

# 如何实现“一次排版、多种媒体发布”？

赵薇，杨丽丽

（《军械工程学院学报》编辑部）

## 1 “一次排版、多种媒体发布”的含义

所谓“一次排版、多种媒体发布”，就是要对期刊数据进行抽象，将抽象出的元数据（可以使用 XML 或 JSON 的数据格式）在不同的媒体形式间共享，也就是将一篇文章抽象为不同的字段和属性，包括文章标题、作者、摘要、内容和参考文献等，进行发布的不同媒体使用统一的抽象数据，并以各自的形式表现出来，实现查询、阅读和共享等功能。

## 2 发布的形式和意义

网络环境下的科技期刊出版是网站、手机 App 和第三方媒体平台等的融合，以多种媒体形式发布，不同的媒体形式对内容的展现方式略有不同：对同一篇文章而言，使用电脑浏览器访问期刊网站时，看到的是网页形式的整篇文章；使用手机 App 阅读时，看到的是适合手机阅读的文字窗口；通过第三方媒体平台浏览时，可能看到的又是嵌入式的内容或信息摘要；针对期刊数据库而言，不同的全文或题录数据库提供的文章摘要、关键词和全文等都是相同的。所有这些多种媒体展现形式的根本内容来自同一篇文章的内容数据，媒体形式只是一种展现手段。后台编辑只要将文章的标题、发布时间、作者、图片和内容编辑等形成一种通用格式，由选择的媒体对内容进行解析和展现，其意义在于：

1) 文章内容被抽象后只是一些元数据，排版时无需考虑这些数据在不同展现形式下如何组织，只要专注于数据录入和内容校对即可。只要对内容进行一次编辑，在后台管理系统或管理软件中录入数字化的内容即可，大大减少了排版的工作量。

2) 同样的数据可以多种媒体形式展现，显示出丰富的展现效果和内容场景，所有推广和发布的渠道均可以使用，读者也能够随心所欲地选择阅读形式，较之于传统的纸质媒体阅读便利性和读者粘性显著增强。

## 3 数据抽象的原则

科技期刊的数据抽象是指将一篇文章的所有相关属性都抽象为字段形式（每个字段有其内容要求），再将其存储为可扩展的数据文件形式的过程，这个过程本质是信息抽取技术的实现。信息抽取技术是指从一段文本中抽取指定的事件、事实等信息，形成结构化的数据并存入一个数据库，供用户查询和使用的过程<sup>[1]</sup>。信息抽取并存储完毕后，读者通过各类客户端媒体访问这些数据文件，媒体内容得以展现。这些抽象出来的数据，即“元数据”是指用来标志、描述和定位网络电子资源的数据，可以描述数据的特征和属性，提供某种资源有关信息的数据结构，是促进网络信息资源组织和发现的数据<sup>[2]</sup>。具体而言，期刊内容可按表 1 和表 2 进行数据抽象和整理。

表 1 整体内容字段表

字段名	数据形式	数据要求	备注
文章 ID	整形	11 位以内整形	一篇文章的唯一标识，不能重复
文章标题	字符型	30 个汉字以内	无
文章副标题	字符型	30 个汉字以内	无
中文摘要	字符型	300 个汉字以内	无
英文摘要	字符型	1 000 个字符以内	无
中文关键词	字符型	20 个汉字以内	无
英文关键词	字符型	20 个英文单词以内	无
文字正文 ID	整形	11 位以内整形	关联到文字对象（另一个表存储）*
作者	字符型	100 个汉字以内	无
联系方式	字符型	15 个汉字以内	无
发布时间	时间类型	按照时间格式填写	无
引用	字符型	500 个汉字以内	无

注：不同科技期刊对于字数和形式的要求可能不同，具体情况因刊而异。\*表示正文存在文字和图片，且可能交叉排版，需要单独的格式存放这些内容（包括格式）。

表 2 正文内容字段表

字段名	数据形式	数据要求	备注
正文 ID	整形	11 位以内整形	一篇正文的唯一标识，不能重复
文字 ID	整形	11 位以内整形	一段文字的标识
图片 ID	整形	11 位以内整形	一张图片的标识
文字内容	字符型	5 000 个汉字以内	对应文字 ID 的文字内容
图片地址	字符型	1 000 个字符以内	对应图片 ID 的图片 URL

注：正文内容字段只是基本排版要求的内容，如果期刊有更丰富的排版，再单独丰富正文表。

根据上述数据抽象原则，一篇文章需要的基本元素均已包含在内，编辑只需填充各字段的内容。针对期刊信息而言，通过对内容信息的编码描述，提供一个中间级别的描述数据<sup>[3]</sup>，各类媒体客户端即可展现这篇文章。由于来源相同，内容也是相同的。

#### 4 可扩展的数据文件格式（XML 和 JSON）

在上述一次内容编辑的基础上进行数据文件生成，使用可扩展的文件格式，通过媒体客户端读取生成的文件，即可实现对内容的展示。

主流的可扩展文件格式主要有 XML 及 JSON 格式。XML 是一个文本式标记语言，用它标记组织的信息单元是一个人工可识别、计算机可解析的文本文件<sup>[4]</sup>。以一篇文章为例，XML 格式的数据文件生成形式如下：

```
<?xml version="1.0" encoding="UTF-8"?>
<Chaper>
  <Title>基于 Landsat TM 数据估算山东菏泽区域杨树人工林碳储量</Title>
  <Author>
    <SubAuthor>
      <Name>刘诗琦</Name>
```

```
<Organization>北京林业大学林学院</Organization>
</SubAuthor>
<SubAuthor>
  <Name>贾黎明</Name>
  <Organization>北京林业大学林学院</Organization>
</SubAuthor>
<SubAuthor>
  <Name>杨军</Name>
  <Organization>清华大学地球系统科学研究中心</Organization>
</SubAuthor>
<SubAuthor>
  <Name>辛建华</Name>
  <Organization>菏泽市林业局</Organization>
</SubAuthor>
</Author>
<Abstract>利用 Landsat TM5 数据，结合地面调查，对我国鲁西黄泛平原典型地区菏泽市的杨树人工林碳储量进行测算。结果表明：单木生物量……</Abstract>
<ContentId>10895</ContentId>
<Recommand>
  <ID>1</ID>
  <Name>国家林业局森林资源管理司……</Name>
  <ID>2</ID>
  <Name>……</Name>
  ……
</Recommand>
</Chaper>
```

如上所示，将文章抽象为 XML 的数据文件格式并保存在数据服务器中，由媒体客户端对数据进行获取和展示。编辑只需维护一次数据，内容即可在各类终端中展现，实现“一次排版、多种媒体发布”。

参考文献：

- [1] 刘迁, 焦慧, 曹惠波. 信息抽取技术的发展现状及构建方法的研究[J]. 计算机应用研究, 2007, 24(7): 6-9.
- [2] 游祯, 赵荣. 我国元数据研究现状与发展[J]. 图书情报工作, 2008(S1): 202-205.
- [3] 徐红. 元数据在期刊信息描述组织中的应用[J]. 现代情报, 2002(6): 57-58.
- [4] 盛梅, 李丰梅. 基于 XML 语言的医院期刊资源的数字化[J]. 医疗设备信息, 2007, 22(8): 54-55.