基于潜在语义的个性化搜索关键技术研究

陈冬玲 著

中国水利水电出版社
www.waterpub.com.cn

基于潜在语义的个性化搜索关键技术研究

陈冬玲 著

中国水利水电出版社
www.waterpub.com.cn

图书在版编目（CIP）数据

基于潜在语义的个性化搜索关键技术研究 / 陈冬玲著. -- 北京 : 中国水利水电出版社, 2013.8
ISBN 978-7-5170-1031-9

Ⅰ. ①基… Ⅱ. ①陈… Ⅲ. ①互联网络－情报检索
Ⅳ. ①G354.4

中国版本图书馆CIP数据核字(2013)第154882号

策划编辑：陈　洁　　责任编辑：陈　洁　　加工编辑：李　燕　　封面设计：李　佳

书　　名	**基于潜在语义的个性化搜索关键技术研究**
作　　者	陈冬玲　著
出版发行	中国水利水电出版社 （北京市海淀区玉渊潭南路1号D座　100038） 网址：www.waterpub.com.cn E-mail：mchannel@263.net（万水） sales@waterpub.com.cn 电话：（010）68367658（发行部）、82562819（万水）
经　　售	北京科水图书销售中心（零售） 电话：（010）88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排　　版	北京万水电子信息有限公司
印　　刷	三河市国源印刷厂印刷
规　　格	170mm×227mm　16开本　9.5印张　170千字
版　　次	2013年8月第1版　2013年8月第1次印刷
定　　价	**36.00元**

前　言

随着网络技术的飞速发展，信息爆炸所产生的个人信息疲劳和信息压力使搜索引擎变得越来越重要，搜索引擎已经成为名副其实的信息枢纽和信息门户，是用户获取网络信息的首选工具。然而，在搜索引擎返回的巨大的结果列表中，只有一小部分信息符合用户的偏好，甚至在 top K 结果中，没有符合用户偏好的信息。面对如此窘境，我们不得不重新审视，究竟如何才能为用户提供符合其偏好的个性化信息？

本文分析其主要原因在于，没有真正理解用户查询背后的潜在语义动机，不清楚用户要做什么，故无法为其提供高质量的个性化服务。

搜索引擎直接面对知识背景及搜索意图各异的用户，因此，不可能有一种普适的查询方式，能弄清楚不同用户输入同一查询词，他们各自的潜在动机分别是什么，他们到底想要得到什么样的信息。例如：用户输入“东北大学”，其可能是想随机了解一些东北大学的普遍信息，也可能是想查询今年的招生政策，还可能是想了解外界对东北大学有些什么评价。由此可见，用户的潜在语义动机理解是个性化搜索的基石，如该环节理解得不够准确，与用户实际需求匹配性不高，那么后续进行的个性化服务工作就有可能误入歧途。在实际查询中，输入“关键词”是用户在搜索中的第一步，代表了用户对于自身的搜索需求的 TAG 化表述，互联网“全息搜索理论”创始人顺风认为：需要深刻的认识到在传统搜索系统中“关键词”在用户心中产生的过程和搜索输出之间的相互关系，发现在用户搜索动机、搜索前思维量与搜索引擎反馈之间的全息联系，用户输入的“关键词”实际上就是一个将心算出的 TAG 引入搜索行为的过程，而且此类 TAG 应该成为最有质量的 TAG，因为其中凝聚了搜索用户第一反映的无意识性的内心智慧。搜索引擎只有准确把握用户的搜索动机，才能有的放矢地为其提供高质量的个性化服务。

基于上述分析，本文从用户潜在语义的用户动机分析入手，并以此为主线，对多种个性化服务关键技术进行了研究，主要工作包括以下几个方面：

（1）在计算机研究领域内，从哲学、心理学角度剖析用户搜索行为，并从认知学的角度，提出了基于概率潜在语义动机分析的用户行为模型，高度概括了各种具体搜索行为，从抽象的角度去理解用户的搜索行为。该模型的提出为进一步

研究个性化搜索提供了新的思路。

（2）在文档潜在语义空间中，应用 Zipf 分布与概率潜在语义分析算法相结合的方式进行文档潜在主题提取，改善了文档潜在主题提取的质量。

（3）以狄氏先验的有限混合模型理论为基础，提出了高效无监督的网页聚类算法。可以有效克服一般的文本聚类算法无法有效应对的高维性、稀疏性文本，以及文本数据之间的相似性函数定义困难，聚类质量和效率低等不足，改善了聚类效果，提高了捕获用户兴趣潜在主题需求的能力。

（4）提出了一种新的基于用户潜在语义分析的查询扩展技术。即将通用搜索中查询扩展的技术与用户动机挖掘技术相结合，而开发出的一种新的查询扩展技术，解决了搜索引擎由于通用的性质而缺乏面向用户的个性化的信息处理的能力，从了解用户的语义上的搜索动机以及了解认知与心理相互作用的角度出发，从根本上解决了查询过程中的一词多义及多词同义等问题，在个性化搜索过程中有效的进行语义消歧。

（5）针对面向查询的排名算法的不足提出了面向用户的重排名算法。即在原有网页排序算法的基础上，根据用户的兴趣偏好而提出的一种局部优化排序算法，既符合用户的个性化需求，又不影响搜索结果的查全率，尽可能做到其排序结果与用户语义动机相符合。

总之，本文从用户潜在语义动机的理解出发，针对个性化搜索各个环节中的关键技术展开研究，如用户建模技术、查询扩展技术、网页局部优化排序技术、聚类技术等，力求达到用户查询与搜索引擎返回结果的高效匹配。

目　录

第 1 章　绪论

1.1　搜索引擎体系结构及功能

随着网络技术的飞速发展，人们在网络上可获取的信息呈 GigaByte 的速度猛增，信息爆炸所产生的个人信息疲劳和信息压力使得人们越来越重视搜索引擎在生活中的作用。据 CNNIC[①]2009 年 1 月最新统计数据表明，网民中遇到问题就会求助于搜索引擎的用户高达 79.6%，搜索引擎（Search Engine）所起到的信息导向作用越来越突出，它已经成为名副其实的信息枢纽和信息门户，是用户获取网络信息的首选工具。

从网络用户的角度看，它根据用户提交的类自然语言查询词或者短语，在一个可以接受的时间内返回一系列很可能与该查询相关的网页信息，供用户进一步判断和选取。搜索引擎的体系结构及功能如图 1.1 所示。搜索引擎工作过程中主要完成三个功能，即信息的收集、预处理和查询服务。

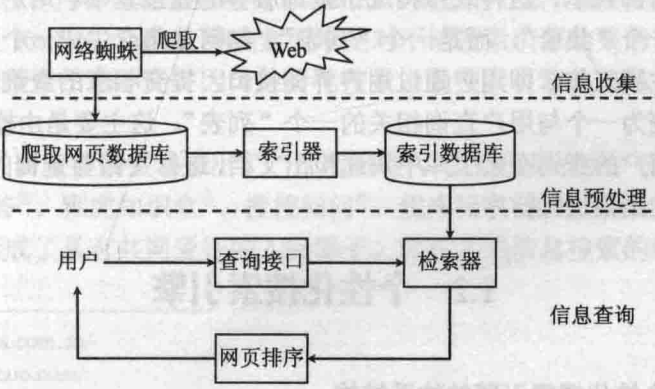

图 1.1　搜索引擎体系结构图

① http://www.cnnic.net.cn/

1.1.1 信息的收集

搜索引擎利用收集器从互联网上抓取网页，收集器又叫爬虫器，其功能是遍历 Web 空间，扫描一定 IP 地址范围内的网站，并沿着网络上的链接从一个网页到另一个网页采集网页资料。它为保证采集的资料最新，还会回访已抓取过的网页。重复这个过程，并把抓取下来的网页信息收集到爬取网页数据库中。

1.1.2 信息预处理

信息收集得到海量的原始网页集合，然后，搜索引擎必须利用索引器程序对收集回来的网页进行分析，索引器可以提取相关网页信息（包括网页所在 URL、编码类型、页面内容包含的关键词、关键词位置、生成时间、大小、与其他网页的链接关系等）形成索引项，然后是“镜像网页”或“转载网页”的消除，最后用这些信息建立网页索引数据库，索引库中存放了用于表示文档以及生成文档库的索引表，以及每一个网页针对页面内容中和超链接中每一个关键词的相关度（或重要性）。

1.1.3 查询服务

信息收集得到一个原始网页集合 S，预处理过程得到的是对 S 的一个子集的元素的某种内部表示，这种表示构成了查询服务的直接基础。用户通过搜索引擎看到的不是一个“集合”，而是一个“列表”。如何从集合生成一个列表，是查询服务系统的主要工作。即用户通过用户界面接口，提交相应的查询需求，系统将网页集合转变为一个与用户查询相关的一个“列表”。这主要是由检索器来完成，它可以根据用户的查询在索引库中快速检出文档，进行文档与查询的相关度评价，对将要输出的结果进行排序。

1.2 个性化搜索引擎

1.2.1 个性化搜索引擎的体系结构

由于网络信息具有动态性及用户兴趣具有迁移性，往往在搜索引擎返回的巨大的结果列表中，只有一小部分信息符合用户的偏好，甚至在 Top n 结果中，没有符合用户偏好的信息。尤其是 Web2.0 的“以人为中心”及“个性化”理念的盛

行，公众对于信息获取除了有量上的增长外，对于信息获取的质量要求也在逐渐提高。而传统的搜索引擎所呈现出来的通用性质，已经很难满足不同用户日趋多样化、复杂化的信息需求。面对如此窘境，人们不得不重新审视，究竟如何才能为用户提供符合其偏好的个性化信息？

在搜索的 2.0 时代，人们想出各种途径为用户提供符合其偏好的信息，但种种途径中，不泛两种技术，一种是个性化搜索技术（本文中特指利用各种搜索算法改进搜索结果的技术），一种是社会化搜索技术，为了下面行文的方便，此处将两个概念进行区分。

社会化搜索到目前似乎没有一个完整的定义，它是于 2004 年和 2005 年间被提出的，是指通过搜索形成一个有共同爱好的人际圈子，在用户收藏的网页或其所在的圈子中进行搜索，并且以标签的形式实现全社会知识的共享，社会化搜索为个性化搜索带来了一线曙光。个性化搜索是指通过改进搜索引擎的底层算法，如用户建模型算法，查询扩展算法或排序等算法，最终实现搜索结果的个性化。

个性化搜索与社会化搜索有着密不可分的关系，如果把个性化搜索理解成利用算法找到用户的所需信息，则社会化搜索可以理解成以个性化搜索为基石，将用户的所需信息进行聚合，虽然二者的根本目标都是为了得到与用户兴趣真正相符合的信息，但却有着本质的不同，个性化搜索强调的是用户心理和行为的深度分析与挖掘，是搜索引擎进行智能分析的种种算法，而社会化搜索更注重不同搜索结果的聚合，聚合的过程和内容都是非推理和非判断的。典型的社会化搜索如下：

（1）如：Rollyo、Swicki 等搜索引擎是对不同搜索引擎结果的聚合，但这种聚合显然与元搜索引擎不同，因为社会化搜索中的搜索引擎的选择是按照用户兴趣进行选择的。

（2）如：百度搜索引擎推出了“百度知道”，类似的还有如：爱问[①]、奇虎问答[②]、天涯问答[③]、雅虎知识堂[④]、搜搜问问[⑤]、优库网[⑥]、猫扑[⑦]等人肉搜索功能，通过搜索，形成了具有共同爱好的人际圈子。即在本地信息搜索的结果中，可以

① http:// iask.sina.com.cn/
② http://www.qihoo.com/
③ http://wenda.tianya.cn/wenda/
④ http://ks.cn.yahoo.com/
⑤ http://wenwen.soso.com
⑥ http://www.ucloo.com/
⑦ http://dzh2.mop.com/

看到他人对于同一搜索结果的评论与评价，用户可以通过口碑相传的方式得到搜索结果。

搜索引擎的社会化是搜索引擎个性化发展过程中的一个步骤，是时间轴上一个连续变化的概念。并且，由于社会化搜索不涉及搜索引擎的底层技术或算法，因而，它永远不能像算法搜索那样全面、有涵盖性，故不是本文讨论的重点。

本文研究的重点是个性化搜索技术。

目前，个性化搜索引擎是新一代搜索引擎研究的热点[1,2]，是针对当前搜索引擎查询手段单一，返回结果不能做到与用户的需求精确相关等缺点提出的改进方案。个性化搜索引擎是通用搜索引擎与个性化技术相结合的新一代搜索引擎，它不仅具备通用搜索引擎的功能，而且还可以根据访问用户不同提供与之相适应的个性化服务。

个性化搜索引擎（如图 1.2 所示）实际上是在通用搜索引擎与用户之间建立了一个个性化技术模块，利用用户的使用记录建立用户兴趣模型，根据用户兴趣模型进行查询的语义扩展，并对检索结果集进行个性化（基于用户兴趣）排序、语义（基于语义概念的相关性）排序或基于推荐列表（基于关联规则）的排序，最终形成一个与用户自身特点相关的检索结果列表给用户。其他还有诸如基于自然语言处理的个性化搜索引擎[3]、基于领域知识的个性化搜索引擎、和基于人类行为学的搜索引擎[4]。个性化搜索引擎极大的提高了搜索引擎的准确性和检索效率，虽然这些搜索引擎还没有实现如通用搜索引擎那样的大规模应用，但随着互联网的发展，它们必然将会成为新一代搜索引擎大代表，是目前搜索引擎一个主要的发展方向和趋势。

1.2.2 个性化搜索关键技术

由图 1.2 可以看出，个性化搜索引擎是在通用搜索引擎与用户之间建立了一个个性化技术模块，通过记录用户点击和浏览行为，并进行相应用户兴趣的语义的分析，获得特定的用户偏好信息，并根据不同用户的偏好信息进行相应的个性化服务，如个性化网页排序或推荐，或者，对用户查询进行语义消歧扩展等，即从各种角度出发满足用户的个性化需求。

但这是一个说起来简单，做起来相当困难的工作。搜索引擎直接面向知识背景，搜索意图各异的用户，由于不同用户具有不同的背景知识和信息需求，不可能有一种普适的查询方式。通常，对于普通网络用户来说，最自然的方式就是“要什么就输入什么”。例如用户输入“东北大学”，可能是他想了解东北大学目前有

些什么信息向外发布，想看看今年的招生政策，也可能是他想了解外界目前对东北大学有些什么评价。这是相当不同的需求。之所以对同一查询词有多种理解，是因为自然语言具有二义性，用户的查询往往很难准确地、完整地描述自己的兴趣需求。情报学认知范式的倡导者贝尔金把这种状态定义为“知识的非常态”（Anomalous State of Knowledge，ASK）[5,6]，即：用户信息需求的产生源于用户认识自己的知识的非常态，用户一般不能解释这种非常状态需要的是什么，因此，检索系统不要奢望能得到用户准确的需求描述，而只能希望处于“知识非常态”下的用户陈述其已知的知识及其目标。因此，个性化搜索时只能要求用户的查询信息描述要尽量完整，关键词用得尽量准确。而主要的精力应放在检索系统中，要求检索系统尽可能描述和存储用户“知识的非常态”，并进行模型化处理，用以推断用户的搜索动机。因此，根据 Calvin W. Mooers[7]对信息检索的界定，有两个亟待解决的问题：其一，将用户信息需求转换为一张文献列表；其二，进行信息的描述、信息的规范化，以及找到检索系统所应用的技术和方法。因此，个性化的关键技术都是围绕这两方面展开的。

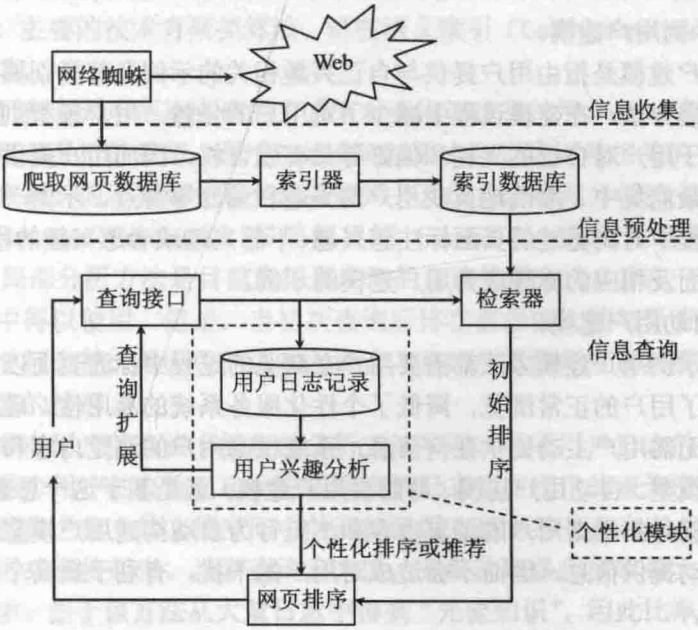

图 1.2　个性化搜索引擎结构

1. 用户建模技术

如果说搜索引擎保证了结果量的提供，那么用户搜索行为与偏好的分析保证了结果质的提升。因此，个性化搜索的核心和关键技术是用户建模，用户建模是指从和用户相关的信息（如浏览内容、浏览行为、背景知识等）中归纳出可计算的用户模型的过程。用户模型质量的好坏直接关系到个性化服务质量的好坏以及搜索引擎性能的优劣。根据建模过程用户的参与程度，用户建模技术可以分为用户手工定制建模、示例用户建模和自动用户建模[8]。

（1）用户手工定制建模。

用户手工定制建模是指用户模型由用户自己手工输入或选择的用户建模方法，如用户手工输入感兴趣信息的关键词列表，或者是选择感兴趣的栏目等。在个性化服务发展的早期，用户手工定制建模是用户建模的主要方法。MyYahoo 是手工定制用户建模的典型代表。但由于完全依赖于用户，容易降低用户使用系统的积极性，即使用户乐意手工输入用户模型，用户也难以全面、准确地罗列自己感兴趣的栏目或关键词。此外，当用户兴趣发生变化时，用户必须重新输入用户模型。

（2）示例用户建模。

示例用户建模是指由用户提供与自己兴趣相关的示例及其类别属性来建立用户模型的建模方法。在建模过程中减少了对用户的依赖，用户模型则通过学习算法得到。由于用户对自己的兴趣和偏好等最有发言权，因而用户提供的有关自己兴趣的示例最能集中、准确地反映用户的兴趣和偏好等特点。示例一般通过用户在浏览的过程中对浏览过的页面标注感兴趣、不感兴趣或者感兴趣的程度来得到。浏览过的页面及相应的标注成为用户建模的示例。

（3）自动用户建模。

目前的示例用户建模方法都需要用户在浏览的过程中标注页面以得到示例，严重地干扰了用户的正常浏览，降低了个性化服务系统的易用性。理想的用户建模方法应该无需用户主动提供任何信息，系统根据用户的浏览内容和浏览行为自动构建用户模型。自动用户建模（即隐含用户建模）就是基于这一思想提出来的。自动用户建模是指根据用户的浏览内容和浏览行为自动构建用户模型。建模过程无需用户主动提供信息，因而不会造成对用户的干扰，有利于提高个性化服务系统的易用性。

显然，用户模型建立过程中，基于隐式反馈的自动用户建模成为目前研究的热点，但用户模型的建立必须依赖于对用户行为动机的分析前提之下，即了解用

户的行为，明确用户动机的情况下，建立的用户模型才具有真正的指导搜索的意义。本文也正是着眼于此，从认知的角度进行了用户行为分析，并提出基于概率潜在语义分析方法建立用户模型的方法，此方法从用户搜索行为中获取用户的潜在语义动机，进而生成用户模型。本文提出的用户模型建立方法是对传统用户模型研究的继承与扬弃。

2. 查询扩展技术

查询扩展指的是利用计算机语言学、信息学、控制论等多种技术发展起来的查询优化方法。它通过在原来查询的基础上加入与用户查询相关联的词或词组的方式组成更长、更准确的新查询，从而达到概化或细化当前查询的目的。这在一定的程度上弥补了用户提交的查询信息不足的缺陷，也有助于改善检索系统的查全率和查准率。传统的查询扩展主要有全局分析（Global Analysis）、局部分析（Local Analysis）以及基于用户查询日志[9]和基于概念的查询扩展[10-12]。

全局分析方法是对全部文档中的词或词组进行相关分析，计算每对词或词组间的关联程度，如共现（Co-occurrence）的概率。当一个新的查询到来时，根据预先计算的词间相关关系，将与查询词关联程度最高的词或词组加入原查询以生成新的查询。主要的技术有聚类算法、潜在语义索引（Latent Semantic Indexing）、相似性词典等。

局部分析则是采用两次查询的方式来对查询词进行扩展。Atter 和 Fraenkel[13] 最早提出了局部分析的思想：利用初始查询结果中与查询词最相关的 N 篇文档作为扩展词的来源，而不需要使用全局关键词关系词典。局部分析的主要方法有相关反馈（Relevance Feedback）和局部反馈（Local Feedback，也称为 Pseudo Feedback）。局部分析方法是目前应用最广泛的查询扩展方法，并在一些实际的信息检索系统中得以使用。但是，当初次查询后排在前面的文档与原查询相关度不大时，局部分析会把大量无关的词加入查询，从而严重降低查询精度，甚至低于不做扩展优化的情形。

和基于单个用户的查询文档集的分析方法不同，基于用户日志的查询扩展方法考虑的是整个用户的查询日志。它根据查询日志建立用户的查询空间，将查询空间和在文档集上建立的文档空间中的词按照用户提交某个查询所点击的文档以条件概率的方式连接起来。当新查询到来时，系统选取条件概率最大的文档用词加入到查询中。由于该方法从大量日志中得到“先验知识”，因此比单个用户的临时判断或系统在没有考虑用户的情况下得到的结果更准确。

上述三种查询扩展方法主要是以查询词为中心，机械式的语词符号扩展，是

在符号层次上进行的查询扩展。基于概念的查询扩展是近几年的研究热点。在基于概念的语义查询扩展中，用户提出的查询以自然语言的形式来体现，因此，如何找到语义相关的概念集描述查询主旨成为一个主要问题。因此，首先要建立概念语义空间，然后从基于概念的语义空间中提取用户查询语义及其语义关联，实现语义概念扩展。目前研究方法主要是根据概念间的各种关系，利用一定的技术（如词语—概念矩阵，即以词语（Word/Term）和概念（Concept）语义相关程度为元素的矩阵）构建概念知识库、语义概念网络、概念语义词典或者概念语义树等概念语义空间，最终找到与查询主旨相关的概念。此时的查询，要看作一系列的概念，然后从建好的概念语义空间提取查询语义及其语义关联，实现语义概念扩展。

3. 对搜索结果进行个性化排序的技术

在搜索结果中，为了找出最符合特定用户的网页，完成搜索过程中的结果呈现是关键的步骤之一，返回的结果与用户的真实需求是否相符是衡量搜索准确性的唯一标准。目前，著名的搜索引擎网页排序算法有：

（1）PageRank 算法，其基本思想是网页超文本文档之间的连接可以看成是引用，如果一个页面被其他许多网页引用，则此页面很可能是重要页面；一个页面尽管没有被多次引用，但被一个重要网页引用，则此页面很可能也是重要页面；一个页面的重要性被平均分配并传递到它所引用的页面。具体来说，假设某用户跟随链接进行了一系列网页浏览，则该用户浏览的起始网页的重要程度就由该网页被此用户向下浏览的其他网页的重要程度所决定。因此，这种引用关系体现出文档的重要性能够较好地符合人们主观意识中的文档的重要性。

（2）HITS 算法，其基本思想是在网页中识别出一个子图，子图的选择依赖于用户的查询，然后对该子图进行链接分析，从中找出权威（Authority）和目录（Hub）网页。其前提假设是一个好的 Hub 网页指向很多 Authority 网页，一个好的 Authority 网页有很多好的 Hub 网页指向它。Authority 网页是相对某主题来说权威的网页，Hub 网页是链接度权威的网页。HITS 算法的目标就是通过一定的迭代计算最终得到针对某个检索提问的最具价值的网页。因为内容权威度与网页自身提供内容信息的质量相关，被越来越多网页所引用的网页，引用越多高质量的网页的网页，其链接权威度越高。

4. 聚类分析技术

聚类分析是在没有先验知识的情况下，将物理或抽象的对象集合划分成为由类似的对象组成的多个簇，使得处于同簇中的对象具有最大的相似性，处于不同

簇中的对象具有最大的差异性，属于无监督学习，也是个性化搜索服务的基础技术之一。应用在搜索引擎的聚类技术主要指文档聚类，其主要应用有两个方面。

第一个方面，对通用搜索引擎的结果进行聚类。对于某个通用搜索引擎而言，如果用户给定一个搜索关键词，返回列表非常巨大，对于用户来说，很难在短时间内找到有效信息，此外，由于阅读惰性问题，很少有用户会沿着搜索结果列表向第二面翻或翻向更多页的结果列表，这导致搜索引擎的效率低下。所以，有很多搜索引擎都提供了返回结果聚类的类界面，即将返回结果中相关信息聚为一类呈现给用户，提高用户找到准确信息的效率。比如，Google，Baidu 等是大家常用的搜索引擎，但众所周知，它们的搜索结果动辄几十页、上百页的罗列出来，而人们通常只对其中第一、二页有兴趣，没有耐心向下继续翻。人们一直设想将搜索的巨大网络资源予以分类，并追踪每个类别里最好的网站，目前主要有人工和自动两种方式。典型的聚类式搜索引擎有：http://www.vivisimo.com, http://news.google.com 等。

第二个方面，对用户日志进行聚类分析。对于非聚类式搜索引擎而言，根据用户点击的日志进行挖掘、聚类分析，找出用户关心的主题所在。即聚类技术应用不止在搜索结果上的应用，随着 Web2.0 技术的推广，无处不在的博客、空间、RSS、Wiki、网摘、社会网络、IM 及 P2P 的流媒体的发展，使用户随时随地可以按自己的兴趣与网络上的其他用户进行交互。这也为搜索引擎提供了更大的可利用空间和更坚巨的挑战，更大的可利用空间是搜索引擎完全可以将用户的博客，空间，及收藏的网摘等进行分析汇总，找出用户的兴趣所在，即用户动机分析；后者也正是本文的着眼点，找到一种最行之有效的聚类算法来进行用户日志的分析和挖掘，获取用户浏览和点击的真正目的。

1.2.3 个性化搜索研究现状

个性化搜索问题是一个经典而又富于挑战的问题，其追求的最高境界是，搜索引擎具有智能，能自动为每个用户从网络上搜索到最符合其偏好的信息。这其中涉及多种关键技术（详见 1.2.2 节），在计算机不能像大脑那样思考、在计算机的自然语言理解技术还未十分成熟的情况下，要实现这样的搜索境界还是一个梦想。但个性化搜索是一个向用户学习的过程，在现有基础上，各大搜索引擎公司在进行各种关键技术研究的同时，纷纷推出自己的个性化搜索服务，尽管很多还是对个性化搜索的尝试，但个性化搜索已经成为搜索技术的新战场。

1. 用户建模与个性化搜索结果排序

个性化搜索的原理是根据搜索及访问记录，来预测用户做新的搜索时的真实