

共享信息的第二类网络

李幼平

(中国工程物理研究院, 四川绵阳 621900)

[摘要] 基于 Internet 的万维网是当代共享信息的原创网络 (primary web), 它是一种隐含冲突的系统, 难以摆脱带宽瓶颈的困扰, 不支持用户数量的自由增长。提出的第二类信息共享网络 (secondary web), 通过广播镜像把许多网站的完整内容收存于家庭, 然后按内容的语义直接提取满足个性需求的读物。第二类网络支持用户数量自由增长, 是一种低成本、无冲突共享知识的新传媒, 它不仅没有带宽瓶颈, 而且避免了信息垃圾。作者认为, 第二类网络将同万维网互补互助, 共同构成二元化信息共享架构。

[关键词] 万维网; 第二类共享网络; 内容的语义; 带宽瓶颈

[中图分类号] TP393.03 **[文献标识码]** A **[文章编号]** 1009-1742(2002)08-0008-04

1 万维网难以摆脱带宽瓶颈

基于互联原理的万维网 (WWW), 支持信息的按需取用, 拥有全球丰富的内容资源, 是共享信息的当代工具, 是人类在 20 世纪末的一项伟大发明。创始人 T. B. Lee 的意图: Web 或 WWW 是“信息进出自如、体现人类知识的网络天地”^[1]。这种仅仅依托互联网单一网络的共享结构, 当网民、网页急剧增长时, 出现了进出经常拥挤堵塞 (带宽瓶颈)、展现庞杂无序 (信息垃圾) 的局面。

带宽瓶颈 把带宽分配给用户是万维网实现按需取用的基础, 网民增长到某一程度, 带宽不够分配, 发生相互妨碍的共享冲突, 要么大家交换信息的速度都变慢, 要么有人上不了网, 这就是带宽瓶颈。万维网是一种潜在共享冲突、不支持用户数目自由增长的网络结构。

信息垃圾 网页增长到天文数字, 无用网页远多于有用网页, 仅依靠网民找网页的被动待取机制, 很难在浩如烟海中很快找来确实有用的内容。

寻求一种非 Internet 的、无共享冲突的、网页也能主动找网民的第二类共享网络 (secondary

web)^[2], 来解决全民低成本共享知识与文化的问题。

2 信息共享本该无冲突

从基础研究的角度看, 信息共享同物质共享存在本质差异。物质共享的本质是母体的分配或瓜分, 信息共享的本质则是母体的复制或再生, 即重新产生许多一模一样的信息。由于复制的份数没有上限, 享用的人数也应当无上限, 共享信息本当无冲突。然而, 当前在信息界中占主导地位的是信息公路的思维, 这是一种物流的思维, 它把信息共享同带宽分配等同起来, 认为车多路堵、人多网堵是天经地义的。实际上, 带宽是一种物质资源, 不管怎样分配, 顾此失彼的冲突是难免的。互联网中把带宽分配给用户, 实现了按需取用, 但用户数目受到限制; 广播网把带宽分配给内容, 用户数目可以自由增长, 但内容品种受到限制, 无法实现按需服务。可见, 只有摆脱物流的思维, 接受信息共享本当无冲突的理念, 才能找到一种用户数目和内容品种都各尽人意的、全新的信息共享结构。

现代科学技术已经提供了两种复制比特的途

[收稿日期] 2002-03-26; **修回日期** 2000-04-05

[基金项目] 中国工程物理研究院重点基金资助项目 (97ZD044)

[作者简介] 李幼平 (1935-), 男, 福建泉州市人, 中国工程院院士, 中国工程物理研究院研究员, 博士生导师

径：利用点对面的电磁辐射，在空域上复制比特；另一种是利用非易失的记忆存储，在时域上复制比特。这两种途径复制 1 个比特（1 b）所需的理论能量极限 E_{bmin} 都是 $0.693 kT$ ，其中 k 为波兹曼常数， T 为温度。在常温下，复制 1 b 只需能量 $< 10^{-20}$ J，若给十几亿中国人每人复制 100 MB 的信息，加起来也只需能量 < 1 J。创造无冲突的信息共享系统，在技术上的可行性也正日益显现。

微波低噪声接收技术正在向理论极限逼近，在 11~14 GHz 波段上，采用 DVB-S 数字广播技术，用 1 个百瓦级的卫星转发器，就可以完成对 960×10^4 km² 的中国大地每天撒播 100 GB 以上的文化信息。光、磁、电非易失性存储密度距离理论极限虽远，但年年都有成倍增长，一个能存储几万份报刊、书籍的 40 GB 硬盘售价仅 700 元人民币，在家庭营造海量文化仓储的年代已经来到。

人们早就单独使用信息的空域复制或时域间复制，只不过尚未把两者整合，用于解决带宽分配顾此失彼的两难问题。我们的解决方案：

1) 用广播办法把带宽分给内容，营造无冲突共享的“信息阳光”^[3]，回避用户数目受限问题；

2) 借助家庭数据仓储，无中断地从数量上积累不同时刻播出的内容，既解决广播方式内容受限问题，又解决双向互动缺少中介介质的问题；

3) 把家庭仓储当成许多网站的代理服务器，它的出口带宽只供 1 个用户使用，带宽瓶颈问题也就自然消失了。

这就是两种复制并举、通过家庭仓储实现信息无冲突共享的理论设想。

3 存网站于家庭日趋现实

上述理论设想能否发展成一种商业运行的应用方案，主要取决于家庭仓储能否通过空域和时域的复制，积累出数量足够丰富的文化内容，以满足用户个性化的自由选择。为做出判断，有必要对国家的数据广播能力和家庭的数据存储能力做出评估。

新华网、新浪网之类的大型新闻网站，每天都要增加几千个新鲜网页，但加起来的日产量很难超出十几兆字节。杂志、期刊网站的平均日产量更低一些。全国拥有最多读者的 1000 个网站加起来，当天日产量不会超出 10 GB。无论是 DVB-S 或 DVB-C 的一个频道，比特率都超过 32 Mb/s，即每天推送的字节量超过 300 GB。10 GB 只是区

区小数，一天之内可以重复广播几十遍。我国已经利用 DVB-S 技术实现贫困山区的村村通，并拥有全球最大的 CATV 资源，但是当前只用来传播电视与音频节目，应该通过技术创新，把这些宝贵的国家资源发展成为存知识于民间的强大工具。

每个家庭在 1000 个网站中选择二三十个网站作为镜像下载的对象，把这些网站播出的内容不加选择地收存在硬盘中，日积月累形成规模宏大、内容丰富的家庭文化仓储。30 个网站相当是 30 种报刊，足以满足一般家庭文化需求，而日存量最多不超过 300 MB，累积 100 天也不会超过 PC 机主流硬盘的容量。即使超过了，还可以吐故纳新。

整合广播与存储，不难形成一个内容十分齐全、网页自动推陈出新的家庭文化仓储。

4 并行播存的改造与升级

中国有条件率先实现家庭文化仓储的理想。国家广电总局 DVB-S 村村通平台、国家教育部 IP-DVB 平台和国家数据广播中心 DVB-C 平台已为营造第二类共享网络提供了主要的物质基础，只要进行大规模并行广播和家庭镜像收存（简称并行播存）的改造，就可实现从传统广播到新型共享网络的升级，参见图 1 所示的实例。

首先在带宽分配上创新。不同于万维网中把带宽灵活分配给用户，也不同于广播电台把带宽灵活分配给内容，改造后的新网络把带宽永久固定地分配给各个网站（ICP），用这种办法把上千种报纸、刊物、课程、新闻网站整合起来，形成有相当规模的文化群体，实现大规模并行广播。把 32 Mb/s 的带宽均匀分配给网站，每个网站将独立拥有 32 kb/s 数据传播能力，即每天可推出超 300 MB 的内容，直接到户、永不中断。带宽永久地分配给网站，广播电台失去了干预节目的能力，无形中取消了节目时间表，捆绑电台与读者的时间绳索也顿然消失。掌握各子信道带宽的 ICP，根据推陈出新的原则，以一定的周期地重复轮播一批又一批新鲜网页，每个网页尽管没有明确播出时间，但都有多次重复播出的机会。终端接收装置 24 小时永远在线，若一次播出没有被正确收存，可在下一次重播时补上，在用户端做到：凡曾广播，我都拥有。

独立并行的运作将把 ICP 培育成独立的经济实体，它不仅拥有自己的子信道地址（PID）、独立确定播出内容和轮播顺序，而且拥有独立定价和

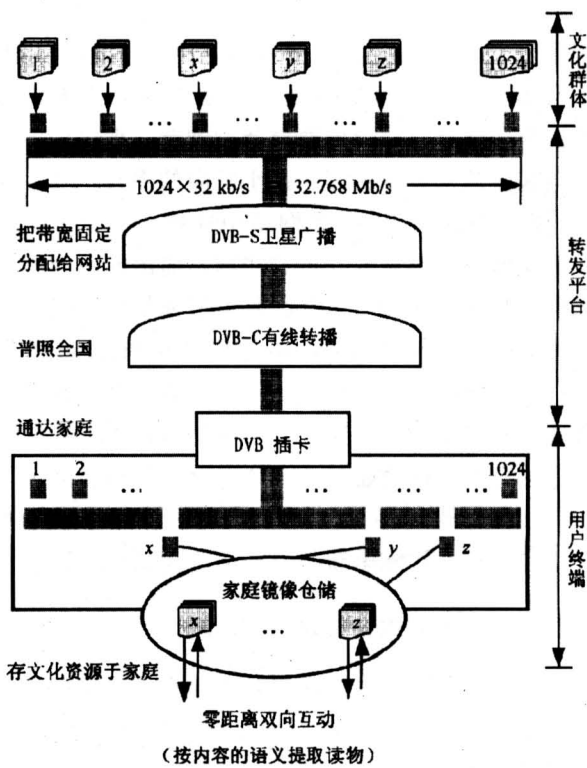


图 1 包括 1024 网站的文化平台示意图

Fig.1 A cultural platform containing 1024 Web sites

获取正比于用户数目的经济回报渠道，营造一个公平价值竞争的社会主义文化市场。

DVB-S 卫星广播和 DVB-C 有线转播作为国家信息基础设施，它的任务是用最低的成本把知识和文化撒播到城乡的每一个家庭，用最快的速度把最新鲜的信息存入家庭仓储。

边远山区用 11~12 GHz 波段的村村通接收机通过 DVB-S 数据接入卡在家庭范围复现文化群体，城市则采用 DVB-C 或 DVB-T 接入卡接入。

家庭镜像收存也是一种创新。其用意是让家庭收存的内容同网站拥有的内容保持动态一致，就像平面镜前后的实物与虚像，始终是一模一样的。它不是逐个网站、逐个网页串列下载，而是同时监视多个子信道，不管那一个子信道出现尚未存入的新网页，都会自动存入。接收设备不漏掉指定网站的任一网页，完成存网站于家庭的使命。

从我国的村村通工程的经验和中央教育电视台 IP/DVB 实验的结果推测，用大规模并行广播和家庭镜像收存的思想改造现有广播设施，营造若干个普照全国的国家文化平台，把几千种报纸、杂志、

网站、课程和音频节目直接送达千家万户，只需要十几亿人民币。然而这种存知识于民间的战略效果是难以用金钱估量的。

5 按内容的语义提取网页

家庭仓储把网页事先收集于用户身边，给人网界面的变革提供了契机。人们希望能按内容的语义(涵义)表达需求，一次拿到成批有兴趣的网页。万维网做不到这一点，需要靠搜索引擎先找来一批网页的 URL，然后到远处一个个取来读物。为了实现按内容的语义直接拿到读物，需要在广播网页上附加一个代表网页涵义的统一内容定位标引(UCL)，说明该网页讨论的话题，属于类别，栏目、标题、关键词是什么^[4]。UCL 作为编者与读者之间的一种语义学的约定，让供求双方在家庭仓储中双向选择，判断是否符合有用网页的标准。读者只要按 UCL 格式事先在 PC 机中设定自己的意图，软件将对源源流入的网页一一做出甄别，是否应该主动提交主人阅读。这样，既可把信息垃圾拒之门外，又解决出现了用户有兴趣的网页苦于没有主动通知的渠道问题，做到网页也能主动找网民。

UCL 的格式至少包括 6 个字段，前 2 个属于外延，表达网页的出处与投入时间；后 4 个属于内涵，分别表达大类、栏目、标题和关键词。出处与入网时间是客观数据，用简单的程序便可自动形成；栏目、标题和关键词则采自原万维网站的主页和页面属性，通过专用的映射程序自动转换形成。需要各网站统一的只是大类的语义规范(像政治、经济、科技、体育、文艺等类的划分)，需要得到国家标准机构的推荐。

UCL 的格式是一种相对稳定的俗成约定，大类维持长年不变。栏目的设置和候选关键词由 ICP 自行定义和修改维护，保证跟进客观形势和用户热点的变化。用户只需在 ICP 提供的选项上做出设定，PC 机将自动编辑出纯属个人爱好的电子读物。文化程度较低的用户可以请他人代为设定，一次设定长期有效。

6 信息共享结构的二元化趋势

今后的信息共享系统，除了基于互联原理的原创网络外，还将增添一种基于并行播存原理的第二类网络，形成二元化的结构，见图 2。

第二类网络用于全民低成本共享文化知识，是

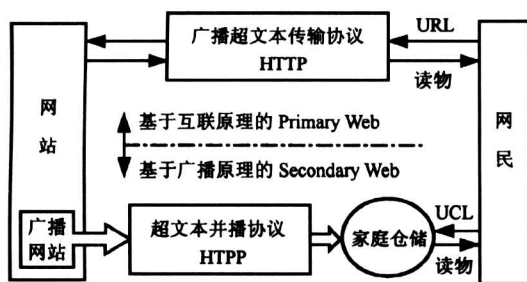


图 2 二元化的信息共享结构

Fig.2 A dual-way structure for information-sharing

新生的第五传媒。它继承纸张传媒存信息于家庭的优点，广播传媒、电视传媒支持网民自由增长和互联传媒按需取用的优点。第二类网络还通过以下两种方式，同互联网实现互助互补：

1) 分流网页 网民对热门网站的访问，占互联网流量相当的比重，如果这部分流量能被广播方式有效分流，互联网的带宽负担将减轻许多。在热门网站周边的地段和网民集中访问的时段，拥塞问题可望明显改善。

2) 推送标引 没有加入广播的网站，可以利用广播来发布 UCL 标引。UCL 是有关内容的内容，

比内容本身短小得多，很容易通过广播存入家庭，让万维网中语义相似的网页也加入被主人遴选的行列，用户可以得到一份更完整的网页清单。内容已被家庭仓储收存的可直接提取；没有被收存的则从互联网索取。两网以最节省资源的方式自动互助互补，用户无需过问某网页是通过何种渠道取来的，网络的业主也不必为瓜分市场而大伤脑筋。

21 世纪的网络不可能被互联网一种网络所垄断。在信息共享领域，今后 10 年的趋势是互联网和广播网在新的创意下整合。整合的呼声已经很高，但创意太少。笔者提出无冲突、低成本共享知识的第二类网络，试图在创意上抛砖引玉。

参考文献

- [1] Lee T B. The World wide web: past , present and future [A], Proceedings of the Fourth International WWW Conference [C], Boston, 1995. 1~7
- [2] 李幼平. 思考万维网的下一步 [A]. 中国工程院信息电子学部 2000 年学术报告汇编 [C], 2000. 66~72
- [3] 李幼平, 信息阳光: 无冲突共享结构 [J]. 中国工程科学, 2000, 2(1): 24~27
- [4] 高扬, 李幼平. 面向对象的信息链接系统 [J]. 计算机科学, 2001, 28(50): 30~33

The Secondary Web of Knowledge Embodiment

Li Youping

(China Academy of Engineering Physics, Mianyang, Shicuan 621900, China)

[Abstract] World wide web (WWW) has become a primary web across the Internet, which has been widely accepted as an information-sharing platform. Due to the nature of the Internet, i. e. accessing with potential collisions, however, bandwidth bottleneck has become a serious issue of WWW, which significantly restricts the growth of web users. This paper proposes a novel information-sharing mechanism, secondary web, which automatically downloads the information that users need or have subscribed into end users' PCs via a broadcast mechanism. End users can retrieve information based on its content semanteme from their local storage instead of Internet. Therefore, due to the avoidance of the information access collision, the secondary web eliminates the bandwidth bottleneck that the primary web is facing and efficiently filters out information garbage as well. The secondary web is expected to be a new information medium, which has low cost and the ability to support the unlimited growth of users.

[Key words] world wide web; secondary web; semanteme of content; bandwidth bottleneck