同构化信息温度与热点发现应用初探*)

周启海 黄 涛 张元新 吴红玉
(西南财经大学经济信息工程学院 成都 610074)

摘 要 本文对信息在生活中的受关注程度进行研究,给出了一种度量信息重要性的标尺——同构化信息温度,并将它与计算机、互联网技术相结合,分别构造了单文本热点挖掘系统、文本数据库热点挖掘系统和 Web 网页热点挖掘系统模型框架。

关键词 同构化信息温度,单文本热点挖掘系统,文本数据库的热点挖掘系统,Web 网页热点挖掘系统

An Initial Research for Isomorphic Information Temperature and its Application in Discovering Information Focus

ZHOU Qi-Hai HUANG Tao ZHANG Yuan-Xin WU Hong-Yu YANG Xiang-Mao
(School of Economic Information Engineering, Southwestern University of Finance and Economics, Chengdu 610074)

Abstract In this paper, the degree which people pay attention to different information are studied, a measurement of information significance——isomorphic information temperature is raised which takes the lead in doing it. And then combining with the technology of computer and Internet; the frameworks for single text focus mining system model, text-base focus mining system model and Web page focus mining system model are constructed respectively.

Keywords Isomorphic information temperature, One text focus mining system, Text-base focus mining system, Web page focus mining system

1 引言

随着科技的发展,信息成为日常生活中的重要组成部分,充斥着生活的每一点。计算机、Web 技术不断发展再次加快了信息的传播速度,并拓展了传播范围。面对巨大的信息数据源,如何进行有效的数据挖掘;在高度动态的信息源中,用什么标准来度量信息的受关注程度,怎样寻求到不同时期、区域中的"热点、冷点"(即热点信息)等,是目前人们迫切希望解决的重要问题。为此,本文提出了同构化的信息温度概念,以它作为标尺,衡量信息的"热度",并应用于单文本、文本数据库、Web 中的热点挖掘。

2 同构化信息温度的定义与测度

同构,是抽象代数学中的基本概念之一,它指一个代数系统(如群、环、模、线性空间等)到另一同类型的代数系统上保持代数运算的一对一映射。

基于周启海教授 1986 年提出的同构化基本原理,本文率先把同构化理论运用于信息的重要性与受关注程度及其度量的创新研究,提出了同构化信息温度新概念,并找出了通过构建特殊的同构化映射来求得同构化信息温度的新方法,从而使同构化信息温度可用作测度信息重要性与受关注程度的新尺度,进而为热点的探测、发现与挖掘提供了新工具。

2.1 信息记录

人们对一个信息 x 的重要性及其关注,可表示成 x 的一个信息记录,并记为:Information$(x)=(D;x;y_1,y_2,\cdots,y_n)$。其中,$D$ 为信息 x 所在的区域范围;x 为某一特定的信息;y_i 为描述人们对 x 关注程度的第 i 个关注指标($i=1,2,\cdots,n<<+\infty$)。例如,Information("Web 挖掘";《数据挖掘》;次数、段数、节数、章数),把"Web 挖掘"这一信息抽象为一个信息记录,而它的三个指标则分别表示信息"Web 挖掘"在《数据挖掘》一书中出现了多少次,在多少段中出现,在多少节中出现和在多少章中出现。这样,利用该信息记录可计算出信息温度;通过选择其信息温度的最大、最小,就可实现在有效范围内各信息中挖掘其热点。

2.2 信息温度的定义与测量

已知一条信息 x 的信息记录 Information$(x)=(D;x;y_1,y_2,\cdots,y_n)$,则信息 x 在有效范围 D 中的信息温度 $T(x)$ 可表示为

$$T(X)=\sqrt{w_1y_1^2+w_2y_2^2+\cdots+w_ny_n^2}$$

其中,$w_i\geqslant 0$ ($i=1,2,\cdots,n<<+\infty$)为一组权重数:当 $w_i=w_j=1$ ($j=1,2,\cdots,n<<+\infty$)时,$T(x)$ 为平凡信息温度;当 w_i 和 w_j 不全相等 ($i=1,2,\cdots,n$)时,$T(x)$为加权信息温度。显然,在《数据挖掘》一书中的对信息 "Wed 挖掘",只需测度其信息温度 T("Web 挖掘"),就可获知信息"Web 挖掘"在《数据挖掘》一书中的重要性程度。

3 信息温度的应用

通过信息温度的定义可以看出,它能从 y_1 至 y_n 多个角度全面而较准确地考查某一信息的"热度",对人们的信息搜索、汲取有一定的指导意义。本文将信息温度与计算机、Web 技术相结合,分别构建了单文本热点挖掘系统、文档数据库热

周启海 教授,博(硕)士生导师,主要研究方向:同构化信息处理、计算几何、财经计算、算法研究与实现等;**黄 涛** 讲师,主要研究方向:计算机应用;**张元新** 硕士研究生,主要研究方向:信息管理、计算机应用;**吴红玉** 硕士研究生,主要研究方向:计算机应用。

点挖掘系统和 Web 热点挖掘系统的框架。为叙述简便，下面仅以关键词形式的热点为例，论述同构化信息温度的应用。

3.1 单文本热点挖掘系统模型

此时，其信息记录一般形式为 information(关键字)=(文本:关键词;次数,段数)；单文本热点挖掘系统的结构模型如图 1 所示，其“单文本受关注情况及其信息温度一览表”示例如表 1 所示。

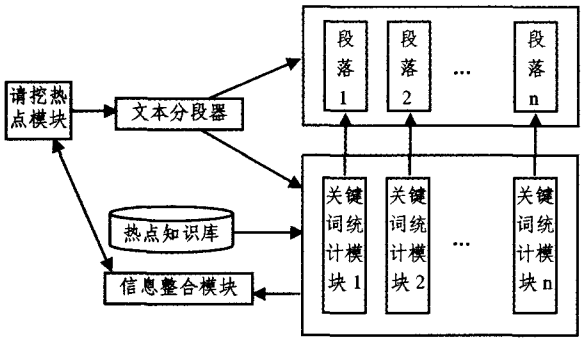

图 1 单文本热点挖掘系统结构模型

表 1 单文本受关注情况及其信息温度一览表(示例)

关键词	次数	段数	信息温度
公平	120	50	130.0000
贫困	40	30	50.0000
…	…	…	…

从单文本热点挖掘系统的结构模型中，可知它由五部分组成：

(1)请(求)挖(掘)热点模块。它是人机交互的接口，主要负责处理：①接受用户的单文本文件及其热点挖掘申请；②接受申请后，进一步调动文本分段器模块、统计模块；③将结果反馈给用户。

(2)文本分段器。它主要负责处理：①接到请求统计模块的挖掘命令后，对文本进行快速扫描，遇到分段符(即回车换行符)就记为一段，并标记该段的开始位置。②按照所统计出的文本段数 n 及其不同段落起始点，分别生成 n 个可对各自关键词进行并行化处理的本段关键词统计模块。

(3)关键词(并行)统计模块。它主要负责处理：①在收到文本分段器处理结果后，就从知识库中提取有关的统计规则，分别从各段自己起始点开始，对本段关键词作并行扫描和并行统计；②并行统计各段相同关键词出现的次数，并将结果提交给信息整合模块。此处，基于并行化思想的 n 个本段关键词统计模块，可显著加快文章关键词的热点挖掘速度。

(4)信息整合模块。它主要负责处理：①将本段关键词统计模块送来的 n 组统计结果，进行归并整合处理；②把归并整合结果统一存放在如表 1 所示的“单文本受关注情况及其信息温度一览表”中(显然，表 1 的四个栏目“关键词、次数、段数、信息温度”可顺次分别记录各关键词的信息记录)；③根据信息记录的定义分别计算表 1 中各关键词的信息温度，并挖掘出其中信息温度相对最大的用户指定个数的若干热点关键词；④把热点关键词提交给请求统计模块，以供用户查询。

(5)热点知识库。它主要负责两部分知识及其处理：①无热点实质的字、词、句处理知识(例如肯定与文档主题内容无关的字、词、句：“是”、“的”、“而”、“但”等，“的确”、“如果”、“而且”、“但是”，“越来越”、“这非常好”等等，即使在文本中出现多次也不是要挖掘的热点)。②有热点实质的字、词、句处理知识(例如必定与文档主题内容有关的字或词，及其同义词和有相同词根的单词，可视为同一单词的不同出现形式：“贫”、“贫穷”、“穷人”、“贫困户”、“家庭贫苦”等等，就可视为“贫困”的不同出现形式)。显然，热点知识库的设立有助于提高热点挖掘的效率和质量。

3.2 文本数据库的热点挖掘系统模型

与普通数据库不同，文本数据库存放的数据都是结构化的数据。如一个文档中包含一些结构化的字段，诸如标题、作者、出版时间、长度等，但也包含大量无结构的文本内容，诸如摘要和内容。为在动态的文本数据库中进行热点挖掘，同时尽可能地缩短耗用的时间和所占数据库的空间，此系统包含了单文本热点挖掘系统除请求统计模块外的各模块，并利用了矩阵理论中的单值分解(SVD)技术。此时，其信息记录一般形式为 information(关键字)=(文本数据库:关键字;次数,段数,文本数)；文本数据库热点挖掘系统结构模型如图 2 所示，其“文本数据库受关注情况及其信息温度一览表”示例如表 2 所示。

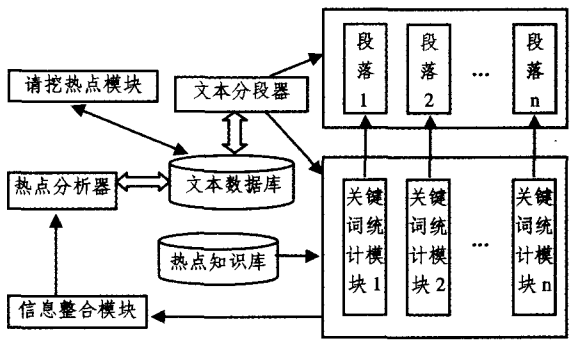

图 2 文本数据库热点挖掘系统结构模型

此时，表 2 与表 1 相比，多了一个用以描述“所论关键词出现在文本数据库中多少文本文件中”的“文本数”属性列。

表 2 文本数据库受关注情况及其信息温度一览表(示例)

关键词	次数	段数	文本数	信息温度
公平	720	150	35	736.2914
贫困	640	123	50	653.6276
…	…	…		…

在文本数据库热点挖掘系统中：①除信息整合模块和请挖热点模块外，凡与单文本热点挖掘系统相同的模块，均具有与其相同的功能。②信息整合模块把待整合数据归并整合成表 2 的处理任务改为交送给热点分析器来完成，故不再进行各记录信息温度的求解和比较；请挖热点模块，改为向文本数据库模块发送热点挖掘请求，故不再向文本分析器发送热点挖掘请求。③当数据库管理员向数据库中存入一个文本文件时，要先调用本系统各相关模块对该新文本文件进行关注指标数据统计处理，并把所得关注指标数据提交给热点分析器；热点分析器用这些关注指标数据，对表 2 的“次数”、“段数”属性数据分别作同类数据的累加刷新，“文本数”属性数据则作加 1 刷新，而“信息温度”属性数据则另作计算刷新、降序刷新。④当数据库管理员向数据库中删除一个文本文件时，基本类同于这里所述的③，即只需把上述“累加刷新、计数刷新”改为“累减刷新、减 1 刷新”。⑤根据表 2 的“信息温度”属性数据，挖掘出文本数据库中信息温度相对最大的用户指定个数的若干热点关键词，以供用户查询。

3.3 Web网页热点挖掘系统模型

当前Web可以看作是一个巨大的数据库，有大量的知识信息在其上表示。它的分布式、动态化和开放性特点，加大了热点挖掘的难度。对此，可应用多层次数据处理和Web服务器日志技术，并结合上述文本数据库的热点挖掘系统来解决。此时，其信息记录一般形式为information(关键字)=(网站：关键字；次数，段数，文章数，网页数，网页点击数，网站数)；Web网页热点挖掘系统结构模型可从略，其“Web网页受关注情况及其信息温度一览表”示例如表3所示。

表3 Web网页受关注情况及其信息温度一览表(示例)

关键词	次数	段数	文章数	网页数	网页点击数	网站数	信息温度
公平	720888	15009	3590	888	6666	44	721084.5254
贫困	55788	9239	1999	555	2587	39	56645.0190
…	…	…					…

在Web网页热点挖掘系统中：①搜索引擎提出热点挖掘请求，由其服务器向其所有能建立链接的站点发出请求，希望对它的“Web网页受关注情况及其信息温度一览表”，提取自己的“关键字、次数、段数、文章数、网页数、网页点击数”基础数据。②各站点内部存储的“Web网页受关注情况及其信息温度一览表”，是由其Web服务器内的文本数据库热点挖掘系统(注意：此时，其挖掘对象由文本变为网页)，并结合Web日志中的点击记录来提供。③收到来自不同站点的记录表后，搜索引擎的服务器负责利用SVD算法再次整合。④对各关键词信息温度进行计算刷新，降序刷新。⑤根据表3的“信息温度”属性数据挖掘出Web网页中信息温度相对最大的用户指定个数的若干热点关键词，以供用户查询。

此外，由于互联网分布存储着大量网页，为节省资源、提高效率，各站点可只对信息记录表进行定期(如一至两日)更新，而不必对新加入的网页进行实时统计处理。

4 文本数据库热点挖掘系统的应用示例

为说明上述系统的应用，作者从“样例文本数据库”中摘取了“2005年全国两会开会期间”的下列5篇样本文章，权作“测算信息温度，挖掘信息热点”的文本对象示例：

【文本1】贫富差距拉大引发社会矛盾　中国居民安全感下降

中新网2月20日电中国青年报报道，零点公司最新发布的一项针对4128名18～60岁常住居民的调查显示，中国城乡居民的社会治安安全感，从2003年开始连续3年呈下降趋势。专家分析认为，各种社会矛盾的集中爆发，在一定程度上是城乡贫富差距持续拉大，以及社会信息开放度增加等因素造成的。

调查显示，2005年居民社会治安安全感得分为3.53分，低于2004年的3.62分，更低于2003年的3.66分。通过对城乡居民的对比，研究人员发现，城镇居民的安全感小幅上升，农村居民却有较大幅度的下降，这是4年中城乡居民治安安全感差距最大的一次。

记者注意到，这是中国城乡居民社会治安安全感自2003年首次回升以来出现的连续下降。此前，无论是国家统计局，还是作为民间调查机构的零点公司，得出的结论都显示，上个世纪90年代后半期开始直到2003年以前，中国公众社会治安安全感一直呈下降态势。2003年后大幅回升，但从2004年又开始呈现下降趋势。

中国人民公安大学治安系教授、硕士生导师王太元就这一问题接受记者采访时表示，这一调查客观地反映了当前老百姓的社会心理状态。

王太元认为，社会信息开放度增加，各种负面报道开始增多，客观上影响了老百姓的心理认知。尤其是去年，国家将重大自然灾害的发生以及死亡人数的报道“解禁”，可以预期的是，未来老百姓了解到的这方面的信息会越来越多，这是社会进步的表现。“社会发展的不均衡，尤其是城乡贫富差距持续拉大，引发了各种社会矛盾的集中爆发，与之相关，各种治安和犯罪现象也有上升趋势。同时我们也应该看到，这种不均衡在客观上加大了公众的心理负担，导致了社会治安敏感度的上升。”

中国社会科学院社会学博士生、华北科技学院副教授、安全社会学研究者颜烨认为，农村居民社会治安安全感下滑严重，反映了村民自治制度下，国家公共权力在农村的消隐。目前，农村缺乏强有力的“替代威权”，村霸势力、地痞流氓等社会恶性势力，在有些地方一度膨胀。另一方面，一些村和乡镇的公共权力对农民的非法盘剥现象，再加上市场经济转轨条件下，农村社会保障制度的缺失，这些都导致了农村居民对自己的财产安全更为担心。

【文本2】2020年城乡收入差距绝无可能扩大到七倍

今天上午，中央财经领导小组办公室副主任陈锡文在国务院新闻办公室举行的新闻发布会上强调，有些专家测算2020年中国农村和城镇的收入差距可能会加大到七倍，是绝对不可能的。

2005年一号文件减轻农民税负

陈锡文说，大家非常关注中国城乡居民的收入差距问题，根据统计局测算的数字，2003年城乡居民之间的收入差距大概是3.23倍，也就是3.23个农民的收入相当于一个城镇居民的收入。去年一号文件中，中央采取了一系列支持农民收入增长的政策，再加上去年天气比较好、收成比较好，粮食价格有了比较明显的恢复，所以去年农民收入增加是比较多的一年。从目前各方面了解的情况看，2004年城乡居民之间的收入差距不会进一步扩大。今年的一号文件一共分九个部分，直接讲了中央要采取各项政策来帮助农民增收，包括会继续减轻农民的税负。

2005年一号文件帮助农民增收

陈锡文说，对中国来说，还有8亿多农民，要想迅速地使农民收入达到城镇居民的水平，这是很难做得到的。但是，从中央的政策方向上可以看到，一是中央政府采取各种措施来减轻农民负担、增加农民收入；二是要对农产品市场进行合理的调控，保证供求总体平衡，价格基本稳定；三是不断地减少农民的数量，使得继续务农的农民逐步地扩大经营规模；四是在财政方面应该更多地向农业、农村倾斜，让农民更多地享受到政府提供的公共服务。

陈锡文强调，通过这些措施，农民的收入会逐步地加快增长，但是这是一个很长的过程，但绝不会出现有的专家测算农

民收入和城市居民收入扩大到七倍以上的现象，这是绝对不可能的。

【文本3】厉以宁：收入分配不能"均贫富"

本报讯"当前最大的社会问题是收入分配的两难境地。"针对时下大家关注的贫富差距问题，在政协小组讨论会上，著名经济学家厉以宁发表了对当前中国发展形势的看法。

"为什么把收入分配说成是两难问题呢？"厉以宁解释，这是因为，一方面，这些年沿海地区发展比内地快，城市发展比农村快，一部分家庭收入增加比较快，贫富差距在扩大，贫富差距的扩大会成为一个社会问题。如不注意这个问题，不设法提高低收入者的收入，社会不易稳定。

"但这样的问题是难以避免的，因为一部分人先富起来，一部分地区先富起来，收入就不可能同步增长。收入差别的扩大不可避免。"

另一方面，外资在进入，外企在以高工资挖人，使国内长时间以来高素质的劳动力拿低收入的现象将不复存在。你如果不给高水平人才以高报酬，他们就都被挖走了。同时，要大力发展高新科技产业，必须鼓励创业投资。这样就要让投资者得到创业利润回报，否则谁都不愿意冒风险去投资。要发展地方经济，缓解就业困难，就必须鼓励民营经济发展。但你如果不让民营投资者得到合理的回报，谁给你投资？

"所以，这样一方面贫富差距扩大，会影响社会的安定；另一方面，对他们不运用激励机制，不给创业者以创业利润，不鼓励民营经济的发展，经济就上不去。这就是我国目前的两难境地。"

怎么办？厉以宁认为，必须在二者之间找出一条平衡的道路。不能采取"均贫富"、"抑富济贫"之类的手段，而是不能让收入差距过分地悬殊，要想办法尽快提高低收入者的收入，"如果采取'均贫富'的做法，顾了这头忘了那头，中国经济起不来！"

【文本4】贫富差距不断扩大　富人出钱实现第三次分配

贫富差距不断扩大已经成为困扰中国的一大社会难题。日前，全国人大常委、上海市人大常委会副主任厉无畏向记者透露，在即将开幕的2005年全国两会上，他将针对于此向全国人大提出自己的建议：积极推进调节收入的社会第三次分配，进一步动用社会力量促进社会和谐、高效发展。

建议为私人捐资成立基金会大开绿灯

2月22日，在上海社科院部门经济研究所所长厉无畏赴京参加全国人大常委会的前一天，他接受了记者的采访。在不久前召开的上海两会上，民生问题便是与会代表关注的焦点，厉无畏代表告诉记者，民生问题也将是他在全国两会上关注的中心议题。

厉无畏认为，处于转型过程中的中国经济，社会收入分配的差距不断扩大，在很大程度上影响到和谐社会的构建。中央也意识到收入差距过大对社会带来的危害，并试图通过二次分配（即税收调节和财政转移支付）来解决公平问题。然而，由于过高的所得税将会影响效率，降低人们创造财富的积极性，同时中央和地方的财力都很有限，财政转移支付能力还远不能解决问题。因此我们应当借鉴发达市场经济国家的经验，推进调节收入的社会第三次分配。

所谓社会第三次分配是从支出上考虑，在一些社会生活领域里如何让富人多出钱，穷人少出钱，也即实行社会收入的转移支付，弥补财政转移支付的不足。厉无畏强调说，由私人捐资建立的各种基金会应该在第三次分配中扮演重要角色。

以教育收费为例，有人认为高校收费过高使很多贫困学生上不起大学，所以要降低收费。其实学费高低并非公平的标准，合理的制度应该是富人多出钱，穷人少出钱。美国大学的学费高达几万美元，但很多中国学生去读书时一分钱都没出，因他们拿到了各种各样的奖学金。穷学生为什么可以不付费、少付费？因为有富人付出了高学费，有大老板捐赠建立的各种各样基金（如洛克菲勒基金、福特基金等），他们在背后支持着美国高校的正常运转。

【文本5】吴敬琏：应向职工划转国有资产以缩小贫富差距

吴敬琏日前在接受媒体采访时提出，向职工划转国有资产以缩小贫富差距，消弭社会矛盾，同时对股市的批评在他过去的基础上又升一级。

向职工划转国有资产

收入不平等、贫富差距过大，是目前我国社会面临的一个严重问题。现在的问题是，在新的一年里，怎样使政府缩小贫富差距的努力更富有成效？

吴敬琏认为，一件眼前能够做、也完全应该做的事情，是划拨部分国有资产来偿还国家对国有企业职工的社会保障隐性负债。这件事情由于种种原因未能实现。2001年再次提出，但是阴差阳错，"划转"演化成完全不同的另一件事情——"减持"。而"减持"由于违反了程序公正的原则也不可能进行下去，于是偿还政府对职工的隐性负债问题也束之高阁了。

吴敬琏认为，向职工划转国有资产，不仅可以缩小贫富差距，消弭社会矛盾，而且有助于解决国有企业国家股一股独大的问题，改善我国大企业的所有制结构。

股市是没规矩的

吴敬琏在接受采访时，对股市当前状况提出了严厉的批评，"股市很像一个赌场，而且很不规范。赌场里面也有规矩，比如你不能看别人的牌。而我们的股市里，有些人可以看别人的牌，可以作弊，可以搞诈骗。做庄、炒作、操纵股价可说是登峰造极。"

吴敬琏说："2001年时还有人说我是'拽着头发却想飞出地球'，很多批评蜂拥而至。现在还有人会说股市没有泡沫吗？我们要做出预测，尤其应尽可能地保护投资者的利益。"

三点建议解决股市危机

解决股市投机性泡沫，吴敬琏提出三点建议：第一用行政的方法限制市场不是解决问题的办法。第二用妥善的方法解决全流通的问题是解决问题的好办法。第三应该设计更完善的社会保险制度和创立更好的金融机构来有效地处理危机，获得好的效果。

问题关键股市制度缺陷

对于记者有关"现在股市已低得不能再低了，2005年是股市关键的一年，真的能带给股民希望吗？"吴敬琏说："高低只是现象，问题的本质还是体制。"

经过对上述5个文本样例的信息温度测算处理，所挖掘出的前30个热度较高关键词及其信息温度如表4所示。基于信息温度发现的前五大热点"农村居民、收入、股市、贫富差距、发展"，恰与2005年全国两会会议期间所证实呼声最高的前五大热门话题相吻合。

表4 多文本样例"信息温度测算与信息热点挖掘"的部分处理结果表

关键词	次数	段数	文本数	信息温度	关键词	次数	段数	文本数	信息温度
农村居民	18	7	2	19.41649	富人	3	2	1	3.741657
收入	18	7	2	19.41649	城镇居民	2	2	2	3.464102
股市	13	9	1	15.84298	社会保障制度	2	2	2	3.464102
贫富差距	11	9	4	14.76482	信息开放	2	2	1	3.000000
发展	7	4	1	8.124038	增收	2	2	1	3.000000
社会治安	6	5	1	7.874008	收入分配	2	2	1	3.000000
安全感	7	3	1	7.681146	和谐社会	2	2	1	3.000000
职工	5	4	1	6.480741	穷人	2	2	1	3.000000
社会矛盾	4	4	2	6.000000	均贫富	2	1	1	2.449490
基金	5	3	1	5.916079	常住居民	1	1	1	1.732051
国有资产	4	4	1	5.744563	社会心理	1	1	1	1.732051
第三次分配	4	3	1	5.099019	社会进步	1	1	1	1.732051
税负	3	3	2	4.690416	村民自治制度	1	1	1	1.732051
收入差距	4	2	1	4.582576	财产安全	1	1	1	1.732051
财政	3	2	1	3.741657	粮食价格	1	1	1	1.732051

结论 在信息爆炸的今天，本文基于同构化基本原理提出了一种可定量化测度信息重要性与受关注程度的新尺度——同构化信息温度，它可用作在众多信息挖掘出重要性与受关注程度最高的热点信息的新方法。同时，本文将信息温度与计算机、互联网技术相结合，分别构造了单文本热点挖掘、文本数据库热点挖掘和Web网页热点挖掘的系统模型框架；从而，为单文本、文本数据库和Web网页的热点探测、热点发现与热点挖掘，提供了一种新工具，开辟了一条新途径。

参考文献

1 夏征农，等. 辞海[M]. 上海：上海辞书出版社，2002
2 朱明. 数据挖掘[M]. 合肥：中国科学技术大学出版社，2002
3 黄海，王儒敬，黄河. 一种基于HornML的web知识表示方法[J]. 计算机工程与应用，2006 (1)：53～55
4 蒋凯，武港山. 基于Web的信息检索技术综述[J]. 计算机工程，2005 (24)：7～9
5 李超锋，卢炎生. Web使用挖掘技术分析[J]. 计算机科学，2006 (2)：220～222

(上接第109页)

国内的理论研究主要是探索如何将新公共管理理论应用于电子政务，而实际应用则主要集中在如何使用具体技术构建应用系统。文[13]对以用户为核心的电子政务发展趋势做出了分析及预测；文[14]对国内电子政务领域1999～2004年的学术文章进行了研究统计，分析了电子政务的各个研究领域，得出标准技术应用和资源整合是电子政务的发展趋势和研究热点；文[15]使用基于工作流的方法设计了行政许可系统；文[16]研究了基于J2EE软件技术框架理论的政务系统框架，并采用Spring技术实现了该框架；文[17]比较了企业信息集成的各种技术，认为业务流程管理是未来企业信息系统体系结构的重要发展方向。

总结 电子政务通用标准顶层设计框架是当前电子政务研究的难点和重点，顶层框架是电子政务建设和实施的前提和关键。本文结合前期电子政务基金项目的工作成果和对发达国家电子政务框架的研究成果，提出了面向公民的电子政务服务框架(CCeGSF)，该框架采用面向用户服务的设计理念，通过SOA和BPM等相关技术标准把任何应用系统都封装成接口性好、重用性强的服务实体。该框架不仅设计灵活、各层之间逻辑性强、具体应用过程形成完整的生命周期；而且具有耦合性松、灵活性强、资源共享性强、可重用性强、投入成本低、实现风险低等优点。CCeGSF在大连市高新技术开发区电子政务服务系统的设计和开发中得到了应用。事实证明，基于CCeGSF的电子政务服务系统操作方便、简洁，能够快速响应用户需求，基本实现了以公民为核心的服务性电子政府特征。

下一步需要对CCeGSF中业务流程层的业务流程动态组合性进行更深入的研究，同时进一步调研、探索更加符合用户行为特征的COI分类算法。

参考文献

1 Erl T. Service-oriented Architecture (SOA): Concepts, Technology, and Design. USA, Prentice Hall, Pearson Education Inc, 2005. 25～56
2 Simith H. Business process management-the third wave: business process modeling language (BPML) and its pi-calculus foundations. Information and Software Technology, 2003, 45: 1065～1069
3 Renner S A. A "Community of Interest" Approach to Data Interoperability. In: Federal Database Colloquium, San Diego, CA, August 2001. 1～6
4 OASIS: Business Process Execution Language for Web Services (BPEL4WS) [EB/OL]. http://xml.coverpages.org/bpel4ws.html, 2003
5 OWL-S: Semantic Markup for Web Service [EB/OL]. http://www.w3.org/Submission/OWL-S, 2004-11
6 The DAML Services Coalition. DAML-S: Web service description for the semantic Web. In: Proceedings of the 1st International Semantic Web Conference (ISWC), Sardinia, Italy, June 2002. 348～363
7 Stoltzfus K. Motivations for implementing e-government: an investigation of the global phenomenon. DG. O, Santa Barbara, CA, January 2004. 333～338
8 Forman M, et al. E-Government Strategy [EB/OL]. http://www.firstgov.gov/Topics/Includes/Reference/egov_strategy.pdf, 2002
9 UK e-Government Interoperability Framework[EB/OL]. http://www.govtalk.gov.uk/documents/e-GIF_version_3_approved.pdf, 2003
10 Santos I, Madeira E, Tschammer V. Towards dynamic composition of e-government services: a policy-based approach. In: Challenges of Expanding Internet: E-Commerce, E-Business, and E-Government, Poznan, Poland, October 2005. 173～185
11 Tsai W T. Service-oriented System Engineering: A New Paradigm. In: Proc. of IEEE International Workshop on Service-Oriented System Engineering (SOSE), Beijing, October 2005. 3～8
12 杨吉江，邢春晓. 美国电子政务通用框架模型研究. 电子政务，2006，4：12～21
13 胡小明. 以公众为中心的电子政务趋势研究. 信息化建设，2006，8：12～15
14 孙玉伟，李勇. 中国电子政务研究现状及趋势分析. 现代情报，2006，4：7～11
15 吴慰娜. 基于工作流的"一站式服务"行政许可系统. 计算机工程，2006，18(30)：267～292
16 刘和洋，王健华，黄永红，等. 基于Web的政务系统通用框架的研究与实现. 计算机工程，2006，14(32)：263～265
17 谭伟，范玉顺. 业务过程管理框架与关键技术研究. 计算机集成制造系统-CIMS，2004，10(7)：737～743