

基于无尺度网络的互联网信息播存体系研究

王恩海^{1,3,4},李幼平²

(1.中国科学院计算机网络信息中心,北京 100190; 2.北京应用物理与计算数学研究所,北京 100088;
3.中国互联网络信息中心(CNNIC),北京 100190; 4.中国科学院研究生院,北京 100049)

摘要: 本文提出并论证了一个相对宏大的工程——构造互联网信息播存体系,即:将互联网上的优秀文化集合打包,定时将这些优秀文化发射到卫星上,再通过广播系统将这些文化下发到各地的本地信息存储服务器,并将最终用户选择的信息主动下载到用户的信息终端上。

关键词: 无尺度网络;播存体系;互联网;文化传播;三网融合

中图分类号: TP393.03 **文献标识码:** A **文章编号:** 0372-2112 (2011) 04-0737-05

Study on Internet Information Broadcast-Storage System Based on Scale-Free Network

WANG En-hai^{1,3,4}, LI You-ping²

(1. Computer Network Information Center, CAS, Beijing 100190, China;
2. Institute of Applied Physics and Computational Mathematics, Beijing 100088, China;
3. China Internet Network Information Center, Beijing 100190, China;
4. Graduate University of the Chinese Academy of Sciences, Beijing 100049, China)

Abstract: The paper brought forward and demonstrated a great engineering project: to build an Internet Information Broadcast-Storage System (IBSS). That is, packaging the excellent information on the Internet, sending them to the satellite periodically, broadcasting them to many local storage servers from the satellite, finally pushing them to the end users who have ordered them.

Key words: scale-free network; broadcast-storage system; internet; culture diffusion; network convergence

1 引言

目前的互联网已发展成为内容无所不包的巨大信息库。据中国互联网络信息中心(CNNIC)统计,到2009年底仅中国就有323万个网站,336亿多个网页,这些网页上存有近1060TB的信息量(不包括后台的数据库及音视频内容)^[1]。在近几年,Web2.0的兴起使得网上信息量增长迅猛,内容也异常丰富。这些信息在不同地点、不同时刻被3亿多网民访问着。随着网民、网页规模以及访问量的急剧扩张,互联网逐渐显现出一些问题:因网络带宽问题带来的网速慢、不稳定,因中国互联网普及程度不高、地区经济发展程度和教育水平参差不齐而带来的数字鸿沟问题,等等。这些问题仅通过在互联网领域采取各种补丁措施是不能根本解决的。而对于另一种信息传播系统广播网,在信息下播时不存在带宽瓶颈问题,也无需顾及受众的多少,也无需顾及受众是在农村还是城市。于是本文提出了互联网信息播存体系,将

互联网上的信息通过这个“播存体系”达到终端用户。这个体系将是互联网与广播网的融合,发挥他们各自的优点,弥补各自的缺点,见表1。

表1 目前的互联结构与本文的播存结构的比较

	互联结构	播存结构
用户规模	受限	不限
内容种类	不限	受限
服务类型	全面服务	文化共享服务
服务范围	城镇为主	城乡全部
信息服务方式	被动待取	主动到家
信息通道	多级路由	无路由
角色扮演	信息公路	信息阳光

互联网信息播存体系的大致工作原理如图1所示:根据互联网的无尺度特性^[2],把受众大、点击量高的网站上的信息整合起来,添加一些信息属性,然后将这些信息发射上星,再通过广播把信息根据需求自动发送到遍布全国的本地存储服务器,最后由用户直接从离自己就近的本地服务器上检索自己需要的信息。

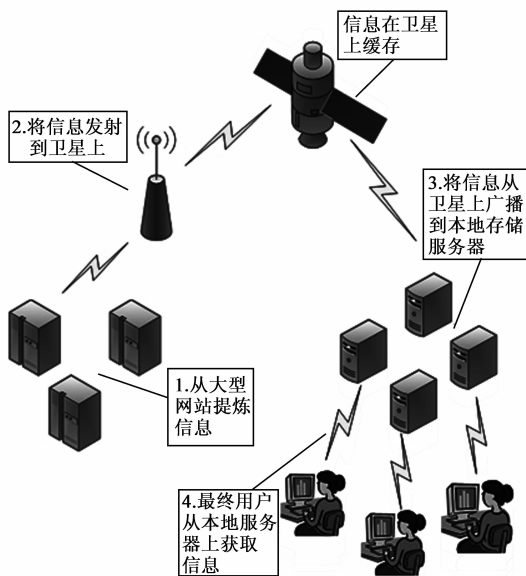


图1 互联网信息播存体系示意图

通过这个播存体系,每天就可以把互联网上当天新产生的大部分信息通过卫星传递到了用户手中,满足多数人享用多数信息的要求,从而避免了用户通过电信线路和层层路由获取信息的种种不利,比如带宽拥挤、信息质量、信息安全等。

2 互联网信息播存体系的意义

2009 年底我国互联网用户的普及率只有 28.9%^[1],刚刚超过全球的平均水平,与互联网发达国家冰岛、美国、日本、韩国等差距很大,这些国家的普及率均已超过 70%。从长远看,我国互联网还有非常大的发展空间。而一旦我国互联网发展到较高水平,目前的互联网服务就会暴露出很多问题。本文提出的这个互联网信息播存体系就是对将来互联网上大信息量、大用户群提供高质量服务的前瞻性研究。

2.1 提炼出我国的优秀文化

党的十六大报告中指出:“互联网站要成为传播先进文化的重要阵地”。但在传播之前首先要找到这些先进文化,这也是互联网信息播存体系的基础。利用互联网的无尺度特性,将网民访问高的网站、网页看成是信息的集散节点,这些集散节点是通过网民的自发访问行为民主产生出来的,具有高度的代表性和先进性,这就是符合我国人们需求、具有中国特色的优秀文化,我们应倍加珍惜并广为传播。

2.2 解决网络拥堵与带宽瓶颈

这是提出这套体系的基本初衷。随着网民的大规模增加和对互联网访问的日益频繁,作为网络通信的带宽资源将越来越紧张。虽然网络带宽可以不断扩大,但一来带宽不可能无限制的扩大下去;二来随着网上内容的丰富,特别是视频内容的增加,网民对带宽的需

求越来越大,带宽的扩大速度将跟不上网民需求的增长;三是带宽扩充的成本会逐渐增高,因为在目前的物理光纤信道用尽时,必须新铺设通信光纤。这些问题在采用信息播存体系后将自然消解。网络信息资源的发射上星推送能力可轻松满足网民的信息需求,每天只需推送当天网上新增的内容量。再者,从卫星广播到地面的接收器——本地存储服务器,也是无用户数量上限的,而且没有传输速度问题。最后,当用户从本地存储服务器上取信息时,由于是到就近的服务器上取,而且这些本地服务器都是服务于很小范围内的用户,比如一个村,一栋楼,一个社区等,用户数很少,同时采用 P2P 技术,所以也基本不存在传输拥塞问题。这样,从信息的上传到信息到达终端用户,可以说是一路畅通,很好地解决了目前互联网上的网络拥塞和带宽瓶颈问题。

2.3 改变信息服务方式

当前我们对网上信息的获取,主要是用户去网上找。在信息播存体系中,用户首先通过本地服务器上的信息服务平台来选择自己当前或将要感兴趣的内容来进行信息订制。这种订制操作非常简单,就像做多选题一样。保存好自己的订制后,本地服务器就会每天通过对从卫星上接收到信息进行检索和筛选,然后将符合用户订制需求的信息全部主动推送到用户的信息终端上。在整个过程中,用户不需要知道信息具体所在位置,而且能一次就收到定制的所有最新信息,并自动保存在用户自己的信息终端里。这种由用户每次去查询一条信息到用户自动就能接收到多条信息的服务方式的转变,是一种信息服务的进步。

2.4 提高信息质量

目前网上的视频内容开始丰富,将来会更加丰富。但遗憾的是,为了使这些视频便于下载,减少网络流量,缩短下载时长,减小用户的下载成本,内容提供商不得不先将视频质量(包括画质和音效等)大大调低以便减小该视频文件的大小。这样用户得到的信息质量很差,甚至不完整。但使用播存体系后,完全可以保真传输,这样用户得到的信息质量自然高,用户体验会更好。

2.5 降低互联网的进入门槛

据 CNNIC 调查显示,目前影响人们上网的主要因素是没有上网技能和没有上网设备。对播存体系来说,这两方面会有很好的改善。用户上网用的电脑与播存体系中用户的信息终端(比如网络计算机 NC)相比,价格要高出许多。因为任何显示设备(包括家用电视机)只要添加一个机顶盒和鼠标键盘就可以成为一个信息终端,而不一定必须是一台计算机,所以手机、电视机

用户很容易变成互联网用户.播存体系的用户只需要会选择自己感兴趣的话题即可,不需要知道操作系统,甚至不需要技术水平.这种情况下,受教育水平不再是用户获取网上信息的障碍,而购置信息终端的成本也大大降低.

2.6 助力社会主义新农村建设,缩小数字鸿沟

据 CNNIC 统计,城镇居民的互联网普及率是 35.2%,而农村仅为 11.7%,可见城乡间的数字鸿沟非常明显.但从电视、广播普及率来看,2008 年末广播综合人口覆盖率为 96%,电视综合人口覆盖率为 97%.如果通过信息播存体系来传递互联网信息,则不论在信息普及便捷程度还是信息终端的普及程度,还是信息的丰富程度和到达范围,都将大大提高,实现优秀文化的上山下乡,从而有效地缩小数字鸿沟.

2.7 改善广播服务

随着现代化水平和人们生活需求的提高,目前的广播系统面临很多问题.首先,它难以实现按需服务.内容与时间安排完全由电台决定,用户对信息资源不能双向互动,缺乏自主权.其次,提供的资源品种不够丰富,尤其是缺少静态视觉的书刊、网页,不能反映数字文化的博大精深.但在与互联网结合,采用互联网信息播存体系后,这些问题都将迎刃而解,互联网上内容的多样化、丰富性,以及与信息交互性,将大大加强广播系统的服务能力.

2.8 促进三网融合

在 2006 年公布的“中华人民共和国国民经济和社会发展第十一个五年规划纲要”中强调,要积极推进“三网融合”,建设集有线、地面、卫星传输于一体的数字电视网络,促进互联互通和资源共享.2007 年 10 月党的十七大报告中也提出,要“运用高新技术创新文化生产方式,培育新的文化业态,加快构建传输快捷、覆盖广泛的文化传播体系”.本文提出的互联网信息播存体系力图实现互联网、广播网和电信网的互补与和谐整合,是对国家战略规划积极响应和尝试.不仅可以成为和谐的文化传播体系的一个组成部分,而且可以为“三网融合”发挥积极的促进作用.

3 互联网信息播存体系的工作过程

3.1 提炼互联网上的优秀文化

要想传播信息,首先要选择信息、整合信息,这是信息播存体系的第一步工作.由于我国网民对互联网网站访问遵循无尺度网络理论,所以提炼我国互联网上的优秀文化的步骤如下:

第 1 步 考虑到提炼的信息必须丰富全面,先将所有网站分类,比如政府、综合门户、新闻、搜索引擎、即时通信、电子商务、网络游戏、视频分享、财经、房地产、

健康、教育、科研、报纸、期刊、杂志,……等等.由于目前互联网呈现领域细分趋势,所以这些分类可以细些,设分类数为 C (例如令 $C = 100$,就是将我国所有网站分为 100 类).当然这种分类很复杂,本文不展开讲述.

第 2 步 定期对我国所有网站的日均访问量进行统计,并归到这 C 个分类里,即可得出每一类里包含的网站数 $c_i (i = 1, 2, \dots, C)$ 以及每家网站的日均访问量.

第 3 步 按照无尺度网络理论^[3],我们定义大型集散网站占所有网站的比例为 p (初步设 $p = 1\%$,为常数),这些大型集散网站即可认为是我国的优秀网站,则我国优秀网站的数量 $W = N * p$, N 为全国所有网站数.按此计算,到 2009 年底,我们可以提取的优秀网站数 W 为: $323 \text{ 万} * 1\% = 3.23 \text{ 万家}$.

第 4 步 计算出每类网站里需要提取的优秀网站数 $w_i = c_i * p$.由于我国在互联网有些领域里的服务企业数很少,也许按此方法计算出的在该领域里需要提取的优秀网站数会小于 3,甚至小于 1,在这种情况下,我们令 $\text{if } w_i < 3 \text{ then } w_i = 3$.比如:搜索引擎在我国互联网一个非常大的分类领域,但目前我国搜索引擎服务提供商只有几十家,这样算下来,要在搜索引擎类里提取的优秀网站数就不到 1 个,但我们仍会强制选择其中的前 3 家作为这个细分领域里的优秀网站.

第 5 步 根据每一类中的网站按日均访问量的排序,得到该类网站中的优秀网站列表,也就是这类网站中排名前 w_i 位网站.

第 6 步 将所有类中的优秀网站整合起来,就得到了按照日均访问量整理出的优秀网站列表,即: $T = \sum_{i=1}^C w_i$,此时 T 为一个网站列表,这个列表中的网站数量就是 W .

第 7 步 有些内容很偏、或者专业性很强,但内容很不错的网站的访问量很小,不可能进入排名全国前 W 位,这种情况主要是由于这些网站的读者群很小.我们可以单独补充进优秀网站的队列里,假设有 S 家这种网站被补充进来.

第 8 步 这样,我们就得到了我国互联网上的优秀网站 $G = W + S$ 家.这 G 家网站上的信息就是我们要找的互联网上的优秀文化.

由于是按照网站的日均访问量来提炼优秀网站,所以这些被提取出来的网站实际上是我国网民民主选举出来的,是大家公认的好网站,所以称其为优秀网站也是理所当然、顺应民意的事情.这些访问量较大的优秀网站上的信息整合为代表整个互联网的优秀文化.

3.2 优秀文化上星

提炼出互联网上的优秀文化后,就可以将其发射上卫星了.但在此之前,还要略做些准备工作.

第 1 步 如果是整个互联网信息播存工程启动时的第一次上传,则将上节中选取的所有 G 家网站上的所有信息全部提取出来;如果不是第一次上传,则只把这 G 家网站上当天新增的信息提取出来。

第 2 步 将所有提取出来的信息以文件(如网页等)为单位,为其加上一个标签.这个标签我们可以称之为 UCL(Uniform Content Locator),是与网页在互联网上的 URL 相对应的一种信息定位器,用于记录网页的一些属性,比如来源,属于哪类内容,内容的可读性级别(如色情、暴力级别等),版权问题,免费还是收费,生产时间,等等。

第 3 步 将加完标签的所有待发射信息通过互联网或电信网专线瞬间送到广播中心,进入广播中心的优秀文化数据库。

第 4 步 广播中心采用并播技术,从数据库中有序地取出数据,通过单一载波,以适当的速率和适当频次的重播,逐渐将所有待传信息发射上星,并在卫星中缓存,等待向地面分发。

西南科技大学和绵阳九洲公司已经在 2004 年度内完成了大规模码分并播研究的地面实验工作,证明通过统一内容定位器 UCL,可以有条不紊地把数千种传媒整合在一个 DMB(Digital Multimedia Broadcasting)载频上,地面用户可以根据 UCL 按自己的兴趣选择任意数量的部分传媒下载镜像.这种关于并播协议的尝试说明播存结构在技术上是可行的。

2004 年 6 月,清华大学力合公司通过卫星进行差分并播定比实验,判断这种可行性^[4].实验和计算结果令人鼓舞:一个传输速率大于 32Mbps 的卫星转发频道,每 24h 有能力把超过 345GB 的网页直接送达全国城市的每一栋大楼和农村的每一个自然村.而我国网上信息量平均日增量在 2007 年仅为 208GB.这足以说明利用信息播存技术进行互联网新增信息的广播是没有任何问题的,广播的推送能力是绰绰有余的。

3.3 优秀文化落地与存储

广播中心通过同步卫星转发,将互联网上的优秀文化直接落地于事先部署好的、遍布全国城乡的本地存储服务器.中途没有路由转接,没有信道拥堵,也没有用户数的限制;不分先后顺序,瞬间、全部到达.具体步骤为:

第 1 步 卫星定时将缓存的、从地面发来的互联网信息转发向地面接收装置——本地存储服务器。

第 2 步 本地存储服务器通过码分并存技术,按照 UCL 对信息进行分类、过滤,将本地存储服务器上所有用户订制的信息及时接收下来,并存储。

第 3 步 根据本地服务器中的各个用户订制的不同信息,本地存储服务器在用户开机时将用户所需信

息经 UCL 过滤,主动传送到用户的信息终端.传送方式可以是局域网、光纤等,由于这个本地服务器的用户数量很少,所以不论采用哪种方式,都不会存在带宽瓶颈问题,传输速度都会比直接从各网站上访问要快的多、内容也要丰富的多且更具针对性。

用户的信息终端一般可使用网络计算机(NC, Net Computer).网络计算机(NC)是基于网络环境的跨平台的网络终端设备,其明显优势是价格低.通过远程桌面显示协议(RDP, Remote Desktop Protocol)与运行多用户的服务器通信.NC 不装任何操作系统、应用软件等,也没有硬盘、光驱之类的存储设备,只是在启动 NC 时选择所需要的操作系统并启动,然后通过鼠标、键盘调用相应的应用软件,访问已经存储在终端服务器里的信息资源,服务器接到请求后把处理结果传递回客户端显示.众多 NC 可以同时登录到服务器上工作,它们之间是相互隔离的^[4]。

目前网络计算机已被广泛应用于电信、银行、税务、教育、办公自动化、电子政务、图书馆管理/电子阅览、窗口行业、医疗等诸多行业,也成为信息播存体系的用户终端设备首选.通过这个终端,用户无需感知信息资源的具体所在,便可以得到多种信息资源的按需服务^[5]。

4 现实中的播存体系应用实例

鉴于信息播存体系的种种优点,国内已经有企业开始尝试了.北京市科委已在资助北京歌华有线数字媒体有限公司开展“大平台”的工程研发,力图将几百种电视栏目、几百种热门网站,自动组织在单一频道上,每天播出超过 300GB 的内容.北京星线空间信息技术有限公司(以下简称星线空间)则已利用这种播存技术,成功开发了多种产品,并且取得了不俗的业绩.以下关于星线公司的业务及技术原理,大部分根据星线空间网站(www.n8vip.com)上的信息整理。

星线空间是国内一家卫星宽带网络电视平台运营商,向社会提供建立在“播存双结构”理论上的卫星宽带数据接入业务.通过长期租用中国卫星通信集团公司的亚太 2R 卫星的 10M 广播带宽,依靠自有专利的“移播”技术,24 小时不间断地为全国网吧传输视频和数据等内容,为网吧等客户提供全新的互动视听新体验.用户不仅可以在电脑上观看无延时、长连续、高品质的直播视频画面,还可以随意点播想看的节目,快进快退,自由自在.遍布全国的网吧服务器经过逐日积累,形成自己的视频库,在低成本的前提下实现没有带宽瓶颈的视频点播(VOD).目前,星线空间已在全国 100 多个城市发展网吧用户 5000 余家。

星线空间目前每天可向网吧分发超过 80G 的数字

娱乐内容,作为播存结构的起点,星线空间非常注重这些数字内容的丰富和取精,主要分发内容有电影、电视、电子杂志等.在将这些数字内容整理好之后,通过 10M 的卫星带宽发射上星,由亚太 2R 卫星转发向各网吧服务器.网吧通过高频头、卫星 IP 接收天线、卫星 IP 数据接收机等设备将卫星信号经过解频、放大、去除杂讯等,变成计算机可以识别的信号.再通过星线空间开发的软件将这些信号接收、存储、转发、编入点播菜单等后续工作.

星线空间作为国家文化创意和技术创新的成功企业,得到了文化部、广电总局、北京市政府、中国工程院、清华大学、中关村高科技园区等政府和学术机构的肯定和支持.如果把星线空间的可行经验,上升成为国家主导的播存结构,不仅播给网吧,还播给全国的 ISP,播给大大小小的局域网服务器;不仅播送电视内容,还播送全部主流网页,这就形成了完整的互联网信息播存体系.这个体系形成后,互联网上流量的一半甚至大半将由播存结构分流,视频分享引起的网络带宽冲突随之化解,中国数亿网民将由此受益.

5 互联网信息播存体系的缺陷

虽然本文提出的互联网信息播存体系有很多优点,但也不可避免地存在一些缺陷.与当前的互联网相比,至少存在两方面的明显不足.

(1) 缺乏互动性

互联网有一个非常突出的特征就是互动性,这也是互联网活力的表现.但在这个播存体系中,信息基本上是单向流动的,用户只能接受信息,用户对信息的评论、反馈以及与全国、全球用户的交流,都难以实现.

(2) 内容不能满足全部需要

这个体系的设计思路出发点就是信息取精,所以被传上卫星的信息只是互联网上的大部分主流信息,并非全部.也就是说,互联网信息播存体系只能是满足多数人的多数需求.显然,有少数人没有被服务到.我们只上载了比例极小的一部分网站(也许只有 1% ~ 2%)上的信息.

6 结束语

本文在提出了构造互联网信息播存体系的思路之后,围绕互联网上优秀文化在信息播存体系中的流动,对该体系中的各个环节展开可行性论述和具体操作步骤分解.虽然互联网信息播存体系目前只是一个前瞻性学术研究课题,但是这个课题已经引起各方面研究机构的重视,比如中国工程院信息学部启动了“播存网

格工程构思”的咨询研究;清华大学、中国科技大学、西南科技大学等高校和一些企业也开展了这方面的理论研究和相关技术的试验测试,并取得了一些重要进展.

当然,这个播存体系的推广应用将是一项艰巨的工程.从理论研究到技术设计、测试,都决定了这将是一个长期的、逐步完善的工程.在这个过程中,需要人才队伍的锻炼,需要国家政策的支持,需要研究资金的支持,需要商业模式的探索.

参考文献

- [1] CNNIC. 中国互联网络发展状况统计报告(2010/1)[R]. 北京:中国互联网络信息中心,2010.
CNNC. Survey Report on The Internet Development in China (2010/1) [R]. Beijing: China Internet Network Information Center, 2010. (in Chinese)
- [2] Barabasi A L, E Bonabeau. Scale-free networks[J]. Scientific American, 2003. 288(5): 60-9.
- [3] Albert R, A L Barabási. Statistical mechanics of complex networks. Reviews of Modern Physics[J]. 2002, 74(1): 47-97.
- [4] 车晴,王京玲. 数字卫星广播系统[M]. 北京:北京广播学院出版社,2000.
Che Qing, Wang Jing-ling. Digital Satellite Broadcasting System[M]. Beijing: Publishing House of Beijing Broadcast Institute, 2000. (in Chinese)
- [5] 张尧学. 透明计算:概念,结构和示例[J]. 电子学报, 2004, 32(12): 169-174.
Zhang Yao-xue. Transparent Computing: Concept, Architecture and Implementation[J]. Chinese Journal of Electronics, 2004, 32(12): 169-174. (in Chinese)

作者简介



王恩海 男,1972 年出生于山东沂水,高级工程师,博士研究生,主要研究方向为信息统计与发布技术.

E-mail: wangeh@cnnic.cn



李幼平 男,1935 年出生于福建泉州,中国工程院院士,通信技术专家.

