

# 加入 PubMed Central 的实践及其对期刊的积极影响

翟自洋 林昌东\* 林汉枫 伍秀芳

收稿日期:2007-01-04  
修回日期:2007-05-15

《浙江大学学报(英文版)》编辑部, 310027 杭州市玉古路20号, E-mail: jzus@zju.edu.cn

**摘要** PubMed Central(PMC)是由美国国家医学图书馆(NLM)下属的国家生物技术信息中心(NCBI)创立的开放存取(open access)的生物医学和生命科学全文数据库。PMC只接收采用国际同行评审制度的期刊,对收录的期刊有科学及编辑质量、数据文件质量两方面要求。《浙江大学学报(英文版)B》(JZUS-B)是中国第一家通过PMC评审并被收录的期刊。PMC要求期刊提供XML(可扩展标记语言)格式或SGML(通用标记语言)格式的全文文档。我们在熟悉PMC“期刊出版文件类型定义”(Journal Publishing DTD)的基础上,开发相应程序,并辅以手工操作,实现了准确快速制作符合PMC要求的数据文件。借助NCBI这个极具影响力的平台,可以促进作者与国际同行的交流,为期刊的进一步发展创造良好环境。

**关键词** PubMed Central(PMC) 《浙江大学学报(英文版)》B辑(JZUS-B) 数据文件 可扩展标记语言(XML) 评估 编辑质量 国际影响力

2000年2月,美国国家医学图书馆(NLM)下属的国家生物技术信息中心(NCBI)创立了PubMed Central(PMC)。PMC利用最新的可扩展标记语言(eXtensible Markup Language—XML)/通用标记语言(Standard Generalized Markup Language—SGML)技术,将传统的印刷本期刊以及近年兴起的在线出版期刊全文转换为XML/SGML格式的文件,一方面起到备份的作用,长久的保存生命科学领域的研究成果,更主要的是建立一个开放存取(open access)的平台,供全世界生命科学领域的研究人员免费浏览、检索和下载<sup>[1]</sup>。

PMC成立之初只有两份期刊: Molecular and Cellular Biology 和 Proceedings of the National Academy of Sciences of the USA (PNAS)。近年收录期刊数量增长较快,截至2007年3月21日,已收录316种<sup>[2]</sup>,其中绝大部分为发达国家的期刊,如BMC,ASM,PLoS系列,等;中国目前只有《浙江大学学报(英文版)B》(JZUS-B)被收录。收录的期刊中有261种出版后即时发布,其余55种延迟1到36个月不等<sup>[2]</sup>;194种期刊全部文章开放存取(open access, OA),另有7种期刊部分文章开放存取,其他115种为自由存取(free access, FA<sup>[3]</sup>)。PMC中不乏在国际上享有盛誉的期刊,除最初的两份期刊外,其他如The EMBO Journal, Molecular Biology of the Cell, American Journal of Human Genetics等,影响因子均在9以上<sup>[4]</sup>。PMC以其开放性、高水平,对生命科学领域研究人员的影响在日益扩大。

为确保数据库的权威性和影响力,PMC选刊时提出了严格的科学和编辑质量要求。与PubMed只提供文摘<sup>[5]</sup>不同,

PMC是全文数据库,它要求编辑部将期刊全文制作成XML/SGML文件(须符合PMC所接受的“文件类型定义(DTD)”<sup>[6,7]</sup>)。

XML为管理大量数据而诞生,是依托互联网的划时代新技术<sup>[8]</sup>。它采用纯文本格式,分类标注数据的各部分内容。其优点在于:(1)“标记”语言,对文档的各部分内容分类标记,便于检索;(2)可扩展,并且便于与数据库进行数据交换;(3)文本文件,格式简单,跨平台。XML文件基于文件类型定义(DTD),后者指定了具体应用的格式要求。

为促进与国际同行交流,提高期刊影响力,《浙江大学学报(英文版)》B辑(JZUS-B)于2004年10月提出申请,经严格的评估,最终于2006年3月15日正式加入PMC。本文将结合我们的实践介绍PMC评估收录过程,另外谈谈加入PMC对期刊的影响。

## 1 JZUS-B 加入 PMC 的实践

PMC申请过程在其网站<sup>[9]</sup>有详细介绍,文献[10]进行了详细译介。这里结合我们的实践对每个步骤分别加以阐述。

### 1.1 提出申请

PMC只接收采取国际同行评审制度的期刊,目前只收录英文期刊或者双语期刊中的英文文章。满足PMC科学和编辑质量要求(详见[10])的生物医学类期刊可提出申请。申请的期刊需将期刊的基本信息,如期刊全名、ISSN、期刊网址(如果有的话)等通过E-mail发送到pmc@ncbi.nlm.nih.gov<sup>[9]</sup>。

\* 通讯作者: 310027 浙江杭州玉古路20号, E-mail: zupnon@zju.edu.cn

## 1.2 接受评估

PMC 从科学和编辑质量以及数据文件质量方面对提出申请申请的期刊进行评估<sup>[9,10]</sup>。如期刊已被 MEDLINE 收录,或经 NLM 的文献筛选技术评议委员会评审合格,则满足前一要求;此后,期刊需提供三期或大约 50 篇文章的数据文件(包括 XML、PDF、高精度 tiff 格式图片等)样品,特殊情况下数量可适当降低。详细要求可参考[9,10]。

《浙江大学学报(英文版)》(JZUS)于 2004 年被 MEDLINE 收录。2005 年,JZUS 按专业类别划分为《浙大应用物理和工程(英文版)》(JZUS-A)和《浙大生命科学(英文版)》(JZUS-B)两个新的月刊。PMC 主要考察 JZUS-B 对 JZUS 的延续性。

JZUS 从 2004 年第二期开始采用 Word 排版<sup>[11]</sup>,并转成 PDF 文件上网。将排版文档转换为 XML 文件有三种途径:(1)外包。有专门公司提供制作 XML 的服务,期刊只需提

供文章的 PDF 文件和 XML 的 DTD。(2)购买软件,自己制作数据。如 Inera 公司基于 MS Word 二次开发的 eXtyle<sup>[12]</sup>。(3)期刊自行转换。我刊在准备 XML 数据前进行过查询,发现前两种方式当时只有国外公司提供,但其费用极为昂贵,每页至少 3 美元以上,无法承担。因此,只能决定采用第 3 种方式。

我们选取了三期文章,首先按照 PMC 的 XML 定义规则(JP DTD<sup>[6]</sup>),手工生成 XML 文件。但纯手工编辑极为烦琐,不仅费时费力,且容易疏漏、出错。因此,我们前两次用纯手工方式转换的数据文件都未能通过评审。表 1 给出了 PMC 第二次评估报告(节选)。随后,我们开发了一些软件工具,尽可能正确地判断文档的格式及内容,将文档自动转换成 XML。目前程序已承担了绝大部分的转换工作,避免了许多手工误操作。由此生成的数据文件终于通过了 PMC 的评估——如果第三次数据评估仍未通过,申请将被拒绝。

表 1 PMC 对 JZUS-B 三期样刊 XML 文件的第二次评估报告(节选)

Vol/Iss Pg. No.	File Name	Error	Context	Fixed?
<b>Required Fixes</b>				
<b>Compliance with the PMC File Requirements Specifications</b>				
05(10)	JZUS05-1298.xml	All file names must match in name and case.	<b>File names:</b> JZUS05-1298.xml.xml jzus05-1298.pdf	Yes
<b>Missing Files/Parsing Errors</b>				
6B(1)	Missing files	Missing PDF file: JZUS06B-0033f		Yes
<b>Systematic Errors</b>				
5(10) 6B(1) 6B(2)	All files	Do not include emphasis tagging such as superscript, in the cross-references (<xref>) in front matter.  <b>Multiple instances</b>	<b>Examples of emphasis in the &lt;xref&gt;:</b> <xref ref-type="author-notes" rid="FN01"><sup>&dagger;</sup></xref> <xref ref-type="aff"><sup>1</sup></xref> <b>Change to:</b> <xref ref-type="author-notes" rid="FN01">&dagger;</xref> <xref ref-type="aff">1</xref>	Yes
5(10) 6B(1)	<b>Multiple files including:</b> JZUS05-1270.xml JZUS06B-0061.xml	The n-dash entity is incorrectly tagged as a hyphen keystroke.  <b>Multiple instances.</b>	<b>Example from JZUS05-1270.xml:</b> <p>Clinical researches... the bone and the prosthesis-the femoral ... <b>Change to:</b> <p>Clinical researches... the bone and the prosthesis&ndash;the femoral ...	Yes
<b>Unique Errors</b>				
5(10) 6B(2)	JZUS05-1298.xml.xml JZUS06B-0079.xml JZUS06B-0100.xml	Article's author names are misspelled.	<b>Exmample from JZUS05-1298.xml.xml:</b> <contrib contrib-type="author"> <name><surname>ZHANG</surname> <given-names>Ren-ya</given-names> </name>	Review #2 - Not all instances were fixed.
5(10)	JZUS05-1298.xml.xml	The section title is misspelled.	<b>XML (JZUS05-1298.xml.xml):</b> <sec> <title>MATERAL AND METHODS</title> </sec><title>Animal and cells</title>	Yes

说明:我们从完整的 PMC 评估报告中挑选了几个有代表性的错误组合在一起,按照他们的格式列出。PMC 报告开头有一页文字对表格中各项进行说明,此处从略。第二次报告是在第一次报告基础上作出的,并补充新出现的错误。第一列为错误出现的卷(期),第二列指出错误出现的频率——是全局性的(all files),还是存在于多个文件(Multiple files),抑或单个文件(Unique Errors)中;第三列是对错误的说明;第四列举例指出具体的错误并提出修改方法;最后一列,第二次提供的数据是否修改了第一次评估中出现的错误,蓝色背景表示错误未全部纠正。值得注意的是,如果某类错误出现次数较多(包括存在于多期文件),PMC 并不会将其出现的场合——指明,所以第二栏给出的文件范围是不具体的,期刊需自行找出错误的所有拷贝。

### 1.3 签署协议

通过评估后,期刊需要和 NLM 签署一份正式的 PMC 参与协议(PubMed Central Participation Agreement)<sup>[9,10]</sup>。协议列出 22 条声明,规定了期刊和 NLM 各自的义务——期刊需按要求提供数据文件,NLM 提供相应的服务。期刊方面需在协议中写明:版权持有人;提供给 PMC 的最早过刊的卷期;何时上传新的数据文件(出版后几周?);何时在网站发布全文(若延迟发布,延迟几个月?);是否将文章列入 OA 名单。PMC 将书面的参与协议邮寄给期刊,以此,JZUS-B 与 PMC 签署了最后的协议。

### 1.4 追加过刊

与 PMC 签署协议后,期刊除按合同定期提供最新数据外,还需追加过刊的数据文件。期刊可将所有过刊均转换上传,也可自行指定从哪一期开始提供 XML 文档(需在协议中指明)。可先提供一整年的数据,待审核无误后进入下一步。

我们追加了 2005 全年的数据。

### 1.5 正式上网

PMC 为每种期刊设立了一个预览站点和一个正式站点。期刊提交第一部分数据(通常为一整卷过刊),PMC 审核无误,将内容发布到预览站点。编辑部浏览后,如发现错误,可及时修改;如确认无误,则选择网站上发布的样式。随后,数据在网站上正式发布。其余过刊可随后追加。

JZUS-B 于 2006 年 3 月 15 日发布到 PMC 的网站,正式被 PMC 收录。

表 2 列出了 JZUS-B 加入 PMC 的进度。由于分刊的缘故,JZUS 在 PMC 收录期刊列表上有 Journal of Zhejiang University SCIENCE 和 Journal of Zhejiang University SCIENCE B 两个刊名。前者只收录第 5 卷第 10 期,后者从分刊后第 1 期开始。

表 2 《浙江大学学报(英文版)》B 辑加入 PubMed Central 的进度

关键步骤	时间
提出申请	2004-10
期刊内容评估通过*	2004-12
数据文件的评估	
第一次评估报告返回	2005-03-23
第二次评估报告返回	2005-06-03
评估通过	2005-08-28
签署协议	2005-10
追加 2005 年过刊,修改直至无误	2005-09 ~ 2006-02
全文在网站发布,正式被 PMC 收录	2006-03-15

\* 因分刊缘故,刊号有变动,PMC 主要考察 JZUS-B 对 JZUS 的延续性

## 2 加入 PMC 对期刊的影响

从申请到被收录,约一年半时间,编辑部接受了挑战,同时为期刊带来了一些变化。下面围绕编辑质量、编辑业务素质、期刊国际影响力三个方面谈谈加入 PMC 对期刊的影响。

### 2.1 提高期刊的编辑质量

期刊在编辑过程中难免出错。XML 文件的制作以及 PMC 的独立审核,有助于减少错误,提高编辑质量,是三审三校的延伸和补充。

#### (1) 编辑在制作数据文件的过程中发现错误

图 1 是 XML 文件制作、上传的流程图。在制作 XML 文件过程中,我们结合了编辑手工处理和程序自动转换两种方式,每个步骤都可说是对 Word 文档的一次全文检查,而且各有侧重;除责编外,另有多位编辑参与其中。因而,能够检查出 Word 文档中存在的个别问题。

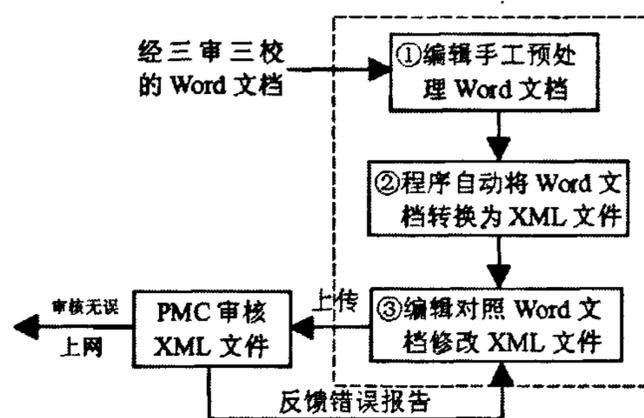


图 1 XML 文件制作、上传过程  
虚线框为制作 XML 文件的三个步骤

下面对制作 XML 文件的三个步骤中可能发现的问题分别加以简单介绍。

① 预处理是在 Word 文档缺少格式标记(如字体、标点等)的地方手工添加特定标记,帮助计算机作出判断。对一期文章的连续处理,在添加标记的同时,不同篇章的相关部分相互参照(譬如文献部分),更容易发现信息的错漏或前后不一致。

② 程序严格利用期刊格式特征或者手工添加的标记作出判断,如果格式有出入,相应部分就无法正确转换。如文献中标点漏用、误用,文献引用缺漏或与文后列出的不一致,特征部分(如摘要、关键词等)的格式错误,等等。

③ 是对照 Word 全文对 XML 的一次彻底检查。②中提到的错误经过这一步检查得以发现;另外,还可纠正图表公式编号错误、正斜体不一致,单词拼写错误、符号误用等。

根据我们的经验,一期 80 页左右的数据文件可以在三四天内制作完成;之后,PMC 审核、返回以及上网预览大概需要 10~15 天。若能调整出版周期,将期刊付印时间推迟到数据文件制作完成之时,可以有效地减小差错率,提高期刊的编辑质量。国外出版商将制作 XML 文件作为出版过程的一部分,推测有这样的考虑。

### (2) PMC 审核数据文件,有助于发现错误

PMC 要求期刊提供的数据文件在内容上必须与文章的印刷版(或网上出版的电子版)完全一致。对每次上传的数据进行程序、人工两方面的彻查,并返回详细的检查报告,指出 XML 编辑的技术性错误,有时包括原始文档遗留的问题。

### (3) 用户浏览时发现错误

期刊在网上正式发布以后,全世界的用户都可以通过 PMC 数据库查询和浏览。用户如发现错误,可以反馈给作者或期刊,或由 PMC 汇总后反馈给编辑部。

编辑部如发现过刊中存在错误,也应当及时更新 XML,重新提交给 PMC,PMC 会及时更新数据。

## 2.2 提高编辑的业务素质

PMC 工作人员严谨的作风、一丝不苟的态度体现在每一期数据的审核过程中,感染并激励着我们,我们受益匪浅。

在制作 XML 文件的过程中,编辑得到了锻炼,开阔了视野,提高了能力;进一步熟悉了一些常用的编辑软件,如 Word, Excel, Photoshop, Emeditor 等,提高了工作效率。新编辑在 XML 制作中熟悉了期刊的格式、编辑规范。通过 PMC 的审核报告,以及制作 XML 过程的查错,在提高文章编辑质量的同时,编辑的业务素质也得到了提高。

如果将制作 XML 文件外包,则难以起到此种作用。

## 2.3 促进作者与国际同行的交流,提高期刊影响力

保证期刊成功的关键是报道和出版最新的优秀科研成

果。SCI 影响因子被视为期刊学术质量的一个关键指标。尽管 SCI 影响因子在学术界和科研评估部门仍存在争议,但作为科学引文检索和评估指标,它是公正客观的,可作为期刊分析、研究的有力工具。

文章发表后能否被 SCI 期刊引用,取决于文章是否真正有价值,但前提是让相关领域的研究人员检索到这篇文章。文献查新离不开 Google 等检索工具。PMC 网站的 PR 值为 8 [PR 值全称 PageRank(网页级别),是 Google 用于评测网页“重要性”的一种方法。Google 通过 PR 值来调整结果,使“更具重要性”的网页在搜索结果中排名提升。PR 值范围 0~10,数值越大,重要性越高<sup>[13]</sup>],收入 PMC 的文章,检索时会在显著位置列出,从而第一时间被世界各地的同行看到。对于没有建立自己网站的期刊,相信加入 PMC 对提高其影响力能起到明显效果。

我们从 ISI Website<sup>[14]</sup> 获得了 JZUS 文章在 SCI 收录期刊中的引用记录。为便于纵向对比,将分刊前 JZUS 中属于医学、生物、环境等专业的文章的引用记录分离出来。图 2(a) 显示 JZUS(2000~2004) 中属于以上专业的文章以及 JZUS-B(2005~2006) 的所有文章在各年度(2000~2006) 的 SCI 总引频次变化;图 2(b) 则是各年度的文章迄今的总引频次变化。统计截至 2006-12-13。

由图 2(a) 可以看出,2004 年以前引用频次极小,到 2003 年为止,总引频次仅为  $0+0+3+7=10$  次。2004 年我们建

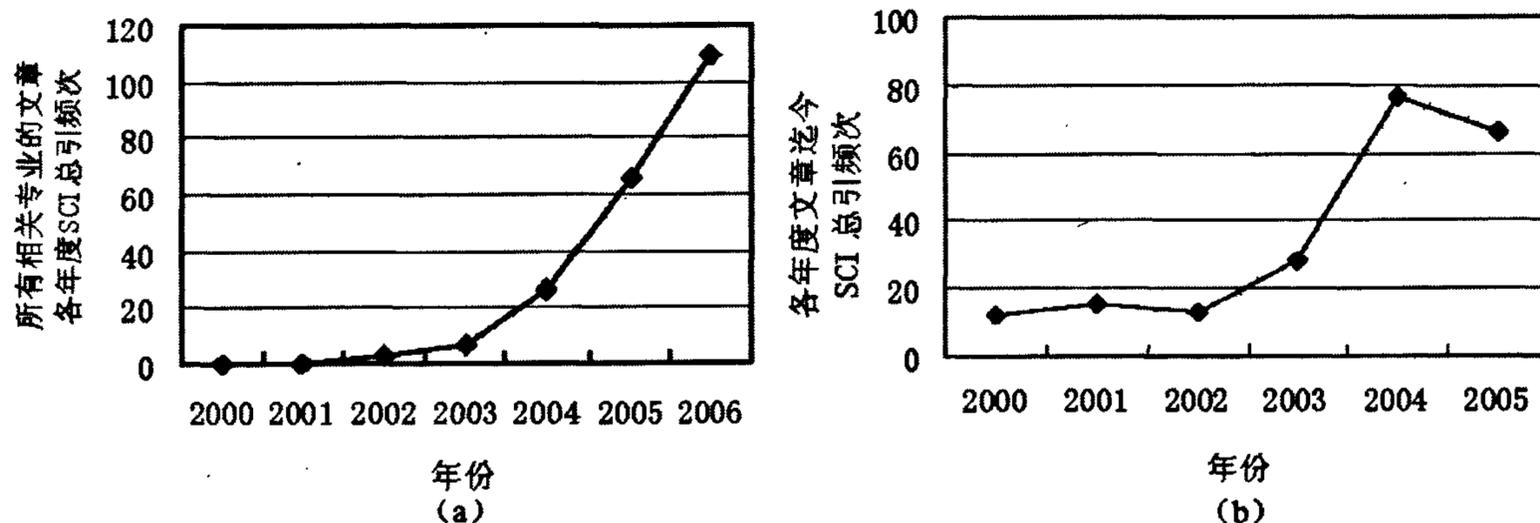


图 2 JZUS(2000~2004 年)中属于医学、生物、环境等专业以及 JZUS-B 的所有文章在各年度的 SCI 总引用频次(a)以及各年度发表的相关专业文章迄今的 SCI 总引用频次(b)

立了网站<sup>[11]</sup>(PR 值为 6),世界各地的科研人员可以通过搜索引擎找到感兴趣的文章,点击、下载。因此,2004~2005 年的引用有了一个飞跃:2004 年当年引用为 26 次,2005 年总引频次达到 66。2006 年被 PMC 收录,当年总引频次为 109。从图 2(b)看出,网站建立以及加入 PMC 之后,2000~2003 年度发表的文章总引频次达到  $12+15+13+28=68$  次,与前面的 10 次相对照,说明网络化对提高影响因子起了很大作用。再看图 2(b),2005 年的文章总被引 66 次,2004 年的文章则为 77 次;我们另外统计了 2004 年的文章在 2006 年总

被引 40 次,表明到 2005 年底,2004 年的文章总被引 37 次。PMC 收录 2004 年第 10 期以及 2005 年以后的文章,2004 年相关专业的文章共 126 篇,2005 年为 201 篇(包括快报、综述及研究论文,不计其中 7 篇会议报告),文章数  $201/126=1.60$ ,引用频次  $66/37=1.78$ ,略大于 1.60。单纯从数据来比较,说明 PMC 对影响因子的提高有一定影响。

以上只是就加入 PMC 对期刊 SCI 引用频次的的影响作粗略分析。因加入 PMC 时间较短,对期刊的影响有待进一步观察。

虽然绝对的引用数字仍然较小,但近两年 SCI 引用记录上升幅度较大,引用频次的变化鼓舞人心。高水平的文章是期刊的灵魂。我们希望借助 PMC 的影响力,通过编辑素质、文章编辑质量的提高,提高期刊声誉,吸引更多高水平的作者和审稿人,接收更多高水平的文章,为国内和国际上的生命科学研究人员提供一个更加有利的交流平台。

### 3 结语

加入 PubMed Central 是我们期刊国际化的一次尝试。除了用英文出版、采用国际同行评审制度之外,期刊还必须吸引国际作者的目光,吸收更多高水平的国际稿件。走出去,了解世界的同时让世界了解我们,在这样的基础上,我们的努力才会取得最大的效果。

截至 2006 年 4 月,我们国家(包括港台地区)被 MEDLINE 收录的英文期刊有 32 种<sup>[15]</sup>,已被 MEDLINE 收录或者已接近 MEDLINE 要求的同行期刊,或许可以进一步考虑加入 PMC。我们写下加入 PMC 的经历和一些感受,希望大家增进交流,共同提高。

**致谢** 加入 PMC 的过程中,我们得到了 PMC 工作人员 Dr. Ed Sequeira 和 Dr. Brooke Dine 热忱无私的帮助,Dr. Jennifer Jentsch 和 Dr. Wayne Jack Logue 为审核数据付出了辛勤劳动,在此对他们表示诚挚的谢意。

张月红老师审阅全文,并提出宝贵修改意见,我们深表感谢。

此外,感谢参与数据制作的浙江大学研究生戴鸿君、张帆、汪莎、梁蛟和沃静静。

#### 参考文献

1 PMC Overview. [2005-01-07]. <http://www.pubmedcentral.nih.gov/about/intro.html>

- 2 PubMed Central Journals—Full List. [2007-03-21]. [http://www.pubmedcentral.gov/fprender.fcgi?cmd=full\\_view](http://www.pubmedcentral.gov/fprender.fcgi?cmd=full_view)
- 3 Open Access Publishing. [2007-03-08]. <http://www.pubmedcentral.nih.gov/about/openaccess.html>
- 4 SCI 生命科学杂志影响因子. [2005-01-06]. <http://life.sjtu.edu.cn/news/show.aspx?id=58&cid=7>
- 5 李友仁,刘松岩,黄敏. PubMed 检索概述. 医学信息,2002,15(7):441-446
- 6 Journal Publishing DTD. <http://dtd.nlm.nih.gov/publishing/>
- 7 Rosenblum B, Golfman I. A Decade of DTDs and SGML in Scholarly Publishing. What Have We learned? Extreme Markup Languages 2002: Proceedings. <http://www.mulberrytech.com/Extreme/Proceedings/html/2002/Rosenblum01/EML2002Rosenblum01.html>
- 8 World Wide Web Consortium. Extensible Markup Language (XML). [2006-09-11]. <http://www.w3.org/XML/>
- 9 How to Join PMC. [2006-09-05]. <http://www.pubmedcentral.gov/about/pubinfo.html>
- 10 沈锡宾,吕小东,郝秀原,孙静,汪谋岳,郭利邵. PubMed Central 简介及其对期刊的评估和收录. 中国科技期刊研究,2006,17(5):866-868
- 11 林昌东,伍秀芳,黄洁卉.《浙江大学学报(英文版)》建设网站走向世界. 中国科技期刊研究. 2004,15(3):294-295
- 12 Order out of Chaos—Editorial and XML tools for Microsoft Word—Product Overview for Scholarly Publishers, 2004. <http://www.inera.com/>
- 13 GOOGLE PR 值查询. <http://link.kuww.net/pr.asp>
- 14 ISI Web of Knowledge. [2006-08-04]. <http://portal.isiknowledge.com/portal.cgi?DestApp=WOS&Func=Frame>
- 15 2006 年 MEDLINE 收录中国期刊名单. [2006-04-07]. <http://www.omschina.org.cn/news/info.asp?newsid=057221138107778>

· 读者 · 作者 · 编者 ·

编辑老师:

您好!

感谢您对我文章提出的宝贵意见,经您们指点之后,我觉得自己的思路更清晰了,通过给您们投稿,的确提高了我的写作能力,是您们培养了我,在此十分感谢!

阅读完审稿意见后,我十分佩服专家,能准确地指出我文章的不足,而这恰恰是我最需要的和自己根本想不到的,经过专家的指点,我重新梳理了思路,阅读了近期有关这方面的文章,补充了新的内容。但不知是否合适,还想请专家帮助修改,提出宝贵意见,使这篇文章能够达到发表的水平。

顺祝编安

辽宁工程技术大学学报编辑部 史庆华 2007-05-14  
123000 辽宁省阜新市中华路 47 号

尊敬的各位老师:

你们好!

寄来的第 4 期期刊及稿费均已收到,谢谢各位老师真诚的帮助!恩师难忘,我将铭记你们曾给予我的帮助。滴水之恩,当涌泉相报。

我永远是您们忠实的读者,我也将尽自己努力做好工作,回报人民,回报社会。

此致

崇高的敬礼!

《南方国土资源》杂志社 陆小芸 2007-08-06  
530023 南宁市建政路 1 号 广西壮族自治区国土资源宣传中心