JSTリンクセンターの新機能

—Googleとの連携とJ-STAGEにおける論文の被引用関係表示

久保田壮一¹; 荒川紀子¹; 和田光俊¹; 近藤裕治²; 小久保浩²; 山崎 匠²

著者抄録 :JSTが運営する電子ジャーナルサイトJ-STAGEのリンク機能を担うJSTリンクセンターが関係する新機能を2つ紹介する。1つはJ-STAGE上の論文本文を検索エンジンGoogleやGoogle Scholarで検索できるようにクロールさせる機能, もう1つは論文間の被引用関係表示を行う機能である。これまでJ-STAGE記事内同士の被引用関係は実現されていたが, 2005年5月からCrossRefのForward Linking機能を利用して, J-STAGE外の記事からの引用関係を取得し, これを被引用リンクとして表示することができるようになった。

キーワード : JSTリンクセンター, J-STAGE, 電子ジャーナル, Google Scholar, CrossRef Forward Linking, 被引用リンク

New services with JST Link Center

—Crawling J-STAGE by Google and citing articles on J-STAGE

KUBOTA Soichi ¹; ARAKAWA Noriko¹; WADA Mitsutoshi¹;
KONDO Yuji ²; KOKUBO Hiroshi ²; YAMAZAKI Takumi²

Author Abstract: We will introduce two important features of JST Link Center, which provides an electronic journal site J-STAGE operated by Japan Science and Technology Agency (JST) with a linking feature. One is a mechanism, which enables Google to crawl J-STAGE articles via the Internet. This enables users to retrieve fulltext keywords of J-STAGE on Google, or Google Scholar. The other is showing a link to citing articles with CrossRef Forward Linking service since May 2005. This enables users to reach the articles, which cite J-STAGE articles beyond J-STAGE journals while we have shown those within J-STAGE so far.

Key words: JST Link Center, J-STAGE, electronic journal, Google Scholar, CrossRef Forward Linking, links to citing articles

原稿受理 (2006-03-14)

(情報管理 49 (2): 69-76)

1. はじめに

JSTリンクセンター¹⁾は, 電子ジャーナルサイトJ-STAGE²⁾のリンク機能を担当するサブシステムである。J-STAGEは, 検索機能や電子付録, 早期公開など冊子体では提供不可能なさまざまな機能³⁾を提供し, 論文などのコンテンツに付加価値を加えている。JSTリンクセンターはJ-STAGEに登録された

論文などのコンテンツと他サイトのコンテンツをハイパーリンクで結ぶことにより, 利用者の求める情報へのアクセスを容易にし, また, J-STAGE掲載論文へのアクセス増加に寄与する。具体的には, J-STAGE掲載論文の引用文献リストから引用文献自体, あるいはPubMedやCAPLus, JDreamなどといった文献データベースの該当レコードへのリンク, 逆にこれらデータベースの検索結果からあ

¹ 独立行政法人科学技術振興機構 (JST) 文献情報部 電子ジャーナル課 (〒102-0081 東京都千代田区四番町 5-3) Tel.03-5214-8455 E-mail: kubota@jst.go.jp, narakawa@jst.go.jp, wada@jst.go.jp

² Electronic Journals Division, Department of Literature Information, Japan Science and Technology Agency (JST) (5-3 Yonbancho, Chiyoda-ku, Tokyo 102-0081)

³ 株式会社アトラス (〒103-0014 東京都中央区日本橋蛸殻町1-39-5水天宮北辰ビル2F) Tel. 03-5642-9300 E-mail: kondo@atlas.jp, kokubo@atlas.jp, yamazaki@atlas.jp

² Atlas Corporation (Suitengu-Hokushin Bldg. 2F 1-39-5 Nihombashi Kakigara-cho, Chuo-ku, Tokyo 103-0014)

るいは他の電子ジャーナルサイトの引用文献からJ-STAGE掲載論文へのリンクを実現している。また、J-STAGEに掲載されたジャーナル論文のほとんどに付与されている識別番号であるJOI (JST Object Identifier) の登録・案内システムとしての役割も担っている。本稿では、このJSTリンクセンターが関係する2つの新サービスについて紹介する。

2. GoogleおよびGoogle Scholarとの連携

Googleは簡易なキーワード検索で広く利用されている検索エンジンである。Googlebotというプログラムがインターネット上のコンテンツを自動収集し（以下、「クロールする」という、またこの自動収集プログラムのことを「クローラ」という）、索引化する。検索結果をPageRank⁴⁾と呼ばれる方式を含む100以上の要素から順位付けを行い、順位の高い順に並べて表示を行う。Google Scholar⁵⁾はGoogleが2004年に始めたサービスで、対象を学術文献に限定した検索サービスである。サービス開始当初から2006年3月現在に至るまでBETA版としてサービス提供されている。Googleとの違いは、①検索結果表示順について、当該論文の他論文からの引用回数、検索キーワードが本文のどこに現れるかなど論文に特化したさまざまな要素を勘案して優先度が決まる。②検索結果に書誌情報の一部や引用回数が記載される。③正式コピーの他に利用可能な別バージョンが存在する場合、それぞれへのリンクをまとめて表示する。④検索オプションに著者名、論文タイトル、発行年など学術論文検索に特化した項目が指定できる。などである。J-STAGEはこれまで、検索エンジンのクローラによるアクセスにより通常のアクセス統計が乱されることへの懸念や、論文の著作権への配慮などから掲載論文へのクロールを許可していなかったが、J-STAGEでジャーナルを公開している学協会からは、Googleで検索できるようにして欲しいという要望を頂いていた。このたび、Googleとの間で検索について合意し、J-STAGE上の論文をGoogleおよびGoogle Scholarから検索可能にすることとなった。

2.1 J-STAGEの論文公開とGoogle連携の概要

J-STAGE上には日本の学協会等が発行したジャーナル（2006年3月3日現在で273誌。誌名変更も含む）の論文が掲載されているがその公開方法はさまざまである。論文本文はPDF形式で（オプションで

HTML形式での公開を付加することも可能）公開されているが、すべて認証なしで公開しているものもあれば、何らかの認証を経ないと閲覧できないようにしているものもある。認証機能には購読者ID・パスワードによるものと、機関のIPアドレスによるものがある。また、認証機能設定した論文をクレジット決済で1部単位で購入できるサービスも提供している。認証設定を行うかどうか、およびその設定内容はジャーナルを提供している学協会等が自由に変更できるようになっている。J-STAGEに掲載された論文は著者抄録までは認証なしで公開している。著者抄録を表示するページから論文全文あるいは引用文献リストへのリンクが張られており、前述の認証機能はこれらのリンクをクリックした際に機能する。

次に今回のGoogleとの連携の概要を挙げる。

- (1) 学協会がジャーナル単位でGoogle連携に参加・不参加が選べる
- (2) Google連携に参加する場合、認証設定の有無によらずそのジャーナルで公開された全論文の全文情報（PDF形式）をGoogleにクロールさせる。また、その書誌抄録、引用情報も別途Googleに提供する。
- (3) 連携参加の場合、Google、Google Scholarともに検索対象となる。
- (4) 検索結果表示からのリンク先は書誌抄録ページとし、全文に直接リンクはしない。
- (5) 検索結果表示に通常のGoogle検索結果で見られる「キャッシュ」へのリンクは表示しない。

上記の(4),(5)により、論文全文に認証設定を行っているジャーナルの論文全文をユーザがGoogleから取得してしまうことはない。

2.2 Google連携におけるJSTリンクセンターの役割

さて、今回のGoogle連携におけるJSTリンクセンターの役割を次に述べる。一見、無関係に見えるが先に挙げた(2)、(4)の実現のためにJSTリンクセンターの機能が利用されている。J-STAGE掲載ジャーナルの論文には、JOIが付与されている（JSTリンクセンターを利用していないジャーナルを除く）が、JOIの前に<http://joi.jlc.jst.go.jp/>を付けることにより当該論文のURLとなる。こうして作られたURLでアクセスすると当該論文の抄録ページへ転送されるが、この転送処理はJSTリンクセンターが行っている。GoogleはJ-STAGE上のコンテンツを収集する際にあらかじめ指定したURLにアクセスす

る。このURLにはJ-STAGE上で、学協会が収集を許可したジャーナルのみのコンテンツの目次に当たるリストを置いておく。Googleのクローラはこの目次をたどりコンテンツを収集するがそのときに論文へのリンクに前述のJOIを使ったURLを使用する。JSTリンクセンターはGoogleからのアクセスかどうかを識別して通常当該論文の抄録ページへ案内するところを全文PDFへ案内する。この振り分け処理により、Googleの検索結果URLをJOI使用のものとすることができ、連携の要請(4)を実現している(図1)。Googleがクロールするサーバを通常のものとは分けているが、これは(i)アクセス統計データへのクローラの影響をなくせる、(ii)クローラに収集させるコンテンツを許可したもののみ限定できるというメリットのためである。

2.3 画面表示例

GoogleでJ-STAGEの論文を検索した場合の表示例を図2に示す。ご覧の通り、検索結果にキャッシュへのリンクは表示されていない。検索結果URLをクリックすると論文の抄録ページへ移動する。

2.4 試行運用でのアクセス状況

現在は、発行学協会の許可を得てJ-STAGEに登載された4ジャーナルを連携のための準備として、2005年8月より試行運用を行っている。2006年2月末時点で2,000弱の論文がクロール対象論文となっている。試行対象となっている論文のGoogle経由でのアクセス状況を報告する。図3に示すとおり、登載論文の増加とともに月ごとのアクセス数もほ

ぼ増加している。アクセス元の国の数も増加傾向を示している。2005年9月から2006年2月までの6か月間の延べで、82か国および地域(アクセス元IPアドレスによる)から12,195件のアクセスがあった(図4)。

2.5 Google連携における今後の課題

Googleとの試行連携を行う中で見えてきた課題をいくつか次に挙げる。

(1) Google Scholarの更新頻度

Google Scholarは通常のGoogleに較べてクロール頻度が少なく、情報の更新が若干遅い。通常のGoogleの索引化に較べて引用解析などに時間がかかるためと思われるが、この問題はGoogleも認識しており、今後改善されると思われる。

(2) ランキング向上

Googleのランキングでは、他のサイトからリンクされている数が重要な要素となる。現在の連携の方式は、Google Scholarでの利用を重視したものであり、通常のGoogleでのランキングは低くなる傾向がある。今後、より上位に表示されるよう検討を行う。

(3) 画像のみの論文全文PDF

Google Scholarでは画像のみのPDFから全文情報を抽出する試みが行われているが、将来的には日本語などへの対応も行われるものと思われる。

3. 被引用リンク表示機能

電子ジャーナルサイトでは、論文を公開する際

図1 JSTリンクセンターによる振り分け処理

図2 J-STAGE登録論文のGoogle検索結果表示

図3 Google経由アクセス数の推移

に、その論文 (Aとする) が引用している別の論文 (Bとする) へのリンクを論文Aの引用文献リストに埋め込むことは広く行われてきた。この引用文献リンクに加え、論文Bを公開する際に、論文Aを始めとする論文Bを引用している文献へのリンク

(本稿では、このようなリンクを被引用リンクと呼ぶ) を張ることも一般的になりつつある。J-STAGEにおいては、J-STAGEで公開された論文同士の間で被引用リンクを2005年1月から提供してきた。

図4 Google経由のJ-STAGEアクセス状況
(国別、2005年9月～2006年2月)

3.1 被引用リンクの仕組み

図5にあるように、論文Aが論文Bを引用している関係をJ-STAGE登録データの中の論文Aの引用文献リストから取得し、JSTリンクセンター内の引用－被引用関係テーブルに登録する。ある論文Cを引用している論文を探す際にはこのテーブルを逆にたどればよい。

3.2 CrossRef Forward Linking

CrossRefでは2004年にForward Linkingサービスを開始した。CrossRefでは、これまで論文の書誌情報とDOI (Digital Object Identifier: インターネット上のデジタルコンテンツの識別子)⁶⁾とその所在URLのセットを出版社等から預かり、DOIの登録機関であるIDF (International DOI Foundation) へのDOI登録および、書誌情報によるDOI検索サービスを行ってきた。Forward Linkingでは新たに、論文とその論文が引用している別論文の関係も預かる。CrossRefにDOIで検索するとそのDOIに対応する論文を引用している論文のDOIを回答する。この機能を使い、J-STAGE登載の論文を引用しているJ-STAGE外の文献も検索できるようになった。ただし、この機能を利用して検索できるためには、出版社が引用情報をCrossRefに登録している必要がある。CrossRefのWebページによれば⁷⁾ 2006年3月7日現在、Forward Linkingに参加しているのは表1の34機関である。出版社は自分が出版している論文の引用情報しか登録できないので、この表に含まれない例えばElsevier社のジャーナルからの引用などは検索できない。Forward Linking機能を使った検索も、自分が出版している論文を引用している

図5 被引用リンクのしくみ

表1 CrossRef Forward Linking
参加出版社一覧 (2006年3月時点)

American Geophysical Union (AGU)
 American Institute of Physics (AIP)
 American Meteorological Society*
 Ammons Scientific
 Annual Reviews
 Ashley Publications*
 Baywood Publishing Co., Inc.
 Biomed
 Blackwell Publishing*
 Cambridge University Press
 CFA Institute
 CSIRO Publishing
 Ecological Society of America
 Emerald
 Future Drugs
 Future Medicine Ltd
 Geological Society of America*
 Hindawi
 Institute of Organic Chemistry & Biochemistry
 Institute of Physics Publishing (IOP)
 Japan Science and Technology Agency (JST)*
 John Wiley & Sons Inc.
 Lawrence Erlbaum Associates
 Mary Ann Liebert*
 Monash University ePress
 National Institute of Informatics
 Nature Publishing Group
 Professional Engineering Publishing Limited
 Public Library of Science*
 Research Foundation of the CFA Institute
 Royal Society of Chemistry*
 S. Karger AG
 Springer SBM
 Wolters Kluwer Health

(*を付けたものは自サイトで被引用リンクを行っている)

論文しか検索できない。この表に含まれている機関が発行しているジャーナルであってもすべてのジャーナルで引用情報をCrossRefに預け入れているかどうかは不明である。従って、この被引用リンクは網羅的とは言えないことに注意する必要がある。とはいえ、このサービスに参加する出版社も増えており(2005年3月時点では約20機関であっ

図6 J-STAGEにおける被引用リンク表示

た), 引用情報も時間を追うごとに増加していくものと思われる。

J-STAGEに登載されたCrossRef参加学協会のジャーナルの論文については, CrossRef Forward Linking機能を利用し, 引用論文を検索している。ここで得られた引用－被引用の関係もJSTリンクセンターのテーブルに登録する。

3.3 被引用リンクの表示例

図6にJ-STAGEのある論文の書誌抄録画面を示す。ここに見られるJST Link Center Citedボタンが被引用リンクである。このボタンは引用している論文が見つかった時にのみ表示される。このボタンをクリックすると被引用文献リストのページへ移動する。一番上に元論文の書誌情報が表示され, その下に元論文を引用している論文の書誌情報とその論文へのリンクのリストが並ぶ。右下に表示される「Linked via」は引用－被引用関係をどこから取得したかを示す。J-STAGEロゴが表示される場合は, JST リンクセンター内で取得したことを示し, CrossRefロゴはCrossRef Forward Linking機能により取得したことを示す。

3.4 被引用リンク統計データ

2005年5月にCrossRef Forward Linking機能を利用した被引用リンク機能の提供を開始してからほぼ

1年となる。J-STAGE登載論文の被引用リンクの付加状況を以下に報告する。図7はJ-STAGE登載ジャーナルの論文数とJST Link Center Citedボタンが付いた(つまり, 他論文から引用された)論文の数のグラフである。論文数の増加とともにボタンの数も増加しているが2005年4月から5月の間に急激な伸びが見られる。これはCrossRef Forward Linking対応によりJ-STAGE外のジャーナルからの引用が多数見つかったことを示している。図8はJ-STAGE登載ジャーナルの論文から他出版社への被引用リンク数を示したものである。前述の通り, このリンク数は他出版社がどの程度自社が出版しているジャーナル論文の引用文献をCrossRefに預け入れしているかによって変化するのでそのまま全被引用数と読むことはできないことに注意が必要である。

3.5 被引用リンクにおける今後の課題

いくつかの課題を以下に挙げる。

(1) CrossRef Forward Linkingへの出版社の参加状況

3.2.でも述べたが, より多くの引用－被引用関係を取得するには多数の出版社がCrossRef Forward Linkingサービスを利用することが必要である。

(2) 引用文献解析精度

通常の引用文献リンクにおいても同様であるが, より正確なリンクのためには引用文献記述の分析

図7 J-STAGE掲載論文数と被引用論文数
(1つでも引用されれば1論文と数える)

図8 J-STAGE論文からの被引用リンク先出版社 (2006年2月末)

精度を高めることが必須である。J-STAGEでは、引用文献記述を自動解析して、ジャーナル名、著者名、巻番号、開始ページ番号などの書誌情報を自動的に取得することをかねてより行っているが、記述によっては、資料名などの特定に失敗することがある。現在、J-STAGE全体ではほぼ7割から8割程度の精度で自動解析によりリンク先が取得されているが、文献確定精度が低いジャーナルにつ

いて原因の検討を行っている。CrossRefにおいても最近、出版社が預け入れした情報の品質向上にかなり力を入れている。著者名、ジャーナル名などの項目ごとに分けられていない平文の引用文献記述を自動的に解析してDOI検索が行えるサービスも最近開始した⁸⁾。J-STAGEでこのサービスを利用するかはこれからの検討である。

(3) 被引用アラート機能

自分の書いた論文が他の研究者から引用されるということは論文著者にとって、興味のあることである。現在は、My J-STAGE機能の拡張として提供予定である。My J-STAGEは2004年1月から提供を開始した機能³⁾で、ユーザがよく閲覧する資料や検索条件、ブックマークしておきたい論文などをシステムに登録することができる。この機能と連動してあらかじめ登録しておいたJ-STAGE掲載論文が他論文から引用された場合にそのことを電子メールで通知する機能を開発中である。今後はJ-STAGEの投稿システム⁹⁾と連動させ、著者はいち

いち登録作業を行わなくとも自分が投稿し掲載された論文が他論文から引用された場合に電子メールが送付されるようにする機能拡張も検討中である。

おわりに

本記事は第2回「情報プロフェッショナルシンポジウム」(2005. 11. 10-11, 東京) における発表「JSTリンクセンターの新機能—J-STAGEとGoogleの連携と電子ジャーナルにおける論文の引用・被引用関係表示」に加筆したものである。

参考文献

- 1) 佐藤恵子. JSTリンクセンター. 情報の科学と技術. Vol.53, No.4, (2003), p.211-215.
- 2) 吉田幸二; 時実象一; 尾身朝子. J-STAGE:「科学技術情報発信・流通総合システム」電子ジャーナル作成とインターネットによる流通. 情報管理. Vol.42, No.8, (1999), p.682-693.
- 3) 植松利晃. J-STAGEの現状とJ-STAGE2の開発. 情報管理. Vol.46, No.8, (2003), p.536-545.
- 4) Page, L; Brin, S. "The anatomy of a large-scale hypertextual web search engine". Proceedings of the 7th World-Wide Web Conference, 1998. (online), available from <<http://www.cis.upenn.edu/~yicn/cse591f05/anatomy.pdf>>, (accessed 2006-03-10).
- 5) Google Scholar BETA. (online), available from <<http://scholar.google.com>> (accessed 2006-03-10).
- 6) 内田尚子. デジタル・オブジェクト識別子 (DOI). 情報管理. Vol.42, No.1, (1999), p.32-46.
- 7) CrossRef forward linking. (online), available from <http://www.crossref.org/02publishers/forward_linking_howto.html>, (accessed 2006-03-10).
- 8) CrossRef Press release. "CROSSREF LAUNCHES FREE-TEXT QUERY TRIAL SERVICE". (online), available from <<http://www.crossref.org/01company/pr/press021406.htm>>, (accessed 2006-03-10).
- 9) 荒川紀子; 和田光俊; 時実象一. J-STAGE投稿審査システム—電子投稿・電子査読システムの現状と課題. 第2回情報プロフェッショナルシンポジウム予稿集. (2005), p.95-99.