンテンツとして新たに映像と論文を追加し、引用する映像と論文の部分要素に関するメタ情報を蓄積している TDAannotator (論文), Synvie (映像) などの各システムから、コンテンツの部分要素を読み込み、必要な部分を引用したメモを作成できる。ユーザが同じメモにコンテンツの部分要素を引用 (これを共引用と呼ぶ) したという情報を用いることで、部分要素間の関係を獲得する。

部分要素を管理する各システムには、引用アノテーションシステムにコンテンツの部分要素を読み込むために、部分要素一覧画面が用意されている。引用アノテーションシステムから「映像シーンの検索」「論文部分の検索」メニューを選択すると、ブラウザが起動して各システムの一覧画面に遷移する。一覧は、コンテンツのタイトル、部分要素に付与されたアノテーション本文などのテキストを対象とした全文検索によって絞り込むことができる。そして、部分要素ごとに「引用」ボタンが設置されており、このボタンを押すと、クライアントアプリケーションが立ち上がり、部分要素が読み込まれノードとして表示される。

図4に引用アノテーションシステムで表現されるノードの種類を示す。あるコンテンツの部分は、コンテンツを表すノードとそれに接続された部分要素を表すノードによって表現される。これは、一つのコンテンツから複数の部分を引用した場合に、それらが元は一つのコンテンツから引用されたことを明示するためである。映像シーンは、その映像シーンを含む映像ノードに接続された部分要素を表すノードとして表現される。映像ノードをダブルクリックすると Synvie で再生され、映像シーンノードの場合は、映像中の特定のシーンのみが再生される。論文部分も映像シーンと同様に、論文ノードに接続された部分要素を表すノードとして表現され、論文ノードをダブルクリックすると該当論文が TDAannotator で表示され、論文部分ノードの場合には TDAannotator 上で論文部分がハイライトされる。

次に、図5に、読み込んだコンテンツの部分要素を引用してメモを作成する手順を示す。ユーザは、引用アノテーションシステム上で引用したい映像シーンおよび論文部分ノードを選択して (図5①)、コンテキストメニューを開き「メモ」を選択する (図5②)。すると、メモのタイトルを入力するフォームが表示され (図5③)、ユーザがタイトルを入力し、決定することでメモノードが作成される (図5④)。作成されたメモノードをダブルクリックすると、内容を編集でき、必要に応じて映像シーンや論文部分ノードの追加・削除を行うことができる。

データベースには、メモのテキスト、関連付けられた部分要素の URI の集合、それぞれの部分要素を包含するコ

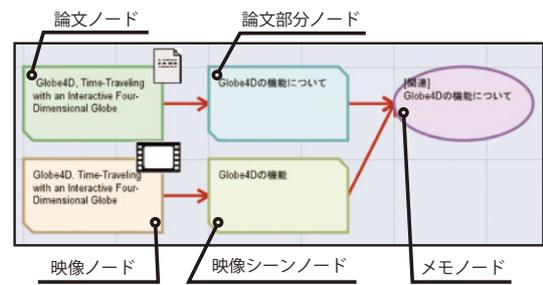
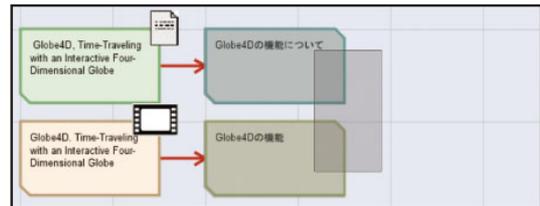
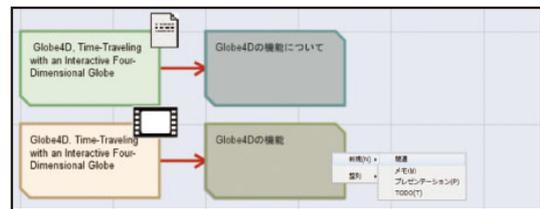


図4 引用アノテーションシステムにおけるノードの種類  
Node types of content-quotation system.

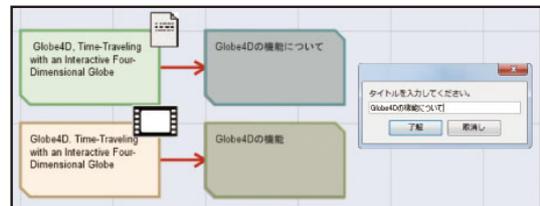
① 引用するノードの選択



② コンテキストメニューで「メモ」を選択



③ メモノードのタイトルを入力



④ メモノードが作成され、引用関係が記録される

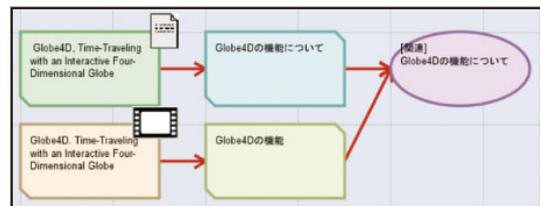


図5 引用アノテーションシステムにおける関連付け手順  
Adding relationships with content-quotation system.

ンテンツの URI が全システム共通のアノテーション作成者の ID と関連付けられて保存される。引用アノテーションシステムは個人の研究活動を支援するためのシステムであるため、作成されるメモは個人的な内容を含む可能性があり、メモのテキスト内容は共有するのに適さない。そのため、メモに引用されたノード間の関係情報のみを他のユーザと共有する。

4. 映像を用いた論文読解支援システム Docvie

これまでに説明した仕組みを利用して取得した、映像の



右側に論文に関連するコンテンツの一覧が表示される。この画面では、主として閲覧する論文を、関連コンテンツを参照しながら読解することで理解を深めることを支援する。

論文上には、TDAnnotatorでアノテーションが付与された論文の部分要素を矩形で表現している。TDAnnotator上での表示と同じく、アノテーションが付与された論文部分は赤い枠線の矩形（論文読解インタフェース上の細線の矩形）で、映像シーンが関連付けられている論文部分は青い枠線の矩形（同、太線の矩形）で表示される。これらの部分は、ユーザによって情報が付与されている部分であり、付与されている情報を読解の手がかりとして参照することで、論文に対する理解を深めることができると考えられる。論文上の矩形を選択すると、関連コンテンツ一覧の中にある、選択した論文部分に関連するコンテンツにフォーカスし展開表示される。

関連コンテンツ一覧には、閲覧している論文に関連するコンテンツが、コンテンツの種類を把握しやすいように種類ごとに色分けされて表示される。映像が青色（関連コンテンツ一覧の上側）で、論文は緑色（同、下側）である。

関連コンテンツ一覧に表示される関連する論文の情報は、初期状態では、それぞれの論文のタイトルが表示される。タイトルを選択すると、その論文の部分要素に付与されたアノテーションの書き出しが一覧表示される。さらに部分要素を選択すると、その部分要素の内容とそこに付与されたアノテーションが表示される。またそれぞれに「詳しく閲覧」ボタンが設置されており、このボタンを押すと部分要素を含む論文が新たなタブとして追加される。

また、関連する映像の情報は、初期状態では、それぞれの映像のタイトルが表示される。タイトルを選択すると、その映像に含まれるシーンのタイトルとサムネイルが一覧表示される。一覧の中から映像シーンを選択すると、その映像シーンのタイトルとシーンを再生するためのプレイヤーが表示される。プレイヤー下部の「シーン再生」ボタンを押すことで映像をシーンから再生することができる。論文の部分要素と同じく「詳しく閲覧」ボタンが設置されており、このボタンを押すことで、そのシーンを含む映像全体を閲覧する画面が新たなタブとして追加される。

このように、関連コンテンツ一覧を段階的に展開できるようにすることで、論文に多くの情報が関連付けられた時に、関連コンテンツを俯瞰し、どの情報を閲覧するかが選択しやすくなると考えられる。論文閲覧時に論文に関連付けられた情報を効率よく閲覧できるようにすることで、関連する情報を探す負担を軽減し、論文読解の効率を向上させることができる。

#### 4.3 映像中心モード

映像中心モードの画面を図9に示す。画面左側に映像閲覧インタフェースが表示されるほかは、論文中心モードと同様である。画面右側の関連コンテンツ一覧には、映像に関連した情報が一覧表示される。この画面では、主として



図9 Docvieの映像中心モードの画面例  
Video centered mode of Docvie.

閲覧する映像を、関連コンテンツを参照しながら視聴することで理解を深めることを支援している。

映像閲覧インタフェースは、映像のプレイヤー、その直下のシーンシークバー、最下部の映像に含まれるシーン一覧で構成される。シーンシークバー上には映像の再生時間が表示される。また、映像のどの部分がシーンとして定義されているかをシーンシークバー上に表現することで、映像のどの部分がシーンとして定義されているかを俯瞰することができ、ユーザが映像全体の中から映像シーンを選択する手掛かりとなる。

シーンシークバーの下部には、映像に含まれているシーンのリストが表示される。リストには、Synvieで映像シーンを作成する際に付与したシーンタイトルとシーンの代表サムネイル、シーンの開始・終了時間が表示される。サムネイルを選択することで、対応するシーンがプレイヤーで再生される。また、サムネイル画像を表示することで、ユーザは、映像を再生する前にシーンの情報を閲覧することができるため、映像へのアクセスを効率的に行うことができると考えられる。ユーザは、シーンシークバーとシーン一覧を用いることで、映像にどのようなシーンが含まれているかを俯瞰することができる。

論文中心モードでは、論文上に表示された矩形を選択することで関連する情報が自動的に展開されるが、映像中心モードでは、映像の再生時間に応じて、その時間を含むシーンに関連するコンテンツが展開される。映像の再生に同期して関連コンテンツを自動的に提示することで、ユーザは視聴中のシーンについての理解を深めることができると考えられる。

## 5. 評価実験

Docvieが提供するインタフェースの有効性を確かめるために、比較対象としてTDAnnotatorで映像を閲覧できるようにした。図2に示すように、3.3節の方法で映像シーンが関連付けられた論文部分は青色の枠線（図2中では、論文左上の太線の矩形）で表示される。通常の論文部分と

- 1, クラスタリングした写真をタイル状に表示するために制約充足法 (CSP) を用いている  
○正しい ×間違い  
根拠となるパラグラフ :
- 2, ROI とは Region Of Interest の略である  
○正しい ×間違い  
根拠となるパラグラフ :
- 3, ROI 検出法にはボトムアップとトップダウンのふたつの手法を用いている  
○正しい ×間違い  
根拠となるパラグラフ :

図 10 実験に用いた大問の例  
Example questions.

表 1 グループ, ソフトウェア, 実験の順番, 問題セットの組合せ  
Sets of group, software and part-set for experiment.

	GoodReader	TDAnnotator	Docvie
グループ 1	1, 問題セット 1	2, 問題セット 2	3, 問題セット 3
グループ 2	2, 問題セット 2	3, 問題セット 3	1, 問題セット 1
グループ 3	3, 問題セット 3	1, 問題セット 1	2, 問題セット 2

同様に, この部分をクリックしてアノテーションを開くと, 映像シーンと論文部分の情報が表示される. この提示方法は単純なもので, 提示された映像シーンに対して別の論文の部分要素が関連付いていたり, その映像シーンを含む映像の他の映像シーンに, 別の論文部分が関連づいていたりしても, それをユーザが知ることはできない.

例えば, 図 6 に示すように, ユーザが論文 1 を閲覧しているときに, 論文 1 の部分要素に関連付けられた映像 1 の部分要素を閲覧することはできるが, 同じく映像 1 に関連付けられている論文 2 を, 映像 1 との関係を進んで閲覧することはできない. Docvie では提示された関連コンテンツを進んで他の映像や論文を閲覧することができ, ユーザが注目するコンテンツの種類に応じた適切なインタフェースを提供している.

本実験では, TDAnnotator と Docvie, 動画と映像の閲覧に利用される一般的なツール (本実験では, GoodReader\* と呼ばれる PDF と MPEG-4 形式の動画を別々に閲覧できるソフトウェアを用いた) の三つのソフトウェアを被験者実験を通して比較することで, Docvie で実現した機能が有効に働いているかどうかを, 読解効率 (回答時間および正答率) の観点から検証した.

### 5.1 実験の内容

本実験は, 被験者に論文と映像を閲覧してもらい, その内容に関する問題を解いてもらう形式である. 論文と映像の閲覧にはタブレット型デバイスを用い, GoodReader, TDAnnotator, Docvie のいずれかのソフトウェアを用いた.

図 10 に被験者に提示した実際の問題を示す. 小問 3 問を含む大問 3 問でひとつの問題セットとした. 小問は回答に際して, 被験者に提示される複数の論文と映像の内容に対して「正しい・間違い」で答えることができ, 提示された論文の中に答えの根拠となるパラグラフが必ず含まれる

ように作成した. また, アノテーションが付与されている部分だけを読めば答えられる問題とならないように, 根拠となるパラグラフが TDAnnotator と Docvie で表示される, アノテーションが付与されていることを表す矩形の文章中には含まれないようにした.

また, 被験者に提示する論文と映像は内容的に関係するもので, 大問ごとに異なる論文と映像を用いた. 論文は情報科学系の Human-Computer Interface に関する英語論文で, 大問ごとに 2 本以上で合計のページ数が A4 サイズで 10 ページとなるように選択した. 映像は論文に書かれている内容に関するデモ映像や発表映像で, 大問ごとに複数の映像を提示する. 映像の長さは平均 5 分 35 秒 (3 分 35 秒~9 分 21 秒) であった. また, 論文と同じく映像の言語も英語である (ただし, 問題文は日本語).

被験者は, 実際に当該分野の論文を読解する必要がある情報科学系の大学 4 年生 5 名, 修士課程 3 名の計 9 名で, TOEIC スコアの平均は 585.5 (465-715) であった. また, 実験で使用するタブレット型デバイスは被験者全員が日常的に利用しており操作に慣れている. 9 名を TOEIC スコアと学年を参考に 1 グループ 3 名の 3 グループに分けた.

また, 利用するソフトウェアの順番, および問題セットを表 1 に示すように組合せた.

### 5.2 実験の手順

実験はグループごとに, 表 1 に示す順番と問題セットで行った. それぞれのソフトウェアを用いた実験の前に, 実験に利用しない問題セットを用い, 5 分間, ソフトウェアの使い方をレクチャーした.

その後, 事前知識をできるだけ均一化することと, 動画の閲覧中に問題に回答することを防ぐ目的で, 出題した問題に関する映像シーン (平均 46 秒, 31 秒~1 分 28 秒) を最初から最後まで閲覧してもらった.

次に, 大問 1 問について 10 分の制限時間を設定し, それぞれのソフトウェアを用いて問題に回答してもらった. これは, 被験者に可能な限り迅速に問題に回答する事を促すためである. また, 小問への回答は「正しい・間違い」の答えと同時に, その回答の根拠となるパラグラフを示してもらった.

三つの手法それぞれに, 上述の方法で問題セット, 大問 3 問を解いてもらい, その回答とかかった時間を記録した.

### 5.3 実験結果と考察

上述した条件の下で行った実験の結果について述べる. 本実験では, Docvie の有効性を読解効率の観点から評価するために, 大問の回答にかかった平均時間と, 問題セットごとの正答率を算出した.

それぞれの結果を示す前に, 3 手法の間で大問の回答時間と正答率の平均値に差があるかどうかを分散分析で確認した結果を表 2 に示す. 結果から, 大問の回答にかかった平均時間と正答率の両方で手法間に有意差 ( $p < 0.001$ ) が認められた. そこで, Tukey の方法を用いて多重検定を行っ

\* <http://www.goodiware.com/goodreader.html>

**表 2** 分散分析の結果  
Result of variance analysis

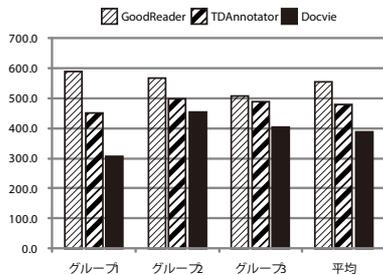
データ名	要因名	F 値	p 値 <sup>1</sup>
平均時間	手法 <sup>2</sup>	19.243	p<0.001 ***
	グループ <sup>3</sup>	2.420	0.096 n.s.
	問題セット <sup>4</sup>	1.276	0.285 n.s.
正答率	手法	16.000	p<0.001 ***
	グループ	0.514	0.601 n.s.
	問題セット	0.757	0.473 n.s.

<sup>1</sup> F 検定による有為確率 (\*\*\*)p<0.001, n.s. 5%水準で有為でない  
<sup>2</sup> GoodReader, TDAnnotator, Docvie  
<sup>3</sup> グループ 1~3 <sup>4</sup> 問題セット 1~3

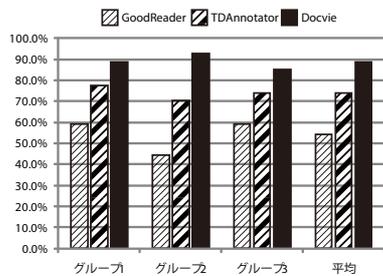
**表 3** 多重検定の結果  
Result of post hoc comparisons.

データ名	手法の組合せ		p 値 <sup>1</sup>
	手法 1	手法 2	
平均時間	GoodReader	TDAnnotator	0.017 *
	GoodReader	Docvie	p<0.001 ***
	TDAnnotator	Docvie	0.003 **
正答率	GoodReader	TDAnnotator	0.005 **
	GoodReader	Docvie	p<0.001 ***
	TDAnnotator	Docvie	0.047 *

<sup>1</sup> Tukey の方法による有為確率 (\*p<0.05, \*\*p<0.01 \*\*\*p<0.001)



**図 11** 大問の回答にかかった時間 (単位:秒)  
Average response time to a part (unit:seconds).



**図 12** 正答率  
Average percentage of questions answered correctly.

**表 4** 大問の回答にかかった時間 (単位:秒)  
Average response time to a part (unit:seconds).

	GoodReader	TDAnnotator	Docvie
グループ 1	588.3	451.0	307.0
グループ 2	568.0	497.9	453.4
グループ 3	507.3	488.1	404.3
平均	554.6	479.0	383.3

**表 5** 時間内に回答できなかった大問の個数と割合  
Average percentage and count of parts overtime.

	GoodReader	TDAnnotator	Docvie
グループ 1	7 (77.8%)	1 (11.1%)	0 (0.0%)
グループ 2	6 (66.7%)	1 (11.1%)	1 (11.1%)
グループ 3	2 (22.2%)	2 (22.2%)	1 (11.1%)
平均	5 (55.6%)	1.3 (14.8%)	0.7 (7.4%)

**表 6** 正答率  
Average percentage of questions answered correctly.

	GoodReader	TDAnnotator	Docvie
グループ 1	59.3%	77.8%	88.9%
グループ 2	44.4%	70.4%	92.6%
グループ 3	59.3%	74.1%	85.2%
平均	54.3%	74.1%	88.9%

図 11 および表 4 に、大問の回答にかかった平均時間を示す。大問の回答にかかった時間は、映像または映像シーンを閲覧し終わって大問への回答開始の合図とともに計測した。結果を見ると、GoodReader を用いた場合に比べて、TDAnnotator では 75.6 秒、Docvie では 171.3 秒早く回答することができている。表 3 に示した多重検定の結果から、各手法間に有為差が認められた。なお、10 分の制限時間内に大問に答えられなかった場合には、600 秒として計算している。制限時間内に答えられなかった大問の割合を表 5 に示す。GoodReader を用いた場合、制限時間内に大問に回答できなかった割合が 5 割を超えている。

このような結果が現れたのは、GoodReader では論文部分がハイライトされていないため、何も手がかりのない状態でどこが重要な部分かわからず、論文を最初から読んだためだと考えられる。また、TDAnnotator と Docvie で差が出たのは、Docvie ではどの部分が重要かというだけでなく、常に関連コンテンツが表示されている状態で閲覧できるようになっていたため、複数の論文を行き来しながら答えを探ることができ、答えに到達するまでの時間が短くなったと考えられる。

また、表 2 よりグループ間に有為傾向 (p=0.097) が見られた。この傾向は、グループ 1 が Docvie を用いて問題セット 3 に取り組んだとき、制限時間内に回答できなかった問題がなかったためだと考えられる。表 2 では問題セット間の大問の回答にかかった平均時間の有為差は見られなかった (p=0.285) が、問題セット 3 が他の問題に比べて容易であった可能性がある。問題セットの難易度、個人の能力差についてより詳細に分析するためには、実験の規模を拡大する必要があると考えられる。

図 12 および表 6 に、問題セットごとの正答率を示す。正答率は、ある問題セットに対する被験者の正答数を問題セットに含まれる小問題の数で割ったものである。ここで、正答としたのは、制限時間内に回答が行われ、問題に対する答えが合っていて、かつ被験者が示した回答の根拠となるパラグラフ中に適切な記述が含まれている場合とした。これは、誤読している場合や勘で答えている場合を不正解と

た。その結果を表 3 に示す。以降、この分析結果に触れながら結果について考察する。なお、分散分析、多重検定には R 2.15.0\*\*を用いた。

\*\* <http://www.R-project.org>

するためである。

正答率も、Docvie を用いた場合が最も高く、続いて TDAnnotator, 最後に GoodReader の順であった。表 3 から 3 手法の正解率の平均値に 5%水準で有為差が認められた。GoodReader を用いた場合の正答率が 5 割程度と低いのは、制限時間を超えてしまうことが多かったこと、手がかりが少ないために根拠となるパラグラフを探すことが難しかったことが挙げられる。TDAnnotator と Docvie の間で差がみられたのは、Docvie では関連コンテンツにアクセスしやすく、映像中心モードで論文に比べて把握しやすい映像を閲覧し、それを手がかりに関連する論文部分を読むことができたためだと考えられる。

以上の結果から、関連するコンテンツを利用して論文読解を支援するシステムとして、Docvie は TDAnnotator と比較して、効率的に内容を把握できると考えられる。

## 6. まとめと今後の課題

本研究では、論文と映像の部分要素に対するアノテーションの獲得と蓄積を行うための仕組み、部分要素間の関係情報を取得するための仕組みを、先行研究を拡張することで準備した。そして、獲得したメタ情報を利用するシステムの一例として、論文読解支援システムを提案した。提案システムは、メタ情報を用いてユーザが閲覧中のコンテンツに関連する他のコンテンツの一部を提示し、その理解を促進する仕組みを実現した。さらに、評価実験を行って、提案システムを用いることで効率的にコンテンツの内容を把握できることを確認した。

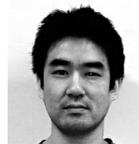
論文読解支援をさらに充実したものにするためには、ユーザが論文の内容をどの程度まで理解したかを量る指標の提案や、論文の内容を適切に理解するためのアノテーションを偏りなく獲得する仕組みを実現する必要がある。

そのために、我々は、論文閲覧システムや映像閲覧システムだけでなく、論文執筆支援システムに部分要素の作成や部分要素間の関連付けを獲得する仕組みを埋め込むことで、論文に対する有用なメタ情報を獲得、蓄積できる仕組みを実現することにも取り組んでいく予定である。

## (文 献)

- 1) 山本大介, 増田智樹, 大平茂輝, 長尾確: “映像を話題としたコミュニティ活動支援に基づくアノテーションシステム”, 情処学論, **48**, 12, pp. 3624-3636 (2007)
- 2) Jane Hunter and Ronald Schroeter: “Co-annota: A system for tagging relationships between multiple mixed-media objects”, IEEE Multimedia, **15**, 3, pp. 42-53 (2008)
- 3) James R. Davis and Daniel P. Huttenlocher: “Shared annotation for cooperative learning”, In Proceedings of the 1st International Conference on Computer Support for Collaborative Learning, CSCCL '95, pp. 84-88. L. Erlbaum Associates Inc. (1995)
- 4) Chunyuan Liao, François Guimbretière, and Ken Hinckley: “Piercraft: A command system for interactive paper”, In Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, UIST '05, pp. 241-244. ACM (2005)
- 5) Craig S. Tashman and W. Keith Edwards: “LiquidText: A Flexible, Multitouch Environment to Support Active Reading”, In Proceedings of the 2011 Annual Conference on Human Factors in

- Computing Systems, CHI '11, pp. 3285-3294. ACM (2011)
- 6) Andrew Crossen, Jay Budzik, Mason Warner, Larry Birnbaum, and Kristian J. Hammond: “Xlibris: An automated library research assistant”, In Proceedings of the 5th International Conference on Intelligent User Interfaces, pp. 49-52. ACM Press (2001)
- 7) Nitin Sawhney, David Balcom, and Ian Smith: “Authoring and navigating video in space and time”, IEEE Multimedia, **4**, pp. 30-39 (1997)
- 8) 長尾確, 大平茂輝, 山本大介: “オンラインビデオアノテーションの現状と課題”, 映情学誌, **64**, 2, pp. 173-177 (2010)
- 9) Matthew Kam, Jingtao Wang, Alastair Iles, Eric Tse, Jane Chiu, Daniel Glaser, Orna Tarshish, and John Canny: “Livenotes: A system for cooperative and augmented note-taking in lectures”, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, pp. 531-540. ACM (2005)
- 10) Richard C. Davis, James A. Landay, Victor Chen, Jonathan Huang, Rebecca B. Lee, Frances C. Li, James Lin, Charles B. Morrey, III, Ben Schleimer, Morgan N. Price, and Bill N. Schilit: “Notepals: Lightweight note sharing by the group, for the group”, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '99, pp. 338-345. ACM (1999).
- 11) Teresa Chambel, Carmen Zahn, and Matthias Finke: “Hypervideo design and support for contextualized learning”, In Proceedings of the IEEE International Conference on Advanced Learning Technologies, pp. 345-349. IEEE Computer Society (2004)
- 12) Bill N. Schilit, Gene Golovchinsky, and Morgan N. Price: “Beyond paper: Supporting active reading with free form digital ink annotations”, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '98, pp. 249-256. ACM Press/Addison-Wesley Publishing Co. (1998)
- 13) 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”, 情処学論, **51**, 6, pp. 1357-1370 (2010)



いしとやけんたるう  
**石戸谷 顕太郎** 2009年, 名古屋大学大学院情報科学研究科メディア科学専攻修士課程修了。2009年, 名古屋大学大学院情報科学研究科メディア科学専攻博士課程。マルチメディアコンテンツに対するアノテーションとその高度利用に関する研究に従事。



やまもと けいすけ  
**山本 圭介** 2010年, 名古屋大学大学院情報科学研究科メディア科学専攻修士課程修了。2010年, 株式会社デンソー ITソリューションズ。マルチメディアコンテンツに対するアノテーションとその高度利用に関する研究に従事。



おおひら しげき  
**大平 茂輝** 2000年, 早稲田大学大学院理工学研究科情報科学専攻修士課程修了。2003年, 早稲田大学大学院理工学研究科情報科学専攻博士課程単位取得退学。2001年, 早稲田大学理工学部情報科学助手。2004年, 名古屋大学エコトピア科学研究所助手。2009年, 名古屋大学情報基盤センター助教。マルチメディアコンテンツの知的処理とそれに付随するテキスト・音声/音響・画像処理に関する研究に従事。



ながお かつし  
**長尾 確** 1987年, 東京工業大学大学院総合理工学研究科システム科学専攻修士課程修了。1987年, 日本アイ・ビー・エム株式会社東京基礎研究所。1991年, 株式会社ソニーコンピュータサイエンス研究所。1995-1996年, 米国イリノイ大学アーバナ・シャンペーン校客員研究員。1999年, 日本アイ・ビー・エム株式会社東京基礎研究所。2001年, 名古屋大学工学研究科助教授。2002年, 名古屋大学情報メディア教育センター教授。2009年, 名古屋大学情報科学研究科メディア科学専攻教授。コンテンツとエージェントを基盤とした人間の知識の共有と再利用に関する研究に従事。