

## アノテーション機能を備えた 文献評価システムの構築

王 森<sup>†1</sup> 大塚 隆弘<sup>†1</sup> 榎原 博之<sup>†2</sup>

近年、書籍の電子化が進み、ソーシャルリーディングと呼ばれるサービスが普及し始めている。ソーシャルリーディングとは読書体験の共有することであり、電子時代の読書会と呼ばれる。電子化された書籍に対して、いかにうまくソーシャルリーディングという概念を実現するかが課題となる。本研究では、研究室向けに特化した、学術論文に対するソーシャルリーディングを実現できる文献評価システムの構築を行う。本システムはソーシャルリーディングでよく用いられるアノテーション機能を PDF ファイルに対して実装する。本システムはウェブ技術で実装するため、ユーザはウェブブラウザ上で文献を参照したり注釈を加えることができる。アノテーション情報は PDF ファイルとは独立に XML ファイルで保存されている。

### Implementation of a Literature Reputation System with an Annotation Function

OU BYO,<sup>†1</sup> TAKAHIRO OHTSUKA<sup>†1</sup>  
and HIROYUKI EBARA<sup>†2</sup>

Recent years, the service of social-reading has become popular with the development of E-book. The social-reading is used to share the reading experience with others. It is called a reading party in electronics age. The key point is that how to make the social-reading for E-book. This research develops a reputation system implements literature social-reading service for a laboratory. This system installs an annotation function for a PDF file. Because the system is made of the web technology, user can read a literature and make annotations in it by a browser. Annotation information is stored by a XML file.

### 1. はじめに

研究活動において、たくさんの論文を読むことが重要である。研究室などでは、同じ研究グループのメンバーがお互いに論文を紹介する勉強会も多く、他のメンバーの論文を参照する機会がよくある。そのような場合、文献の共有化と文献評価が必要となる。近年、多くの論文は PDF 化されており、文献の共有化と評価を実現するウェブシステムが望まれる。

一方、電子書籍に関しては、電子化が進み、ソーシャルリーディングと呼ばれるサービスが普及し始めている。ソーシャルリーディングとは読書体験の共有することであり、電子時代の読書会と呼ばれる。読書会とは、読んだ書籍の評価や感想を発表し、また自由に他人の感想を参照することである。評価や感想の手段として、書籍に注釈やマーカを記述する。電子化された書籍に対して、いかにうまくソーシャルリーディングという概念を実現するかが課題となる。

ソーシャルリーディングという概念は一般の文学作品だけではなく、学術的な分野においても価値がある。理由は研究者が論文に対して、マーカで記述することがよくあるためである。また、論文には専門用語が多く出るため、研究者は習慣的に専門用語の注釈を行う。そこで、ソーシャルリーディングを実現し、注釈やマーカなどのアノテーション機能を実装した文献評価システムは、文献の情報の共有において便利なツールであると考えられる。

本研究では、研究室向けに特化した、学術論文に対するソーシャルリーディングを実現できる文献評価システムの構築を行う。本システムはソーシャルリーディングでよく用いられるアノテーション機能を PDF ファイルに対して実装する。本システムはウェブ技術で実装するため、ユーザはウェブブラウザ上で文献を参照したり注釈を加えることができる。また、ユーザ間でアノテーションの内容を共有できる。ユーザが他のユーザのアノテーション内容を参照して、その文献の理解をより深めるだけではなく、広い視点から文献を読むことができる。さらに、公開された文献に対して、全てのユーザが、コメントや評価を行うことができる。評価を確認することにより、自分の研究に役立つ文献を探し出すことができる。

本稿の構成は以下のようになっている、本章を第 1 章とし、第 2 章ではソーシャルリー

†1 関西大学 理工学研究科  
Graduate School of Science and Engineering, Kansai University

†2 関西大学 システム理工学部 電気電子情報工学科  
Department of Electrical and Electronic Engineering,  
Faculty of Engineering Science, Kansai University

ディングの現状と課題について述べ、第3章ではアノテーションツールの現状と課題、第4章では文献評価システムの開発、第5章ではアノテーションの構築について述べる、第6章でまとめについて述べる。

## 2. ソーシャルリーディングの現状と課題

ソーシャルリーディングサービスの主な応用分野は電子書籍である。ソーシャルリーディングという概念を実現するのは2つの形式がある。

1つめは専用電子端末とウェブサイトを連携させる方法である。専用電子端末にはアプリでハイライトと注釈を付けて、ウェブサイトでその内容を共有する。代表的なアプリとして、Amazonの「Kindle」やiPadを電子端末としての利用する「Qlippity」がある。

例としてKindleについて述べる。Kindleでは同名の電子端末があるが、それ以外にパソコン、スマートフォン、タブレットでも利用できるという利点がある。ユーザはKindleのアプリを通して、電子書籍にハイライトや注釈を付けて、ウェブサイトでハイライトや注釈の内容のみを共有できる。ユーザがAmazonのウェブサイトで他者からのコメント、感想及び電子書籍の評価点数を見ることができる。Qlippityなどの他のアプリでも同様の形式のものが多い。しかし、この形式のソーシャルリーディングには欠点もある。

- ユーザが特定の文庫の電子書籍がしか読めない(例えばKindleはKindle Storeのみ)
- 大部分のアプリは専用の端末が必要である
- ウェブ上ではハイライト、注釈の内容のみを表示し、書籍を読むことができない

2つめの方式はオンライン閲覧サービス(ウェブ上で書籍を読むことができる)を提供して、ユーザは電子書籍を読みながら、他の人のコメントと評価を確認することができる。例えば、スクリプト\*1やブックワーム\*2、シェルフアーリ\*3というサイトがある。これらはソーシャルリーディングをサポートしているサイトであり、オンライン閲覧が可能である。しかし、専用端末やアプリを利用しないため、ハイライトや注釈を付けることができない。

## 3. アノテーションツールの現状と課題の実行環境

電子化された文献に対するアノテーションツールが多数発表されている。しかし、それ

\*1 <http://www.scribd.com>

\*2 <http://www.bookwormr.com>

\*3 <http://www.shelfari.com>

表1 表1 主なアノテーションツールの比較  
Table 1 Comparison with annotation tools

	Adobe Reader*4	iAnnotate*5	LORE*6	iPapers*7
長所	豊富なアノテーション機能	豊富なアノテーション機能	ウェブ対応	ウェブ対応
短所	1. アノテーション機能がウェブ上で利用できない 2. ソフトをインストールする必要がある	1. iPadの端末でのみ動作する 2. ローカルでのみ動作する	1. PDFファイルに対応していない 2. Firefoxでのみ動作する	1. マーカを記述できない 2. Macでのみ動作する 3. ソフトをインストールする必要がある

らのツールはローカルで動作するため仲間との情報共有が難しいなどの問題点がある。表1に、主なアノテーションツールを示す。

表1から見るとAdobe ReaderやiAnnotateなどは豊富なアノテーションを表示できるが、ローカルで動作するため情報共有が難しい。また、LOREやiPapersなどはウェブ上で動作するが、LOREはPDFファイルに対応していない。また、iPapersでは注釈のみに対応しており、マーカには未対応である。

## 4. 文献評価システムの開発

### 4.1 設計目標

提案システムではソーシャルリーディングの特徴であり、豊富なアノテーション機能の実装を目指す。

システム設計の目標を以下に示す：

- (1) ウェブブラウザ上でアノテーション機能を利用できる。

\*4 <http://www.adobe.com/jp/>

\*5 <http://app.if.journal.mycom.co.jp/apps/363998953/iAnnotate/>

\*6 <https://addons.mozilla.org/en-US/firefox/addon/lore/>

\*7 <http://ipapers.sourceforge.net/iPapers.html>

- (2) 豊富なアノテーション機能を実現させる。
- (3) 注釈やマーカを共有できる。
- (4) 評価とコメントを共有できる。

## 4.2 概要

システムは文献評価、ソーシャルリーディングを実現するため、ユーザが電子化された文献に対して、アノテーションだけではなく、コメントや評価の機能も備える。

本研究で構築したシステムでは開発言語に JAVA, ウェブサーバに Tomcat、データベースに MySQL を使用している。アノテーション機能は、注釈とマーカの2つ部分を分けて保存しておき、Flex と JSP を統合して表示する。注釈部分はデータベースに保存して、マーカの部分は XML で保存する。

## 4.3 文献評価システムの構築

本節では、文献評価システムの構築手法について述べる。

### 4.3.1 Flex の導入

システムはウェブブラウザの上で PDF ファイルに対するリーダーとアノテーション機能を実現するため、Flex を使用する。Flex は表現力豊かなウェブアプリケーションを構築できる。Flex とは Flash をベースにしたリッチインターネットアプリケーションの統合開発環境である。デザインには MXML(Macromedia Flex Markup Language)、プログラムには ActionScript を利用し、Flash Player の上で実行する SWF ファイルを生成する。

### 4.3.2 Flex と JSP の統合

Flex の MXML を JSP ページに統合するため、FlexTagLib という JSP タグライブラリを使用する。タグライブラリを利用すれば、ハイブリッドアプリケーションが開発できる。形式は図 1 のようになっている。例えば MainApp.mxml という MXML ファイルを JSP ページに埋め込む。

### 4.3.3 Flex と Java の通信

「BlazeDS」という Adobe のオープンソースプロジェクトを通して、クライアント側の Flex アプリケーションとサーバ側の Java のオブジェクトの通信を実現する。「BlazeDS」は、クライアント側の Flex アプリケーションからサーバ側の Java オブジェクトのメソッドを呼び出すことができるリモートキャッシング機能がある。

### 4.3.4 SWFtools の導入

Flex は SWF ファイルしか識別できないので、まず PDF ファイルを SWF ファイルに変換する必要がある。SWFtools を使って、SWF ファイルに変換する

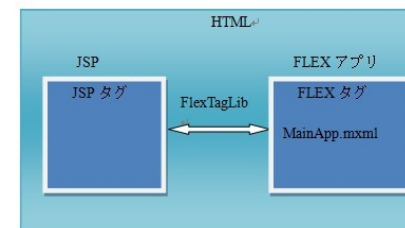


図 1 ハイブリッドアプリケーション  
Fig.1 Hybrid application

SWFtools はオープンソースの Flash を使う上で便利なコマンド集である。PDF などフォーマット内容を SWF に変換すると特に専用のプレイヤーをインストールしなくても内容を見ることができる。

### 4.3.5 システムの構築

システムのトップページを図 2 に示す。本システムではユーザごとにアカウントを作成する。ユーザは PDF ファイルのアップロード、文献に対するコメント、アノテーションの作成編集、また他人の文献やアノテーション内容を見ることができる。ユーザはアップロードする際、アノテーション内容を公開するかどうかを決めることができる。ユーザがアップロードした後、他のユーザがコメントや評価に点数を付けることができる。

本システムの流れは図 3 のようになっている。ユーザは PDF ファイルをサーバにアップロードする。アップロードされた PDF ファイルは SWFtools を使って自動的に SWF ファイルに変換される。変換された SWF ファイルをクライアントに送信する。クライアントは送信された SWF ファイルの内容を見ることができる。ユーザは文献に対してコメント、評価、アノテーションを加えることができる。このコメント、評価、アノテーションなどの情報を BlazeDS を通して、サーバに送信する。

## 5. アノテーションの構築

現段階では、「注釈」と「マーカ」の2つのアノテーション機能を実装している。図 4 ではアノテーションを追加している。左側のフリーハンドの部分はユーザが自由に線を描くことである。右上の注釈の印(?マック)は注釈が付いている場所である。マウスを注釈の印に合わせると注釈の内容を表示する。右下のマーカの部分はユーザがアンダーラインを引いている。

Reputation System>>Literature List

Upload Papers

Thesis List

<input type="checkbox"/>	Thesis Name	Author	University	Laboratory	Upload User	Upload Date	Download Times	Thesis Reputation Value
<input type="checkbox"/>	<a href="#">Metaheuristics From design to implementa</a>	Talbi, E.G.			ww	2011年08月04日 14:14:41	3	3
<input type="checkbox"/>	<a href="#">ECMAScript</a>	Brin, S. and Page, L.			ww	2011年08月04日 14:13:49	7	4
<input type="checkbox"/>	<a href="#">Overview - Phusion Passenger</a>	Muendler, M.A. and Ribeiro, T.			ww	2011年08月04日 14:13:49	5	3
<input type="checkbox"/>	<a href="#">Ajax: A new approach to web applications</a>	Garrett, J.J. and others	Kansai University	AL	ww	2011年08月04日 14:02:27	11	2
<input type="checkbox"/>	<a href="#">Improving scalability in MMOGs-ScalaMo</a>	David, P. and Vardi, A.	Kansai University	AL	ww	2011年07月03日 12:30:20	2	4

TOTAL PAGE 2 PageNo.1

図 2 トップページ  
Fig. 2 Top page

### 5.1 アノテーション権限

文献をアップロードするユーザがこの文献を他のユーザに公開するかどうかを決めることができる。公開された文献に対してすべてのユーザがアノテーションを追加できる。文献を注釈するユーザごとに XML ファイルを生成し、アノテーション情報を保存する。ユーザ自身のアノテーション情報は閲覧、修正、追加ができるが、他のユーザのアノテーション情報は閲覧のみ可能である。

例えば、図 5 のように、ユーザ A と B はお互いに相手のアノテーション情報を閲覧可能である。ユーザ A は A のアノテーションを修正、追加可能であるが、B のアノテーションを修正、追加することはできない。

### 5.2 マーカの表示

マーカの部分は直線とフリーハンド 2 種類がある。図 6 のように、Flex アプリケーションがマーカを作成して、マーカと文献の情報を文字列パラメータとして、サーバ側に送信する。サーバ側はクライアント側から送られるパラメータを解析して、文献ごとに XML ファイルを作成する。XML ファイルには文献のマーカの位置とカラーなどの情報を保存する。

XML ファイルの内容とデータベースの注釈の内容を合わせて、文献のアノテーション内容を表示する。複数のユーザが作成したアノテーション情報は重ねて表示する。

### 5.3 マーカの保存

本システムでは JDOM を用いて、XML ファイルの作成や解析などの動作を行う。PDF ファイルをサーバにアップロードすると、サーバ側は PDF ファイルに対して自動的に XML ファイル作成する。JDOM(Java Document Object Model) は Java で XML を容易に扱うため、使いやすい API を提供することを目指して設計された軽量 XML ライブラリである。JDOM の設計は、構文からセマンティクスまで Java 言語を取り入れている。本システムではすべてのマーカ情報を XML ファイルに記録する。マーカが直線の場合、起点と終点の座標を記録すれば充分である。しかし、フリーハンドの場合は線が通るすべての点の座標を記録する必要がある。図 7 は XML ファイルの基本例である。

XML ファイルの解析の効率を向上するため、マーカのすべての点を 1 つ文字列に記録する。そこで解析する際、マーカのタグまで解析する必要がある。抽出された文字列をそのままクライアントに送信する。クライアントでは送られた文字列の座標の情報を Flex アプ

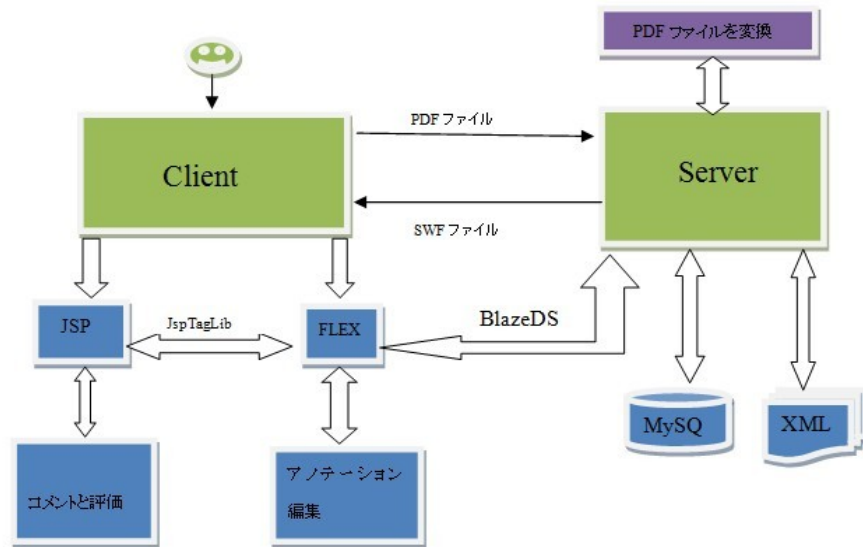


図3 システムの流れ  
Fig.3 System flow

リケーションを用いて読み出す。こうするとXMLファイルの解析時間と送信されたデータの量が減らすことができるため、システムの高速化を図ることができる。

例えば、直線の場合、line というタグの下に style と path という2つプロパティがある。style には straight, path には "174 634 547 634" の値を与える。XMLファイルを解析する際、path の文字列をクライアントに送信する。"174 634 547 634" の前の2つ数値174と634は起点の座標 X1 と Y1、後ろの2つの数値547と634は終点の座標 X2 と Y2を表す。フリーハンドの場合はデータの基本的な形式が同じであるが style は freehand を設定し、path の文字列の長さはマーカの長さ(点の数)によって異なる。フリーハンドは現在折れ線で表示しているが今後スプライン曲線で表示するように変更する予定である。FLEXアプリケーションが path の文字列を読み込む際、文字列をスペースで分割して、点の座標を読み出す。

以下はXMLファイルの操作について説明する。

図4 アノテーションの表示  
Fig.4 Annotation view

- XMLファイルから情報の抽出  
クライアント側から文献名とページ数をパラメータとしてサーバに送信する。サーバでは受け取ったパラメータからアノテーションの内容を抽出する。そして、抽出したアノテーション内容をクライアントにマーカの配列として送信する。
- マーカの追加  
Flexアプリケーションを用いてマーカとページ数の情報をサーバに送信する。サーバは受け取ったページ数とマーカの内容をXMLファイルから抽出する。そして、抽出し

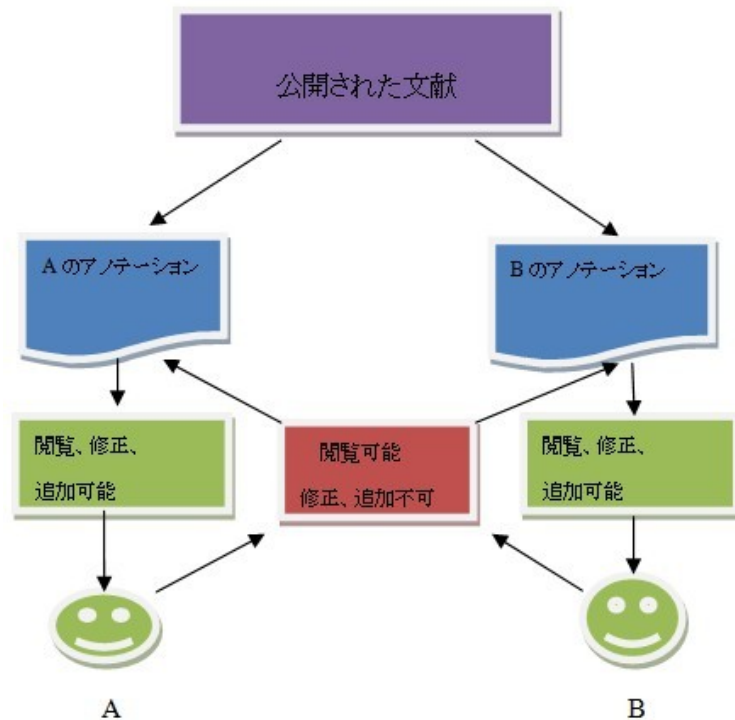


図5 アノテーション権限  
Fig.5 Annotation authority

たXMLファイルにマーカの情報を新しいタグとして加える。

● マーカの削除

Flexアプリケーションで削除するマーカの座標とページ数の情報をサーバに送信する。サーバは受け取ったページ数とマーカの内容をXMLファイルから抽出する。クライアントから送られた座標とXMLファイルから抽出した座標を比較して一致したマーカを削除した後、XMLファイルのページの内容を更新する。XMLファイルの内容を更新後、Flexアプリケーションがページの情報を読み込み、ウェブページを更新する。

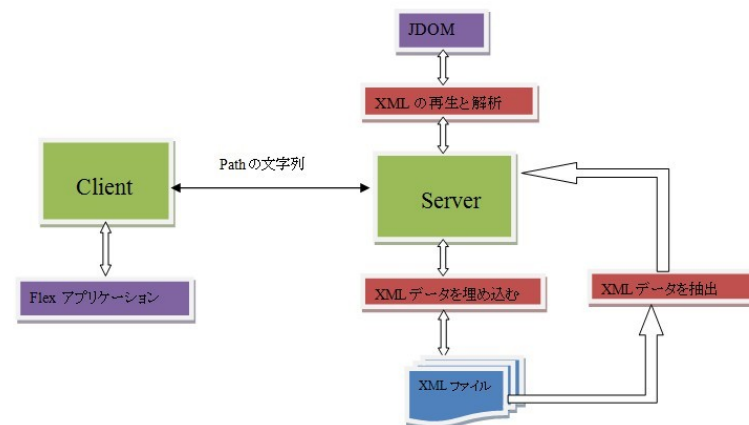


図6 XMLの流れ  
Fig.6 XML flow

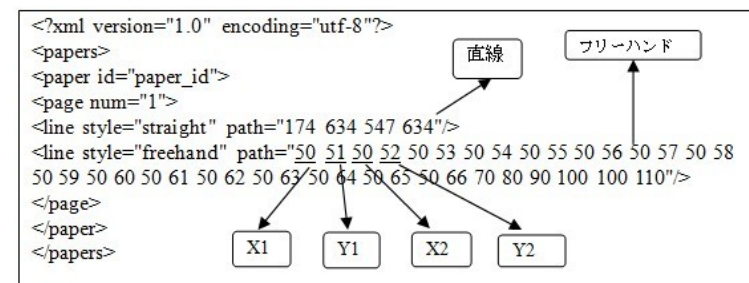


図7 XMLファイルの基本例  
Fig.7 Example of XML file

5.4 注 釈

自分がほしいところに注釈文を付けることができる(図4の注釈の印)。マウスを注釈の印に合わせると、注釈画面を表示する。マーカの内容をアクセス効率を向上するためXMLファイルに保存する。また、注釈の位置や内容の情報はデータベースに保存する。ユーザが自分の注釈の内容を閲覧する時、図8のように編集することができる。他のユーザの注釈内

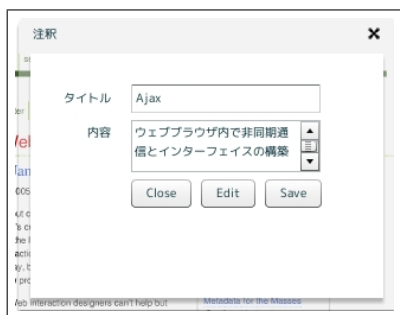


図 8 注釈編集可能な画面  
 Fig. 8 Annotation edit screen

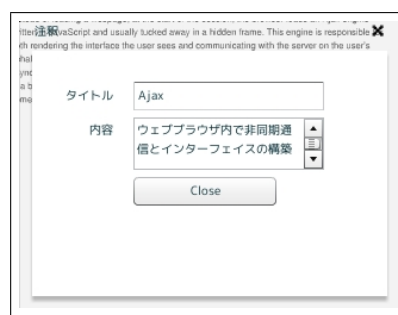


図 9 注釈編集不可の画面  
 Fig. 9 Impossible to edit screen

注釈履歴	
タイトル	内容
Ajax	エイジャックス、アジャックス、アヤックス
XML	エクステンシブル マークアップ ランゲージ
J2EE	Java Platform, Standard Edition

図 10 注釈履歴  
 Fig. 10 History of annotation

内容を閲覧する時、図 9 のように編集することができない。ユーザは注釈履歴 (図 10) を見ることができる。ユーザは他のユーザのアノテーション内容を見る時、注釈履歴を見ることができる。図 10 は注釈履歴例である。文献に対して、ユーザの注釈記録を表示する。図の左側がタイトルで、右側が注釈内容である。ユーザは自身の注釈履歴に対して、すべての注釈項目の閲覧や注釈の編集をすることができる。

## 6. おわりに

様々な研究を行う上で電子化された文献の処理は重要である。研究者として、論文に対して意思疎通を図る上で一番直感的な方法は、コメントを付けることではなく、アノテーションの内容を共有することである。しかし、ウェブ上でアノテーション機能を実現するのはローカルなソフトと比べて困難である。原因としては、まずウェブブラウザは機能の制限があり、ローカルなソフトのような豊富 API がない。また、アノテーション内容の保存に

おいては、マーカの内容がデータベースに保存できない。

そこで本研究では Flex と Java を統合して、ウェブ上で動作するアノテーション機能を備えた文献評価システムを構築した。本システムはソーシャルリーディングを実現する上で、研究者は文献に対して注釈やマーカといったアノテーションを記述することができる。本システムを利用することにより、研究活動の効率化を進めることが期待できる。

今後の課題を以下に例挙する。

- (1) スプライン曲線でフリーハンドを表示するなどの機能を実現する
- (2) システムの性能を評価するため、システムの利用者を対象にアンケートを行う

## 参 考 文 献

- 1) 宮田 亮 : Flex4 プログラミング入門, 秀和システム (2010).
- 2) Colin Moock, 永井勝則 : 詳説 ActionScript 3.0, オライリージャパン (2008).
- 3) 公門和也, 大谷晋平, 堀越悠久史 : Adobe Flex 3 and AIR ではじめるアプリケーション開発, インプレスジャパン (2008).
- 4) 河村 嘉之, 川尻 剛, 福沢 知海 : Java 開発者のための Ajax 実践開発入門, 技術評論社 (2007).
- 5) 清野 克行 : Ajax による業務アプリケーション開発, 秀和システム (2008).
- 6) Ginsburg, M. and Kambil, A. : Annotate: A web-based knowledge management support system for document collections, System Sciences, 1999. HICSS-32, pp. 10-pp (1991).
- 7) Allen B. : Annotate: How to Think Like a Computer Scientist: Java Programming, Createspace, (2009)
- 8) Fain, Y. and Rasputnis, V. and Tartakovsky, A.: *Rich Internet Applications with Adobe Flex & Java: Secrets of the Masters*, SYS-CON Media, (2007).
- 9) Volkmer, T., Smith, J.R. and Natsev, A.P.: *A web-based system for collaborative annotation of large image and video collections: an evaluation and user study*, Proceedings of the 13th annual ACM international conference on Multimedia, Vol.1, chapter2, pp.892-901, ACM, (2005).
- 10) Garrett, J.J. and others: Ajax: A new approach to web applications (online), available from (<http://www.robertspahr.com/teaching/nmp/>) (accessed 2005).