

信息检索丛书

VISUALIZATION
FOR INFORMATION RETRIEVAL

信息检索可视化

〔美〕 Jin Zhang 著
夏立新 陆伟 沈吟东 等译

 科学出版社
www.sciencep.com

信息检索丛书

信息检索可视化

Visualization for Information Retrieval

〔美〕 Jin Zhang 著

科学出版社

北京

内 容 简 介

本书从理论和实际应用方面,对信息检索可视化领域最新进展进行了系统的阐述,集中介绍了信息检索可视化的基本原理及其结构、方法和算法,较为系统地比较了多重参考点可视化模型(MRPBM)、基于欧几里得空间特征的可视化模型(ESCBM)、探路者关联网(PFNET)、多维尺度模型(MDS)、自组织图模型(SOM)等五种主要的信息检索可视化模型,讨论了信息可视化系统的评估及信息可视化系统设计中人的行为,分析了信息可视化中的一些特有的歧义现象,并提出解决办法。

通过本书,读者可以了解和掌握信息检索可视化的现状,获悉主要学者的一些技术、理论发现和前沿进展,信息检索可视化系统实现的一些充分的、有实用价值的细节,以及研究这个领域所必须注意的问题。

本书可作为信息科学、计算机科学及相关学科的研究人员、从业者、系统设计者和分析者、网络管理员、图书管理员、教育者和研究生的参考资料。

Translation from the English language edition:
Visualization for Information Retrieval by Jin Zhang
Copyright ©2008 Springer-Verlag Berlin Heidelberg
Springer is a part of Springer Science+Business Media
All Rights Reserved.

图书在版编目(CIP)数据

信息检索可视化 = Visualization for Information Retrieval/(美)张进(Jin Zhang)著;夏立新等译. —北京:科学出版社,2009
(信息检索丛书)
ISBN 978-7-03-024743-8

I. 信… II. ①Jin… ②夏… III. 计算机应用-情报检索-研究
IV. G354.4

中国版本图书馆 CIP 数据核字(2009)第 094515 号

责任编辑:余 丁 潘继敏 / 责任校对:宋玲玲
责任印制:赵 博 / 封面设计:耕者

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009年6月第 一 版 开本:B5(720×1000)

2009年6月第一次印刷 印张:16

印数:1—3 000 字数:295 000

定价:45.00 元

(如有印装质量问题,我社负责调换〈双青〉)

译者序

本书的作者张进博士现为美国威斯康星大学密尔沃基分校信息研究学院终身职位的教授。1979年到1986年的7年时间里,他在武汉大学信息管理学院完成了情报学本科专业以及计算机信息检索方向硕士研究生专业,获理学学士学位及理学硕士学位。1994年,张进只身前往美国,在匹兹堡大学图书情报学院攻读计算机信息检索方向博士研究生,从此走入信息检索可视化研究领域,在国际学术期刊发表学术论文30多篇,曾经获得美国福布莱特基金奖、美国情报技术协会Pratt-Severn奖、美国情报技术协会国际数字图书馆论文奖等奖项,在信息检索领域产生了重要的学术影响。最近的一项研究表明,张进博士的SCI和SSCI论文在全美所有的图书情报学院排名第六。

本书是张进于2008年出版的有关信息检索可视化方面的力作,是施普林格出版社推出的最负盛名的信息检索系列丛书之一,而该丛书主要由在情报学和计算机科学领域享有盛誉的学者撰写。本书主要介绍了信息检索可视化的方法和技术,集中介绍基本原理及信息检索可视化的结构、方法和算法,较为系统地比较五种主要的信息检索可视化模型:多重参考点可视化模型(MRPBM)、基于欧几里得空间特征的可视化模型(ESCBM)、探路者关联网(PFNET)、多维尺度模型(MDS)、自组织图模型(SOM)。本书还讨论了信息可视化系统的评估及信息可视化系统设计中的行为,这对设计一个用户友好和有效的界面起着至关重要的作用,而界面是信息检索可视化的中心环节。本书重点关注网络通信,分析了信息可视化的原理和应用,这可以使网络信息搜索和浏览更为便利。本书还分析了信息可视化中的一些特有的歧义现象,并提出解决办法。最后,本书对未来的探讨方向给出了一些建议。

众所周知,网络信息资源等各类信息系统正急剧地向着多样化、复杂化发展,现有的信息检索方法已无法满足日益增长的检索需求,传统信息检索的方法和原理面临着空前的挑战,这就需要寻找出一种新的方法来帮助人们检索相关信息。信息可视化通过一种透明的、形象的、交互的环境来揭示数据之间的关系,这正能应对信息检索所面临的挑战。然而,现有的研究没有系统全面地将信息检索和信息可视化联系起来,信息检索的可视化没有得到很大发展。本书就是应当前学科发展需求和科学研究的缺陷而生,从理论和实践上系统地说明了信息检索可视化的最新进展。通过本书,读者可以深入地了解可视化在信息检索中的地位和作用。本书在理论上和技术上为信息检索可视化系统的执行作了充分的准备,必将信息

检索可视化带领到一个新的高度,而现存的一些问题更应引起广大研究者及专业人员的关注。本书的面世,必然会引起更多人对信息检索可视化的关注,促进信息检索方法的发展,为人们信息检索创造更为便利的环境。

本书的内容涉及:信息检索的基础理论与框架,适用于多控制点的可视化模型,基于可视化模型的欧几里得空间特征,自组织图,网络导航链接,多维排列法,层次结构可视化,网络信息可视化,可视化中的歧义,信息可视化系统设计中人的行为,信息检索可视化评估,总结以及未来发展等。

从学术思想上来说,本书从信息检索的角度来看信息可视化,作者以崭新的视角将二者联系起来,探询二者之间的联系并提出新的观点,这在现有书籍中是非常少见的。从本书的内容来说,除了包含先前作者已出版的信息检索和信息可视化方面的研究外,还提出了一些未出版的新观点,如系统评价标准等。本书介绍了在不同信息检索可视化系统中产生歧义现象的原因和联系,这在以前的信息可视化研究中是少有涉及的。作者对信息检索可视化还提出了新的分类系统。另外,在本书中,作者对他以前的一些研究成果作出了一些创新。本书的内容大多都是作者的创新观点,是在现有信息检索和可视化基础之上的融合,具有很强的理论和实践价值。从结构上来说,本书在结构的编排上逐层深入,从信息检索和可视化的一般特征揭示逐层深入到对信息检索可视化模型、特征和各种方法的讨论,重难点突出,层次分明。

作为一名海外华人学者,张进教授凭借自己扎实的功底和富于创新性的思维在信息检索领域,尤其是信息检索可视化领域独树一帜,论著颇丰,影响广泛。他从信息检索的角度来看信息可视化,抑或是从可视化的角度看待信息检索,视角新颖,研究深入,论述严密,代表了国外在此领域最新的进展。特别需要一提的是,张进博士早在20世纪90年代在武汉大学工作期间就在国内学术界崭露头角。1997年他荣获中国教育部跨世纪人才,同年破格晋升为教授。虽然身在海外,张进博士多年来一直关心支持国内的学科发展和专业建设,为国内学者到美国去进行学术交流和开展合作研究提供便利。最近几年,张进博士几乎每年抽时间回国讲学和开展合作研究,是中国教育部/国家外国专家局高等学校学科创新引智计划“教育数字媒体与可视化创新引智基地(批准号:B07042)”的海外学术骨干。

作为译者,能够将这样一位学者全新的观点和理论介绍给学术界同仁,为本书在国内的出版尽一份力,我们感到荣幸之至。

将这样一部内容翔实、观点新颖的英文著作翻译成中文,是对我们专业背景和知识积累的一次考验。本书涉及的领域众多,包括数学、信息科学、信息检索、计算机、心理学、社会学等。为了能够做好作者和读者之间的一座桥梁,将作者要表达的意思最大程度地完整地传递给读者,我们组建了翻译团队,并积极请教所涉领域的权威学者,力求做到准确通达。

在翻译的过程中有几点需要特别指出:首先,在人名的翻译上,为避免中文名与英文名对应上的失误,我们直接使用外国作者的原英文名而未作翻译。其次,文中的脚注,对脚注中援引判例、文献等的部分保留原文不动。再次,参考文献原样保留,便于对原著所引数据、文献等的查询。最后,在附录中,将术语后都附上英文,以便读者对照、理解和把握。

由于本书专业性较强,我们建议读者可先阅读第12章。这一部分对全书有一个提纲挈领的回顾,读者通过阅读本章可以对全文进行整体把握,便于在接下来进行深入细致的阅读理解。此外,书中出现的专业词组也是读者深入理解本书的前提,如自适应学习(adaptive learning),聚集算法(aggregation algorithm),基于角度-角度的可视化模型(angle-angle-based visualization model),人造神经网络(artificial neural network),书目耦合(bibliographic coupling),卡斯尼椭圆评价模型(Cassini oval evaluation model),认知隐喻理论(cognitive metaphor theory),戴斯系数法(Dice co-efficient measure),析取评价模型(disjunction evaluation model),文献参考点向量(document reference point vector),特征分解定理(eigen-decomposition theorem),分级聚类算法(hierarchical clustering algorithm),信息检索可视化(information retrieval visualization),反演转换(inversion transformation),集成角度-距离相似性方法(angle-distance integrated similarity measure),自动关联器(autoassociator),合取评价模型(conjunction evaluation model),同位程度轮廓线(iso-extent contour),非分级聚类算法(non-hierarchical clustering algorithm),特种评价模型(non-traditional evaluation model),基于优先级遍历(priority-based traversal),分割聚类算法(partitioning clustering algorithm),弹簧理论(spring theory),检索轮廓线(retrieval contour),共现分析(term co-occurrence analysis),非准确性的容忍(tolerance of imprecise)向量数据检索系统(vector information retrieval system),视觉信息检索(visual information retrieval)等。

夏立新主持了本书的翻译工作。各章的翻译分工分别是:武汉大学信息管理学院陆伟教授负责第4、5、10章的翻译(陆伟、秦喜艳翻译了第4、10章,陆伟、龙泉翻译了第5章),华中科技大学控制科学与工程系沈吟东教授负责第1、12章的翻译,夏立新、李楠翻译了第2章,夏立新、翟姗姗翻译了第3章,夏立新、吴佳妮翻译了第6章,叶飞翻译了第7章,夏立新、张玉涛翻译了第8章,徐晨琛、杜晓曦翻译了第9章,冯小琴翻译了第11章。金晶翻译了序、前言。夏立新翻译整理了词汇表。译稿的第一轮校对工作由各位译者进行交叉校对;张进博士本人承担了第二轮的校对工作,并为本著作的出版撰写了中文版前言,他的工作使本书在学术性、准确性和可读性上增色不少;第三轮的文字校对工作主要由金晶、徐晨琛、李楠等承担。最后,夏立新对全书进行了修改和统稿,并完成终校。没有我们这个团队的精诚合作,本书的翻译工作也难以顺利完成。本书的翻译出版也得到了教育部新

世纪优秀人才支持计划(项目编号:NCET-08-0788)和高等学校学科创新引智计划(项目编号:B07042)的资助。

由于译者水平有限,本书虽经反复修改但仍无法完全避免纰漏,望各位读者能够不吝赐教,指出错误与疏漏,我们将不胜感激。

夏立新

2009年2月于武汉·桂子山

中文版的前言

能够为《信息检索可视化》中文版撰写前言,我感到非常荣幸。中国有句谚语,“十年磨一剑”。这本书是我在美国历时十年进行信息检索可视化研究的成果。1994年,我首次踏上美国匹兹堡大学攻读博士的征程,幸运之神眷顾了我,被称作信息检索可视化之父的 Korfhage 博士引我进入信息检索可视化研究之门。作为这个领域的领路人和开拓者,他领导着一个信息检索可视化团队,并且带领这个团队创建了五个信息检索可视化模型。而我就是这个研究团队的一员。起初我对信息检索可视化概念一无所知,但当我沉浸在对它潜心研究之中,我被它的魅力深深地吸引,它激起了我内心强烈的兴趣。我梦想着将来能够出一本有关这个领域的专著。在 Korfhage 博士的鼓励和启发下,我不断努力着。天道酬勤,我提出了两个信息检索可视化模型,从此,我的科研生涯翻开了崭新的篇章。遗憾的是,我还没有从匹兹堡大学毕业,Korfhage 博士因为癌症撒手人寰。

Rasmussen 博士接替 Korfhage 博士,担任我博士毕业论文的导师,指导我完成论文。在匹兹堡大学获得博士学位后,我从 1999 年开始在威斯康星州立大学密尔沃基分校任教。一个新的旅程开始了!作为一个教授,我可以继续集中精力在我钟爱的领域进行研究。得益于美国的大学给教师六年一次的学术假制度,六年工作期满后我可以花一年时间写完这本专著。总结和概括过去的研究成果,听上去是一件非常简单轻松的工作,其实是非常花费心血和时间的,不过非常值得!

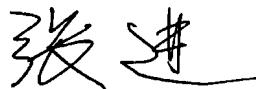
可以毫不夸张地说,如果没有我在国内十年的经历,我的研究不会达到现在的水平。我的学士学位、硕士学位都在武汉大学获得。20 世纪 70 年代末,武汉大学创办了科技情报专业。我有幸进入这个刚刚起步、前景光明的专业学习。但由于创办之初缺乏教员,我们被派往计算机科学系,和其他学科非常优秀的同学一起学习。这实在是一个难得的学习机会。在计算机和数学领域,我们学到了广泛和深入的系统知识,这为我后来的研究打下了坚实的基础。我已经从这些课程中受益,现在并且将来仍然会受益良多。同时,我在情报学系所学到的情报专业课程使我牢牢掌握了信息检索的本质。毕业后,我留在武汉大学情报学系担任教师。当时我遇到一个很大的学术上的挑战,就是将传统的信息检索理论和现代计算机技术结合起来,使其焕发青春。我对这个跨学科的研究领域充满兴趣,所涉猎的内容包括信息检索算法的优化、字符串匹配算法、汉语自动分词、数据压缩,以及人工智能。在国内的研究经历对我后来在美国的研究工作产生了深远的影响,它锤炼了我用数学方法来描述和解决信息检索领域问题的能力,培养了我对信息检索的理

解力,并且使我拥有一种在复杂情况下发现潜在研究课题的良好研究直觉。

现在读者朋友们可以理解为什么我对这本书即将翻译成中文倍感高兴。通过这种方式,更多的国内学者和学生可以分享我在信息检索可视化领域的研究成果。我衷心希望这本书能够为中国信息检索的研究和教育工作做出贡献。

这本书的翻译团队非常优秀,他们不仅是信息检索领域内经验丰富的学者,也是这个领域受人尊敬的教育者。他们对这本书的翻译工作极其重视,我充分相信他们的翻译能力。

同时我还要借这个机会感谢承担《信息检索可视化》中文版出版工作的出版社——在国内久负盛名的科学出版社。能够与这样优秀的、知名的出版社合作,我感到非常荣幸!

A handwritten signature in black ink, consisting of the Chinese characters '张进' (Zhang Jian) in a cursive style.

2009年1月

序

我第一次见到张进的时候,他受武汉大学委派,来到美国进行为期一年的学习研究。那时的我还是匹兹堡大学信息科学学院一位较为年轻的教授,对于和他的会面,我倍感荣幸。张进充沛的精力和对所研究领域的热忱给我留下了深刻的印象,有机会和这样的人一起工作,是我所期望的。他深知自己在美国的时间仅有一年,所以倍加珍惜,在这短暂的一年中完成的工作是别人要花二到三年时间才能完成的。与此同时,这一年的学习也为他在信息科学领域博士学位的获得打下了坚实的基础。尽管一年后他回到中国,但是他仍在继续研究、发展和完善他在信息检索可视化领域的一些思想以及一些以数学为基础的可视化模型。几年之后,张进有机会带着他的家人返回美国完成他的学位,那时他仍然处于“信息检索可视化环境”这一领域研究的狂热阶段。1999年,张进获得匹兹堡大学博士学位,同年前往威斯康星-密尔沃基大学信息学院任教,现在他已经是该校的副教授。

张进在浓厚兴趣的指引下,沉浸在信息检索可视化模型的研究工作中。他的灵感来源于匹兹堡大学 Robert Korfhage 教授。几年前 Bob 教授设计出信息检索可视化的一个最早模型——VIBE(示例可视化)。关于将多维空间导入到二维空间的问题虽然精妙,但可以简单地用 VIBE 解决,不过,这只是众多可能的解决路径的一种。Bob 的数学背景和张进是很好的互补,当他把张进引入信息检索可视化研究领域时,他们的合作开始了。开发有趣的并且支持高效检索的算法模型,研究有大型数据库处理能力的无歧义的数据形式,这项工作无疑是具有挑战性的。张进把 VIBE 模型作为一个起点,不是局限于早些年案例,而是发展出一些新的模型,并坚持将它们加以实现及评估。1998年,在张进即将取得博士学位的六个月前,Bob 去世了,他们的合作也随之结束。尽管张进日后的研究兴趣已经拓宽到其他领域,但他仍然继续从事信息可视化数学模型和原型的开发工作,包括开发出一些 Web 环境下的模型。《信息检索可视化》是他对这个领域研究了十年的结果。在该书里张进展示了一些模型,也指出了信息检索可视化领域的局限和挑战,为此领域的后来者提供了重要的科研资源。

Edie Rasmussen
大不列颠哥伦比亚大学

前 言

飞速增长的互联网上动态、多样、异质、复杂的信息和其他信息检索系统给传统的信息检索技术和理论带来了史无前例的挑战。这些挑战使人们对信息检索提出了加强交互性、直观性、有效性等要求,也使专家们必须集中精力寻找满足用户检索相关信息的新方法。信息可视化技术可以在一个视觉的、明晰的、交互性的环境下论证数据之间的相关性,它已成为我们应对挑战的最大希望。信息可视化和信息检索有着天然的联系。事实上信息检索就像一条线一样贯穿着整个信息可视化系统。信息可视化向用户提供独特的方式去检索信息,它可以把隐藏的信息用视觉的形式呈现,并允许用户以视觉形式进行检索。浏览作为一种强有力的信息检索方式将在可视化环境中被充分应用。可视化技术给信息检索带来了希望,从信息检索的角度来看待信息可视化,最终将使信息检索和信息可视化都受益。

本书从理论和实际应用方面,对信息检索可视化领域最新进展进行了体系化的阐明,评价了这个领域主要的技术和方法,说明了信息检索和信息可视化的理论关联,并介绍了信息检索可视化的主要算法和模型。本书也说明了信息检索可视化的一些重要的和普遍的问题,如在深度应用中出现的难以评价性、不精确性以及隐喻性。此外,它深度探讨了信息检索可视化在网络流量分析、网络信息搜寻、浏览等领域的理论和实际应用。在本书的结束部分,作者多角度比较了他所介绍的几个信息检索可视化模型。最后,本书探讨了信息检索可视化的重要问题和检索领域以后探索的方向。

本书的读者将通过此书更好地理解信息检索可视化的现状,获悉主要学者的一些技术、理论发现和前沿进展,信息检索可视化系统实现的一些充分的、有实用价值的细节,以及这个领域的研究者所必须注意的问题。

本书的结构如下:

第1章介绍了信息检索可视化的一些基础问题,如为什么信息可视化技术对于信息检索非常关键;如何从查询和浏览两个方面加强信息检索;什么是基础的信息检索可视化范例;在信息检索中信息可视化的潜在应用和影响是什么;什么是信息检索可视化模型的基础程序。

第2章包含了信息检索中一些基础的概念和理论。这些概念和算法,如相似性方法、信息检索模型、语词加权算法等,都是后续章节中关于信息检索可视化模型介绍的必要前提。把这些概念和理论放在一起组成一个章节不仅仅是为了避免在后续的章节中重复介绍,更是为了打好一个理论基础,方便读者更好地理解信息

检索可视化模型。

第3章到第7章分别介绍了多参考点模型、欧几里得空间分布特征模型、自组织图模型、探路者关联网络模型、多维尺度模型。深入探讨了这些模型的来源、概念定义、分类、算法描述、算法程序,以及这些主要信息检索可视化模型在信息检索中的应用和影响。这些章节构成了本书的中心内容。

第8章介绍了信息检索可视化模型在网络上的应用。因特网不仅给信息检索可视化带来了史无前例的挑战,也为它的应用提供了非常多的机会。信息检索可视化技术可以减轻让人烦躁的网络空间综合征和迷航,使导航更畅通舒适。除此之外,信息可视化可以应用到其他相关领域,如超链接的层次结构、主题目录、浏览历史、视觉搜索引擎结果演示,网络用户信息搜寻行为模式,网络安全,同用户进行在线讨论。

第9章内容是讲述将一个概念不明晰的情况放在视觉空间里。在不同的信息检索可视化环境里,对不明晰现象的原因加以解释,对于信息检索积极和消极的影响都加以说明,不明晰的类型被定义,如何解决这个问题的方法也一并提出。

第10章探讨了用户、系统开发商、系统设计师三方沟通和交流的隐喻与认知。分析了信息检索可视化各种情况和不同层面上隐喻的应用,同时说明了隐喻应用过程中的原则。

第11章关注的是评价问题。对信息检索可视化的评价既重要又困难,需要对可视化环境评价和可视化检索评价进行区别和分析。本章提出了一个包括信息探索、查询搜索、视觉信息表示、可控性等内容的信息检索可视化评价指标体系。

最后一章“后记”概括了本书的主要观点,从视觉空间、语义框架、投影算法、模糊性、信息检索的角度比较了五种信息检索可视化模型,最后指出了信息检索可视化的重要观点、挑战和以后的研究方向。

这本书所选取的信息检索可视化模型是基于以下几种原则:

(1) 它们是信息检索可视化领域主流的、成熟的算法或模型。这些模型被广泛应用并重组。

(2) 它们代表着各种各样的信息检索可视化。每个引入的模型都是复杂精细的,足以得出一个集群的相关模型。

(3) 它们必须反映信息检索的特点,同样也包括信息可视化语境中信息检索的特点。

(4) 它们可以挖掘出所显示对象的深层次语义以及全方位的关系。

尽管这本书只介绍了这五个模型,仍然有许多模型包含在各种各样的语境中,如隐喻的应用、书中的信息检索可视化评价。每个章节提供了一个完整的可视化模型,并且说明了它们对信息检索的影响。网络信息可视化是一个独立的章节,因为网络提供了一个理想的信息可视化技术应用的舞台,并且大部分信息检索可视

化技术要应用到网络。

我想借用这个机会感谢 Dr. Edie Rasmussen,感谢她为本书写序,以及她一直以来的鼓励和支持;感谢 Dr. Robert Korfhage,当我还在匹兹堡大学攻读博士学位时,就带我进入信息检索可视化这个神秘的研究领域;还要感谢对本书进行评论并给出宝贵建议的 Dr. Dietmar Wolfram;此外还要感谢那些匿名的评论以及审稿人和手稿最后评审人员精辟的评价;感谢 Ralf Gerstner 以及 Springer 全体工作人员,他们杰出的、专业的工作也对本书做出了巨大的贡献;感谢 Ms. Lynda Citro 对本书进行了编辑;在此也一并感谢其他对本书做出了贡献的人。同时,我也对那些允许我使用他们的图解的出版社(Elservier, Wiley, IEEE 等)表示感谢。本书有一部分由中国教育部/国家外国专家局高等学校学科创新引智计划“教育数字媒体与可视化创新引智基地”(批准号:B07042)赞助。此外,美国威斯康星大学密尔沃基分校一直非常支持这项工作。

最后,感谢我的家人对我工作的支持。

张 进

目 录

译者序

中文版的前言

序

前言

第 1 章 信息检索与可视化	1
1.1 可视化	2
1.1.1 定义	2
1.1.2 科学可视化和信息可视化	3
1.2 信息检索	4
1.2.1 浏览与提问搜索	4
1.2.2 微观信息与宏观信息	6
1.2.3 信息空间的空间特性	7
1.2.4 浏览的空间特性	8
1.3 可视化的感知与认知	9
1.3.1 感知	9
1.3.2 认知	10
1.4 信息检索的可视化.....	11
1.4.1 基本原理.....	11
1.4.2 三种信息检索可视化模式.....	13
1.4.3 建立一个信息检索可视化模型	13
1.5 小结.....	16
第 2 章 信息检索基础	18
2.1 向量空间模型.....	18
2.2 词语加权法.....	21
2.2.1 停用词	21
2.2.2 逆文献频率	21
2.2.3 萨尔顿词语加权法	22
2.2.4 另一个词语加权法	22
2.2.5 概率词语加权法	22
2.3 相似性方法.....	23

2.3.1	内积相似性方法	23
2.3.2	戴斯系数法	23
2.3.3	加卡德相似性方法	24
2.3.4	重合系数相似性方法	24
2.3.5	余弦相似法	24
2.3.6	距离相似法	24
2.3.7	集成角度-距离相似性方法	26
2.3.8	皮尔森参数 r 相关性法	26
2.4	信息检索(评价)模型	27
2.4.1	基于方向的检索(评价)模型	28
2.4.2	基于距离的检索(评价)模型	28
2.4.3	椭圆检索(评价)模型	28
2.4.4	合取检索(评价)模型	29
2.4.5	析取检索模型	29
2.4.6	卡斯尼椭圆检索(评价)模型	30
2.5	聚类算法	31
2.5.1	非分级聚类算法	32
2.5.2	分级聚类算法	33
2.6	检索结果评价	34
2.7	小结	35
第3章	多重参考点可视化模型	37
3.1	多重参考点	38
3.2	多重固定参考点模型	39
3.3	可移动多重参考点模型	41
3.3.1	原始 VIBE 算法描述	41
3.3.2	关于模型的讨论	47
3.4	自动旋转参考点模型	51
3.4.1	视觉空间的定义	52
3.4.2	参考点的旋转	53
3.5	信息检索的含义	55
3.6	小结	56
第4章	基于欧几里得空间特征的可视化模型	57
4.1	欧几里得空间及其特征	57
4.2	信息检索评价模型介绍	58
4.3	基于距离-角度的可视化模型	61

4.3.1	视觉空间(visual space)的定义	61
4.3.2	信息检索评价模型的可视化	62
4.4	基于角度-角度的可视化模型	67
4.4.1	视觉空间的定义	67
4.4.2	信息检索评价模型的可视化	68
4.5	基于距离-距离的可视化模型	73
4.5.1	视觉空间的定义	73
4.5.2	信息检索评价模型的可视化	74
4.6	小结	78
第 5 章	Kohonen 自组织图——一种人造神经网络	79
5.1	神经网络介绍	79
5.1.1	神经网络的定义	80
5.1.2	神经网络的结构和特性	80
5.2	Kohonen 自组织图	82
5.2.1	Kohonen 自组织图结构	83
5.2.2	SOM 算法的学习过程	85
5.2.3	特征图标识	88
5.2.4	SOM 算法描述	89
5.3	SOM 于信息检索领域的意义	91
5.4	小结	92
第 6 章	探路者关联网	94
6.1	探路者关联网的特性/性能和描述	94
6.1.1	探路者关联网的定义及说明	94
6.1.2	算法的描述	97
6.1.3	图布局法	100
6.2	信息检索的含义	101
6.2.1	作者同被引分析	101
6.2.2	语词关联网	103
6.2.3	Hyperlink 超级链接	104
6.2.4	探路者关联网中的搜索	104
6.3	小结	105
第 7 章	多维尺度法	107
7.1	多维尺度法简介	108
7.1.1	经典 MDS 算法	108
7.1.2	非计量 MDS	114

7.1.3	计量 MDS	118
7.2	MDS 技术对信息检索的启示	119
7.2.1	显示对象和对象之间的距离的定义	119
7.2.2	MDS 显示空间的开发	120
7.2.3	讨论	121
7.3	小结	123
第 8 章	因特网的信息可视化	124
8.1	引言	124
8.1.1	因特网的特征	124
8.1.2	因特网的信息组织和显示的方法	125
8.1.3	因特网信息的利用	127
8.1.4	因特网的挑战	128
8.2	因特网的信息可视化	129
8.2.1	因特网信息结构的可视化	130
8.2.2	因特网信息查询的可视化	138
8.2.3	网络交通流量的可视化	140
8.2.4	联机讨论历史记录的可视化	144
8.3	小结	144
第 9 章	信息可视化中的歧义性	146
9.1	信息可视化的歧义性及其影响	146
9.1.1	信息可视化中歧义性的原因	146
9.1.2	歧义信息可视化的影响	147
9.2	信息检索可视化模型中的歧义性分析	148
9.2.1	欧几里得空间特征为基础的信息模型的歧义性	148
9.2.2	基于多参考点的信息可视化模型中的歧义性	154
9.2.3	探路者网络中的歧义性	157
9.2.4	SOM 中的歧义化	159
9.2.5	多维排列的歧义化	160
9.3	小结	160
第 10 章	隐喻在信息可视化中的意义	162
10.1	隐喻的定义、基本要素和特征	162
10.2	隐喻的认知基础	164
10.3	心智模型、隐喻和人机交互	165
10.3.1	人机交互中的隐喻	165
10.3.2	心智模型	165