

面向对象的网络信息空间模型与信息传播^{*}

马卫东^{1,2} 李幼平¹ 袁宏春²

(中国工程物理研究院电子工程研究所 绵阳 621900)¹ (电子科技大学计算机学院 成都 610054)²

The Object Oriented Model of Cyberspace and Information Dissemination

MA Wei-Dong^{1,2} LI You-Ping¹ YUANG Hong-Chun²

(Institute of Electronics Engineering, China Academy of Engineering Physics, Mianyang 621900)¹

(School of Computer Science, UESTC, Chengdu 610054)²

Abstract The Cyberspace is composed of a set including all electronic information elements and its information dissemination networks. In this paper, a kind of Cyberspace Model that four-dimensional entity described is submitted, and the classification and characteristic of the information element and information node are studied. The Object Oriented Model of Cyberspace that anthropomorphized is discussed and the mechanism of the information dissemination is analyzed. Finally we put forward to a Personalizing Information Storage Model.

Keywords Information element, Network node, Grid, Information dissemination, Cyberspace, Personalizing information storage

1. 引言

向人们传递信息的手段称为信息媒体, Internet 是信息媒体发布传输的典型构架。以计算机、电子与光通信、网络为代表的信息技术(IT)在二十世纪后半叶取得了惊人的进步, 信息时代的大门开始向我们打开, 信息理念充斥社会的每一个角落, 其意义、影响之深远是不言而喻的。

作为电子信息的理想载体, Internet 无疑发挥了至关重要的作用。事实上现在的 Internet 世界早已超出了原本定义的范畴。电子信息的物质载体, 从以电信、科研教育网为骨干的基础网络世界, 到以 DVB 和 MPEG 为核心的数字直播卫星系统(Direct Broadcast Satellite, DBS)^[6]; 从有线电视系统数字化改造到宽带 ISDN 与综合布线; 从数字蜂窝式移动电话到 3G 通信; 各种数字传输技术都与以 TCP/IP 协议为代表的 Internet 技术进行了深入的融合, 逐步形成了一个整合的网络信息世界。

信息是物质世界最基本的概念。信息就是事物运动的状态和方式(本体信息), 就是关于事物运动的千差万别的状态和方式的认识(主体信息)^[12]。也就是说, 事物运动的状态和方式构成了事物的本体论层次的信息, 认识主体在某种约束条件下对本体信息的测度, 构成了认识论层次定义的主体信息。

全球电子信息的集合及其构成网络组成了网络信息空间(Cyberspace)。从人文角度来说, 信息是以人为中心的一种概念, 而不仅仅是限于通信中的载荷与编码效率。以人为本的信息发布、信息处理、信息获取乃至主动信息服务需要在一个概念清晰的网络信息空间模型。不仅在信息的传输方面, 在信息元素的语法上以至内容语义上都要能够方便地表述与分析。

2. 网络信息空间模型

2.1 网络信息空间模型的形式化描述

^{*} 本课题获中国工程物理研究院重点基金 97Z044 资助。马卫东 副研究员, 研究生, 主研方向: 信息系统共享与主动内容服务技术。李幼平 研究员, 中国工程院院士, 博士生导师, 主研方向: 飞行器电子学系统、宽带信息系统。袁宏春 教授, 主研方向: 网络多媒体技术。

网络信息空间是指世界上所有电子信息的集合, 我们采用四元组形式化描述网络信息空间模型 Γ :

定义 1 网络信息空间模型 Γ 为一个形式化四元组:

$$\Gamma = (V, I, E, \sigma)$$

其中 $V = \{x | x \in \text{节点}\}$, 信息节点的集合; $I = \{i | i \in \text{信息元素}\}$, 信息元素的集合; $E = \{\langle x, y \rangle | (x \in V) \wedge (y \in V)\}$, 信息节点之间连接关系的集合; $\sigma = \{\langle x, i \rangle | (x \in V) \wedge (i \in I)\}$, 信息节点与信息元素关系的集合。

这里, 拥有存储、发布、转发、消费等计算处理能力的基本信息处理单位称为一个信息节点; 信源服务器、路由器、客户计算机、移动便携式设备都是信息节点。

有独立意义的信息最小单位称为信息元素(Information Element); 从数字通信意义上讲, 最小的信息元素是比特; 从信息处理意义上讲, 最小信息元素决定于其语义编码方案, 如 ASCII 编码、Unicode 编码、图像与声音编码等。

信息节点之间的连接关系称为边。正如人类社会丰富多彩的网络信息世界构成模式一样, 边的形式也是多种多样的。PSTN 中的双绞线、局域网中的 3 类或 5 类双绞线、CATV 中的 Cable、蜂窝式移动通信中的微波等都属于某种类型的边。

网络信息空间模型 Γ 实际上表述了网络中信息节点之间的拓扑结构以及信息元素与信息节点之间的依赖关系。

2.2 信息节点的分类

信息节点 V 有四种基本分类形式:

1) 按信息的存储性质来分: ① 存储节点—带存储的信息节点; ② 非存储节点—不带存储的信息中转节点。

2) 按信息传播的性质分类: ① 信源节点—信息的出发点; ② 中转节点—中转信息(包括带 Cache 与不带 Cache 两种类型); ③ 宿信节点—信息变迁的终点。

3) 按信息生产与消亡的特点分类: ① 制造节点—按照某种给定的随机过程制造信息元素; ② 消耗节点—按照某种给

定的随机过程消耗信息元素。

4)按节点是否具有计算能力划分:①网格(Grid)一具有信息处理能力的信息节点;②路由节点一仅具有信息转发能力的信息节点。

上面四种性质的组合可以描述信息节点的典型特性。

2.3 信息的层次描述与特性

对信息本身的研究,通常从两种角度出发:信息的通信特性与信息的内容特性。C. E. Shannon 最早针对有噪声信道信息的传输特性作出了突出贡献,而针对信息内容特性的研究目前处于起步阶段。关于信息内容语义特征的形式化表述常见的是矢量空间模型^[4],即用矢量化的多维空间描述信息的

语义特征,进而方便信息的语义处理。

由信息元素按照某种语法规则可以组成人类智能标志的文字、词语、句子、段落、篇章,最后可形成一步完整的作品。按照信息出版者的组织形成的信息作品群,构成了信息传媒。文[13]中从信息逻辑语义上将信息内容划分为页面、作品、媒体三个层次。对语言(语音)、图像、文字等媒体典型的表现方式来说,上面三个层次具有不同的内涵。论文将具有独立意义的信息逻辑单位统称为信息单元(Information Unit)。

任何一种信息单元,从其以人为本的内容本质来说,均具有如表1所示的一些性质。

表1 信息单元的基本性质

描述种类	性质	内容	示例
结构描述	编码	信息的编码方式	ASCII、Unicode、二进制字节码等
	语法	信息的组织形式	HTML、XML、JPEG、MPEG-II 等
	冗余度	信息有效载荷在全部信息存储空间中所占的比例	体现编码的压缩效率
内容描述	热度	人们对信息兴趣的一种统计量	非常感兴趣、比较感兴趣、不感兴趣等
	类属	信息的内容类别与标识	汉语语音、足球信息、好莱坞电影等
	源标	信息制造的源头、作者、时间	(新浪网;记者名;2002/2/1 等)
动作方法	执行特征	信息的可执行情况	Java 小程序、VB Script 等
	被解释性	信息某种描述的被解释情况	
	自解释性	信息是否具有可以自解释的特性	自我定义的、或规定的一组动作

信息单元的结构描述、内容描述以及动作方法均具有相关的性质。如信息单元 IU 的结构描述涉及信息的编码方式、组织形式等信息的结构特征;信息单元的内容描述涵盖了信息的内容类别与标识、信息的出处、制作者、制作时间以及人们对信息的兴趣的某种统计量;信息的动作方法说明信息的执行特征、被解释性或自解释性等等。信息的这些描述通常隐含在信息的内容之中,将之从信息中提出并显示的表达出来,对信息单元的语义分析有着重要的帮助。也就是说,每个信息单元 IU 可以由这些典型性质定义其信息的基本特征,进而构成一个以人为本的面向对象的信息处理单元,从而方便信息的处理、分发、获取与消费。

2.4 以人为本的网络信息空间模型

前面描述的网络信息空间模型 Γ 暂且没有考虑人文因素。事实上,没有人参与的信息系统是毫无意义的。任何信息系统的构建,都本着为人类服务为最终目的。因此,以人为本的网络信息空间模型是分析网络信息空间信息演化效果的最终评价手段。信息生产与消亡的过程就是由人直接参与的、或者是间接参与的信息处理、分发与消费的过程。

以人为本的网络信息空间模型中,一组相关的信息元组成一个有独立意义的、采用面向对象方法描述的信息单元。每一个信息单元都与其信息属性、操作方法封装在一起,具有直接的自解释行为能力或被解释表达能力。面向对象的网络信息空间模型 Γ_0 定义如下:

定义2 面向对象的网络信息空间模型 Γ_0 是一个四元组:

$$\Gamma_0 = (V, I_0, E, \sigma)$$

其中: $I_0 = \{i_0 | i_0 \in \text{面向对象的信息单元的集合}\}$; $\sigma = \{(x, i_0) | (x \in V) \wedge (i_0 \in I_0)\}$; 信息节点 V 与边 E 的定义同上。

面向对象的网络信息空间模型 Γ_0 的核心是将信息的内容、属性与操作封装在一起,形成真正的信息智能发布、处理与消费,进而对信息传播的多种方式给与合理的支持。

2.5 面向对象的信息单元模型

面向对象的信息单元(Object Oriented Information Unit, OOIU),是面向对象的网络信息空间模型 Γ_0 的核心。事实上,采用面向对象的信息单元 OOIU 描述网络信息空间的一个基本信息组成单位是合理的、也是有力的描述手段。

OOIU 是信息存储与传播的、具有完整语义特征基本单位。有了 OOIU 这个概念,既可以把我们从通信协议与数据分发的 Bits 流中解放出来,又便于我们定义信息传播的逻辑单元及其动作特征。以面向对象的手段,将逻辑信息单元与它的结构描述、内容描述和动作方法紧密地封装在一起,形成了以人为本的 OOIU 概念。

面向对象信息单元 OOIU 可以用类 C++ 语言伪码表征,本文中定义如下:

定义3 面向对象的信息单元 OOIU 是如下一个类:

```
class OOIU{
//Structure Description
CODE Code; //编码方式:CODE 类
FORM Form; //描述形式:FORM 类
//Content Description
ORIGIN Origin; //信息源标:ORIGIN 类
GROUP Group; //信息类属:GPOUP 类
INTEREST_VECTOR InterestVect; //信息兴趣矢量:INTEREST_VECTOR 类属
//High-level Action Method
virtual int Running(); //OOIU 的运行特征
virtual int MoveTo(); //OOIU 的可移动性及其移动方法
virtual int Showing(); //OOIU 的显示绘制
//Low-level Action Method
virtual int open(); //OOIU 的 Open 操作
virtual int read(); //OOIU 的读/写或发送/接收操作
virtual int write();
virtual int close(); //OOIU 的 Close 操作
virtual int ioctl(); //OOIU 的行为控制操作
};
```

从 OOIU 的类定义中可以看到,其属性描述用一系列子类及其相关的定义表达,操作既有高级操作形式,也可以提供低层的调用接口。将 OOIU 的属性定义为子类,可以封装其属性与操作。将其操作方法用虚函数定义,便于派生与继承。采用 OOIU 可以很好地描述网络信息空间信息的基本特性。

3. 网络信息空间的信息传播

3.1 边与信息变迁

网络信息空间中信息的传播有着多种方式。Client/Server 结构、Client/Middleware/Server 结构、Source/Client/Agent 结构、Broadcaster/Receiver 结构等都代表着典型的信息传播结构。

信息节点之间的连接关系—即边,包括有向边和无向边两种基本形式,其物理属性涉及到边的带宽、迟延特性以及传输代价等,是信息传输信道的数学抽象。

在网络信息空间模型 Γ 中,信息节点按照一定的产生率与消失率制造与消耗信息元素;同时,信息元素在信息节点之间按照某种要求流动,产生信息的变迁过程。

从图论出发考察网络信息空间内信息 I 的流动特性,应该包括三种基本形式:

- 1) 信息 I: $P_1 \rightarrow P_2$ 单播(带反馈时:信息交换);
- 2) 信息 I: $P_0 \rightarrow (P_1, P_2, P_3, \dots, P_n)$ 信息广播;
- 3) 信息 $(P_1, I_1), (P_2, I_2), \dots, (P_n, I_n) \rightarrow P_0$ 信息汇聚。

对于多点到多点的形式如:

信息 I: $(X_1, X_2, X_3, \dots, X_n) \rightarrow (Y_1, Y_2, Y_3, \dots, Y_m)$;

可以认为是上述三种基本信息流动的组合。

信息元素最基本的构成单位是比特(bit)。对于一个信息元素 I 的变迁过程,可以定义为:

$$\tau = \{(x, i) \rightarrow (y, j) | x \in V \wedge y \in V \wedge i, j \in I\}$$

信息的传播实际上是一个信息动态变迁的过程。也就是说,信息的传播就是一系列比特流从某一个信息节点传输到另一个信息节点的过程。

3.2 信息传播机制

无论从那个角度来划分,信息传播的模式均为从信源到信宿的传输过程。我们这里想要说明的是,有三种明显的信息分发特性^[1-3]:

1) 推送 Push—拉取 Pull: 信息是由源端主动 Push 到用户端还是由用户自己发出请求等待服务器响应

2) 定期的 Periodic—不定期的 Aperiodic: 信息的递送是定期发送的,如股市行情、广告等;还是不定期发送的,如电子邮件、电子期刊等

3) 单播 Unicast (1-1)—多播 Multicast (1-N): 是从一个信源(发送者)到一个信宿(一个接收者)还是到多个信宿(多个接收者)

这三种不同的信息传输形成了八种鲜明的信息递送风格,如图1所示。使用这种区分方法,可以描述现有流行的信息传播的机制类型。

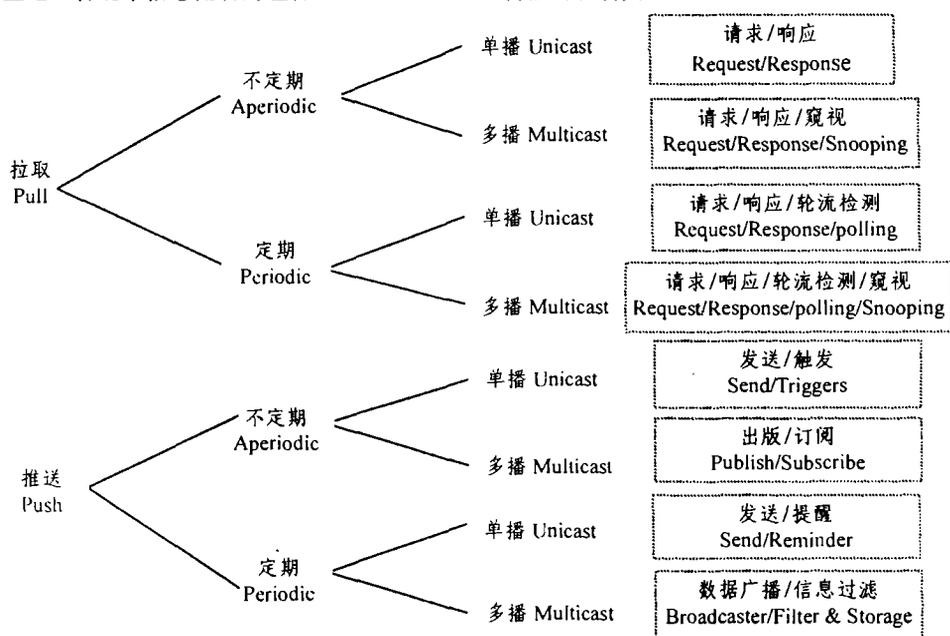


图1 信息递送风格

请求/响应模式 这是传统的客户/服务器模式。对于1-1模式,即一个请求一个响应,在 Internet 上应用得最为广泛。至于1-N 模式,可以理解为信息侦测与捕获过程,即一个窥探程序截获其他用户的连接以获取信息。

轮流检测模式 某个应用程序定期或周期性地检查远程服务器以获得其所需要的内容。

出版/订阅模式 通常采用基于 Push 技术的由信息源端发送的电子刊物的发行方式。在 Internet 网上目前也比较流行。

数据广播/信息过滤模式 在广播信道如卫星、CATV 系统、无线数据通信中开展的新型数据服务业务。信源前端定期地广播其定制的信息,多个接收者按照自己的信息需求寻找与之匹配的信息(即所谓信息过滤)。

3.3 信息传播的层次模型

表2 DBTA 抽象层次模型

	层次名称	内容	Layer Name
信息内容层	文化传媒层	整合文化	Culture-media Layer
	信息作品层	完整信息单元	Works Layer
	页面单元层	逻辑页面	Pages Layer
信息传输层	网络层	网络数据包	Network Layer
	链路层	链路数据包	Data Link Layer
信息调制层	物理层	物理比特	Physics Layer

网络信息空间中的信息传播对不同的信息传播类型有着不同的层次模型。在 Internet 中采用的是 TCP/IP 混合模型;在数据广播中,我们提出了一个数据广播传输体系(Data Broadcasting Transport Architecture, DBTA)信息传播的层次模型^[11]。DBTA 层次模型包括三个主层:信息内容层、信息传输层、信息调制层。每一层又包括若干子层,如表2所示。

4. 个人信息仓储模型

4.1 信息的个性化特征与用户肖像模型

我们说信息尽管具有鲜明的共享性,同时对认识主体来说又具有明显的个性化特征。信息的共享性是指信息可以被所有的认识主体获取和理解,方便每一个人对其共享,不会因为信息的传播而从信息母体中失去它,是典型的拷贝 COPY。信息的个性化是指每个人的认知层次、经历、兴趣存在差异,信息服务对不同的认识个体应有所区别。Internet 既体现了信息的共享性,同时又满足了信息的个性化需求,因而获得了巨大成功。

用户肖像模型(User Profile)是对信息个性化的一种较好的形式化描述手段。文[5]表达了在移动信息服务领域中关于个性化服务的数据交换用户肖像模型。这种模型可以应用到所有信息分发系统(Dissemination-based Information System, DBIS)中。在 DBIS 中,用户肖像模型的核心实现机制包括数据收集器(Data Collector)、用户节点(User Node)、信息代理(Information Broker 或 Information Agent)。

在不同的信息服务领域,如数据广播系统、固定节点的 Internet 用户、移动用户等,用户肖像模型可以有着不同表达实现方法。结合面向对象的信息单元 OOIU 中的兴趣矢量 INTEREST-VECTOR 模型,可以采用兴趣矢量来定义个性化的用户肖像模型。在兴趣矢量中的每一维数量代表一个用户兴趣点,进而可以与 OOIU 进行兴趣匹配计算,从而确定该用户对此信息的关联度,方便信息的采集、获取。

4.2 信息发布技术及其存储整合

对于任何一种信息发布技术来说,用户的个性化需求要求对其感兴趣的信息进行本地存储。对于数据广播技术来讲,用户本地存储整合极为重要。因为信息是由信源主动提供的,信道是按照信息内容分配的,所以用户为了获得其感兴趣的信息不得不将大量的信息首先存下来,有挑选的--信息过滤,或无挑选的--信息存储镜像,以被今后按照自己的需要取用消费。IT 技术的进步使得存储器越来越便宜,数百亿字节的磁盘存储器只需数百元人民币,而且价格还在不断下降。同时半导体存储器也越来越便宜,数百兆字节的半导体存储器已经下降到百元人民币以下。这些存储器技术的进步为基于存储整合的主动信息服务打下了坚实的基础。

4.3 个性化的信息仓储模型

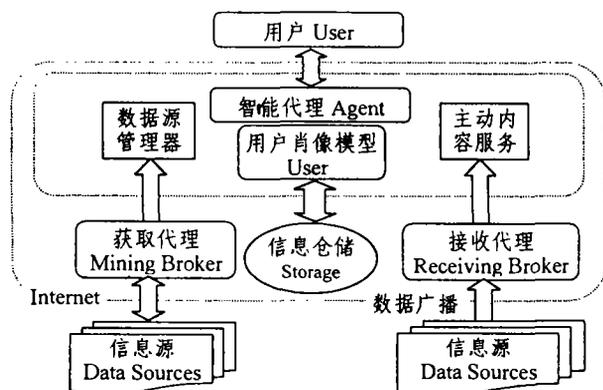


图2 个性化的信息仓储模型

所谓个性化的信息仓储(Personalizing Information Stor-

age, PIS)就是一个属于他的信息服务代理,该代理能够帮助他负责完成信息的获取与存储、个性化分析、主动内容服务等任务。

个性化的信息仓储模型应该既具备 Internet 的常规信息获取与处理能力,又能够适合广播型业务带来的新型的信息服务机制。典型的个性化的信息仓储模型 PISM 必须具备下面特性:①信息的直接获取或间接获取能力;②信息存储、分类、检索、关联、淘汰能力;③形式化的用户兴趣模型及其智能演化;④个性化的主动内容服务能力。

本文给出了一个 PISM 如图2所示。

结束语 以人为本的面向对象的信息单元 OOIU 模型及面向对象的网络信息空间模型对信息的传播发布,尤其是对数据广播业务信息发布提供了一个可计算的形式化模型。我们认为,信息共享系统的本质具有二元性(duality,即信息的被动服务方式,如 Client/Server 模式,与主动服务方式,如 PUSH 技术、数据广播等)。网络信息空间中信息的传播方式与传播机制是多种多样的。数据广播分发系统作为 Internet 互联系统的一个有力的补充机制,具有强大的生命力。用户消费信息的个性化特征与共性化特征同时存在,建立信息共享系统的用户肖像模型与个人信息仓储,既适合 Internet 广为流行的 C/S 信息发布机制,也适合新兴的基于 IP/DVB 构架的数据广播^[7-9,13,14]信息共享系统。

参考文献

- 1 Acharya S, Alonso R, Franklin M, Zdonik S. Broadcast Disks: Data Management for Asymmetric Communication Environments. ACM SIGMOD Conference, San Jose, CA, May 1999
- 2 Aksoy D, Franklin M J, Zdonik S. Data Staging for On-Demand Broadcast. The 27th VLDB Conference, Roma, Italy, 2001
- 3 Juvva K, Rajkumar R. A Real-Time Push-Pull Communications Model for Distributed Real-Time and Multimedia Systems. CMU-CS-99-107, School Of Computer Science, Carnegie Mellon University
- 4 Lee D L, Chuang H, Seamons K. Document Ranking and the Vector Space Model. IEEE Software, 1997, 14(2)
- 5 Cherniack M, Franklin M J, Zdonik S. Expressing User Profiles for Data Recharging. IEEE Personal Communication, August, 2001
- 6 Mead D C. Direct Broadcast Satellite Communications--An MPEG Enable Service. Prentice Hall PTR, 2000
- 7 European Telecommunication Standard. Digital Video Broadcasting (DVB). DVB specification for data broadcasting, EN 301 192 V1.2.1 (1999-06)
- 8 European Telecommunication Standard. Digital Video Broadcasting (DVB). Implementation guidelines for Data Broadcasting, TR 101 202 V1.1.1 (1999-02)
- 9 Evain Jean-Pierre (EBU Technical Department). The Multimedia Home Platform (MHP), 2000
- 10 Carl Fenger, Michael Elwood-Smith. The Fantastic Broadband Multimedia System, Software Platform Description, 19 May 2000
- 11 马卫东, 李幼平. 数据广播传输体系结构研究. 北京: 计算机工程与应用, 2000, 37(24)
- 12 钟义信. 信息科学原理. 北京邮电大学出版社, 1996
- 13 马卫东, 袁宏春, 李幼平. IP over DVB 技术及其应用. 微电子学与计算机, 2001(10)
- 14 马卫东, 李幼平, 袁宏春. 并行广播技术及其应用前景. 计算机科学, 2002(2)