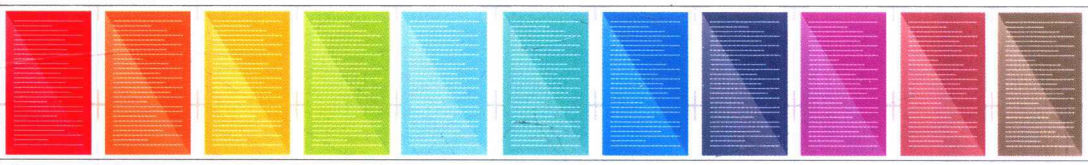
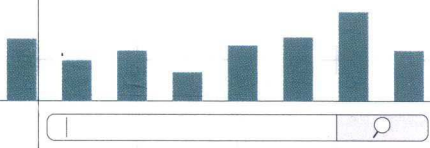
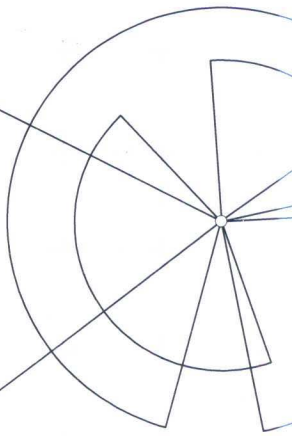
国家科技支撑计划课题“文化遗产知识本体构建存储可视化技术研究”
（2012BAH33F03）资助出版

查询意图自动分类与分析



张晓娟　著





WUHAN UNIVERSITY PRESS
武汉大学出版社

国家科技支撑计划课题“文化遗产知识本体构建存储可视化技术研究”(2012BAH33F03)资助出版

查询意图自动分类与分析

张晓娟◎著

WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

查询意图自动分类与分析/张晓娟著. —武汉:武汉大学出版社,2015.11

ISBN 978-7-307-16754-4

Ⅰ.查… Ⅱ.张… Ⅲ.互联网络—情报检索 Ⅳ.G354.4

中国版本图书馆 CIP 数据核字(2015)第 209503 号

责任编辑:路亚妮 孙 丽 责任校对:徐 纯 装帧设计:吴 极

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)

(电子邮件:whu_publish@163.com 网址:www.stmpress.cn)

印刷:虎彩印艺股份有限公司

开本:720×1000 1/16 印张:11.5 字数:217 千字

版次:2015 年 11 月第 1 版 2015 年 11 月第 1 次印刷

ISBN 978-7-307-16754-4 定价:60.00 元

前　言

当今时代，信息呈指数级增长，信息社会给用户带来丰富信息的同时，也使得用户在信息海洋中容易迷失方向。从海量的信息资源中准确、快速地获取所需信息成为信息服务者不断努力的方向，在此背景下，搜索引擎成为帮助用户快速定位到互联网资源并获取相关信息的重要工具。然而，用户向搜索引擎输入的简短查询存在模糊性和歧义性，通常只能粗略地表达用户信息需求，因此，用户迫切希望搜索引擎能自动识别查询中包含的用户意图，直接返回与其信息需求相关的文档。于是，查询意图（即查询中应包含的用户信息需求、目标等）识别是当前学界和业界的一个研究热点。

其中，给定类目体系下的查询意图分类是查询意图识别的重要研究方向。当前的此类研究大多基于 Broder 提出的分类体系（即将查询意图分为信息类、导航类和事务类）进行，且主要工作是探讨如何对信息类和导航类进行有效区分，而对如何实现信息类、事务类和导航类三者自动分类的研究甚少。此外，查询意图的分类信息需最终用于指导搜索引擎的性能优化，而当前对如何利用查询意图分类信息来指导搜索引擎优化的探讨较少。

基于此，本文首先实现信息类、导航类与事务类的自动分类，且在此基础上，首次尝试从搜索引擎稳定性、个性化潜力和网络动态三角度来对查询意图进行分析，以期为搜索引擎性能优化提供相关建议。全文共分 7 章，主要内容如下：

引言。首先交代了本文的研究背景和研究意义。在回溯大量中英文文献的基础上，首先从查询意图识别和查询意图分析两个角度对查询意图进行了文献回顾。其中，查询意图识别的文献综述又包含给定类目体系的查询分类和不给定类目体系的查询识别两部分。其次，对搜索引擎稳定性、个性化潜力和网络动态相关研究进行了文献梳理和综述。在以上工作的基础上，确定了本文的研究内容和思路，接着介绍了本研究的创新之处。

相关理论基础。厘清了在认知检索模型中用户信息需求表达存在的问题，以及用户的意图在用户相关性判断中的作用；对查询意图的概念、理解维度、分类体系与分析维度进行了归纳总结。

查询意图自动分类。首先在人工标注数据集合的基础上，借用文本分类思想实现信息类、导航类和事务类三种意图的自动识别。其中，本文在已有查询意图特征的基础上，提出了查询表达式、点击网页 URL（统一资源定位符，Uniform Resource Location）信息、查询结果表单信息与重构查询词四层面新特征。另外，实验还探讨不同层面特征以及单个特征对分类效果的影响，并分析了不同层面特征在稀有与非稀有查询以及模糊与非模糊查询中的查询意图分类效果。

查询意图的搜索引擎稳定性分析。以百度、雅虎和搜狗三搜索引擎为研究对象，以两个月为观察期，分析同一搜索引擎或者不同搜索引擎之间针对不同查询意图随时间变化的稳定性。

查询意图的个性化潜力分析。首先，通过显式指标和隐式指标来衡量不同查询意图的查询个性化潜力。然后，通过对显式指标与隐式指标之间的相关性分析，获得有效的隐式指标，在此基础上，再分析针对不同查询意图哪些查询特征能有效地表征其个性化潜力。

查询意图的网络动态分析。分别从查询动态、文档内容动态和信息需求变化三角度出发，分析不同查询意图随时间变化所呈现的特征。然后，针对不同查询意图，分析了在不同查询流行度特征中，其文档内容以及信息需求变化情况。

研究总结与展望。对本书主要内容进行了总结，并在此基础上提出了本研究存在的不足以及对后期研究工作的展望。

著　者

2015 年 7 月

目　录

0 引　言

0.1　选题背景与研究意义

0.1.1　选题背景

自1990年以后，互联网呈现快速发展的趋势，其规模几乎以每年翻一倍的速度增长，逐渐改变着用户的生活方式，因此，网络在人们生活中扮演着越来越重要的角色，如根据CNNIC(中国互联网信息中心)于2014年1月发布的《中国互联网络发展状况统计报告》显示(图0-1)，截至2013年12月底，我国网民规模已达到6.18亿，全年共计新增网民5358万人，互联网普及率为45.8%，较2012年底提升了3.7%，由此可见，互联网已成为人们获取信息和进行事务的重要平台。

因互联网上的信息具有规模庞大、动态、非结构化等特点，信息社会给用户带来丰富信息的同时，也使得用户在信息海洋中容易迷失方向。从这样的信息资源中准确、快速地获取需要的信息成为信息服务者不断努力的方向。于是，传统的信息检索技术在网络资源面前已显得无能为力，在此背景下，搜索引擎作为一种能对网络资源进行有效提取、组织、搜索等的工具，能帮助用户快速定位到互联网资源并获取相关信息，已经成为人们日常生活和社会发展中必不可少的因素。但是，目前主流搜索引擎还停留在基于关键词匹配的全文检索阶段。虽在网络发展的初期阶段，基于关键词的全文检索理念极大地提高了搜索引擎的检索性能和效率，然而随着网络资源规模的剧增以及检索需求的提高，完全基于关键词形式匹配的传统搜索引擎也开始暴露出自身存在的一些问题：

图 0-1　中国网民规模与互联网普及率

（来源：CNNIC 中国互联网络发展状况统计调查。）

(1)用户对搜索结果的满意度不佳。信息检索的本质乃是从用户检索需求出发，在大量的文档库中匹配相关资源或者文档的过程。然而在现实中，基于关键词的全文检索技术缺乏对文档内容的语义，也不曾考虑用户检索式背后的用户意图，这在很大程度上影响到搜索结果的用户满意度。

(2)检索结果对词汇过度敏感。因为语言使用习惯问题以及汉语一词多义现象的广泛存在，对于同一事物，用户通常有多样的表达方式。所以，内容语义极度相关的文档因为使用了检索关键词不一样的描述方式而未能成功检索的现象极为普遍。

(3)过分孤立地看待用户的检索行为，忽略了检索过程的情境性因素。目前的搜索引擎尽管在智能性方面有所提高，但是基本还是以关键词词形匹配技术为核心。这种技术固有的一些缺陷导致用户单词检索的成功率并不是很理想，通常情况下，面对某种信息需求，用户与检索系统之间存在多次检索交互过程。目前的检索系统通常忽略这一系列检索的情景关联性，孤立地响应每一次检索行为，由此导致某些重要的情景相关信息被忽略，并间接影响了检索结果的准确性。

导致以上现象的最根本原因在于对复杂检索问题进行了以下几个方面的假设：① 用户检索式能够完全反映用户真实的信息需求；② 语法层面上基于检索式词形匹配的手段能够达到语义层面上面向用户意图的语义匹配；③ 用户的每次检索请求具有情景独立性。这些假设在 Web 发展的初期阶段简化了信息检索的复杂性，提高了检索的效率，但随着索引资源的剧增和用户需求的提高，这些假设开始暴露出其自身的许多弊端，并发展成为制约信息检索性能提高的一个瓶颈。基于关键词匹配的信息检索技术存在以下问题：

(1)用户针对同一事物往往使用不同的检索词。Furnas[①]通过实验发现,两个人使用同一关键词表达同一事物的概率小于20%。因选择关键词过程包含了用户自身的知识背景以及使用习惯等,如何选择关键词没有统一的标准,这对基于关键词匹配的检索系统来说是一个严峻的挑战。

(2)中文检索词比较短。余慧佳等人[②]通过对Sogou查询日志中的查询进行统计分析发现,检索词小于3的查询占总查询的93.15%,且查询的平均长度为1.85个;王继民等人[③]基于北大天网搜索日志进行了分析研究,统计出中文检索词多为一个或两个。短查询通常意味着不同明确地表达用户信息需求,从而导致搜索引擎返回的结果常常难以令人满意。

(3)多义词、近义词问题。多义词与近义词现象普遍存在于自然语言中,搜索引擎在接收到此类查询词时较难准确地判断用户真实的检索需求,也就难以返回用户感兴趣的检索结果。

总之,网络上蕴含的海量信息和查询中包含的极少信息,使得基于关键匹配的搜索引擎难以满足用户的信息需求,用户还需在返回结果列表中寻找符合自己真正意图的结果。也就是说,通用搜索引擎在理解用户的真正意图方面显得比较薄弱,不能针对查询的隐含意图返回更加相关、准确的答案。虽当前垂直搜索引擎(如图片搜索、视频搜索、音乐搜索等)在一定程度上满足了用户的特定信息需求,但由于其针对性强、通用性弱,在面对具有多个查询意图的情形下,还需提交给多个垂直搜索引擎,这必然会增加用户负担。因此,查询意图的识别与分析对于搜索引擎来说具有重要的意义。比如,包含不同意图类别的查询,其返回的相关文档类别不同。对于大多数查询来说,维基百科将在查询结果中的第一页中返回。对于信息类查询来说,通常情况下是适用的,因维基百科主页中包含了相关主题的详细描述;而对于导航类和事务类查询来说,维基百科并不适合于在其查询结果中返回。如用户键入导航类查询"武汉大学 主页"是想到达其相关主页,而对其相关描述信息不是很感兴趣,则返回维基百科中与"武汉大学"相关描述,与用户信息需求不相关。同样,用户键入查询"QQ 下载"则对维基百科中描述QQ产品相关历史信息不感兴趣。

---

① G W Furnas, T K Landauer, L M Gomez, et al. The Vocabulary Problem in Human-system Communication[J]. Communication of ACM, 1987, 30(11): 964-971.

② 余慧佳,刘奕群,张敏. 基于大规模日志分析的搜索引擎用户行为分析[J]. 中文信息学报,2007(1):231-239.

③ 王继民,陈种,彭波. 大规模中文搜索引擎的用户日志分析[J]. 华南理工大学学报:自然科学版,2001(1):1-5.

目前，查询意图识别的解决方法主要是将其简化为查询意图的分类问题，先确定几个意图类别，然后试图将查询归类到某个或者某些意图类目中。Broder[①]的查询意图分类体系最受学界青睐，即将查询意图划分为信息类、导航类和事务类。许多研究者以此分类体系为基础，尝试选取相关特征对查询意图进行自动区分。而当前研究大多集中在探讨信息类与导航类之间的有效区分，而如何对导航类、信息类与事物类三者进行有效区分的探讨较少。另外，因搜索引擎识别用户意图的真正目的是针对不同的用户意图提供不同的信息，而获取和分类用户意图仅仅是手段，当前的相关研究大多停留在对查询意图进行识别，而如何对识别出的查询意图进行分析并以此为搜索引擎优化提供相关依据的探讨较少。

鉴于此，本书研究内容主要包括：首先对查询意图自动分类，即对信息类、导航类、事务类三者进行有效区分。在此基础上，再分别从搜索引擎稳定性、查询个性化潜力以及网络动态对识别出的查询意图进行分析，以期为搜索引擎优化提供相关建议。需要指明的是，本书中的查询特指网络查询，即提交给网络搜索引擎，用以满足某些特定需求的查询。具体来说，本书中查询是指用户提交给搜索引擎的那一串字符。除此之外，考虑到不同查询针对不同用意有不同的查询意图[②]，而本书所探讨的查询意图并非指用户的个性化意图，而是指一般用户意图。

0.1.2 研究意义

随着网络信息呈指数级增长，面对搜索引擎用户日趋多元化以及个性化的检索需求，深入研究搜索引擎用户查询信息的行为习惯和意向，根据用户的使用习惯，挖掘用户的信息需求，加强搜索结果与用户搜索意图之间的匹配度的相关性，是进一步提高搜索引擎准确度的关键。因此，深入分析和挖掘海量检索用户的行为和需求，是未来搜索引擎致力研究的重要方向。具体而言，对用户查询意图识别的研究意义体现在以下几个方面：

(1)增加用户体验。用户的查询一般都非常短小，因此对用户意图进行识别，可以给简短的查询提供用户意图类别特征，在搜索引擎的结果排序的过程中，可用意图来作为参考，提高满足查询意图的结果的排名，以此可提高用户检索体验，是解决“信息过载”“信息迷航”的有效途径，可大大节约用户检索时间，提高检索效率。

---

① A Broder. A Taxonomy of Web Search[C]. SIGIR Forum, 2002, 36(2): 3-10.

② J Teevan, S T Dumais, E Horvitz. Personalizing Search via Automated Analysis of Interests and Activities[C]. In: Proceedings of 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2005.

(2)广告推荐。广告业务是搜索引擎的主要收入之一,为正在使用搜索引擎的用户提供或者推荐与用户搜索相关、符合兴趣的广告,是搜索引擎研究的重要方向。对用户查询进行分类,根据用户意图类别信息推荐相应广告是搜索引擎一很好的策略。

(3)有助于搜索引擎组织信息。对搜索引擎来说,将 Web 上的信息资源按照不同类别进行存储,可以梳理和有序化网络上的专业信息资源,使得搜索引擎能更有效率地检索结果,减轻了用户查询信息的负担,从而为建立下一代信息服务系统提供了新思路。

因此,查询意图的自动分类研究对提高搜索引擎以及其他信息检索系统的性能有着重要意义,在未来个性化、智能化搜索引擎的发展中起到重要作用。另考虑到当前对识别出的查询意图如何应用的探讨较少,本书尝试从搜索引擎稳定性、个性化潜力以及网络动态三方面来对查询意图进行分析,为搜索引擎优化提供依据。其中,分析查询意图的搜索引擎稳定性,可让搜索引擎获得针对哪类查询在稳定性方面存在着缺陷,并通过提高其稳定性来进一步满足用户信息需求;基于个性化潜力的查询意图分析,可让搜索引擎获得哪类意图类别查询应该采用个性化排序算法,哪些意图类别查询应该采用趋众性排序算法;分析查询意图的网络动态,可使得搜索引擎获得哪类意图类别查询的检索结果需不断融合具有新颖性的网页内容。

0.2 国内外研究现状分析

本节首先从基于分类的查询意图识别、基于排序的查询意图识别与查询意图分析三方面对查询意图的相关研究进行文献综述。另考虑到本研究将从引擎稳定性、查询个性化潜力以及网络动态三维度对查询意图进行分析,对这三方面的相关研究也进行了文献综述。

0.2.1 查询意图研究现状

在 Broder 提出查询意图之前,一些用户意图的研究主要通过实证和调研搜索引擎使用情况来进行。一些研究者尝试通过各种控制实验或者直接观察的方式来研究用户与检索系统之间的交互行为中所包含的用户意图元素,即根据不同用户意图对用户的一些交互行为进行归类。其中,在网络多媒体环境中,用户浏览行为受到大量关注,如 Carmel 等[①]将用户浏览行为分为面向搜索

① E Carmel, S Crawford, H Chen. Browsing in Hypertext: A Cognitive Study[J]. Journal of IEEE Transactions on Systems, Man and Cybernetics, 1992, 5(22): 865-884.

的浏览、回顾浏览与扫描浏览；Marchionini[①] 将用户浏览行为分为导向浏览、半导向浏览以及非导向浏览。另外一些研究者尝试观察用户搜索行为并对其进行归类，如 Byrne 等[②]定义了用户进行信息查询的任务层级图；Day 等[③]罗列出三大宽泛的搜索策略：监督、遵循计划、探索；Morrison 等[④]将用户的搜索行为分为查找、探索、监督与搜集；Choo 等[⑤]提出了用户网络搜索的行为模型，并将其任务定义为正式搜索、非正式搜索、监督以及无指导性浏览。另一些研究者的注意力从用户搜索行为转移到搜索目标进行归类，如 Rozanski 等[⑥]将用户目标划分为单一任务、重做、及时问答、信息请求、事实查找、冲浪；Sellen 等[⑦]将用户目标划分为信息查找、信息搜集、浏览与交易。

以上研究大都基于用户使用搜索引擎的意图是浏览或查找信息。而随着搜索引擎的发展，其除能为用户提供查找和浏览信息的功能外，也是用户的导航工具与从事相关活动（如游戏、购物等）的场所。在此背景下，Broder 认为用户的信息需求已不仅仅停留在信息类信息，也有可能是想获得某种服务等。于是，Broder 在用户调查和对查询日志中查询进行手工分类的基础上将查询这一交互行为中所包含的用户意图分为信息类、导航类和事务类。自此之后，有关查询意图的相关研究引起了学术界的广泛关注。当前查询意图的研究主要归为以下三类：基于分类的查询意图识别、基于排序的查询意图识别以及查询意图分析。

---

① G Marchionini. Information Seeking in Electronic Environments[M]. Cambridge: Cambridge University Press, 1995.

② M D Byrne, B E John, N S Wehrle, et al. In the Tangled Web We Wove: A Taskonomy of WWW Use[C]. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1999: 544-551.

③ V O Day, R Jeries. Orienteering in an Information Landscape: How Information Seekers Get From Here to There [C]. In: Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems ACM, 1993: 438-445.

④ J B Morrison, P Pirolli, S K Card. A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions[C]. In: Proceedings of Extended Abstracts on Human Factors in Computing Systems, 2001: 163-164.

⑤ C Choo, B Betlo, D Turnbull. A behavioral Model of Information Seeking on the Web: Preliminary Results of a Study of How Managers and IT Specialists Use the Web[C]. In: Proceedings of the 61st Annual Meeting of the American Society for Information Science, 1998: 290-302.

⑥ H D Rozanski, G Bollman, M Lipman. Seize the Occasion! The Seven-segment System for Online Marketing[J]. Strategy and Business, 2001, 6(24): 42-53.

⑦ A J Sellen, R Murphy, K L Shaw. How Knowledge Workers Use the Web[C]. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2002: 227-234.

0.2.1.1 基于分类的查询意图识别

考虑到此类问题将查询意图识别问题转化为给定意图类目体系下的查询分类问题，此小节将分别从查询意图类目体系构建、查询意图特征识别、查询意图分类方法、评价方法四个方面对其加以评述。

(1)查询意图类目体系构建。

Broder 等[①]通过用户调研与对 Alta Vista 查询日志分析将用户查询意图分为信息类、导航类和事务类。信息类是指用户以一种静态方式去查询被认为能在网络上获取到的信息，除阅读之外无其他交互信息，查找内容可以是数据、文档或多媒体，信息需求既可以是精确的又可以是模糊的；导航类是指用户查找某个特定网站(网页)，该网站(网页)可以是个人网站(网页)也可以是组织网站(网页)等，即用户在执行检索时已在头脑中形成了查找意向，知道或者认为存在网址可以满足自己的信息需求；事务类是指用户通过查找获取一些资源或网络服务，比如购买、下载等。另在 Broder 的基础上，Rose[②] 认为 Broder 的事务类不足以概括网上的所有资源，提出以资源类将其取代，指出资源类不再局限于一般的 Web 活动，而是包括网页上可获取的任何资源(而非信息类)，并在此基础上提出了更细致的层次结构。以上研究虽是通过人工分类方法展开，但证明了查询的确存在可被分类的性质且大部分查询拥有可以被预测的类别。因以上分类体系都是基于学者的经验知识构建的，故并非所有学者都同意上述观点，如 Marchionini[③] 将导航类和事务类归为查找搜索类(look up search)；Baeza-Yates 等[④]将用户意图分为信息类、非信息类与模糊类；Kang 和 Kim[⑤] 将查询分为话题查询、主页查询和服务查询；Waller 等[⑥]认为搜索引擎除了是获取信息的接口和到达某网站的通道，也是休闲的场所，故将查询意图分类体系扩展为信息类、导航类、事务类和休闲类，但目前缺乏相关的实证研究。除在文本检

① A Broder. A Taxonomy of Web Search[C]. SIGIR Forum,2002,36(2):3-10.

② D E Rose,D Levinson. Understanding User Goals in Web Search[C]. In:Proceedings of the 13th International Conference on World Wide Web,2004:13-19.

③ G Marchionini. Exploratory Search:From Finding to Understanding[J]. Journal of Communications of the ACM,2006,49(4):41-46.

④ R Baeza-Yates,L Calder'on-Benavides. The Intention Behind Web Queries[C]. In:Proceedings of the 13th International Conference on String Processing and Information Retrieval,2006:98-109.

⑤ I Kang,G Kim. Query Type Classfication for Web Document Retrieval[C]. In:Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,2003:64-71.

⑥ Waller ,Vivienne. Not Just Information:Who Searches for What on the Search Engine Google [J]. Journal of the American Society for Information Science and Technology,2011,62(4):761-775.

索中研究用户的查询意图类目外，另一些学者也尝试探讨非文本检索中的查询意图类目体系。如 Lux 等[①]通过研究发现，图像检索很少包含意图结构，并且 Broder 和 Rose 提出的意图分类体系不适合对图片检索的查询意图进行分类。基于此，Klofer 等[②]提出了图像检索意图类目体系主要包含以下四大类：面向知识类(knowledge orientation)、导航类(navigation)、事务类(transaction)和意识图像类(mental image)。段焕中[③]将事务类意图进一步细化为下载、娱乐、交互、获取以及购物。Chen 等[④]从社区问答服务角度出发，将用户的提问意图分为主观、客观与社交。任豪栋等[⑤]将导航类意图进一步细分为：直接 URL 类和间接 URL 类，其中，这两类意图的共同点都是在用户头脑中存在着一定的网页地址，而前者能用准确查询词描述，如用户键入“刘德华的博客”，是为了获得相关博客地址；而后者利用 Web 链接结构来导航到自己需要的 URL 地址。

González-Caro 等[⑥]认为查询只能表达用户意图的冰山一角，并指出从以下维度理解用户的意图和构建查询类目体系：信息题材(genre)、主题(topic)、任务(task)、目标(objective)、专指度(specificity)、范围(scope)、权威敏感性(authority sensitivity)、地理敏感性(spatial sensitivity)、时间敏感性(time sensitivity)。其中，在这些维度中主题和任务是两个最重要维度，且 Jansen 等[⑦]在此基础上探讨了查询主题与查询任务之间的关系。

(2)查询意图特征识别。

因查询文本包含的词汇和信息量较少，因此需要在特征提取阶段对查询进行词汇或者语义上的丰富从而获得更丰富的查询特征。从目前已有方法来看，可将特征选取方法分为以下两类：第一类方法称为事先方法，即在提交给搜索

---

① M Lux, C Kofler, O Marques. A Classification Scheme for User Intentions in Image Search[C]. In: Proceedings of the 28th International Conference Extended Abstracts on Human Factors in Computing Systems, 2010: 3913-3918.

② C Kofler. An Exploratory Study on the Explicitness of User Intentions in Digital Photo Retrieval [C]. In: Proceedings of the 9th I-KNOW and I-SEMANTICS, 2009: 208-214.

③ 段焕中. 事务类搜索意图分类模型研究[D]. 北京：北京邮电大学，2012.

④ L Chen, D Zhang, M Levene. Understanding User Intent in Community Question Answering[C]. In: Proceedings of the 21st International Conference Companion on World Wide Web, 2012: 823-828.

⑤ 任豪栋，贾年. 基于用户相似度计算的导航类意图分类研究[J]. 西华大学学报，2011，3(30)：101-106.

⑥ C González-Caro, L Calderon-Benavides, R Baeza-Yates. Web Queries: The Tip of the Iceberg of the User's Intent [C]. In: Proceedings of the 2011 the International Conference on Web Search and Web Data Mining, 2011: 282-291.

⑦ B Jansen, D Booth. Classifying Web Queries by Topic and User Intent[C]. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, 2010.