

論文アノテーションの共有システムの構築

青山 俊弘

鈴鹿工業高等専門学校 電子情報工学科

aoyama@info.suzuka-ct.ac.jp

山地 一楨

国立情報学研究所 学術ネットワーク研究開発センター

yamaji@nii.ac.jp

概要

近年、論文は電子ファイルによって公開されることが一般的となってきた。本研究では、電子ファイル論文に対するアノテーション付けを支援し、ウェブサービスを用いて共有するシステムの構築を行った。

キーワード

アノテーション、電子論文、ウェブサービス

Sharing system of annotation for paper publication

Toshihiro AOYAMA

Department of Electronic and Information Engineering, Suzuka National College of Technology

aoyama@info.suzuka-ct.ac.jp

Kazutsuna YAMAJI

R & D Center for Academic Networks, National Institute of Informatics

yamaji@nii.ac.jp

Abstract

The sharing the annotation of article is useful among co-researchers and students in same laboratory. We developed the system to create annotations to PDF articles and to share these annotations using Web-service.

Keywords

Annotation, PDF paper, Web service

1. はじめに

近年ではインターネット上に多くの論文検索サイト [1, 2, 3] が存在し、論文自体も電子化 (PDF 化) された [4]。すなわち、ネットワークを通じた文献検索、論文入手という流れは一般的なものになっている。論文の書誌情報がデジタル化され、検索方法が API として公開されることにより、API を使った元の検索サイトとは異なる検索インターフェースなども提案されている [5]。これにより論文検索、入手、整理までは電子化されることとなり、研究者にとって便利な環境になりつつある。

ところで論文を読む際には、読者は単に読むだけではなく、疑問点、自分の理解、他の文献や知識との関連などをテキストや図などの形でメモすることとなる。現在はこれらの元のコンテンツに対する付加的な情報 (アノテーション) は論文の欄外やノートなどに書かれることが多い。従って、論文を入手するところまでは電子化されているものの、その後の利用では従来と同じ紙ベースとなってしまっている。書誌情報、論文自体は電子化されているため可用性があがった論文管理システムであるが、アノテーションは紙媒体に書かれてしまうため一体となって管理することが難しい。

本研究では PDF となった論文へのアノテーションを電子化し、それを共有するシステムを構築した。論文アノテーションの共有は大学などの研究室や共同研究者間などにおいて、論文から得た知識を共有する際に有益であると考えられる。

2. アノテーション

2.1 デジタルアノテーション

デジタルライブラリに対するアノテーションは、Web が発達したことからさまざまな形で利用されている。例えば、一般的な Web サイトに対するアノテーション付けシステムである Annotate[6]、教育目的のアノテーションツールである Digital Library for Earth System Education(DLESE)[7] がある。これらは Web コンテンツに対して利用者が付加的な情報を付加することができるシステムである。

一方、論文の電子媒体としても普及が進む PDF に関しても、Text, Link, FreeText, Line, Circle, Polygon, PolyLine, Highlight, Underline, Stamp など、文章に対するマーク、メモ、校正などを PDF 上に残すことができる仕様となっている [8]。また、EMC Documentum PDF Annotation Services など製品もあり、PDF のレビューやコメント機能を統合するサービスも存在する [9]。

2.2 文献アノテーション

文献に書かれたアノテーションは、文章の関係を強調する、他と文章と比較する、文章関連する考えを記録する、などの目的で利用される。これらのアノテーションのうちのいくつかは、共有されることで他者の読みの質を変えることがある [11]。

また、不特定多数のアノテーションやレーティングが製品を購入するときの指針になっていることからも、デジタルライブラリに対する社会的なアノテーションは有益であると考えられる [10]。しかし、紙媒体に書かれた私的なアノテーションは、再入力をする必要があることもありあまり公開されることはほとんどない。そのため、アノテーションの公開を促すにはアノテーションを効率よく電子化する必要がある [11]。ラボ内でアノテーションを共有するためのシステムとして SmartCourier[16] があるが、アノテーションの収集を広く行うには Web サービスとして実装するのが利用しやすいと考えた。

そこで、論文管理ソフトに PDF 閲覧機能と、PDF に対するアノテーション付け機能を統合し、論文の入手、整理、閲覧、アノテーション付けの一連の作業をシームレスに行う環境を整備するとともに、Web サービスを通じて同じ論文に対する公開されたアノテーションも参照することができるシステムを開発する。

3. PDF アノテーション共有システム

3.1 構成

本研究では電子化された論文を対象にシステムを構築した。システムはアノテーションデータを保存するアノテーションサーバと、アノテーションを PDF 論文に付加するクライアントアプリケーションから構成される。サーバ、クライアント間の通信は SOAP(Apache Axis2[12]) を用いた。

3.2 サーバ

論文アノテーションはクライアントアプリケーションである PDF 閲覧ソフト上で作成され、ユーザによって選択されたアノテーションのみサーバに送信され、共有可能となる。

各アノテーションは、該当する論文の ID にひも付けされることで論文に対する関係性を保持することができる。ある論文は出版社による ID、各文献検索サイトでの ID など、複数の ID を持つことがある。そのため、論文検索サイト (DOI, PMID, CiNii, etc.) と、各サイトにおけるユニークな ID を組み合わせることで、論文との関連性を保持することとした。サーバ上で同一論文の複数の ID 同士を相互変換することで、異なる検索サイト ID を持つ同一論文の名寄せも可能となる。

サーバ上ではクライアントから送られたアノテーションを SQL データベース内に保持する。表 1 に示すデータを SQL データベース内に保持し、クライアントからの問い合わせに対して対応するアノテーションリストやアノテーションを送信する。

page	PDF 論文上のページ
position	PDF のページ上での位置 (x,y)
date	アノテーション作成日時
note	アノテーション本文
contributor	アノテーション制作者
title	タイトル
type	アノテーションのタイプ (予約)

表 1 サーバ上に保管されるデータ

3.3 クライアント

クライアントの一つとして MacOSX 上で動作する文献管理ソフト iPapers2[13, 14] に、アノテーション編集とサーバからのダウンロード機能を追加した (図 1)。読んでいる文献にテキストアノテーションを追加することができるほか、論文 ID をクエリとし、サーバへ問い合わせすることで、対応する文献のアノテーション入手でき、PDF 論文上に F アノテーションとして参照することができる。

また、赤外線を用いたデジタルペンによるクライアントやタブレット PC 上で動作するクライアントも開発している (図 2)。これらは紙媒体からデジタルデータへの橋渡しをする上で重要な役割を果たすものと考えられ



図1 iPapers2

る。今後、開発中であるWindows版の文献整理ソフトと連携することで、印刷した論文に対するデジタルペンを使った手書き入力によるアノテーション付けや、タブレットPC上に表示された論文に対して手書き入力でアノテーション付けが行える環境を構築する予定である。

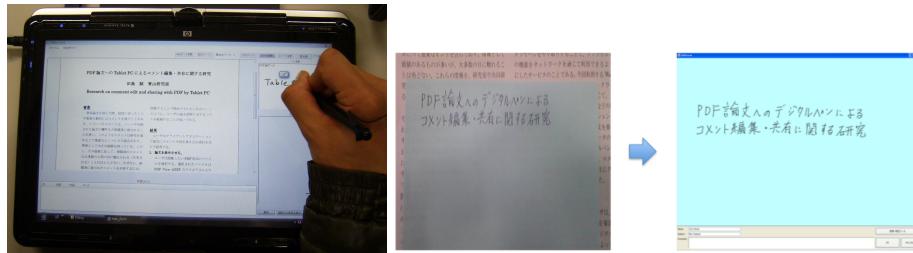


図2 TabletPC/Digital Pen

4. おわりに

本システムにより、研究者は、自身の文献を読んだ記録をデジタル情報として保存、管理することができる。また、他の研究者が公開している論文コメントを、文献検索時に参照することも可能である。これらは学生や新しい分野に参入した研究者にとって、圧倒的な量の論文から必要な論文を選択する際に有益な情報を与えるものとなると考えられる。また、共同研究のグループや研究室などでは、文献に関するアノテーションが蓄積されることで、グループ内で分野への理解や、知識を広げていくきっかけになるものと考えられる。

参考文献

- [1] CiNii, <http://ci.nii.ac.jp/>
- [2] Google Scholar, <http://scholar.google.com/>
- [3] PubMed, <http://www.ncbi.nlm.gov/pubmed/>

- [4] Delamothe, T. and Smith, R. PubMed Central: creating an Aladdin's cave of ideas. *BMJ*, 322, pp.1-2, 2001.
- [5] Muin, M., Fontelo, P., Liu, F. and Acherman, M. SLIM: an alternative Web interface for MEDLINE/PubMed searches - a preliminary study. *BMC Medical Informatics and Decision Making*, 5(37), doi:10.1186/1472-6947-5-37, 2005.
- [6] Annotate, <http://www.annotate.org/>
- [7] Arko, R.A., Ginger, K.M., Kastens, K.A., and Weatherley, J., Using Annotations to Add Value to a Digital Library for Education. *D-Lib Magazine*, 12(5), 2006.
- [8] Document management - Portable document format - Part1:PDF1.7, First Edition, http://www.adobe.com/devnet/acrobat/pdfs/PDF32000_2008.pdf
- [9] PDF Annotation Services, <http://japan.emc.com/products/detail/software/pdf-annotation-services.htm>
- [10] Gazan, R., Social Annotations in Digital Library Collections. *D-Lib Magazine*, 14(11/12), 2008.
- [11] Marshall, C. C., Brush, B. AJ., Exploring the relationship between personal and public annotations, International Conference on Digital Libraries archive Proceedings of the 4th ACM/IEEE-CS joint conference on Digital libraries, 349-357, 2004
- [12] Apache Axis2, <http://ws.apache.org/axis2/>
- [13] 青山俊弘, 堀浩之, 白井支朗, メタデータに基づく文献管理ソフトウェア: iPapers, 医学図書館, 54(3), 243-247, 2007.
- [14] iPapers2, <http://ipapers.sourceforge.net/>
- [15] Agosti, M., Frerro, N., Frommholz, I. Thiel, U., Annotations in digital libraries and collaboratories: Facets, models and usage., *Research and Advanced Technology for Digital Libraries.*, Lecture Notes in Computer Science, Heery, R., Lyon, L.(Ed), Springer, 3232, 244-255, 2004.
- [16] 伊藤禎宣, 角康之, 間瀬健二, 國藤進, 人工知能学会論文誌, 17(3), 301-312, 2002