

基于云端的个人文献管理解决方案

刘 颖

(沈阳师范大学图书馆 辽宁沈阳 110034)

摘 要:[目的/意义]随着云计算技术的发展,多种终端设备的出现,科研人员的个人文献管理方式面临变革,云存储、云同步、云分享、云备份等文献管理方式渐成趋势。[方法/过程]对比分析云盘、云笔记、搜索引擎的云端应用、文献管理软件 4 种类型的云端文献管理工具,指出这些工具的功能既有共性,又各有侧重。[结果/结论]指出科研人员应根据实际需求和意向,选择适合的文献管理系统,提出帮助科研人员克服使用障碍的对策。

关键词:个人文献管理;云存储;云分享;个人云;云图书馆

中图分类号:G250.1

文献标识码:A

doi:10.3969/j.issn.1005-8095.2016.04.018

Cloud-based Solutions to Personal Document Management

Liu Ying

(Shenyang Normal University Library, Shenyang Liaoning 110034)

Abstract:[Purpose/significance] Along with development of cloud computing technology and emergence of multiple terminal devices, researcher's personal document management method faces change, document management methods such as cloud storage, cloud synchronization, cloud sharing, cloud backup are becoming trends.[Method/process] The paper compares and analyses four types tools of cloud document management including of cloud disk, cloud note, cloud application of search engine, document management software, points out these tools' functions have commonality and also have each focus.[Result/conclusion] The paper points out that researcher should base on actual demand and using intension to select suitable document management system, and puts forward measures to help researcher overcome using obstacles.

Keywords:personal document management; cloud storage; cloud sharing; personal cloud; cloud library

0 引言

随着互联网、云计算技术的发展,网络学术资源的日益丰富,各种智能终端设备纷纷出现,人们对云存储、云同步、云分享等云端文献管理需求日益增多,科研人员的文献管理方式面临变革,基于各种云端存储工具的个人文献管理渐成趋势。目前,网上推出的各种云存储工具,种类繁多,包括各种系统、软件、平台等,使基于云端的个人文献管理成为可能。借助云端工具,人们不再需要把文献资料存放到实体的存储设备上,也无须将文献资料在各种实体存储设备间拷来拷去,通过互联网,可以随时随地安全存放数据和重要资料,轻松从云端读取自己所存储的信息,不但降低了文献存储的成本,还可以更加方便安全地管理文献,大大提升科研工作的效率。正如美国福布斯杂志网络版所断言:随着移动设备和云计算服务的发展,个人云将在未来几年逐步取代个人电脑,成为数字时代的核心^[1]。开启个人云时代,

构建文件同步、工作协同、多应用汇聚^[2]的私人数据中心,应当是每个科研人员梦寐以求的。然而,当前真正熟悉并使用云存储工具管理文献的人并不多,可谓“身在云端不识云”。因此,选择和利用适合的云存储工具,设计符合个人文献管理需求的“云”解决方案,打造私人专属的“云图书馆”^[3],是当前图书馆员和科研人员进行信息修炼和从事科研工作的必修课。

1 云存储的定义、功能与特点

1.1 云存储的定义

云存储是指通过集群应用、网络技术或分布式文件系统等功能,将网络中大量各种不同类型的存储设备通过应用软件集合起来协同工作,共同对外提供数据存储和业务访问功能的一个系统。简而言之,云存储就是将储存资源放到云上供人存取的一种新兴方案。使用者可以在任何时间、任何地方,通过任何可联网的装置连接到云上方便地存取数据,

收稿日期:2015-12-16

作者简介:刘颖(1971—),女,本科,副研究馆员。

实现移动访问和多设备互通^[4]。

1.2 个人云的定义

个人云是指可以借助智能手机、平板电脑、电视和个人电脑,通过互联网无缝存储、同步、获取并分享数据的一组在线服务。个人云是云计算在个人领域的延伸,是以因特网为中心的个人信息处理,即通过因特网对个人的各种信息进行组织、存储、分发和再加工^[5]。云计算的发展,使得个人文献的存储与管理由个人计算机终端向云端迁移,云时代的每个人都可以拥有自己专属的个人云。随着多屏时代的到来,所有的电子产品都有可能变成云的终端,个人云服务将打破不同品牌设备之间的局限,真正实现相互之间的内容访问和共享。

1.3 云存储的功能与特点

云存储是突破了传统存储方式的性能和容量瓶颈,提供拥有灵活性和按需扩展功能的新一代存储服务,可以满足当下由于信息过载,数据总量的指数增长而带来的存储难题,启动成本低,同时能够满足用户个性化及多样化需求。云存储的主要功能与特点可归结为:

(1) 提供超大存储空间,配备基础存储相关功能,随着容量增大,可线性地扩展性能和存取速度。

(2) 提供必备的数据安全保障,确保用户数据的隐私安全,结束颠覆式的技术升级和数据迁移工作。

(3) 云端与终端之间保持实时同步、备份、互操作,用户可以将数据存储按需迁移到分布式的物理站点。

(4) 支持账户平台全局搜索功能,快速定位目标文件,深入挖掘存储信息。

(5) 易于团队成员之间分享,通过提取码或邮件,轻松实现与他人资源分享。

(6) 跨终端数据管理,多终端数据互通,使大量终端协同工作成为可能。

科研人员借助云存储,可以很方便地建立起知识管理和信息发现的在线科研环境,建立云端工作室,进行科研协作,形成相关学科领域的学术圈子,大大提升科研工作效率,促进学术交流。

2 云存储工具

当前为个人提供云端数据管理功能的工具很多,但尚未见有明确的分类。笔者根据它们文献管理功能特点的异同,将其归纳整理为4种主要类型。

2.1 网盘

网盘是互联网存储工具,是由互联网公司的网

络服务器提供空间,通过互联网为企业和个人提供信息的储存、读取、下载等服务,实现随时随地安全存取数据和重要资料,分享一些重要文件资源,无需U盘,不用担心丢失问题,内容时时触手可及,具有安全稳定、海量存储等特点。现在网盘服务有很多,有些还支持在线播放、备份通讯录等功能,当前比较知名且好用的网盘服务商有百度云盘、360云盘、金山快盘、新浪微盘、腾讯微云等。网盘相对于传统的实体磁盘来说,更为方便,用户不需要把储存重要资料的实体磁盘放在身边,却一样可以通过互联网,轻松从云端读取自己所存储的信息。

以百度云盘为例,它提供文件的网络同步、备份和分享服务,具有空间大、速度快、安全稳定等特点。百度云盘目前有Web版、Windows客户端、Android手机客户端等多种版本,支持多个平台,用户通过注册,建立个人帐户,就可以轻松把自己的文件上传到网盘上,集中管理,并向不同设备推送数据,实现跨终端随时随地查阅和分享数据。百度云盘新用户注册的初始容量为5GB,通过各种活动和任务可以提升网盘空间,永久容量直至2T,若容量仍不足,可通过购买月费与年费的方式来获得更大网盘容量。百度云盘除了可以快速上传文件,智能分类,云端推送外,还支持免费离线下载,以及闪电互传等功能,充分展现云端存储优势,节省用户时间,极大方便了用户的学习和科研活动。

2.2 云笔记

云笔记是专注于简单高效的个人记事工具^[6],是个人云存储产品中的一类,具有云端同步功能。云笔记它偏重于文本,主要是以文字化的形式进行记录和存储。对于知识的再利用、组织和创新具有重要意义^[6]。用户可以在任何一台电脑和手机上使用,可随时随地编辑、搜索笔记内容,保存笔记,摘录网页、文件和图片。目前,主流应用的云笔记有为知笔记、有道云笔记、印象笔记、麦库记事等。

以为知笔记为例。为知笔记可以帮助科研人员将日常漫游网络、阅读所获得的有用信息碎片实时记录保存,比如有价值的网页或片断、随机的灵感笔记、重要文档、图片等,快速建立笔记、便签等,并进行条理化保存,设立多级目录、标签,实现多终端随时浏览、永久安全云存储以及群组分享等功能。为知笔记同样可以实现不同电脑、PAD、智能手机等终端之间云同步,相互之间便利地查看个人知识库。为知笔记在同步完成以后,即使在离线状态,也可以直接

打开客户端直接阅读笔记。为知笔记还有插件和模版机制,可以实现更多的功能;如创建日志、发布博客、订阅 RSS 等。对于使用者来说,为知笔记具备界面简洁舒适、运行速度快、安全稳定、全中文界面、友好度高、基础功能免费等优点。

很多人评价为知笔记为国内最好的个人知识管理笔记软件,为知笔记的特长是网页捕捉,能帮你记住一切想记的内容。有了它的辅助,科研人员可以做到将大脑清空,腾出更多的空间去思考和创造。

2.3 学术搜索引擎的云端应用

很多学术搜索引擎为用户提供了“我的图书馆”云服务功能,帮助用户将检索结果进行云存储,免去经常需要下载文献并使用 U 盘保存等步骤。如 Google 学术搜索、读秀等的“我的图书馆”或“个人数字图书馆”等。以 Google 学术搜索为例。Google 学术搜索可以免费搜索世界上绝大部分的学术文献,帮助科研人员在整个学术领域中确定相关性最强的研究,它的“我的图书馆(英文 My Library)”,是 Google 学术开发的一个可以在线收藏管理文献的应用,它可以帮助科研人员将搜索结果中筛选出来的有价值的学术文献,保存到“我的图书馆”,同时也可以将“我的著作引用情况”“我的更新”“订阅的‘快讯’”“统计指标”和设置保存到个人收藏,方便引用查找,这一功能,既可以收集、整理、维护自己发表的所有著作,还可以订阅自己著作的被引用情况,同样还可以订阅和跟踪感兴趣的最新发表的论文。谷歌学术快讯对于追踪最新论文非常好用,可以任意定制检索词,然后定期收到符合检索词的论文。此外,Google 的多语言支持,更是为用户提供了范围更为广泛的中外文文献搜索和文献管理保障。

同 Google 学术“我的图书馆”类似的,还有读秀的“我的图书馆”,它也是为用户开辟了一个属于个人的网上收藏空间,帮助用户将读秀等网站上的各种学术信息收藏进来,并可在这里发表文章、发布图片、上传文件等,创建各种专题知识库,还可与朋友交流,共享信息。

2.4 支持云端功能的专业文献管理软件

文献管理软件是一种具有文献检索与整理、引文标注、论文写作自动生成参考文献列表等强大功能的软件,可嵌入文字处理软件中使用,还可以直接通过在线数据库下载文献题录并对其进行统计分析,文献管理工具推出后迅速得到推广应用^[7]。文献管理软件主要分为 2 大类,一种单机版,也称桌面

版;另外一种是在线版(Web-based)。随着用户需求的不断深化,文献管理的重心正在慢慢由桌面版(Desktop)的软件向在线版发展^[8],或者拥有云端和客户端的共同设计,实现在线检索、下载、云同步、云备份、互动分享等在线功能,可以非常方便地管理全文、题录,可以边阅读边做读书笔记,随时记录思想火花,并对笔记进行管理等等。

支持云端功能的网络版文献管理软件,比较知名且应用较多的国外软件如 Endnote、Mendeley、PaperBox,国内软件如 Notefirst 等。这些文献管理软件基本上都具备网络检索、文献管理、文献引用、知识管理、文稿模板和文献分析等功能,帮助用户管理数目繁多的文献资料,加速科学发现与创新,辅助科研写作,借助在线文献管理,还可以实现云端和终端实时同步,随时随地访问,群组资源共享,同时成为专家学者展示自己的科研成果的有效平台,促进学术交流。

文献管理软件既能在日常文献的组织管理、论文写作中具有强大而灵活的功能,又具有网络功能,可以实现资源的共享,以及具备社会化的功能,让管理文献的用户能够建立起联系,促进沟通与合作。

除了以上 4 种类型存储工具外,还有一些云办公平台或系统,也同样具备专业数据管理、多终端任意共享、大文件快速分发、多屏互动协作办公等功能,如 Gleasy 格子云办公平台等。还有微软的 OneDrive 云存储服务,提供免费云空间,具有相册的自动备份功能、在线 Office 编辑功能、分享指定的文件、照片或者整个文件夹链接功能等。超星公司新近推出的云舟域空间知识服务系统,更是一款功能强大的云服务工具,集空间、工具、平台、资源、服务五位于一体,帮助用户创建独具特色的资料域,打造具有个人风格的学术域空间。

3 科研人员对不同类型文献管理工具的需求与选择分析

由以上介绍可以看出,目前可供科研人员选择的文献管理工具众多,4 种主要类型云端文献管理工具都能够实现文献资源的云存储、跨终端访问、群组分享等功能。它们在功能上除了这些共性,也各有侧重,有着不同的应用场景,适用不同的用户群体。

下面从存储容量、文献类型支持、文献收集范围、文献管理方式、文献使用目的等方面,对比云端文献管理工具的功能差异(见表 1)。

表 1 4 种云端文献管理工具的功能对比

工具名称/功能	存储容量	文件类型	文献收集方式	文献管理方式	文献使用目的
云盘	容量足够大	多媒体	上传	分类	存储/备份
云笔记	有一定限制	文字图片	上传/网页捕捉/桌面便笺	分类/标签/星标	笔记/隐性知识挖掘/知识创新
学术搜索引擎的云端应用	有一定限制	文字	检索结果收藏	分类/标签	学术文献搜集/引文管理
文献管理软件	有一定限制	文字图片	导入/在线检索	分类/标签/星标	论文写作/笔记/题录管理/引文管理

由表 1 不难看出几种云端工具的功能差异,科研人员要根据个人实际需要来选择和利用这些工具。

云盘:适合大型文件的长期保存,特别是适合音视频等多媒体资料的存储等。云盘更偏重于“藏”,在线搜集资料能力不强。

云笔记:适合碎片时间随时随地记录一些碎片信息、网页捕捉、个人记事,个人电脑和移动终端都能随时捕捉,导入文字、图片等各种类型文件,对格式要求不高,更适合知识管理,相对于云盘等其它管理工具,更加垂直细分,运行速度快,灵活性更强。科研人员在浏览网页时,及时地将所看到的有用的信息保存和管理起来,并加上一些简单的批注将当时的一些心得和感受及时的记录下来,对于知识的再利用、组织和创新具有重要意义。^[6]

学术搜索引擎的云端应用:适合在特定平台中收藏重要的学术资源,它的文献引用功能,能够帮助科研人员很方便地生成参考文献,不足之处是文献来源受到限制,只能在搜索引擎检索结果中选择。谷歌学术的“我的图书馆”,除了可以让用户即时便捷的分类保存自己需要的文献结果,还可以根据需要进行导出以扩展文献的使用。^[9]由于学术搜索引擎与其它文献管理软件的兼容性好,它可以很方便地将搜集来的文献资源导出转入文献管理软件,再进行整理排序,因此它可以作为文献的中转站来使用。

文献管理软件:对学术文献管理的专业性更强,特别适合参考文献管理及专业写作,提供专业应用模板,自动生成规范的参考文献格式,能够提高题录信息的处理速度,适合集中使用某一专题文献,但相较云笔记,对文献类型限制较严格,对多类型文献的整体综合呈现性差,对导入文献的格式要求较高,有些文献类型不支持导入。

在科学研究中,科研人员借助文献管理工具,全方位获取、保存、利用和交流文献资料,首先要弄清楚自己的文献管理意向和要求,如待管理文献的类型、文件的大小、文献语种、文献搜集的范围等,其次还要考虑个人的使用偏好、文献使用目的以及工具

使用成本、易用程度等等。

不同类型文献管理工具各有所长,优势互补,科研人员在选择时还可以将它们“混搭”,比如日常收集网页或其它零散信息可以选择云笔记,如果想集中一段时间利用相关文献进行论文写作,可以选择专业文献管理软件,谷歌学术的“我的图书馆”既可以收藏搜索结果,也可以作为文献收集的临时中转站,将收藏的文献导入 Endnote,与其它文献进行整合。如果需要存储或分享一些大型文件,特别是音视频文件,那么云盘工具是比较理想的选择。

4 科研人员使用文献管理工具的障碍

4.1 科研人员自身信息素质欠缺

科研人员缺乏对科学文献管理工具的了解及使用意识。有的科研人员习惯了原有的传统文献管理模式,不愿尝试新兴的云端管理工具,或者对云端管理方式信心不足,存在安全隐私方面的顾虑等。

4.2 文献管理工具服务障碍

首先是费用问题。虽然一些文献管理工具提供基础功能免费,但还有很多功能收费,甚至有些文献管理工具需要完全付费,造成用户使用上的成本负担,如果用户所在机构没有付费购买,用户根本就无法使用这些工具。此外,还有一些工具存在功能不够完善,如数据格式兼容性差,以及数据迁移障碍等。以文献管理软件使用为例,导入数据通常需要进行格式过滤,一旦格式不匹配或过滤器选择不当,就无法正常导入数据。还有当用户更换新的文献管理系统时,旧系统中的数据无法正常迁移到新系统中,使用户选择新系统进退两难,这些都是较常见的问题。

5 克服科研人员文献管理工具使用障碍的对策

5.1 充分调研,积极引进先进的文献管理系统

图书馆要向上级部门争取经费的支持,积极引进适合本机构科研用户的文献管理工具,搞好调研,了解本机构用户的使用需求,尽量选择用户使用范围较广,与其它系统兼容性好,运转相对稳定,安全性高,功能即时更新的系统。同时注意搜集各种免费系统,向他们推荐那些应用较广、获得好评较多的云端文献管理工具,帮助他们建立起自己的个性化文

献管理系统。

5.2 提升科研人员信息素养

高校和科研机构的图书馆员要对科研人员开展基于工具的科研支持服务,加强对文献管理工具的宣传推广,并有效开展使用培训,设置专门的咨询馆员深入到科研人员当中,开展专题培训,介绍文献管理工具的功能与使用方法,设置并演示现实使用场景,让他们了解这些工具的优越性,教会他们熟练使用,及时帮助他们解决在使用过程中遇到的各种问题。

5.3 持续追踪文献管理工具的发展,将最新相关信息及时推送给用户

当前,各种云文献管理系统发展很快,不断更新升级,各种新功能新应用频频推出,因此,图书馆应定期开展文献管理系统使用交流,将实时最新的相关信息传达给科研人员,让他们与时俱进,不断拓展和完善个人的“云图书馆”^[3]。

6 结语

“工欲善其事,必先利其器。”云存储、云同步和跨系统,已大大超出传统文献管理的范畴,是网络数字文献管理发展的大趋势。新兴云端文献管理工具,不但能够实现文献资源的云端存储与管理,随时获取,按需使用,还可以提速资源共享,加强科研协作,日益成为学术交流与共享的社会化平台,有着广阔的发展前景。开展深度有效的科研支持服务,帮助科

研人员占据在文献管理能力上的制高点,建立基于云端的“私人数据中心”,提升科研效率,是未来图书馆创新信息服务的重要内容,期待图书情报领域同行在此方面做出更加深入的研究和探讨。

参考文献

- [1] 李诗平. 基于 NAT 穿透技术的个人云图书馆构建[J]. 图书馆学研究, 2014(24): 2-4.
- [2] 佚名. 多屏开启个人云时代 [EB/OL]. IT 专家网. (2012-12-31) [2015-12-02]. http://m.ctocio.com.cn/cloud/139/12499139_m.shtml.
- [3] 驱动中国网. 万本好书随身带 爱阅读首创私人“云图书馆” [EB/OL]. e 线图情. (2014-08-01) [2015-12-15]. <http://www.chinalibs.net>.
- [4] 百度百科. 云存储 [EB/OL]. 百度百科 [2015-12-02]. <http://baike.baidu.com/view/2044736.htm>.
- [5] 百度百科. 个人云 [EB/OL]. 百度百科 [2015-12-02]. <http://baike.baidu.com/view/8112579.htm>.
- [6] 杨志萍, 杜瑾, 李红培, 等. 个人知识管理工具综述[J]. 知识管理论坛, 2013(3): 9-15.
- [7] 孙铭媚, 刘航, 戎誉. 常用文献管理工具的比较浅析[J]. 科技信息, 2014(5): 142-143.
- [8] 小木虫. 常用文献管理软件介绍 (桌面版 vs 在线版) [EB/OL]. 小木虫. (2009-09-20) [2015-12-02]. <http://e-much.net/html/200909/1555709.html>.
- [9] 林睿. Google 学术搜索的云端个人文献管理应用[J]. 农业图书情报学刊, 2014, 26(8): 46-48.

(上接第 87 页)

3 结语

本文提出了基于共享池的高校信息集成环境下的共享池精简框架。针对校园遗留数据中存在不含有主键的数据需要进行精简, 以及含有主键的数据需要进行字段精简、集成的现状, 本文构建了 2 套精简流程框架。

在高校信息系统中, 不仅存在着数据库重复记录的冗余数据, 还存在不同部门存储系统间的重复文件, 以及在数据备份的过程中由于全局备份产生的大量冗余数据。这部分重复数据的识别与精简, 被称为文件级的精简, 这将是下一阶段研究的课题。

参考文献

- [1] 郭志懋, 周傲英. 数据质量和数据清洗研究综述[J]. 软件学报, 2002, 13(11): 2076-2083.
- [2] 唐懿芳, 钟达夫, 张师超. 数据清洗前的预处理方法[J]. 广西科学, 2005, 12(2): 118-122.
- [3] 梁文斌. 数据仓库中数据清洗的研究与设计[D].

苏州: 苏州大学, 2005.

- [4] Sudipto Guha, Nick Koudas, Divesh Srivastava. Reasoning About Approximate Match Query Results[C]. Proceedings of the 22nd International Conference on Data Engineering 2006.
- [5] HuiXiong, Enhancing Data Analysis with Noise Removal [J]. IEEE Transactions On Knowledge and Data Engineering, 2006, 18 (3): 304-320.
- [6] 刘芳, 何飞. 基于聚类分析技术的数据清洗研究[J]. 计算机工程与科学, 2005, 27(6): 70-77.
- [7] 韩京宇, 徐立臻, 董逸生. 一种大数据量的相似记录检测方法 [J]. 计算机研究与发展, 2005, 42 (12): 2206-2212.
- [8] 戴东波, 汤春蕾, 熊冀. 基于整体和局部相似性的序列聚类算法[J]. 软件学报, 2010, 21(4): 702-717.
- [9] 胡洪宁. 聚类算法的分析与研究[D]. 西安: 海军工程大学, 2005.
- [10] Zhou Zhengda, Zhou Jingli. High availability replication strategy for deduplication storage system. Advances in Information Sciences and Service Sciences. May, 2012(8): 115-123.