

doi:10.3772/j.issn.2095-915x.2015.01.001

自助式情报管理与服务系统的研究与实现^{*}

吴广印¹, 徐飞², 冯丹¹

(1. 中国科学技术信息研究所 北京 100038; 2. 江苏省广电有线信息网络股份有限公司 南京 210008)

摘要: 本文首次提出了“自助式情报管理与服务系统”的概念,旨在利用 Web 3.0 的技术为用户构建一个集数据采集、编辑、评价、发布、沟通一体化的企业内部情报管理与服务系统。本文主要内容包括概念提出、需求研究、目标设计、系统架构及实现等内容。

关键词: 自助式情报管理与服务, 竞争情报管理, Drupal

Implementation and Research of Self-information Management and Service System

WU Guangyin¹, XU Fei², FENG Dan¹

(1. Institute of Scientific and Technical Information of China, Beijing 100038;
2. Radio and Television Cable Information Network Inc of Jiangsu, Nanjing, 210008)

Abstract: This paper first proposed the concept of "self-information management and service system", designed to take advantage of Web 3.0 technology for users to build an enterprise information management and service system, integrated data collection, editing, review, publishing and communication. The main contents include concepts put forward, the demand study, target design, system architecture and implementation, and so on.

Keywords: Self-information management and service, competitive intelligence management, Drupal

^{*} 基金项目: 本项目得到国家科技支撑计划项目科技文献动态数字出版技术研发与应用示范(项目编号: 2012BAH90F00)支持。
作者简介: 吴广印(1965-),男,中国科学技术信息研究所研究员,北京万方软件股份有限公司董事长,研究方向:大数据管理与应用;徐飞(1983-),男,安徽滁州人,硕士,编辑中级,主要研究方向为:影视动画,信息分析;冯丹,女,中国科学技术信息研究所,硕士研究生,研究方向:知识重组。

1 引言

目前多数大型科技型企业,越来越重视机构内部情报服务系统的建设,甚至提出“大情报”的发展战略。各类图书情报服务系统为了提高自己的专业服务能力,提出了加强“学科服务能力建设”要求。现有多数情报系统的建设是构建一个系统平台,然后向专业服务商购买所需信息,并将信息灌入自己的系统对内提供服务,也有一些自建内容,则依靠的是机构内部的信息服务机构负责收集、加工和发布,受人力数量、质量、资金等方面的影响,收效甚微。为此作者提出“一个机构的情报系统的内容建设,一方面是拥有更多更有用的情报内容,更重要的是培养树立机构内部用户的收集、共享自己私有情报信息的情报意识,让每一个情报需求人员成为情报内容的收集、甄别、发布者”。

2 自助式情报管理与服务的研究

下面将从自助式情报管理与服务的概念提出、系统特点以及系统所应具备的功能架构三个层面来对自助式情报管理服务进行研究:

2.1 自助式情报管理与服务的提出

2.1.1 情报管理与服务的现状

经过多年的情报系统的建设与发展,特别是经过20多年互联网技术与应用的发展各图书、文献情报服务机构都建立自己的服务系统,利用这些系统为自己用户提供情报检索、查新、咨询、原文传递等信息服务。现有服务系统在资源、系统和服务三个层面总结如下:

资源层面:以购买专业服务机构提供资源(二次文献和全文)为主,自建信息为辅,进行整合服务,整合包括数据整合和业务整合(比如:本地原文传递)。自建服务主要是针对自己的特殊

文献资源,由图书馆、情报室等专业人员建设完成。资源服务的主要类型是图书、期刊、会议、专利、标准等传统文献。

系统层面:主要是对购买或自建的各类服务系统进行整合(国内外各类服务厂商提供的功能和服务没有统一标准,用户使用不便)。整合分为两种情况,一是对各类系统进行虚拟整合,虚拟整合主要是用户同意认证和功能层面的整合,实现用户统一认证和检索功能的虚拟整合(各系统检索结果的整合),实际功能的实现由各子系统完成,系统存在很多局限性。另一种是整合各系统的元数据(或直接要求服务商提供元数据),构建一个新的检索服务系统,实现机构所需的情报相关服务。后一种整合服务模式是当前主要整合服务模式。

服务层面:现有文献服务系统基本上都是以整合服务为基础的,通常提供情报检索服务和资源发现系统(全方位的资源定位系统包括本地馆藏、远程定位以及常规互联网定位),同时为查新、咨询、学科服务提供基础服务。在“大数据”服务背景下近两年一些系统提供了分析、预测、评价功能,比如北京万方软件股份有限公司提供的“中国学术搜索网”。

综上所述是目前国内外图书、情报服务系统主要特征,总结具有如下特征:

数据资源以传统的科技文献服务为主,追求的是科技文献的覆盖率(检索层面),缺少个性化数据,资源利用率较低,更缺乏具有针对性的情报数据。可以说以资源服务为中心,考虑机构用户的个性化需求不够。就数据资源而言科技文献资源仅仅是情报服务的一部分。从而导致现有机构情报服务系统大都拥有庞大的数据资源,而真正有价值的资源非常有限,无法满足用户获取情报的实际需求。

在服务层面仅仅以检索和获取文献资源为主,

没有考虑文献实际需求和应用价值。情报服务包括情报获取、共享、交流、评价、应用等多个环节。显然现有文献服务系统是无法满足这些情报服务需求的。

总之, 作者认为现有机构建设的文献服务平台仅能满足常规文献资源获取需要, 而且具有资源针对性不强, 服务功能不足等缺点^[1]。总之不能满现代网络环境下用户对获取、共享、交流、评价、应用等多个环节的服务需求。也就是针对上述问题作者多年情报服务系统研究、构建实践基础上提出了“自助式情报管理与服务系统”建设思路并开发完成。

2.2 自助式情报管理与服务系统的功能设计

自助式情报管理与服务系统(以下简称“本系统”)是针对现有机构所用情报服务系统的不能满足当前情报管理与服务的需求而提出的, 主要功能设计如下:

2.2.1 一般情报管理与服务功能

本系统首先要具备现有情报服务的数据管理与检索服务功能这些功能具体如下:

数据库定义、数据记录的增、删、改功能。使每一授权用户可以方便管理自己所需资源, 包括分享自己情报资源, 用户可自主决定是否与其它用户共享。另外系统的记录维护可支持文字、图像、声音、视频等混合富媒体。

数据库索引字段、索引方式的定义及索引管理。授权用户可根据需要灵活定义可检索字段、索引方式、检索结果格式、排序方式、聚类字段等。

外部数据库数据导入, 包括其他数据库数据和特定格式的数据导入。授权用户可购买、整合所需数据, 比如可以从其他系统检索并下载检索结果, 然后导入本系统, 每一授权用户都可以使用这一功能。

互联网数据获取与管理, 互联网数据已经成

为一种庞大而重要的情报资源。系统应支持 RSS 源订阅、网页自动采集等互联网资源的获取功能。

支持术语管理, 用于系统分类等用途。

2.2.2 情报速递与订阅

情报的及时获取、发布、使用是情报服务系统的核心支撑。本系统除了满足用户快速获取并管理信息外, 还应提供快速传递的时段, 使情报能够及时传递到每一个情报需求者。同时每一个情报人员也可以自主订阅所需情报内容。比如通过邮件自动将系统的信息快速传递到每一个需求者的邮箱, 这里也包括一些信息的推荐等功能。

2.2.3 情报评价

传统情报服务系统, 对用户而言只是提供信息资源, 用户只能根据自己的需要通过检索获取信息, 信息是否真正有价值只能凭表面文字来做简单分析。本系统需要为每一个授权用户提供一个信息评价功能, 用户可对任意情报源做实名评价, 为其它用户提供参考。这一功能为用户自主发布信息提供了一个信息交流的通道, 通过对评价者身份的识别可以更加准确的获取信息的价值。

2.2.4 用户沟通

本系统是机构内部的情报管理与服务系统, 用户既是情报的提供者又是情报的使用者, 用户之间的交流沟通也是一种情报服务的重要功能, 系统应具备多样化的沟通渠道, 主要包括:

邮件沟通: 为系统内嵌一个邮件服务系统, 便于用户推荐系统信息、交流等。

私信: 用户可通过系统平台随时为其它用户发送一对一的即时信息, 进行交流。私信是在系统内进行的, 具有方便、简洁、高效、安全等特点。

公共会话: 为在线用户提供一个公共会话间, 用户可点对点在线交谈, 方便沟通。

FAQ: 系统应提供一个常见问题应答系统, 为用户分专题进行专家咨询服务。

博客: 博客系统是当今网络信息交流的重要

手段。对于情报服务系统而言更是尤为重要。

2.2.5 系统整合

Web1.0 的本质是联合,那么 Web2.0 的本质就是互动,它让网民更多地参与信息产品的创造、传播和分享,而这个过程是有价值的。Web3.0 更多的不是仅仅一种技术上的革新。而是以统一的通讯协议,通过更加简洁的方式为用户提供更为个性化的互联网信息资讯定制的一种技术整合。将会是互联网发展中由技术创新走向用户理念创新的关键一步^[2]。本系统应具备和其他图情服务系统整合的能力,比如和图书馆的 OPAC 系统进行整合等。

2.2.6 用户管理

情报服务系统的用户管理和安全性十分重要,本系统的用户角色多样、权限管理复杂。要求系统具有灵活的角色和权限管理功能和安全防护功能。

2.3 自助式情报管理与服务系统的框架

自助式情报服务系统作为“一站式”情报管理与服务平台,集上述传统情报检索服务、情报速递与订阅、情报评价、用户沟通、系统整合、用户管理等功能为一体的企业级平台系统。这一系统总体上是由资源层、系统层、应用层三个层面构成的。

(1) 资源层

资源层主要是系统可管理的情报资源它包括:外购各类文献数据、机构内其它需要整合的数据资源、用户分享信息、互联网网页信息资源等。

(2) 系统层

系统层是本系统的管理核心,由数据库系统、索引服务系统、内容管理系统三部分组成,其中:

数据库系统是整个系统基础,用来存储和管理系统的所有内容信息,这些信息包括情报信息、用户信息、日志信息等。数据库系统提供了对内容信息增、删、改等基础管理功能。

索引服务系统是情报检索服务的核心,大家知道传统数据库管理系统提供查询功能无论在性能还是功能上都无法满足情报检索的需要。通常搜索引擎是把数据存储在常规数据库系统下,另外在附加一个专门索引搜索系统来完成系统的检索服务,开源系统的 Lucene^[3] 和 Solr 是典型的索引服务系统。情报检索所需的各种功能都是由索引服务系统完成的。

内容管理系统负责整个系统的内容组织、关联和展示,内容管理系统是构架在数据库系统和索引服务系统之上的服务系统,系统的各项具体服务功能都由它来完成的。

(3) 服务层

服务层是直接呈现在用户面前的,用于情报服务、内容的交互,为用户提供一个一站式的服务界面。它主要包括如下功能:

①用户管理:它包括用户注册、用户角色定义、用户角色授权等功能。

②资源类型定义:它完成系统需要管理的资源类型定义(资源包含的字段名、字段类型等)、资源显示格式等。这一功能通常只有系统管理员才能使用。

③资源索引定义:这一功能用来定义资源的索引方式、检索字段、聚类字段、检索结果展示方式,包括多种排序方式等。这一功能由索引服务系统完成,通常由管理员使用。

④资源维护:这一功能由授权角色用户完成,是特定的注册用户。使用这一功能可以构建不同类型的资源内容,可以在线编辑,可以从外部导入数据,甚至可以直接采集互联网上的各类信息资源。

⑤搜索服务:为各类用户提供搜索服务,用户提交检索请求,系统返回检索结果,检索结果可以聚类展示(分面搜索),可以按多种条件进行排序等。对于注册用户可对检索结果的内容进行评价和推荐。

⑥用户沟通：为用户提供交流互动功能。

⑦统计分析：包括检索热词分析、最新内容、检索日志、活跃内容、索引日志、订阅源日志、系统运行日志等服务功能。

⑧其他：系统各项管理功能配置及其他服务

3 自助式情报服务系统实现

经过大量的调研和实验分析，我们最终选择开源内容管理系统 Drupal 系统为基础开发平台。Drupal 作为一款强大的内容管理系统，在 Web2.0 时代以开源的方式发展壮大，相比 DSpace 和 Fedora，Drupal 具备更明显的面向对象思想和 Web2.0 特征。Drupal 拥有活跃的社区支持，较多的第三方模块的贡献，它不仅仅是一个 CMS，而被当做是一款内容管理框架（CMF）。其实早在 2006 年，图书馆人员就已经开始研究 Drupal 在图书馆服务中的应用^[4]。至今为止，已经有大量的案例，其中较为有名的有美国图书馆学会、斯坦福大学图书馆、哈佛大学图书馆等，国内的有北京大学图书馆、外语教学与研究出版社等^[5]。文中将采用 Drupal 来构建自助式情报服务系统。关于 Drupal 系统的详细介绍可参见 <https://www.drupal.org/>。下面介绍系统几个核心功能的实现。

3.1 情报资源采集子系统

在上节中已经提到，自助式情报服务系统资源分为内部资源和外部资源，内部资源即系统已有的数据库系统、外购期刊、图书、标准、专利等文献数据，外部资源则需要从各大门户网站、智库网站去采集。Drupal 中 Feeds 模块可以满足系统资源采集的需求，该模块拥有强大的数据采集能力，还包含有 FeedsCrawler、FeedsSpider、Feeds SelfNode processor、Feeds Xpathparser、Feeds Smart Parser、job scheduler 等多个插件。通过插件可以编写 xpath 规则来解析 XML、HTML，抽

取内容并导入到规则的数据库中；提供一个 RSS feed，设置好 RSS 源之后，执行 cron，系统将按照之前的设定从指定的网站上抓取数据，抓取的信息包括标题、正文和它的原始链接。采集得到的资源再根据用户的阅读习惯分类归档发布。

上述功能在系统中都已经实现，用户不需了解 Drupal 本身的开发技术，直接使用即可，比如只需提交一个 RSS 订阅源地址系统即可自动完成采集、分类功能，可以直接去重。对于其他格式数据系统提供了相应的导入功能直接导入即可。对于用户自建资源可以选择一个资源类型直接输入即可，所需资源类型的字段、输出格式等都已建好，系统为每一个注册用户提供了一个“我的工作空间”管理模块。当然用户如果需要，也可以通过管理员创建新的资源类型。

3.2 情报资源发布子系统

通过自建、导入、采集得到的情报资源需要及时发布到平台上，系统根据不同内容类型预先设定了不同的发布模板。发布子系统中集成了学术文献在线发布模块，当用户发布一篇学术文献时，需要选择出版类型，可供选择的出版类型有：Web 服务、专利、会议录、图书、学位论文、法案、法规、期刊、数据库、研究报告等多种类型，系统根据选择的出版类型调用不同的发布模板，以图书为例，发布时需要填写的模板字段包括标题、作者、摘要、关键词、全文内容、出版信息（出版年、系列标题、卷、卷数、版次、页数、再版次数）、出版者（出版社、城市）、标示符（PMID、PMCID、ISBN 编号）、URL、DOI、是否有参考文献等信息。所有这些 Drupal 中有现成的模块 Bibliography 可以完成，Bibliography 支持的导入格式有：PubMed, BibTex, RIS, MARC, EndNote tagged and XML，导出时支持的格式有：BibTex, EndNote tagged and XML，同时还可以将内容链接到 Google Scholar、PubMed 等外部搜索

引擎。注册用户登录后根据授权还可以创建 Web 表单、博客、网页内容、新闻速递等。系统为用户提供了一个富媒体编辑器，用户可直接建立自己的多媒体信息，甚至可以直接粘贴任意内容到编辑器，支持内嵌动态地图。总之系统为授权用户提供了丰富的情报资源的管理与发布功能。

3.3 情报资源导航子系统

资源导航是一种大型信息资源的浏览与检索策略，能迅速帮助用户在众多资源中快速、准确的找到目标资源。如果把结构和组织比喻成房间，那么导航就是给为房间增加窗户和门^[6]。导航系统的主要作用就是为用户提供标识，告诉用户在哪、这个地方有什么以及这附近有什么。系统采用 menu 模块来构建一级导航栏，superfish 模块构建二级下拉菜单，同时系统还提供 sitemap 等补充导航系统，让用户直接获取所需要的内容。系统还是用了 views_data_export 模块，用来处理从外部渠道抓取得到的资源，将它们以 views 的形式展示。自助式情报服务系统具有主动性、交互性和一站式的特点，所以需要构建导航系统来帮助用户分析，提高用户对资源的感知度。同时系统还可以利用索引服务器提供的强大的分面搜索功能，提供系统的分类导航功能。对于分类导航，系统提供了一个强大的术语管理系统，便于用户定制管理分类词表等。

3.4 情报资源检索子系统

自助式情报服务系统中资源检索分为站内检索和相关站点检索。相关站点检索直接调用相应数据库的接口即可实现，而站内搜索则采用 Drupal+Apachesolr 来实现。虽然 Drupal 核心自带搜索 (Search) 模块，但由于其主要是针对英文的搜索，对中文的支持不太友好，所以系统没有采用核心自带的 Search 模块。Apache solr^[7] 是一个独立的企业级搜索应用服务器，其开放性、

跨平台性等优势被众多的企业所应用，而 Drupal 也可以与 Solr 完美整合。在 Drupal 中有 2 个整合了 solr 的搜索模块：Apachesolr 和 search api。可以通过 Apachesolr 设置搜索服务器，再配置 schema 文件来确定需要索引的字段 (field)；相比 Apachesolr，search api 则更加灵活，它可以新建多个搜索服务器和多个索引 (统称为索引服务器)，然后分别为每个索引指定不同的搜索服务器。自助式情报服务系统分别使用上述两种搜索模块，对文章的标题、内容、作者等字段分别建立索引，为不同的类型指定不同的索引服务器，还加入了中文分词器来解决中文分词的问题。本系统利用 Solr 作为索引服务器，并对其做了大量扩展，Solr 作为一个支持云计算的分布式企业级索引及搜索服务系统，可在互联网任意 IP 下进行部署，并和本系统进行整合，目前有不少机构提供 Solr 的云搜索租赁服务。总之，利用 Solr 可以构建企业级搜索服务，并提供高性能的云搜索服务。另外系统对用户的检索日志进行记录和分析提供热门搜索、最受欢迎的网页等分析功能。

3.5 用户互动子系统

让用户亲自参与到系统的构建中是自助式情报服务的一大特色，用户不仅作为资源的消费者，同时也担任着资源创建者的角色。系统提供多种方式增加用户之间的互动，充分利用 Drupal Web2.0 特性，拉近用户之间的距离。首先，利用 Drupal 框架发布的资源，都可以直接在资源的下方评论，其他用户也可以针对某一个用户的评论回复，这样就形成了一个小的交流网络。其次，利用 Drupal 中的 Webform 设计表单，用户可以在线提交问题，类似于传统的 FAQ。系统中还使用了 DrupalChat 模块，用户可以在公共会话室跟所有的在线用户聊天，也可以选择跟单个的在线用户交流。这些都使得用户与服务平台、用户与用户之间的交流变得实时，也从传统的一对一的

服务模式变成一对多、多对多的服务。另外系统还提供了脱机“私信”沟通方式和用户在线邮件推荐方式，可以方便系统内容的推广和交流。

3.6 与其他系统的应用集成

虽然本系统是一个独立的自助式情报管理与服务系统，但它除了整合企业内部其他系统外还可以集成外部任意系统。本系统提供了一个相关站点频道，授权用户可以增加自己喜爱的互联网站点并与其他用户共享。相关站点资源包含站点名称、站点简介、站点分类、站点 URL、站点关键词等字段，用户输入相关信息后就可以通过本系统直接链接相关站点。同时系统还提供了对相关站点的分面检索，可方便按分类、关键词进行导航。

同时该系统通过学术文献频道提供了与机构 OPAC 系统的统一认证集成，这是 Web 3.0 的一个重要技术特征。

3.7 系统优化设计

本系统是企业级信息管理与服务系统，为了确保系统的高性能并发运行和系统安全性，系统设计中采用了如下核心技术：

①集成业界认可的 Solr 系统为系统的索引和检索服务器，Solr 的最新版本支持分布式云计算

服务，可以将 Solr 单独部署为索引服务集群，同时为了更好地支持中文信息的检索服务，系统增加了我们自主开发完成的中文分词模块。另外为了提高系统的索引和检索效率，系统集成了 Solr 并行运算模块。

②为了提高匿名用户的访问效率，系统提供了针对匿名用户的缓存策略，而对授权用户采用的是权限 + 网页的特殊缓存机制，目前系统经过测试在 32G 内存的单服务器下可支持上万用户的并发访问。另外市场上有多种第三方解决方案用于 Drupal 架构系统的性能优化方案。

③采用人工 IP 限制技术和自动识别非法入侵检测相结合的安全方案，确保系统的安全性。Drupal 系统的安全性是业内公认的安全架构^[8]，美国军方和安全部门很多采用 Drupal 系统作为系统基础平台进行二次开发。

④为了确保系统内容的安全性，系统所有资源进行的动态 URL 编码链接，可以防止其它搜索引擎及爬虫软件的内容获取。

总之在自助式情报管理与服务系统中采用了很多先进的技术和方案，尤其在后台管理和系统日志方面还有很多功能，由于篇幅有限本文不再做详细介绍。



图 1 系统主页，系统信息和所有栏目的相关内容都是可配置的



图 2 系统学术文献的搜索加结果页面，搜索结果可导出多种格式

4 结语

伴随着计算机技术的迅速发展,网络渗透到各行各业和千家万户,然而传统的情报服务机构的网络化程度却并不高,当用户面对海量的情报资源却不知道如何选择时,构建自助式情报服务系统就显得尤为重要。自助式情报服务系统提供一站式的资源集散地,通过合理的导航子系统,资源检索子系统,快速、准确的帮助用户定位到所需资源,还设计用户互动子系统,让用户可以

实时的与其他用户交流沟通,拉近用户之间的距离,形成小型的学术圈。而 Drupal 作为业界公认的一款强大的 CMS,加之其在图情领域的成功应用,以及其自身 Web2.0 的特征,采用它来构建自助式情报服务系统是一次成功的尝试。作者认为本系统同时还集成了机构仓储管理,专家管理,成果推广等专业模块,由于篇幅有限本文不再做介绍。作者认为利用该系统还可以用于机构学科服务系统建设等情报服务系统。

参考文献

- [1] 吴广印 .RMS 系统架构与情报检索系统的功能需求研究 [J]. 数字图书馆论坛 ,2013(6):31-28.
- [2] Web 3.0[EB/OL]. [2014-05-05]. <http://baike.baidu.com/subview/851883/13220838.htm>.
- [3] Apache Lucene[EB/OL].[2013-08-01]. <http://lucene.apache.org/>.
- [4] 吕秀云 . 大学图书馆自助式信息服务模式 [J]. 图书馆界 ,2003(3):27.
- [5] 马建玲 . 开放源代码的内容管理系统 Drupal[J]. 现代情报 ,2007,27(2):50-52.
- [6] 马凌, 漆颖 . 基于信息构建的自助式信息服务平台建设研究 [J]. 新世纪图书馆 ,2012(5):41.
- [7] Apache Solr[EB/OL].[2013-12-05]. <http://lucene.apache.org/solr/>.