

# 学術論文の電子化の技術とコンクリート工学における現状

多田眞作

## 概要

研究論文の電子化について、その制作に必要な基礎技術を解説した。さらに電子化されてネット上で流通する電子論文がどのように検索され、評価されているかについても、コンクリート工学分野の現状に即して概観した。電子化された論文には印刷体として公開される場合とは違った、いくつかの新しい性格がある。2003年から JCI が刊行している国際英文誌 *Journal of Advanced Concrete Technology* を例に紹介する。

## 1はじめに

研究者や技術者が技術文書、研究論文の作成やその情報の検索に費やす時間は多く、その合理化に関心を持つ方は少なくないであろう。情報処理インフラの整備に伴い、技術文書や研究論文の電子化とネット上での流通が急速に進展し、いまや製品マニュアルや図面の閲覧、論文の投稿と査読そしてこれらの電子化されたドキュメントの流通と検索はコンピュータ間で行われている。この変化は急速であるため、なかなかその本質を理解する時間がなく、次の変化への対応に不安がつきまとうのが実情である。

本稿は学術論文を対象を絞り、論文電子化の作成技術について基礎的かつ実践的な解説を試みた。我々が明日、投稿しようとしている論文の送り先ほとんどでは、すでに電子投稿システムが稼働している。送り先の学会や出版社が何を要求しているのかを知れば、オンライン投稿時の複雑なやり取りにも見通しが出てくる。電子論文の公開は論文を瓶に入れてネットの大海に流すことに似ている。誰がどのような手がかりで読むことになるのか、論文の評価はどのようになされるのか、等についてもその現状と予測される今後を展望した。

## 2電子論文の要素

### 2-1 マークアップ言語

論文の第1ページ目の上に、ひときわ大きな字で“Simulation of crack propagation and failure of concrete.”と書いてある。これは論文の表題だろうか。またその下には“Zaitsev, Y. B. and Wittmann, F. H.”とある。これはこの論文の著者だろうか。実際、このようなことで悩む人はまずいない。編集者は読者の常識を踏まえ、表題や著者名の位置やフォント工夫して画面を構成するからである。しかし、この様な理解が得られるための暗黙のルールをコンピュータに理解させるのは難しい。コンピュータが読み込んだ論文は、通常はフォントの属性が取り払われた単なるテキストである。第1ページ目の一番上の文字列が論文の表題だと教えたいところだが、ここにはヘッダとして雑誌名、巻、号、ページなどが来ることもある。そこで識別専用の記号(タグ)を使って明示的に示さねばならない。例えばそれは<title> Simulation of crack propagation and failure of concrete.</title>のような形式であり、<title>などをタグと呼んでいる。しかし、このまま読者に見てもらおうとなると、このタグはとても目障りである。従って別のところでこのタグにはこのスタイルという規則を作っておき、このタグ付き文書がその規則を実行するプログラムを通過した結果、一方では美しくレイアウトされた文書として出力され、一方では文書の構造が明確に規定された文書としてコンピュータのデータ処理に供されるというのが合理的である。この様な文書処理ができる一連の規則をマークアップ言語(Markup language)と呼んでいる。すなわち、マークアップ言語は電子文書の基本的な要件といえる。

現時点で代表的な電子文書の形式は次の5つである。

#### (1) SGML(Standard Generalized Markup Language)

1969年にIBMのプロジェクトを率いていたCharles Goldfarbらによって開発され、1986年にISO-8879として制定された電子文書を扱うための言語。日本では1992年にJIS化が行われ、製造業が避けて通れないCAL(S)(Continuous Acquisition and Lifecycle SupportあるいはCommerce At Light Speed)の電子文書規格である。SGMLは電子出版のための電子文書の原本としての性格を持っている。SGMLのソースからLaTeX、HTML、PDF、XML(後述)などの目的に応じた別の電子文書形式に変換できる。SGMLをよりフレキシブルにしたXMLは電子論文の形式として極めて重要である。

#### (2) LaTeX

TeXは当初から、印刷用のデータとなりうる電子文書作成プログラムとしてスタンフォード大学(当時)のDonald E. Knuth<sup>1</sup>により1980年に開発された。このため文字の位置決め分解能が約0.005 $\mu\text{m}$ といわれるほどで、専用のフォントを併用した数式の記述は特に優れている。数式の記述もソースはテキストなので、検索・置換の対象になりうるなど、この言語でなければ出来ない機能も多い。80年代中頃にDEC社(当時)のLeslie Lamportによって機能強化版のLaTeXが公開され、現在使用されているTeXのほとんどはLaTeXである。数物系の学会では投稿のファイル形式としてLaTeXが要求されることが多い。

### (3) HTML(Hyper Text Markup Language)

1990年に当時 CERN (欧州原子核共同研究所) に招聘されていた 英国人の Tim Berners-Lee<sup>2</sup>によって開発された。Web サーバが提供するデータを記述する言語で SGML のサブセットである。データの属性を表現法 (レイアウトなど) に絞っているのがシンプルであり、普及した。最近では WORD などの、論文を作成するためのアプリケーションの殆どがスタイル情報 (レイアウトやフォントの属性などの情報) を HTML に変換して出力できるので、HTML を直接記述する必要性は薄れた。電子論文の形式としては PDF に比べ表現力は劣る。

### (4) XML(Extensible Markup Language)

XML は SGML の包括性と HTML の簡潔性を備えて、Sun Microsystems の Jon Bosak<sup>3</sup>らによって 1997 年に登場した。SGML と同様、ユーザー定義のタグが使い、マークアップ言語を定義できるメタ言語として性格が継承されている。同時に、HTML に代わって Web の言語となるための要件が備わっている。SGML では手を抜かれていた数式の構造化が 1999 年から MathML として進められている。リンクについても、XLINK (<http://www.w3.org/TR/xlink/>) や XPOINTER (<http://www.w3.org/TR/WD-xptr>) という、新しい概念が導入されている。電子論文においては論文本体もさることながら、その中の参考文献、化合物名、実験データなどがネット上のデータベースにリンクすることにより、印刷物ではあり得なかった高度な利用が可能になることを考えるとその可能性を持つ XML の重要性が際だっている。

### (5)PDF(Portable Document Format)

Adobe 社が 1992 年に発表し、事実上の電子出版の標準となったファイル形式。Adobe 社は PDF 形式の文書を出力 (画面表示、印刷) するためのソフトを主要な OS 全てについて用意し、無料で配布した。このため、プラットフォーム (アプリケーションが稼働する環境) に制限されずにこの形式の文書の配信と閲覧が可能になった。画面表示、印刷ともにハードウェアの解像度に依存しない表示が可能で、ハイパーテキスト化や音声、動画の組み込みまでもサポートされている。高度な圧縮技術によりファイル容量は極めて小さい。従って Web サイトに置いてダウンロードさせたり、メールに添付して送ることが容易である。また PDF ではフォントの埋め込みが可能なので、自分のマシンとは異なる環境でフォントが置換されて表示される問題を避けることもできる。

前 4 つの文書形式はいずれもマークアップ言語である。この中で、SGML もしくは XML が電子文書の標準形式と云って良い。しかし、実際に標準的な電子文書の形式として、我々が良く目にするのは PDF である。本学会の年次大会用の論文の形式も PDF である。これは各 OS 間でのポータビリティに優れること、普段使用している文書作成ソフトから、PDF の出力が多面的にサポートされていること、商業印刷のデータの標準となっていること、画面やプリンタへの出力が容易かつ美しいこと等が理由となっている。その利害得失については後述する。

## 2-2 グラフィックス

電子ファイルとなった論文では、そのテキスト部分が上記のマークアップ言語で記述されたとしても、図、表、数式、参考文献などは別の処理が必要となることが多い。コンピュータ上で画像がどのように扱われるかを知るために画像ファイルフォーマットについて見ておこう。

画像ファイルフォーマットとは、画像をディスクに保存する際に用いられるデータ構造である。画像ファイルフォーマットの種類に応じて、そのデータを画像処理ソフトやレイアウトソフトが読み込めるかが決まり、また画像の処理時間や結果にも影響がある。とりあえずビットマップグラフィックスとベクトルグラフィックスを区別しておくと、コンピュータ上の画像の性格を理解しやすい。

#### (1)ビットマップとベクトル

ビットマップグラフィックスとはモザイク (ピクセル) で絵を構築することに似ている。画素 (Picture cell: Pixel) 毎に輝度やカラー等をコントロール可能だが、画像化された対象を全体として構成するための情報を一つの画素は知るよしもない。またビットマップの解像度がそのまま出力の解像度となる。

ベクトルグラフィックスは図形のデータが基準位置と曲線の関数によって保存されている一種のプログラムであり、出力装置 (プリンタやディスプレイなど) の解像度に合わせてラスターライズできる。ラスターライズとはデータを人間の目に見える点の集まりに変換することである。その結果、PDF の中の図形の様に、拡大しても次々に滑らかな曲線が、画面や印刷結果に現れる。後で述べるように、論文にグラフや挿絵を貼り込む際に、論文の著者は極力このベクトルグラフィックスを維持することに留意しなければならない。

#### (2) 画像ファイルフォーマット

画像ファイルフォーマットは特殊なものも含めると実に多様だが、動画フォーマットを除けば、主なものは次の 5 つである

- ・ TIFF

高解像度のカラービットマップで、作られた画像よりはスキャンされた画像に用いられDTP（コンピュータによる印刷データの前処理）に多用されている。代表的な画像処理ソフトはTIFFを標準のファイルフォーマットにしている。TIFFは当初から高解像度のグレースケール画像（階調を持ったモノクロ画像）の標準フォーマットであり、グレースケールもしくはカラー画像を扱う場合に現在もっとも信頼性が高い。TIFFのメリットは各OSで共通であること、グレースケールマップがアプリケーションにより変更できること、デメリットはビットマップであるため拡大には限界があることと、保存した際の容量が大きいことであろう。

- ・ JPEG

おそらくインターネット上で最も数多く存在する画像ファイル形式で、本来は静止画データの圧縮方式である。高圧縮性とフルカラーのサポートによりWebの標準画像ファイル形式となった。カラービットマップの品質はTIFFに及ばない。このため商業印刷や画像解析のためのフォーマットとしてはあまり用いられない。

- ・ EPS

画像をPostScript（Adobe社が提唱し、印刷データの標準となったページ記述言語）のベクトルで保存する。EPSにはASCII（8ビット単位で文字などを表現するデータ形式）とカラービットマップを保存するバイナリ（ASCII以外のデータ形式）のものがある。PostScriptファイルをラスタライズできないパソコンの画面で中身を確認できるように、低解像度のスクリーンプレビュー用のビットマップ画像が高解像度のPostScript印刷用データ（ASCIIまたはバイナリ）とセットになっている（後述の事例3を参照）。このプレビュー画像はレイアウトにおけるアタリをとるためのもので精密な位置決めにはならない。この形式のファイルを画像として読み込むことができるのはIllustratorやPhotoshopなどで、プログラム自体がASCIIファイルのものはエディタで加工することができる。EPSは通常の科学的画像処理では用いられることはないが、画像の印刷用データの業界標準であるため、論文作成の際には特に重量な役割を果たす。

- ・ WMF

線画、フォント等のベクトル、写真などのビットマップを混在させて表示・保存できるWindows OSのネイティブ・フォーマット

- ・ PDF

PostScriptは1985年に登場し、すぐに印刷データの標準形式となっていた。このAdobe社製品の根幹を成す図形記述言語をあらゆるプラットフォームでラスタライズできれば、標準の電子文書形式を提案できることになる、とAdobe社は考えたのであろう。PDFの優位性が際だつのは、フォントの埋め込みである。この機能を使うことにより、そのPDF文書はどのような環境でも元の文書に使用したフォントによる出力が可能となる。フリーで流通しているAdobe Reader 8.0はPDFファイルを画面上で閲覧したり、印刷したりする為の専用ソフトである。PDFを生成・編集するソフトがAcrobatで、後者は有償の商用ソフトである。

[事例1]埋め込みを忘れて問題となるフォントは主に2バイト文字である。日本人がWORDで書いている場合、文中のギリシア文字、℃などの記号にMS明朝などの2バイトフォントを無意識に使用してしまう。そのまま国際会議にPDFで投稿すると、先方からファイルが開かない等のクレームを頂戴することになる。フォントをきちんと埋め込めるかどうかは、PDF生成時の設定にかかっている。AcrobatをはじめとするPDF作成ソフトは、必ずその様な設定ができるようになっている。また、さらに、次の様なチェックをする。できたPDFをAcrobatで開いて、ファイルメニューから文書のプロパティを選択する。その中の項目「フォント」を選択すると、そのPDFで使用されているフォントの種類と埋め込み状況が分かる。もし埋め込まれていないフォントがあったら、文書本体に戻り、それらしいフォントをまず発見する。現在のAcrobatには特定のフォントが使用されている位置を検出する機能がないので、手作業となる。問題のフォントをTouchUpテキストツールで選択し、マウスの右クリック（またはControl+クリック）してコンテキストメニューを出す。「プロパティ」を選択するとダイアログが現れ、そのフォント名が表示されている。その表示の近くに、「埋め込み」、「サブセット」などのチェックボックスがあるので、それらをチェックしておく。この操作によって、埋め込まれていなかったフォントが確実に埋め込まれる。

### (3) 写真の解像度

最近国際会議の論文でも電子化ファイルの提出が多くなった。その場合、著者が準備するWord等の原稿に、解像度が350dpiの写真を貼りこんでおけば写真の解像度の問題は余り著者が気にする必要はない。しかし、現在でもcamera-ready manuscriptの提出を要求されることがまだある。camera-ready manuscriptというのは版下原稿のことで、元来はオフセット印刷の製版カメラの工程にすぐにも渡せる状態の原稿のことである。著者が論文の版下を準備するためには、写真を網点処理しなければならない。手持ちの銀塩写真を納得のいくモノクロの印刷物にするために、著者はどの様にスキャナを設定すればよいのだろうか。

印刷された写真は網点で構成されている。この点の大きさは画像の明るさに応じて大小があり、見えない格

子の中に点が入っている。この場合、印刷された写真の見かけの解像度とも云うべきスクリーン線数(見えない格子の大きさ、単位: lpi)を自分で決めねばならない。網点処理とは印刷物において擬似的な階調(グレイスケール)を得るための処理であり、スクリーン線数、階調数と出力機の解像度との間には関数関係がある。まず階調数を検討すると、人間が見てもっともらしい階調は最低 36 階調、まあまあ見られるのが 64 階調、普通のモノクロ印刷物が 256 階調と云われている。

網点処理ではたとえば 64 階調を出すのに小さな点  $8 \times 8$  個からなる正方形の大きな点をきめ、小さな点を 1 個減らす毎に階調が 1 つ減るといふ様に考える。これをディザ(dithering)と呼んでいる(図 1)。

従って出力機の解像度は 64 段階の擬似階調を出すために見かけ上  $1/8$  にされてしまうわけで、もし 600dpi のレーザープリンタを使っていれば、 $600/8=75$  が lpi と言うことになる。一方、lpi は印刷される用紙の品質に依存し、週刊誌等の紙で 80、書籍などの上質紙で 100、コート紙などで 120 が良いと云われる(これは水を使うオフセット印刷でのことなので乾式のトナーを使う場合はまた異なるだろう)。

ともあれ、600dpi のレーザープリンタに写真を出力すればまあまあ見られる程度に出せると云うことになる。さらに良い品質を追求して、階調数  $G$  を上げる、スクリーン線数  $L$  を上げるということになればレーザープリンタの解像度  $R$  を上げるほかに手はない。以上をまとめれば次のようになる。

$$L = \frac{R}{\sqrt{G}} \quad (1)$$

$$120 \geq L \geq 80, 256 \geq G \geq 64$$

これらの関係を図 2 に示す。

一般ユーザー用のレーザープリンタではスクリーン線数と階調数はプリンタ・ドライバが自動で決めているようである。また、基本的に最終印刷結果の lpi より大きな解像度で元の写真を取り込んででも無駄になる。従って論文の版下原稿に使用する写真を、著者自らがパソコンに取り込んで版下原稿を作成する場合、その取り込みの解像度はモノクロ写真なら 144dpi 程度で十分ということになる。

## 2-3 数式と参考文献

数式と文中のシンボルであるが、数式は LaTeX か MathType (MS-Word の数式エディタのスーパーセット) であれば基本的に問題ない。ただ、MathType では、分数の横棒の太さが問題となることがある。太さの指定はパーセントやポイントなどが選べるが、これを 0.5pt とするつもりが誤って 0.5% になっていると、最終印刷では消えてしまうことがある。

英語論文の場合、文中のシンボルに関して最も推奨されるのは Arial、Times New Roman、Symbol 等の 1 バイトのローマン・フォントの使用に徹することである。Arial や Times New Roman にも  $^{\circ}$  の「 $^{\circ}$ 」や  $\pm$ 、 $\partial$  はあるし、Symbol フォントであればほとんどの記号とギリシア文字が用意されている。後述のように電子投稿・査読システムでは著者から投稿された Word 等のファイルからシステム側が PDF を作成する。間違いのない PDF が出来るためには、著者が使用したフォントと同じものがシステム側にも用意されていなくてはならない。もし MS 明朝などのフォントを使って  $^{\circ}$  や  $\pm$  を記述した場合に、問題を起こす可能性がある。

参考文献のリストは従来から間違いの多発地帯であった。ところが文献データベースと電子投稿によってこの問題は解消されつつある。書誌データが疑わしい参考文献を調べる場合、すぐに使える Google Scholar を用いて、表題の最初の幾つかの単語を "Simulation of crack propagation and failure" の様にクォーテーション・マークで囲って完全一致の検索を行う。Zaitsev と Wittmann の論文は瞬時に検索結果のリストの先頭に現れる。電子投稿システムでは、著者のアップロードした論文の文書構造を自動解析し、参考文献リストが自誌の書き方に則っているかだけでなく、その文献のネット上の所在まで追跡し、正しい書誌データが検討できるようにしてくれる。

## 3 電子論文の投稿

### 3-1 電子投稿・査読システム

電子論文の投稿から出版までの基本的な流れは、学会や出版社に依らずほぼ次の通りである。著者は標準的な論文作成ツール (MS-Word、LaTeX 等) で書き投稿する。学会事務局はこれを SGML や XML 等の文書形式に変換して保管する。オンライン公開用はこの元ファイルから適宜 HTML や PDF 形式に変換する。数式や図形などは部分的に LaTeX データや PostScript データを組み込む。現在提案されている学術論文電子出版システムを表 1 にまとめた。これらの元になっているものとして、「電子投稿規定作成のためのガイドライン」(SIST 14)<sup>4</sup>がある。また、国際的な規格では「雑誌論文の構造に関する ISO 規格」(ISO 12083 1994)があり、のちに JIS X 0804 1996 となっている。

この様に、電子投稿の際に著者自身が論文を書くために SGML を記述する必要はなく、標準的なワードプロセッサのファイルを投稿できるようになっている。RTF(テキスト及び画像の書式情報を含んだ汎用文書フォーマット)から SGML への変換ツールが幾つか存在するので、投稿を受けた学会や雑誌社側で変換を行ってくれることが多い。

タグの定義は DTD (Document Type Definition) と呼ばれている。タグとは文書の論理構造をマークアップするものである。論理構造とは言い方が難しいが、データベースの「フィールド」に該当し、論文なら、表題、著者、パラグラフといったデータの項目である。場合によっては、実験方法、結論、などの文章の意味にまで及ぶことも不可能ではない。従って、DTD は電子論文の記述に際して大きく関わってくるが、この場合も著者が意識する必要はあまり無い。標準的なワードプロセッサのテンプレートとなって供給されると思われる。以上の様に、電子投稿のための標準化は論文のテキスト以外のところではまだ課題を多く残している。

### 3-2 電子論文のグラフィックス

電子投稿の要求に対する著者の対応は出力の品質レベルに応じて異なったものとなる。論文が最終的に印刷体としても流通する場合には、テキスト以外の図形や写真の最終品質は、著者への依存性が最も高い。商業印刷は 2400dpi という解像度で出力される。このため、図に用いられていた細線（ヘアライン）が、著者のパソコンのスクリーンやレーザープリンタには問題なく出力されていたものが、極端に細くなるか消えてしまう。画面上では申し分ない写真が荒くなる、といった問題が発生する。

現在ではほとんどの電子投稿システムで図（Artwork と呼ばれることが多い）の投稿形式として PostScript が指定されている。PostScript はコンピュータが画像データを描画するための汎用的な方法である。その特徴は、ベクトルデータの中で 1 ポイント以下の太さの線、自由曲線、破線、回転文字などをサポートしている等、高品質のグラフィックスを可能にする仕様である。このため商業印刷のためのデータの標準となっており、投稿された論文は最終的に紙に印刷されるならば、論文中の図はこのデータ形式に準拠する必要がある。前述の様に、PostScript 形式のデータは EPS というファイル・フォーマットで保存される。EPS は通常はパソコンでは実行されず、データだけがプリンタに送られ、プリンタが EPS を解釈し、レンダリングして印刷する。このため PostScript 非対応のプリンタで印刷しようとすると、EPS ファイルに付属する解像度のビットマップ画像だけが印刷されてしまう。

グラフ等を作成した場合、そのグラフ書きアプリケーションが EPS をサポートする仕様であればそのメニューを選択するだけで PostScript 形式のデータを保存できる。Excel などのように保存できる形式に eps が無い場合は Adobe 社が無償で提供している PostScript プリンタ・ドライバをダウンロードし、出力先をプリンタではなくファイルとして PostScript ファイルを生成させる。これを Ghost Script などのツールを用いて eps ファイルを作成することになる。WORD や PowerPoint で描いた図に関しても同様である。描画したグラフにさらに手を加えてから eps で保存したい場合は、Adobe Illustrator、Canvas 等の PostScript のベクトルを編集できるアプリケーション・ソフトにグラフを持ち込んで、編集・保存を行う。但し、Windows 環境ではグラフィックスをコピーした時点で wmf 形式となるので、上記編集ソフトに正しく持ち込めるかどうかは分からない。編集ソフトが読み込める形式に一旦保存し、編集ソフトからそのファイルを読み込む必要がある。

[事例 2]2005 年ロンドンで開催されたある国際会議の要項に、最終論文の図の形式は PostScript とし、全体は WORD ファイルで提出と書かれていた。著者の提出したプリントをそのまま研究室のプリンタで印刷するのではなく、印刷所でデータとして扱うことが予想され、高品質の論文集ができそうだと期待しながら、WORD に eps 形式のグラフ等をインポートして提出した。しかしこの会議のセクレタリは出版社が云っていることをあまり理解せずに執筆要項に盛り込んだようで、集まった WORD ファイルを自分のプリンタで印刷して、これを印刷に使うよう指示をしたと思われる。しかもそのプリンタは PostScript プリンタではなかった。パソコンが送ってくる PostScript コマンドを解釈できぬまま、みすばらしい低解像度のビットマップ画像だけを印刷する結果となった。どうやら筆者以外は要項を無視していたらしく、悲惨な目にあった人間は一人で済んだことがせめてもの幸いであった。国際会議の事務局は意味を理解せぬままに執筆要項を作ることがあるという教訓である。

### 3-3 オンライン投稿・査読システムの利用

主要なセメント・コンクリート系学術雑誌で稼働している電子投稿システムは、表 2 に示す様に独自に開発したもの、商用アプリケーションを利用するものなど様々である。しかし、ここ数年で定評ある商用オンライン投稿・査読システムへの切替えが進展し、事実上 Editorial Manager か Manuscript Central だけになっている。Elsevier 社の Elsevier Editorial System も Editorial Manager をベースに開発され 2003 年から運用されているものである。

Web を利用した電子投稿・査読システムに共通する手順は最初の投稿時では 4 段階で、(1)Corresponding author となる人の登録、(2)カバー・レターの内容の入力、(3)本文、図表ファイルのアップロード、(4)システムが作成した PDF による内容確認と修正、の順になる。図に eps が要求されている場合でも、WORD に貼り込んだ形のベクトル図形もしくは高解像度ビットマップ画像であれば受けられる。また表は WORD の表が推奨されていることが多い。結局全ての内容を本文と数式、図、表に分けて WORD ファイルで準備しておけば電子投稿の準備はほぼ完了である。雑誌によっては独自の要求項目をシステムに付加している。著者からみて査読者として相応しい研究者または不適当な研究者は誰かを聞かれたり、投稿する価値がどこにあるのかを簡潔に書くように言われる場合もある。

[事例 3]

Editorial Manager を例にとり、具体的な投稿プロセスを説明する。

最初の頁で Register メニューから情報を入れ ID と Password の発行を依頼し、メールで受け取る。その後は Author としてログ・インできる。ログ・インすると、新規の投稿か、論文の修正に関わる場合か、最終投稿かによって分岐する。以下は新規の投稿につき説明する。図 3(a)

論文タイプを選択する。その後、表題、各著者、所属、アブストラクト、キーワード、論文の扱う分野など、一昔前の投稿におけるカバー・レターの内容に相当する情報を入力させられる。ログ・インしたユーザがデフォルトで First Author となるが、順序は変更可能である。

次にアップロードするファイルを選択する。選択の後 Attach This File のボタンを押すことを忘れない様にする。本文、図、表、等アイテムごとに順番にアップロードリストに入ってゆくが、Web 画面でその表が完成されて行くので分かりやすい。その後、これらのファイルをもとにシステムが PDF を作成する。

ここで一連の作業は一段落し、翌日に回すことも可能になる。PDF 作成には問題がある場合は数分を要する。PDF 作成が完了すると画面上では Building a PDF が Needs Approval に変わる(図 3(b))。また、親切にも、Corresponding author 宛のメールでも通知してくれる。Action Links から View submission を選択すると PDF がダウンロードされてくる。

問題のあるファイルは PDF に変換されず、その理由も PDF に記されている。その問題を解決したら、Action Links から Edit submission を選択する。先ほどのプロセスが繰返される中で、問題解決後のファイルを選択し直して再度 PDF を Build する。先ほどの問題のあるファイルも残っているので削除を忘れない様にする。有難いのは参考文献の Validation まで自動で行い、著者自身の間違いをチェックしてくれる。ACT の論文を引用している場合は CrossRef 経由で J-Stage 中の ACT の文献情報も表示し、著者の確認に供される。

PDF の内容をすべて確認し、了承すれば事務局に送られる。Corresponding author がログ・インすると、自分の仕掛かり中の論文が区別できるようになっているので特に論文番号は発行されない。

## 4 電子論文の公開と評価

### 4-1 文献データベース

#### (1) 文献検索のパラダイム変化

先行する研究を調べるために、1970 年代の研究者は図書館の書庫に入り、金箔の文字が押され合本された雑誌の過去号を片っ端からめくっていた。この時代、通常は定期購読している定番雑誌をチェックすれば十分だったようで、ソファでめくり慣れた雑誌をブラウジングするのはリラックスした時間であったに違いない。

1980 年代には文献データベースがオンラインで公開され、コンピュータ検索が普及した。論文は書誌事項(著者名、論文名、掲載誌名、巻、号、頁、刊行年、アブストラクトなど)だけが部分的に電子化され、有料で公開された。文献探しは 1 分数百円という課金を意識しながら、しかも目の前に見えない情報をコマンドだけで手繰り寄せねばならない緊迫した時間変わった。筆者も当時やっとな漢字で抄録が読めるようになった国産の科学技術文献データベースの JOIS (現在は JDreamII) や、当時世界最大のオンライン文献データベースであった Dialog を使っていた。通信速度が 300bps という信じられない低速の音響カプラの付いた端末と、KDD の国際デジタル回線がまだ ICAS と呼ばれていた実験回線を利用してアクセスしていた。それでもわずか十数秒で研究室の自分の机の上から、パロ・アルト (現在はミシガン州イーガン) にある大型コンピュータの中に入ってしまったのには仰天した。そして自分の指先一つで広大な知識の森を歩き回れるのを知ったときには胸の高鳴りを覚えたものだ。その興奮は二ヵ月後に請求書を見て胸の鼓動がほとんど停止するときまで続いたのである。

1990 年代には世界中のコンピュータがネットワークで繋がり、高速で通信を始めた。データベース化されていない情報までもが容易に公開できる様になり、文献を探すことは書誌事項だけでなく、全文の入手が期待出来る。しかも、その文献の多くは無料である。そしてインターネット上の Web サイトを検索し、閲覧する速度が、自分のパソコンのハードディスクの内部を検索することと体感的には何ら違いがなくなった。この

ため、必要に応じてその都度検索し、それを利用すれば目的は達せられる。必要なのはその情報に到達する手順だけである。閲覧から検索へ、所有からリンクへというパラダイム変化が見て取れる。

### (2) 文献データベースへの登録

書誌事項だけを提供する文献データベースは現在でも依然利用されているが、従来の様に人手による入力は近年の外注先の低賃金国のコスト上昇に伴って困難となってきた。そこで前述のように SGML や XML で記述された論文本体のデータが2次利用され、書誌事項を自動的に切り分けて文献データベースを構築することが期待されている。全ての論文が電子化されるまでは、紙の論文誌を見ながらの入力は続くことになるが、人手によるものだけに見落としも多い。日本語化されている最大の文献データベース JDreamII を例にとると、JCI の「コンクリート工学論文集」の第1巻はこのデータベースに収録されていない。先方が新しく刊行された論文誌に気づくのが遅れたためと思われる。

このため ACT は創刊当初から各データベース制作機関に接触し、献本を行っている。現在 JDreamII、Compendex、Chemical Abstracts などの主要なデータベースに書誌データを、大英図書館のようなハードコピーによる全文提供機関、J-Stage のような電子論文レポジトリには全文を提供している。J-Stage 経由で CrossRef、ChemPort などの電子論文レポジトリ間にはリンクが張られ、Google Scholar にもデータが提供されている。また 2007 年刊行分から Science Citation Index Expanded-SCIE への収録が開始された。SCIE は電子論文本体を所有しない書誌事項だけのデータベースであるが、収録する論文の引用文献も併せて収録している。このため後述するように、ある論文の被引用情報を割り出すことができる。その結果、学術誌の重要性の評価値であるインパクト・ファクターの算定が可能になっている。ACT のインパクト・ファクターは 2009 年分が 2010 年夏に公開される予定である。

### (3) 電子論文の全文提供

文献情報のデータベースによって検索された論文は、従来は所属機関の図書館などからそのコピーを紙媒体で入手していた。しかしこの数年間に、特に学会誌においてフリー・アクセスの考えが広まり電子化された論文全文の無料公開が進展した。国内のセメント・コンクリート系学会誌の全文提供の状況を表 3 にまとめた。提供される電子論文は全てが PDF 形式である。ただし、PDF 内の全文検索が可能かどうかで使い勝手は大きく異なる。MS-Word 等のデータから作成された PDF は文字情報を維持しているため、その PDF ファイルを開いているかどうかに関わらず、また自分のパソコン内か、ネット上のどこかに存在するかに関わらず、全文検索が可能になる。文献の PDF ファイルのファイル名を書き換えたり、キーワードを付与したりすることが実際には時間的に難しい以上、ファイルとしては未整理のままその都度検索して使用できるのは有難い。各学会誌のうち、論文内検索が可能な PDF を提供しているのは、最近の電子化技術によってデータから PDF を作成しているものであることが分かる。一方、過去に刊行された論文を PDF 化する際には、スキャナでビットマップ画像として取り込んだものになるので、通常は論文内の検索はできない。しかし、スキャニングと同時に光学的文字認識を行って文字情報を取得し、画像の背後に文字情報を埋め込んで検索を可能にする方法が一般的になってきた。最近ではどの職場でも手持ちの資料が紙媒体より PDF が量的に半数を超えるようになり、一気に過去の紙媒体の資料の PDF 化が進むと考えられており、そのための PDF 作成に特化した低価格のスキャナが市販されている。付属のソフトには光学的文字認識を行って検索可能な PDF を作成できるものもあるので、今後論文内検索が可能な PDF はさらに増えてゆくだろう。

紙のコピーを提供するだけのセメント・コンクリート論文集が、なぜ電子論文の全文提供といえるのかには理由がある。セメント協会研究所は、実は他誌に先駆けてセメント技術年報の過去の論文を PDF 化し、独自の検索システムによるサービスを試験的に提供していた。筆者はこれを高く評価しているが、現状のサービス形態が実情に合わせて改善されることを期待したい。

## 4-2 評価手法

### (1) インパクト・ファクター

論文の評価については、現在利用可能な定量的な手法として Thomson-Reuter 社 (ISI 社が開始し展開させたものを買収) が独占的に行う格付け手法「インパクト・ファクター」が代表的である。これは被引用文献数を元に学術雑誌を格付けする方法であり<sup>5</sup>、その有効性も検証されている<sup>6</sup>。

ある学術雑誌  $v_i$  のある年  $t$  のインパクト・ファクター (Impact Factor: IF) は、先立つ 2 年間に対象とし、その 2 年間の  $v_i$  誌の掲載論文を引用した学術雑誌  $v_j$  の論文数  $c(v_j, t)$  を求め ( $v_i$  誌の自己引用も含めている)、これをその 2 年間の雑誌  $v_i$  の掲載論文数  $n(v_i)$  で除したものである。すなわち、

$$IF(v_i, t) = \frac{\sum_j c(v_j, v_i, t)}{n(v_i)} \quad (2)$$

つまり、 $v_i$  誌が登載した論文 1 報当たり、他の何報の論文に引用されたか、という意味を持っており、その雑誌の学術的な重要性を示すものとされている。一方、有名誌に登載された論文またはその著者の格付けが間接的に行われる結果にもなっている。Thomson-Reuter 社の文献データベース Science Citation Index Expanded-SCIE を使用すれば、全く同じ手法で研究者個人のインパクト・ファクターに相当する評価値が計算できるはずであるが、現実には難しい。これは同じ分野で論文を発表する同姓同名の人間が区別できないためである。Thomson-Reuter 社がある論文を SCIE に登録する際に、必ず所属期間名を登録したり、著者名をフルネームで登録したりするだけでは解決できない。問題は、その論文が引用している参考文献である。参考文献の書誌事項には、通常所属機関は書かれていないし、著者名をフルネームで書かない論文誌も少なくない。この様な事情も論文の個別評価もしくは研究者個人の評価ではなく、雑誌の評価が行われる理由となっている。表 2 にコンクリート系学術雑誌の IF を示した。

## (2) 被引用の検索

引用情報を有する文献データベースは SCIE に限らず、CiNii (国立情報学研究所：日本語論文主体、無料)、Scopus (主に Elsevier 社の学術誌掲載論文、有料)、Google Scholar (無料) などがあるが、信頼性・公平性において SCIE にまだ及ばない。

SCIE を使う方法は幾つかあるが、多数の商用データベースを統一的なコマンドで検索させるデータベース検索サービス提供者を経由してこのファイルを検索する方法が一般的であろう。筆者が最初にこのデータベースを検索したのは 25 年前で Dialog 経由であった。昨年までは国立情報学研究所 (NII) が web で提供する CiNii において利用可能だったが現在は利用出来ない。このため所属機関が Web of Science のなどの Thomson-Reuter 社のサービスを使用契約する必要があり、一部の研究者のものとなっている。CiNii は未登録でも直ちに使うことができる。自分の論文を引用している文献の書誌の概要は分かるが、その詳細を知るには、やはり所属機関経由の登録が必要となる。登録後は IP アドレスで認証されるので所属機関内部からアクセスしなければならない。また、SCIE で出力される結果に比べ CiNii の結果は極端に少なく、信頼性に欠ける。そこで、直ちに自分の論文の引用論文を調べる必要がある場合には Google Scholar を使うのが良い。[事例 4]

トップページの検索窓の右に「Scholar 検索オプション」とあるので、これをクリックする。詳細検索条件が出るので、上から 6 番目の窓 (「次の著者の記事を検索する」と書いてある) に著者名を例えば「S Tada」と入れる。この場合、必ず半角のクォーテーションマークで囲う。そうでないと、他の Tada も検索されてしまう。さらに、これでも別の S. Tada が混じるので、例えば一番上の検索窓 (「すべてのキーワードを含む」と書いてある) に「concrete」等と入れて、範囲を限定する。以上がすんだら、右端の「scholar 検索」ボタンをクリックすると次の様な結果が出る。

**[引用] Microstructural approach to properties of moist cellular concrete**

**S Tada, S Nakano - Autoclaved Aerated Concrete, Moisture and Properties, ..., 1983**

[引用元 10 - 関連記事 - ウェブ検索](#)

(中略)

[Simultaneous determination of moisture capacity and water diffusivity of cement based materials.](#)

**S TADA, K WATANABE - ... Proceedings of Cement & Concrete (Japan Cement Association), 2001 - sciencelinks.jp**

TOP > J-EAST > List of Journal Titles (J) > JCA Proceedings of Cement & Concrete (Japan Cement Association)(2001) > Simultaneous determination of moisture capacity and water diffusivity of cement based materials. ... Journal Title; JCA Proceedings ...

[引用元 2 - 関連記事 - キャッシュ - ウェブ検索](#)

[Dynamic determination of sorption isotherm of cement based materials - 全 2 ページョン](#)

**S Tada, K Watanabe - Cement and Concrete Research, 2005 - Elsevier**

... Cement and Concrete Research Volume 35, Issue 12, December 2005, Pages 2271-2277, ...

Specimens 3.2. Test apparatus 3.3. Automatic measurement software 4. Results and discussion 4.1. Test results 4.1.1. Autoclaved aerated concrete 4.1.2. Cement ...

[引用元 1 - 関連記事 - ウェブ検索](#)

必要なものは、各レコードの一番下の行の[引用元](#)というところで、ここをクリックするとこの論文を引用した論文レコードを見ることができる。



SCIEの結果との違いは、SCIEがインパクト・ファクター対象誌8500誌のみを対象にして引用文献を表示するのに対し、Google ScholarはWeb上にある全ての学術文献を対象にしており、国際会議提出論文などもカウントする点である。またGoogle Scholarは基本的に論文本体にアクセスできるものをデータベースに収録している。しかし論文本体へのアクセスの可能性がないものでも、引用された論文であれば、検索結果の一番目の論文にあるように[引用]というマークをつけて収録する。2番目3番目の論文は（有料かもしれないが）論文本体まで行き着ける可能性がある。

引用文献のデータベースへの収録は、SCIEが人手で行うのに対し、Google Scholarでは機械が行っていると見られる。Google本体と同様に、ネット上のテキストデータを自動的に収集するのであれば、引用文献の切り分けの信頼性は高くはない。しかしGoogle Scholarは論文本体の収集も積極的に行い、データベースの信頼性を高めている様である。Googleの圧倒的な露出効果の見返りに、多くの電子論文収集機関（レポジトリ）や商業出版社までもが、（Google自体は原文を提供しないという約束のもとに）電子論文や電子書籍本体をGoogleに提供している。日本を代表するレポジトリのJ-Stageも、加盟学会の許可を得てGoogle Scholarに対する全文公開を行っている<sup>7</sup>。

### (3) 電子論文-個別評価の可能性

論文の個別評価をする場合にも、その論文が別の論文に何度引用されたかという、被引用件数が評価の尺度になると思われる。では電子化された論文の被引用情報を系統的に知ることが可能であろうか。

すでに見た様に、電子化された研究論文はPDF形式によりインターネットに公開されることが事実上の標準となっている。Googleに代表されるインターネット上の検索エンジンはWebページにあるテキストだけでなく、PDFファイルの内部からもテキストファイルを収集し、サーバ内に「単語」-「URL」の膨大なインデックスを作成してユーザの検索に供している。しかし、前述の様に、電子論文がPDFや現行のWebページ記述言語であるhtmlで記述される限り、コンピュータにはそれが論文であることが理解できず、その論文の一部が、他の論文を引用している引用文献であることも決して分からない。電子論文が冒頭に述べたマークアップ言語によって記述され、論文の中の引用文献も適切にマークアップされていることが、どうしても必要になる。その上で、これらの電子論文から引用情報を割り出して論文の個別評価、ひいては個人のインパクト・ファクターを算定することが可能になるであろう。検索された電子論文に現時点での被引用情報が付属するようなイメージである。

ここで重要な点は、必ずしも学術雑誌に掲載された論文である必要がないという点である。査読の制度は、学術雑誌の限られた紙面と時間の中で、品質の悪い論文を早期に判断して排除しようという要求から生まれたと思われる。しかし、本来限られた数人の査読者ではなく、世界中の研究者による時間をかけた評価が理想であることはいうまでもない。これまでの学術雑誌の形態では、それは明らかに不可能だったが、電子論文においては決して不可能なことではない。無審査でネットに公開された論文が、世界中のかけ出しからベテランまでのあらゆる研究者によって良いものは引用され、そうでないものは無視され、時間も費用もかかることなく、被引用数によって評価されてゆくと考えられるからである。そしてその被引用情報はその電子論文に付随して流通し、優れた論文を自由に読むことができる。表4に学術雑誌と電子論文のそれぞれの役割を比較した。学会連合や機関レポジトリには、このような電子論文の著者認証、日付認証や永久保存などの役割が期待される。このことは学会、所属機関の図書館等の将来的な役割を考える際に重要な点であろう。

## 5. おわりに

論文の電子化について、その作成のための技術と電子化がもたらす影響について述べた。

論文の電子化は炭酸ガス排出量削減に大きく寄与する。ACTは電子版と印刷版の両方を発行しているので、電子化による環境保全効果は出ていないが、毎号余計なページ出ない様な編集上の努力を行っている。

最後に、論文の電子化は論文の個別評価に繋がる可能性がある。それはCDに収録されて販売されていた楽曲がネット上で個別ダウンロード販売が可能となり、楽曲個別の売上集計が可能になっていることに似ている。電子化がもたらす論文の個別評価は、学術雑誌と学会のあり方にも大きく影響するだろう。

表 1 学術論文電子出版システムの文書形式

システム名	DTD	入稿データ形式	システム内のデータ形式	Web 公開データ形式	組版用データ形式
化学系学会 1999	学術論文汎用 DTD (NACSIS) を一部変更	電子化原稿	SGML->LaTeX	PDF および HTML	LaTeX
科学技術振興機構 J-Stage1997	科学技術論文用 DTD	MS-Word (RTF) LaTeX	SGML	PDF および HTML	FrameMaker+ SGML
国立情報学研究所オンラインジャーナル	学術論文用 DTD	MS-Word, LaTeX, XML	XML	PDF, XML, HTML	PostScript
日本コンクリート工学協会 ACT 2003	J-Stage に倣う	MS-Word LaTeX	SGML	PDF および HTML	PostScript (PDF)

表 2 主要なセメント・コンクリート系学術雑誌の電子投稿システムとインパクト・ファクター

雑誌名	編集長/編集機関	出版元	刊行頻度	電子投稿システム	IF (2007)
Cement and Concrete Research	Karen Scrivener Elsevier	Elsevier	12	Elsevier Editorial System	1.028
Cement and Concrete Composite	John E. Bolander Elsevier	Elsevier	10	Elsevier Editorial System	0.962
Construction and Building Materials	M C. Forde Elsevier	Elsevier	12	Elsevier Editorial System	0.841
Computers and Concrete	Chang-Koon Choi Techno-Press	Techno-Press	6	-	0.351
Magazine of Concrete Research	P C Hewlett Thomas Telford	Thomas Telford	10	Editorial Manager	0.317
Materials and Structures	Jacques Marchand RILEM	Springer	10	Editorial Manager	0.530
ACI Materials Journal	ACI	ACI	6	Manuscript Central	0.670
ACI Structural Journal	ACI	ACI	6	Manuscript Central	0.665
Journal of American Ceramics Society	Lisa Klein* American Ceramic Society	Blackwell	12	Manuscript Central	1.792
Journal of Materials in Civil Engineering	Antonio Nanni ASCE	ASCE	12	-	0.452
Journal of Engineering Mechanics	Ross B. Corotis ASCE	ASCE	12	Editorial Manager	0.787
Journal of Advanced Concrete Technology	Hirozo Mihashi JCI	JCI	3	-	2009 より

\*4 人の編集長の中のセメント系材料担当

表 3 セメント・コンクリート系の全文提供国内データベース

提供機関	データベース名	公開先	収録範囲	無料公開	ファイル形式	PDF の品質
JCI	年次論文集	JCI-Web	1979～	2001 年以前	PDF	高解像度ビットマップ (検索不可)
	ACT	J-Stage	2003～	1 年前まで	PDF	論文内検索可能
AIJ*	日本建築学会構造系論文集	CiNii	1936～	1 年前まで	PDF	低解像度ビットマップ (検索不可)
		eAINA	1995～	1 年前まで	PDF	低解像度ビットマップ (検索不可)
	J. Asian Arch. & Build. Eng.	J-Stage	2002～	1 年前まで	PDF	論文内検索可能
	日本建築学会技術報告集	CiNii	1995～2007	1 年前まで	PDF	論文内検索可能
J-Stage	2007～					
JSCE	土木学会論文集 E	J-Stage	2006～	1 年前まで	PDF	論文内検索可能
CERSJ	日本セラミックス協会学術論文誌 (2008 年より英文表記のみ)	J-Stage	2003～	現在より	PDF	高解像度ベクトル (検索不可)
JACASSOC	セメント・コンクリート論文集	論文 OPAC	1968～2006	有料	印刷体	紙コピー (検索不可)

\*年次大会梗概集もオンライン化されているが省略

表 4 論文の個別評価における学術雑誌の印刷論文と電子論文の役割

	情報の配信	質の確保	評価基準	被引用情報の登録	所在	日付認証	永久保存	著作権	著者認証
学術雑誌	郵送	査読	雑誌の IF	人手	雑誌自体	雑誌自体	図書館	出版社	なし
電子論文	ネット	自然淘汰	個別論文の IF	都度検索	被引用	学会、 機関レポジトリ	学会、 機関レポジトリ	著者	電子認証

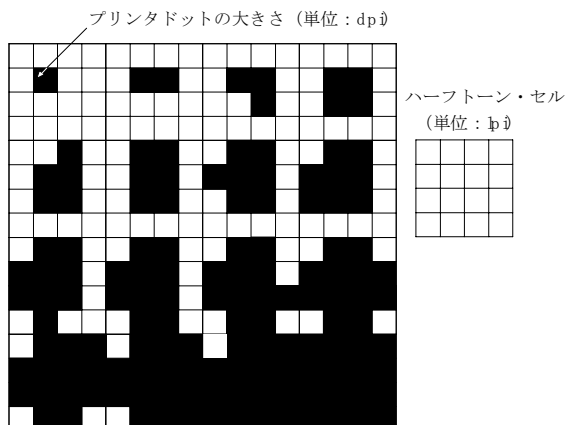


図 1 疑似階調生成の原理 (16 階調の場合)

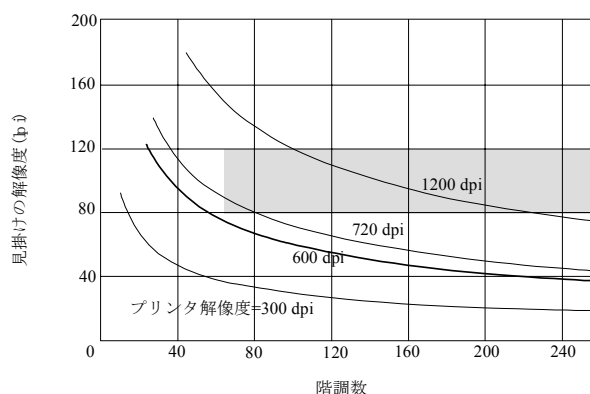
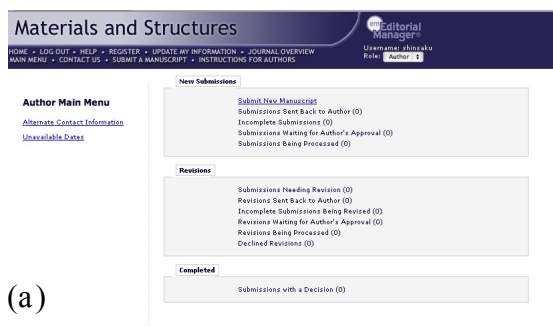
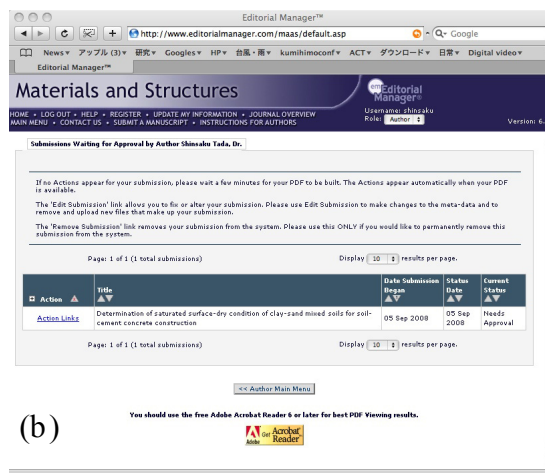


図 2 階調数、プリンタ解像度および見掛けの解像度の関係



(a)



(b)

図 3 電子投稿・査読システムの例 (Editorial Manager が稼働する RILEM の Materials and Structures)

<sup>1</sup> Donald E. Knuth, <http://www-cs-faculty.stanford.edu/~knuth/>

<sup>2</sup> Tim Berners-Lee, <http://www.w3.org/People/Berners-Lee/>

<sup>3</sup> Jon Bosak, 1997, XML, Java, and the future of the Web, <http://www.ibiblio.org/bosak/xml/why/xmlapps.htm>

<sup>4</sup> SIST 14, 2001 <http://sist-jst.jp/perusal/index.html>

<sup>5</sup> Johan Bollen, Marko A. Rodriguez, and Herbert Van de Sompel. Journal status. *Scientometrics*, 69(3), December 2006 (arxiv.org:cs.DL/0601030, DOI: 10.1007/s11192-006-0176-z)

<sup>6</sup> Somnath Saha, Impact factor: a valid measure of journal quality? *J Med Libr Assoc* 91(1) January 2003.

<sup>7</sup> 久保田 壮一, 荒川 紀子, 和田 光俊, 近藤 裕治, 小久保 浩, 山崎 匠. 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. *情報管理*. 49(2) 69-76, 2006.