基于语义相关的信息聚焦数学模型及方法研究

黄宏斌[1] 熊 芳[2] 邓 苏[1] 张维明[1]

(国防科技大学信息系统工程重点实验室 长沙 410073)[1] (中南大学湘雅医院网络中心 长沙 410078)[2]

摘 要 信息服务目前依然是一个研究热点。提出了基于语义关联的信息聚焦概念及数学模型。信息聚焦作为一种新的信息服务方式,利用信息之间的语义关联关系实现信息汇聚。首先给出了基于本体的元数模型以及信息、信息空间的数学模型和信息聚焦的数学描述,并在此基础上提出了基于本体的语义相关及信息聚焦的形式化描述。

关键词 信息空间,信息聚焦,本体,语义相关

中图法分类号 G633.67 **文献标识码** A

Semantic Relatedness-based Information Focusing Service Model and Approach

HUANG Hong-bin[1] XIONG Fang[2] DENG Su[1] ZHANG Wei-ming[1]

(Science and Technology on Information System Engineering Laboratory, National University of Defense and Technology, Changsha 410073, China)[1]

(Department of Informatics, Xiangya Hospital, Central South University, Changsha 410078, China)[2]

Abstract Information Service is still a hot topic for information sharing research. This paper presented the concept and mathematics model of semantic relatedness based information focusing. Information focusing, as a new information service, implements information congregating by using the semantic association relationship between information units. This paper presented the Ontology-based metadata model, and gave mathematics model of information and information space, and presented the mathematics about information focusing. By giving the formally definition of semantic relatedness, this paper also defined the concept of information focusing based semantic relatedness.

Keywords Information space, Information focusing, Ontology, Semantic relatedness

1 引言

信息的接收、获取、组织、管理与分析是个人、部门、行业以及各领域所必须时刻面对的任务。目前在广域环境下存在着大量信息资源,信息资源之间存在大量的语义关系。如何科学、合理地描述信息以及所处的信息空间[1-3],如何解决由海量的、无序的信息造成的"信息泛滥"和"信息缺乏"问题,如何在信息组织基础上提供用户需求的信息服务等,一直是信息科学研究的重要课题。而我们目前获取信息的手段往往还停留在关键词搜索或基于数据库的查询等方式,无论是搜索还是查询都未能充分利用信息之间的语义关系。因此,如何充分利用信息之间显现的或潜在的语义关系,为用户提供一种个性化的精确信息服务,成为当前信息管理与信息服务的一个重要研究问题。

信息空间[4]是描述信息环境的一个重要概念。关于信息空间的定义和结构亦有不少研究成果。文献[4]提出了信息的全息空间拓扑结构;文献[5]提出了信息以及信息空间的一种数学描述;文献[6]基于信息的"差异性"给出了信息的数学描述以及信息空间的公理化体系。在文献[2,3]中把信息空间(Information Space)定义为由信息系统承载的信息单元(对象)以及这些信息单元(对象)之间的关系组成。信息单元可以是知识、文档、结论或者用户表示。

元数据模型是对信息空间中信息的概念建模,是信息空间完成信息组织与信息服务的基础。本文首先给出了基于本体的元数据模型,给出了信息、信息空间及信息聚焦的数学模型与描述,并在此基础上基于元数据模型给出了语义相关及信息聚焦的形式化描述,提出了一种基于语义关联的信息聚焦方法,从而为实现基于语义相关的信息聚焦服务提供了理论基础。

2 基于本体的元数据模型

本体(ontology)是领域知识的概念化说明,它将特定领域有关的对象、概念及其关系以形式化的说明来严格规定。本体由于具有较强的数据描述能力、一定的推理能力以及面向语义的输出能力和相似度的计算能力,因此基于本体论来描述语义信息,提供面向语义的查询等,可以使用户更好地理解数据。利用本体建立面向语义的元数据模型,可以将元数据中实体类的含义、类间的关系更加明确地表达出来,从而支持广域分布信息空间环境下的概念建模、信息搜索与交换、信息资源共享与服务系统建设等研究。基于本体的元数据模型为我们在信息空间以及信息聚焦服务过程中充分利用信息资源语义关系提供了一个良好基础。

到稿日期:2010-09-06 返修日期:2011-01-26

黄宏斌(1975—),男,博士,副教授,主要研究方向为信息系统与智能决策、信息物理系统,E-mail:Hongbinhuang2000@yahoo.com.cn;**熊 芳**(1979—),女,硕士,工程师,主要研究方向为医院信息系统;**邓 苏**(1963—),男,博士,教授,主要研究方向为信息系统与智能决策、信息物理系统;**张维明**(1962—),男,教授,主要研究方向为信息系统与智能决策、指挥信息系统。

假设 T 为领域语义词典(Domain semantic dictionary, DSD),是领域术语词汇的集合,用来规范元数据模型中元素命名。对应领域语义词典中的词汇,主要包括类术语、属性术语、关系术语等。Dbasic 是预定义的基本数据类型集合,Denum 是预定义的枚举类型。可采用《知网》的语义描述结构来组织领域语义词典 T,这是一种开放结构,可以进行扩充。

本文基于信息空间是由信息单元及信息单元之间的关系组成的这一理解,基于本体的思想可以给信息及其相关概念一个形式化描述。采用实体类和实体类之间的关系及相关约束和规范描述信息空间,进而对其数学结构予以研究。具体定义如下。

定义 1 元数据模型是一个八元组 $OMD::=\langle E,A,L,H^C,R,I,F,P\rangle$。其中,

1)E 是实体类集合,$\forall c\in E, c=(\text{Name},A^c)$,Name($c$)表示 c 的命名词汇,Name(c)$\in T$,其中 $A^c=\{x\mid x\in A$,且 $Att(x)=c\}$。Att 是函数集中的属性映射函数。

2)A 是定义到实体类的属性集合,属性又分为基础属性和复合属性。$\forall a\in A, a=(\text{Name},dt)$,Name($a$)表示 a 的命名词汇,Name(a)$\in T$,$dt\in D_{\text{basic}}\cup D_{\text{enum}}$;

3)L 取值域集,$L=D_{\text{basic}}\cup D_{\text{enum}}$;

4)H^x 是实体类间的一种二元层次关系,该关系是一种有向的、传递的关系,是一种偏序关系,包括了实体类之间的继承 is-a 关系和聚合(组合)part-of 关系。$H^c\subset E\times E$,$H^c=H^{\text{Inh}}\cup H^{\text{Aggr}}$,$H^{\text{Inh}}(c_1,c_2)$,$c_1,c_2\in E$,表示 c_2 是 c_1 子类即 c_2 is-a c_1,$H^{\text{Aggr}}(c_1,c_2)$,$c_1,c_2\in E$,表示 c_2 是 c_1 的组成部分,即 c_2 is part-of c_1,如 H^{Inh}(学员,本科生),H^{Aggr}(学院,系)[7]。

5)R 是实体类之间的二元语义关系集(关联关系)。任何所连接的实体类超过两个的关系,都能够转换为一组二元的多对一关系集合,而不丢失任何信息。因此,该模型中的关系设计为二元关系。$\forall r\in R$ 可表示为 $N^r(c_1,c_2)$,$c_1,c_2\in E$,$N^r\in T$ 为关系的名称。如 R(授课)=(教员,学员)[7]。

6)I 实例集,是信息单元实体类对象集合。

7)F 表示函数集,主要包括如下函数:

$Att:A\to E\cup R$,属性函数,将属性分配给某一实体类或关系;

$val:V\to(E,A)$,属性取值函数,属性具有数据类型;

$Inst:E\to 2^I$,信息单元实例化函数,可以写为 $Inst(E)=I$ 或 $E(I)$[8];

$Instr:R\to 2^{I\times I}$,关系实例化函数,可以写为 $Instr(R)=\{I_1,I_2\}$ 或 $R(I_1,I_2)$[8];

$Rela:R\to E\times E$,关系映射函数,$Rela(R)=(C1,C2)$,也可以写为 $R(C1,C2)$;

$dom:R\to E$,由 $dom(R)::=\Pi_1(rel(R))$ 给出关系 R 的定义域[9];

$range:R\to E$,由 $range(R)::=\Pi_2(rel(R))$ 给出关系 R 值域[9]。

定义 2 给定 $OMD::=\langle E,A,H^C,R,L,I,F\rangle$,语义解释为一个三元组,记为 $I=\langle\Delta^I,\Delta^L,\cdot^I\rangle$[7]。其中 Δ^I 是一个非空集合(包含所讨论领域中的所有个体),Δ^L 是一个非空集合(包含所讨论领域中的所有数据值),$\cdot^I$ 是解释函数,它将 E 中的每个实体类 C 都映射为 Δ^I 的一个子集 C^I($C^I\subseteq\Delta^I$),每个具体域 L 解释成一个集合 L^I($L^I\subseteq\Delta^L$),每个关系 R 解释成一个二元关系 R^I($R^I\subseteq\Delta^I\times\Delta^I$),每个属性 A 解释成一个二元关系 A^I($A^I\subseteq\Delta^I\times\Delta^L$),每个个体 c 解释成一个元素 c^I($c^I\in\Delta^I$),每个值 l 解释成一个元素 l^I($l^I\in\Delta^L$)。

3 信息聚焦的数学模型

3.1 信息、信息空间

信息空间中的信息单元可以分为原子信息单元、复合信息单元[4]。

定义 3(原子信息单元) 不可再分的信息单元。

依据定义 1 我们可知只有一个实体对象的单一属性取值的信息称为原子信息单元。

依据定义 1 和定义 3 可以进一步给出如下定义。

定义 4(简单信息) 给定实体对象 $i\in I$,设其属性集 A_i、属性值集 L_i,若存在 $e\in E$,有 $Inst(e)=i$,且 $\forall a\in A_i$ 有 $Attr(e)=a$(将属性 $a\in A_i$ 赋给 e);即 $\forall l\in L_i$,有 $Val(i,a)=l$(将属性值 l 赋给属性 a),则三元组 $Inf_i=\langle\{i\},A_i,L_i\rangle$ 称为关于实体对象 e 的简单信息。

在简单信息单元的基础上,我们可知原子信息单元是简单信息,且为三元组 $\langle i,a,l\rangle$。我们可以进一步定义信息的和与积,称为信息的运算。易证以下性质。

性质 1 信息的和与积仍为信息。

性质 2 和、积运算可推广至有限个或可数个信息的运算。

定义 5(复合信息单元) 其是由具有相互关联关系的原子信息单元和其他复合信息单元组成的信息单元。

复合信息单元是在简单信息单元基础上经过运算得到的信息。

定义 6(全信息空间) 由所有的原子信息单元或复合信息单元的集合以及信息单元之间所有关系的集合,称之为全信息空间。

全信息空间的非空子集称为信息空间。

显然,全信息空间亦为信息空间(在不引起混淆的情形下统称为信息空间)。

则全信息空间的一种形式化定义如下。

定义 7(信息空间) 设 I 是一个信息单元的集合,$i\in I$,M 是一个信息存储器,ξ 是 M 中的坐标或存储单元。符号 $I\Rightarrow M$ 表示将 I 的元素装入到存储器 M 之中,即把 I 中的元素赋予 M,$i_k\to\xi_j$ 表示元素 i_k 被存入存储单元 ξ_j 中,存储器的坐标反映了信息单元之间的关系。那么,在存储器 M 中的信息集合 $\{i\}$ 的全体和 M 的坐标一起,称为一个信息空间,记作:Ω,则

$$\Omega=(I,M)\mid^{I\Rightarrow M}=\{(i_k,\xi_j)\mid i_k\in I,\xi_j\in M,i_k\to\xi_j\}$$

3.2 信息聚焦数学描述

信息与客观事物是不可分割的两个概念。事物的存在是以其属性表现出来的,信息是人们对事物(实体对象)某个属性欲知的需求[5],是刻画实体对象属性的取值。

定义 8 事物为一个二元组:$T=(A,L)$。其中 A 是属性集合,$A=\{a_i\mid i=1,2,\cdots,n\}$;$L$ 是一个值域集合。属性 a_i 在 L 中取值,该值表达了属性 a_i 的具体情况。属性 a_i 的取值域记为 $L(a_i)$。显然有 $L(a_i)\in L$。并且规定 $L(a_i)\neq\phi$。

元数据模型简化实体、属性、值,得三元组,其中关系可通过属性的取值进行描述。

定义 9 设 $a\in A$,u 为某特定信息单元属性,存在一个映射 $f:L(a)\to L(u)$,$\exists l\in L(a)$,使 $f(l)\in L(u)$,则称 a 为 u 的相关信息,记为 $a\Delta u$。否则称 a 为 u 的无关信息,记为 $a\nabla u$。

依据定义 9,针对某一特定信息 u,可将信息分为两类:$\{t\mid$

$t\Delta U\}$及$\{t|t\nabla u\}$。信息聚焦感兴趣的是$\{t|t\Delta U\}$。信息可以直接获取，即定义9所述。也可以间接获取，这里给出信息导出关系，它说明了信息的一种间接获取途径。

设A,B是u的两个相关信息，即$A\Delta u,B\Delta u$。

定义 10 存在映射$f:L(A)\to L(B)$，只要有一个$a\in L(A)$，有$f(a)\in L(B)$，则称由信息A导出信息B，记为$A\xrightarrow{f}B$，称映射f为一个导出方法。

定义 11 若存在信息$A_1\xrightarrow{f_1}A_2\xrightarrow{f_2}A_3,\cdots,\xrightarrow{f_{n-1}}A_n$，则称$A_1A_2,\cdots,A_n$为信息路径。

事物T的属性集$A=\{a_i|i=1,2,\cdots,n\}$中的多个属性a_i之间绝不是相互无关的，而是存在一定的关系。因此，以$V=\{L(a_i)|i=1,\cdots,n\}$作为图的顶点集，而以导出关系$a_i\xrightarrow{f}a_j$作为有向边集合。这样，整个属性之间就形成了一个有向图。信息的获取方法就是在图$G=(V,E)$中找出一系列特定信息单元属性的导出路径。

定理 1 设$a_1,a_2,\cdots,a_n,u$组成一条信息路径，则一定存在一个获知u的获取方法。

证明参见文献[5]。

通过上述定义可在定义7的基础上，进一步将信息空间形式化定义如下。

定义 12(信息空间) 信息空间定义为一个五元组$\Omega=(A,D,F,M,G)$。其中，

A为一个属性集：$A=\{a_i|i=1,2,\cdots,n\}$。

L为a_i取值域。

F为一映射集：$F=\{f_i|f_i:L(a_i)\to L(a_j)\}$，反映信息之间的关系。

M为存储器，存储器内带有坐标系统。

$G=(V,E)$，其中$V=\{L(a_i)|a_i\in A,i=1,\cdots,n\}$，$E=\{f|f\in F\}$是一有向图。反映信息映射到存储器中的坐标之间的连通关系。

聚焦原本是一个光学领域的概念，信息聚焦参照光学中的聚焦概念，是通过聚集形成一个以某个或某些信息单元为中心的相关信息集合的过程，即将相关信息汇聚而忽视无关信息的过程。

依据定义12和定理1可知，在信息空间中获取信息是在信息空间寻找一条最短路径的过程。这里可对信息空间中信息聚焦进行形式化定义。

定义 13(信息邻域) 设$a_o\in\Omega$，点$\xi(a_o)$是其在Ω中的坐标，设$\varepsilon>0$，对于Ω中所有到点$\xi(a_o)$的距离小于ε的点所组成的集合，称为以$\zeta(a_0)$为中心、以ε为半径的$\zeta(a_0)$的ε-邻域，记作$N(\zeta(a_0),\varepsilon)$。

定义 14(信息聚焦) 设$a_o\in\Omega$为用户关注信息点，其在Ω中的坐标为点$\xi(a_o)$，同时a_o为信息导出关系图G的顶点，即$a_o\in V$，设$\varepsilon>0$；则满足图G连通关系的$\zeta(a_0)$的ε-邻域，称为以$\zeta(a_0)$为焦点、以ε为半径的信息聚焦集合，记作$F(\zeta(a_0),\varepsilon)$。

4 基于语义相关的信息聚焦

定义14给出了信息聚焦抽象的概念。这里在定义14的基础上先给出语义相关的概念，并基于语义相关关系给出具体的信息聚焦概念。语义相关度和语义相似度是两个不同的概念[10,11]，但两者之间存在密切联系。语义相似度是两个不同实体具有某些相似性特征的一种度量，是在不同上下文环境中可以互相替换而不改变语义环境结构的程度。语义相关度包含了语义相似度的概念，目前其研究主要集中在自然语言处理和文本检索与分类等领域。文献[10]给出了一个关于语义相似度的定义。

定义 15(语义相关度) 语义相关度是在句法分析中一个短语结构两个词能够组成修饰关系、主谓关系、同指关系的程度。

参考上述定义给出信息语义相关度的定义：在某一领域内的两个信息单元实体满足相关关系的程度。

定义 16(语义相关) 给定两个信息单元实体$E_1,E_2\in E$，我们定义一个语义相关函数$F:E_1\times E_2\to D$，F为E_1,E_2相关程度的度量，其中D为相关度的值域。D可以是连续的，也可以是离散的，F为相关度量函数。

根据信息单元语义关系属性及方式的不同，有很多不同的相关关系，因此在定义16的基础上，依据定义1给出的基于本体的元数据模型，可以定义下列具体化相关函数。

定义 17(时变相关) 给定两个信息单元实例$i_1,i_2\in I$，$e_1,e_2\in E$，$i_1=inst(e_1)$，$i_2=inst(e_2)$，$T\in A$，T为时间属性，$T\in attr(e_1)$，$T\in attr(e_2)$，则称i_1,i_2时变相关，并定义时变相关度函数为

$$F_t(i_1,i_2)=|val(i_1,T)-val(i_2,T)|$$

在研究时变信息的表示和推理过程中，出现了多种对时间的不同观点。这些观点的差异主要体现在时态原语和时态属性两方面[12,13]。基本的时态原语有两种：时间点(instants)和时间段(intervals)。本文采用时间点作为时间表示原语，时态属性考虑时间是离散的和线性的，而时变关系主要是考虑时间间隔和前后两种关系。

定义 18(空间相关) 给定两个信息单元实例$i_1,i_2\in I$，$e_1,e_2\in E$，$i_1=inst(e_1)$，$i_2=inst(e_2)$，$P\in A$，P为空间属性，$P\in attr(e_1)$，$P\in attr(e_2)$，则称i_1,i_2时变相关，并定义时变相关度函数为

$$F_p(i_1,i_2)=|val(i_1,P)-val(i_2,P)|$$

空间属性的理解在不同研究领域是有差异的，这些差异主要体现在空间原语和空间属性两方面[12,13]。空间原语的选择有两种：基于点和基于区域。本文空间相关的研究主要是以点(经度、纬度、高程)为原语进行研究。空间关系主要包括距离关系和方向关系。

定义 19(关联相关) 给定两个信息单元实例$i_1,i_2\in I$，存在一组关系$R_1,\cdots,R_k\in R$，一组信息单元实例$j_1,\cdots,j_{k-1}\in I$，满足$R_1(i_1,j_1),R_2(j_1,j_2),\cdots,R_k(j_{k-1},i_2)$，则称$i_1,i_2$关联相关，并称$i_1,j_1,\cdots,j_{k-1},i_2$为$i_1,i_2$的关联路径，并定义$i_1,i_2$的关联相关度函数如下。

$$F_R(i_1,i_2)=\begin{cases}\max\limits_{l=1,\cdots,k}\{\sigma(i_1\to_l^m i_2)\}, & \text{if } k\geqslant 1\\ 0, & \text{otherwise}\end{cases}$$

式中，k为语义关联网络中i_1和i_2之间路径的总数目，$i_1\to_l^m i_2$代表第L条路径(路径长度为m，$m\in Z$且$m\geqslant 1$)，$\sigma(i_1\to_l^m i_2)=\sigma_{1R}\cdot\sigma_{2R}\cdot\cdots\cdot\sigma_{hR}\cdot\cdots\cdot\sigma_{mR}$为第$L$条路径的信息语义关联强度，$\sigma_{hR}$为该路径中第$h$个语义关系的关联系数。

从函数定义可以看出，语义相关度为i_1,i_2满足关联相关的最短路径中语义关联系数的乘积。

说明：

1)关联相关可以扩展定义到实体类，对于任意一对信息单元实体类E_i和E_j，总有$F_R(E_i,E_j)\in[0,1]$；

2)对于任意一对信息单元实体的元素E_i和E_j，如果在语义关联网络中存在连通E_i和E_j的路径，则E_i和E_j的信

息语义关联度为所有路径中最大的路径语义关联强度值，即最短路径中语义关联系数的乘积；否则，$F_R(E_i,E_j)=0$。

关联相关是基于本体元数据描述的信息单元实体之间的一种非层次语义关系，表述的是两个信息单元实体之间存在某种关联(association relationship)关系。关联关系一般是通过信息单元实体的属性进行关联的。语义关联有直接关联和间接关联两种：如果定义 19 中 $k=1$，则称其为直接语义关联相关，是指两个信息单元实例之间直接通过某个属性进行关联；$k>1$ 称为间接语义关联，是指两个实例之间存在一个实例或属性序列进行关联。

举例："教授"，黄三；"学生"，张五。如果张五是黄三的学生，则两者通过学生-导师关系直接关联。如果张五不是黄三的学生，但黄三参与了"信息聚焦服务"课题，张五也参加了该课题(同课题的两个学生相似或相关)，则黄三和张五通过课题"信息聚焦服务"间接关联。

同时我们将语义相似也定义为一种语义相关，它是关联相关的一个特例，是一种模糊的关联相关，其关联相关度就定义为语义相似度，即 $F_m(i_1,i_2)=\mathrm{Sim}(i_1,i_2)$。

由此，有 A 与 B 关联相关，B 与 C 相似，则 A 与 C 关联相关，并定义 $F_R(A,C)=F_R(A,B)\times\mathrm{Sim}(B,C)$。

由此，我们很容易得出如下定理。

定理 2 如果 A 与 B 关联相关，B 与 C 语义相似，如果 A 与 C 不关联相关，即不存在关联语义相关路径，则 A 与 C 模糊关联相关，且 A 与 C 的模糊关联相关度为 $F_R(A,C)=F_R(A,B)\cdot\mathrm{Sim}(B,C)$。

证明：A 与 B 关联相关，则 A、B 存在一条最短关联语义相关路径，且 $F_R(A,B)=\sigma(A\rightarrow^m B)=\sigma_{1R}\cdot\sigma_{2R}\cdot\cdots\cdot\sigma_{hR}\cdot\cdots\cdot\sigma_{mR}$；$B$ 与 C 语义相似，B 与 C 关联相关且 $F_R(B,C)=\mathrm{Sim}(B,C)$，则可认为 $\mathrm{Sim}(B,C)$ 为 B 与 C 关联系数；则 A 与 C 存在一个最短关联语义相关路径，且 $F_R(A,C)=\sigma_{1R}\cdot\sigma_{2R}\cdot\cdots\cdot\sigma_{hR}\cdot\cdots\cdot\sigma_{mR}\cdot\mathrm{Sim}(B,C)=F_R(A,B)\cdot\mathrm{Sim}(B,C)$。

证毕。

根据以上定义，可以看出一般信息单元实体之间的相关关系包括时变、空间、关联、语义相似等类别。同时我们允许用户相关关系进行自定义，即用户可以根据定义 16 自定义信息单元实体的相关关系和相关函数。用户自定义相关是按照用户需求定义的一种信息单元实体的相关关系，该相关关系可以是定义在两个信息单元实体之间的一种约束及其度量，可反映用户的信息需求关联。

信息聚焦是信息空间中围绕某一核心信息视图将相关关联的信息聚集起来形成一个以核心信息视图为中心的关联信息集合的过程。

对基于语义相关的信息聚焦的形式化定义如下。

定义 20(信息聚焦) 给定一个信息单元实例集合 I_{foucus}，根据信息单元之间语义相关关系，对信息在一定范围内进行汇聚，形成一个信息单元实例集合 I'_{foucus}，对于 I'_{foucus} 的元素 i'，$\exists i\in I_{\text{focus}}$，$\exists f\in F$ 且满足 $f(i,i')>\theta$。其中称 I_{foucus} 为聚焦的焦点；θ 为一阈值，称为聚焦因子。

定义 21(时变聚焦) 如果聚焦是以信息单元之间时间属性的时变关系进行聚焦，即 $\exists f\in F$，$f=F_t$，f 为时变相关函数，则称为时变聚集。

定义 22(空间聚焦) 如果聚焦是以信息单元之间空间属性的空间关系进行聚焦，即 $\exists f\in F$，$f=F_p$，f 为空间相关函数，则称为空间聚集。

时变聚焦和空间聚焦统称为时空聚焦。

定义 23(语义关联聚焦) 如果聚焦是以信息单元之间的关联相关关系进行聚焦，即 $\exists f\in F$，$f=F_R$，f 为关联相关函数，则称为关联聚集。

定义 24(聚焦式服务) 聚焦服务是指以用户信息需求为聚焦焦点、以信息聚焦为手段向用户提供的一种主动信息服务方式。

结束语 (1)信息空间及信息聚焦的数学模型的研究为实现信息聚焦服务提供了更为可靠的理论支撑。

本文的研究结果是一种对信息相关概念的抽象描述。本文将现实世界的客观事物进行抽象形成逻辑世界的基于本体的元数据模型，并在元数据模型基础上对逻辑世界的信息空间及信息聚焦进行数学描述，并给出了基于语义相关的信息聚焦的形式化描述。在结合工程实现时可以适合各种建模方法，有利于对广域分布环境下的信息进行组织、索引，并利用信息和用户信息需求的语义关系形成以用户关注的信息点为核心的、满足用户需求的信息聚焦集合，为用户提供有效的信息服务。

(2)进一步研究方向。

(ⅰ)在元数据模型基础上，进一步定义和完善信息的运算，使信息空间具有更良好的代数结构；(ⅱ)扩充和完善信息之间的关系，使之更有利于信息聚焦服务；(ⅲ)针对具体的语义相关关系计算相关度，为信息聚焦的计算提供基础；(ⅳ)进一步扩展元数据模型的时空要素及有关时空相关度计算方法；(v)与工程实现相结合的研究。

参 考 文 献

[1] 李必祥. 试论信息空间与信息空间结构[J]. 情报理论与实践，1996(6)：23-25

[2] Newby G B. Building Information Space (Unpublished manuscript)

[3] Papazoglou M P, Proper H A, Yang J. Landscaping the Information Space of Large Multi-database Networks[J]. Data and Knowledge Engineering, 2001, 36(3): 251-281

[4] 毕家祥. 全息空间的拓扑结构与人工智能模型[J]. 科学探索，1985(36)：73-85

[5] 杜智华. 信息空间的数学模型[J]. 新疆师范大学学报：自然科学版，2000，19(2)：12-14

[6] Sure M E. Ontology mapping-an integrated approach[C]//Proceedings of the First European Semantic Web Symposium. Heraklion. Greece: Lecture Notes in Computer Science, 2004: 10-12

[7] 黄宏斌，张维明，邓苏，等. 面向语义信息共享的元数据模型的研究与实现[J]. 计算机科学，2008，35(4)：124-128

[8] Maedche A. Clustering Ontology-based Metadata in the Semantic Web[C]//Proceedings of the Jiont Conferences 13th European Conference. 2002

[9] Rodriguez M A. Determining Semantic Similarity Among Entity Class from Different Ontologies[J]. IEEE Transaction on Knowledge and Data Engineering, 2003, 15(2): 442-456

[10] 许石，樊孝忠，张锋. 基于知网的语义相关度计算[J]. 北京理工大学学报，2005，25(5)：411-414

[11] 张运良，张全. 基于 HNC 理论的语义相关度计算方法. 计算机工程与应用，2005，34：1-3，18

[12] 刘大有，胡鹤，王生生，等. 时空推理研究进展[J]. 软件学报，2004，15(8)：1141-1149

[13] 王生生，刘大有. 时空推理前沿研究综述[J]. 计算机科学，2004，31(9)：16-19