

# 查询专指度与检索多样化的关系研究

赵春晖<sup>1</sup> 张俊<sup>2</sup>

1. 武汉大学信息检索与知识挖掘研究所 武汉 430072;  
2. 深圳证券信息有限公司 深圳 518000

**摘要** 为了研究不同查询专指度语句的多样化检索效果,进而为提高检索质量和用户体验提供借鉴。本文基于 TREC(2009-2012) Web Track 检索词,人工进行专指度标注,选用语言模型狄利克雷平滑、语言模型线性差值平滑和 TF-IDF 三种模型,使用 MAP-IA,  $\alpha$ -nDCG@k, ERR-IA@k 评价多样化检索结果,对检索结果进行统计学分析和差异性检验,全面分析了查询专指度与多样化检索效果之间的关联性。实验结果表明查询词的专指度特征与多样化检索的效果间存在一定的关联。强专指度查询的多样化检索效果要明显好于弱专指度查询,特别在靠前的检索结果列表中,这种差异性最大。

**关键词:** 查询意图; 查询专指度; 检索多样化

**中图分类号:** G353.1

开放科学(资源服务)标识码(OSID)



## Research on the Relationship between Query Specificity and Search Diversification

ZHAO Chunhui<sup>1</sup> ZHANG Jun<sup>2</sup>

1. Information Retrieval and Knowledge Mining Laboratory, Wuhan University, Wuhan 430072, China;  
2. Shenzhen Securities Information Co., Ltd, Shenzhen 518000, China

**Abstract** In order to study the effectiveness of search diversification on queries with different specificity and improve retrieval quality and user experience, this paper manually labelled the specificity of queries from the TREC (2009-2012) Web Track, adopted the Dirichlet language model, linear interpolation

**基金项目:** 深圳市技术攻关资助项目证券信息多层网络数据检索与监控关键技术研究(JSGG20160229121006579)

**作者简介:** 赵春晖(1993-), 硕士研究生, 研究方向: 信息检索、知识挖掘; 张俊, 男, 通讯作者, E-mail: zch\_whu@yeah.net, 研究方向: 信息检索、数据挖掘。

language model and TF-IDF model, used the MAP-IA,  $\alpha$ -nDCG@k, ERR-IA@k to evaluate diversified search results and performed statistical analysis and differential test on the results to comprehensively analyze the relationship between query specificity and search diversification. The experimental results showed that there was a certain correlation between the specificity feature of queries and the effect of diversified search. The diversified search results of the queries with stronger specificity were significantly better than those of the queries with narrower specificity, especially in the top search results list.

**Keywords:** Query intention; query specificity; search diversification

## 1 引言

随着互联网环境下信息量的快速增长,搜索引擎成为网络信息查找和获取的必备工具。传统搜索引擎的搜索方式大多基于关键词组合,而用户提交给搜索引擎的有限关键词常常不能完整地表达用户的信息需求<sup>[1]</sup>。因此,根据用户输入的查询信息,研究用户的查询意图(即查询中包含的用户信息需求、查询目标、查询动机等)<sup>[2,3]</sup>,帮助搜索引擎返回与用户信息需求更相关的内容,能够有效地提高搜索引擎检索质量和用户信息获取效率。

除了通过查询检索式发掘用户的查询意图,在检索结果中综合考虑用户的查询意图,返回满足用户信息需求的结果列表,也是近年来信息检索领域研究的重要领域之一。传统的信息检索系统通常只基于文档与查询检索式的相关性返回检索结果列表,然而在现实的检索情境中,用户输入的查询词有限,在需求的表达上通常具有模糊性或歧义性,这种情况下仅仅考虑文档与查询式的相关性不可避免的会有许多相似的文档,甚至有些文档完全是另外一些文档的副本<sup>[4]</sup>。此时需要检索系统能够返回一个多样化的检索列表,降低检索结果的冗余,即

实现检索结果的多样化,以更多地覆盖用户的查询主题范围。

查询专指度(Query Specificity)和检索结果多样化(Search Diversification)都是针对用户查询意图提出的概念,前者关系检索过程中用户查询主题的限制范围程度,后者则关系检索结果的限制范围程度。检索结果是查询主题反映,检索过程也必定会对检索结果产生影响,据此猜测,查询专指度与检索结果多样化之间也存在一定的关联性。针对这一猜测,本文提出以下研究问题:

查询专指度与检索结果多样化之间是否存在关联?如果有,存在怎样的关联?

## 2 研究现状

### 2.1 查询意图研究现状

查询意图是介于用户查询语句与用户真实信息需求之间的一种中间形式,用于表示用户的搜索目的<sup>[5]</sup>。不同用户在不同搜索情境下的查询意图不尽相同,因此学者们试图从分类的角度对查询意图进行深度整合,构建完整的查询意图类目体系。2002年,Broder等<sup>[6]</sup>在进行

用户调研与对 AltaVista 查询日志分析的研究中,将用户查询意图分为三类,即:信息类(I)、导航类(N)和事务类(T)。以 Broder 的分类体系为基础,部分学者结合自己的理解提出一些改进的方案:Rose 等<sup>[7]</sup>认为 Broder 提出的分类体系中的“事务类”无法概括互联网上的所有资源,提出使用“资源类”进行替代,并将“资源类”定义为网页上可获取的任何资源(而非信息类);Marchionini 等<sup>[8]</sup>将导航类和事务类统一归为“查找搜索类”;类似地, Lee<sup>[9]</sup>将事务类和信息类合并为“研究信息类”;Mendoza<sup>[10]</sup>将导航类和事务类归为“非信息类”,并引入“歧义类”。但由于这一分类体系过于简单,难以应对具有复杂信息需求的查询<sup>[11]</sup>。因此, Gonzalez 等<sup>[12]</sup>认为应从信息题材(Genre)、主题(Topic)、任务(Task)、目标(Objective)、专指度(Specificity)、范围(Scope)、权威敏感性(Authority Sensitivity)、空间敏感性(Spatial Sensitivity)、时间敏感性(Time Sensitivity)等维度对用户的查询意图做更深入的理解和分析。本文所讨论的专指度就是在这一分类体系下的维度之一。

## 2.2 查询专指度研究现状

唐祥彬等<sup>[13]</sup>将专指度定义为“用户通过查询语句对自身信息需求或某查询主题的限制范围程度”。章成志等<sup>[14]</sup>认为查询式的专指度包括查询式所属学科领域的专业程度,查询式表达概念的宽泛性和查询词的个数。上述两种定义均提到了专指度在主题、概念等范围上的限制程度,但在侧重点上存在着一定的差异。本文主要基于前者定义,认为专指度侧重于探讨用户对信息需

求范围的界定是否明确清楚,即专指度主要是分析用户查询中使用了哪些限制,如数量限制、名字限制、时间限制、位置限制等。目前针对查询专指度的研究主要分为以下三个方面:

(1) 查询专指度的分类:目前学界关于查询专指度的分类并无统一说法,较为常见的是将查询语句专指度划分为两类或者三类<sup>[13]</sup>。在二值分类体系下, Igwersen<sup>[15]</sup>等从信息需求的角度将“知识陈述明确性(Knowledge State Specificity)”划分为“宽泛的(Generic)”或“具体的(Specific)”两种。Ramiez 等<sup>[16]</sup>使用“广泛的(Broad)”和“狭隘的(Narrow)”去定义不同的专指度类型。Hafernik 等<sup>[17]</sup>则使用9个语义属性将查询专指度分为强和弱两大类。由于二值分类方法较为笼统, Calderón-Benavides 等<sup>[18]</sup>提出按照专指度强弱程度将查询表达式分为“强专指度(Specific)”、“略专指度(Medium)”、“弱专指度(Broad)”三类。唐祥彬等<sup>[13]</sup>基于这一分类方法结合10个查询串特征实现专指度的自动识别和分类。

(2) 查询专指度的相关理论:这一方面的成果主要包括研究专指度本身的特性以及与其他查询式特征之间的关联。Donato 等<sup>[19]</sup>通过研究用户信息搜寻行为,表明专指度是影响用户查询语句可表达性的主要影响因素之一;Phan 等<sup>[20]</sup>研究了专指度与查询式长度之间的关系,发现长查询词通常具有较高的专指度;类似地, Hafernik 等<sup>[17]</sup>发现查询长度和词性可以提高专指度强弱识别的准确性; Najork<sup>[21]</sup>等研究了专指度与其他选择的特征之间的关系,发现基于链接的特征更适用于一般查询,而针对专指度较高的查询使用 BM25F 算法能够取得更好的检

索效果；Arampatzis<sup>[22]</sup>等通过实证研究了查询日志中查询式专指度与平均专指度分布问题，发现对于较短的查询词专指度近似呈 $\gamma$ 分布而较长的查询词专指度则近似呈对数正态分布；任珂等<sup>[11]</sup>分析了不同查询专指度语句的检索效果，发现在最靠前的几条检索结果中，强弱专指度查询语句的检索效果差异最大，强专指度的检索效果要明显好于弱专指度，并提出搜索引擎在专指度维度下，应重点关注最靠前的几条检索效果的准确性。

(3) 查询专指度的应用：专指度在一定程度上反映了用户的查询意图，因此可以用以辅助搜索引擎的检索。Yu<sup>[23]</sup>等在研究地理信息检索中，将专指度定义为度量查询具体或一般程度的一种度量标准，并应用于决定每个查询的排序依据中不同资源的权重；章成志等<sup>[14]</sup>利用专指度对搜索引擎和全文数据库上检索所得的查询词相关词进行分析；Tamine<sup>[24]</sup>等选择专指度作为查询式的特征之一，研究了TREC和CLEF医学检索任务中查询式的相似性和差异性及其对检索性能的影响；Krynski<sup>[25]</sup>等通过评估查询词的专指度，确定提交用户的意图在特定搜索和一般搜索之间的连续性，提供与用户意图更接近的搜索结果；Mu<sup>[26]</sup>等使用查询式专指度和长度两个特征来辅助查询扩展，并应用在统一医学语言系统(UMLS)的检索中；Haiduc等<sup>[27,28]</sup>通过测量专指度来反映查询检索式的质量开展了一系列的研究工作，实现了查询性能的自动评估，并提出了一种新的预检索方法，用于支持软件工程检索任务；类似地，Hauff等<sup>[29]</sup>也将专指度特征用于查询性能的预测研究中，并在TREC数据集上进行评估；He<sup>[30]</sup>等则

借助专指度进行查询词的分类。

### 2.3 检索多样化研究现状

检索多样化要求在传统的信息检索基础上，不仅考虑检索结果的相关性，也考虑检索结果的新颖性，从而满足用户多个方面的信息需求<sup>[31]</sup>。目前，针对检索多样化的研究主要集中在以下两个方面：

(1) 检索多样化的相关理论：这一方面的研究包括多样化的定义、分类以及内在含义的阐述等。Drosou<sup>[32]</sup>等从内容(Content-based)、新颖性(Novelty-based)以及覆盖度(Coverage-based)三个角度对检索结果多样化进行了概念界定；Radlinski<sup>[33]</sup>等从信息需求的角度将多样化分为外在多样化(Extrinsic Diversity)和内在多样化(Intrinsic Diversity)；Croft等<sup>[34]</sup>指出造成外在多样化的原因主要是查询词的歧义性；Santos等<sup>[35]</sup>则将多样化分为显式(explicit)和隐式(implicit)两大类，隐式多样化强调从文档本身角度来判断相互之间的不相似度，显式多样化则考虑的是文档与查询之间的关系，这也是目前普遍认可的分类体系。

(2) 检索多样化的实现方法：在相关理论的基础上，学者们提出了多种有效的方法实现检索结果的多样化。其中最为重要的是Carbonell等<sup>[4]</sup>提出的MMR(Maximal Marginal Relevance)方法，其核心思想是将文档与查询的相关度和文档本身的新颖度相结合对检索得到的文档进行排序。其他学者在这一方法的基础上进一步提出了改进方案，如：Zhai<sup>[36]</sup>等使用语言模型计算相关性和新颖性；Zhu<sup>[37]</sup>等提出了一种基于吸收马尔科夫链的随机游走算法；

张语晨等<sup>[38]</sup>通过构建子查询优化检索结果; Santos<sup>[35]</sup>等提出一种 xQuAD (explicit query aspect diversification) 模型, 通过为原始查询词构建子主题集合实现检索结果的多样化; 类似地, Dang<sup>[39]</sup>等提出了 PM2 模型, 也将子主题作为实现多样化的重要考虑因素之一。

## 2.4 研究现状总结

通过以上研究可以看出, 查询专指度的相关研究还存在着以下的不足:

(1) 目前查询专指度的理论研究中, 主要侧重专指度和其他单个特征之间的关联性分析, 而对于专指度与整体检索效果的分析相对较少;

(2) 目前尚未有对于查询专指度与多样化检索效果之间关联性的相关研究。

因此, 本文在借鉴前人的研究基础上, 对查询专指度和多样化检索之间可能存在的关联性进行全面分析和深入探讨。

## 3 研究设计

### 3.1 研究方法

本文采用定量研究方法, 选取 TREC 公布的查询检索词, 进行专指度标注, 并采用 TREC 中评测多样化任务的三个官方评测指标 MAP-IA,  $\alpha$ -nDCG@k, ERR-IA@k 评价多样化检索结果, 通过统计分析和显著性差异检验, 来研究查询专指度与检索结果多样化的关联。

### 3.2 研究过程

#### 3.2.1 查询检索词选择

选择合适的查询检索词, 并按照科学的分

类标准进行专指度标注, 是本研究首先需要解决的问题。

TREC (Text REtrieval Conference, 文本检索会议) 发布的查询检索词具有普遍适用性和合理性。TREC 自 2009 年在传统的 Web Track adhoc 检索任务基础上新增检索多样化任务 (diversity task), 由 NIST 每年为该任务提供 50 个新构建的查询检索词 (主题)。这些查询从商业搜索引擎的查询日志中产生, 并由专业评估者进行相关子主题、检索词类型的标注。查询检索词的示例如图 1 所示。

如图 1 所示, 查询检索词以 XML 格式进行存储, 单个检索词以 <topic> 标签进行区分。<query> 标签表示具体检索词, <description> 标签是该检索词的基本描述, 单个 <topic> 下包括多个 <subtopic>, 表示属于该主题下的子主题列表。<topic> 被分为 “ambiguous (歧义的)” 和 “faceted (多面的)” 两类, 通过 “type” 属性标识。歧义类查询通常含有多种不同的解释, 假定用户在查询过程中只对其中某一种解释感兴趣。多面类查询指代那些含有多个不同方面的查询, 认为用户可能同时对多个方面的信息存在兴趣。<subtopic> 则分为 “navigation (导航类, 记为 nav)” 和 “informational (信息类, 记为 inf)”。针对导航类子主题通常只含有较少的相关文档, 认为用户通常只寻找含有特定 URL 的页面。相反, 针对信息类查询则通常含有数量较大的相关文档。

本文选用 TREC 自 2009-2012 年四个年份发布的共 200 个查询检索词作为研究对象。

```

<topic number="1" type="faceted">
  <query>obama family tree</query>
  <description>
    Find information on President Barack Obama's family history, including genealogy, national origins, places and dates of birth, etc.
  </description>
  <subtopic number="1" type="nav">
    Find the TIME magazine photo essay "Barack Obama's Family Tree".
  </subtopic>
  <subtopic number="2" type="inf">
    Where did Barack Obama's parents and grandparents come from?
  </subtopic>
  <subtopic number="3" type="inf">
    Find biographical information on Barack Obama's mother.
  </subtopic>
</topic>
<topic number="23" type="ambiguous">
  <query>yahoo</query>
  <description>
    I'm looking for different sites hosted by Yahoo!
  </description>
  <subtopic number="1" type="nav">
    Take me to the Yahoo! homepage.
  </subtopic>
  <subtopic number="2" type="nav">
    Take me to Yahoo! Mail.
  </subtopic>
  <subtopic number="3" type="nav">
    I'm looking for the Yahoo! Messenger homepage.
  </subtopic>
  <subtopic number="4" type="nav">
    Take me to Yahoo! Finance.
  </subtopic>
  <subtopic number="5" type="nav">
    I'm looking for the Yahoo! Music homepage.
  </subtopic>
  <subtopic number="6" type="nav">
    I want to log in to my Yahoo! account.
  </subtopic>
  <subtopic number="7" type="inf">
    Find information about Yahoo!, the company.
  </subtopic>
</topic>

```

图 1 TREC 查询检索词示例图

### 3.2.2 查询专指度特征选择

本文借鉴任珂等<sup>[11]</sup>的研究中选择的 9 个查询式特征，用于对查询式专指度的强弱程度的分类，如表 1 所示。

表 1 查询式属性特征及举例

编号	查询语句属性特征	查询式举例
1	查询语句包括 URL 或者网站名称或 IP 地址	yahoo
2	查询语句包含确切地方名称及其他词项	map of the united states
3	查询语句比较不同事物或同一事物的不同方面	butter and margarine
4	查询语句比较多种不同的观点或话题	keyboard reviews
5	查询语句为包含确切答案的问题	who invented music
6	查询语句包含方向、建议、指导等方面的信息需求	how to build a fence
7	查询语句包含确切时间及其他词项	mother days songs
8	查询语句包含特定数字及其他词项	HP mini 2140
9	查询语句包含事物名称及其他词项	Obama family tree

### 3.2.3 查询专指度标注

本文选取二类分类体系，将查询专指度分为“强专指度”和“弱专指度”两类。

(1) 强专指度查询：用户有较为明确的信息需求，在查询检索词中有一定的条件限制，如地点限制、人物限制等，使得信息需求可以界定在某一特定范围内。如：obama family tree、map of the united states 等；

(2) 弱专指度查询：用户信息需求较为宽泛，在查询检索词中不包含相应的限制条件，无法明确界定用户的信息检索范围，如：map、rice 等。

标注工作由 2 名具备信息检索专业知识的图情专业硕士共同完成。针对选取的 200 个查询词，最终标注出“强专指度查询”103 个，“弱专指度查询”97 个。经 KAPPA 检验，计算出的 KAPPA 系数为 0.91，大于 0.8，表明本研究

标注结果一致性较高。

笔者对标注得到的两类查询的子主题个数

及类别、查询检索词长度、查询检索词长度等

特征进行了统计, 如表 2 所示。

表 2 查询检索词专指度标注结果及特征统计表

	强专指度查询		弱专指度查询		总体	
	faceted	ambiguous	faceted	ambiguous	faceted	ambiguous
查询类别个数	81	22	61	36	142	58
子主题类别个数	“信息类”居多	非“信息类”居多	“信息类”居多	非“信息类”居多	“信息类”居多	非“信息类”居多
	81	22	79	18	160	40
平均查询长度	3.204		1.701		2.475	
平均子主题个数	3.796		4.464		4.120	

### 3.2.4 检索结果多样化评价

在传统的信息检索中, 主要关注结果文档与查询式相关性的评价, 然而在检索多样化的问题中, 除了要考虑文档与查询的相关性, 同时还要关注结果的多样性。Turpin 等<sup>[40]</sup>指出查全率、查准率、MAP、DCG 等这些传统的评价指标并不一定与用户的满意度是相关的。例如在一个返回的结果列表中存在大量的相似文档会提高 MAP 值, 但是这个检索列表是高度冗余的, 对用户来说并不是最佳的结果列表。可见, 传统的信息检索评价方法和指标并不完全适用于多样化检索的评价。因此许多研究者在传统的评价指标基础上进行扩展用来评价多样化的检索结果。

本文选取 TREC 中评测多样化任务的三个官方评测指标: MAP-IA,  $\alpha$ -nDCG@k, ERR-IA@k 用以评价多样化检索结果:

(1) MAP-IA: 该指标是 Agrawal 等<sup>[41]</sup>提出的基于用户查询意图的评价指标。该指标在传统的 MAP 指标基础上进行了改造, 单独计算每个子主题用户意图的 MAP 值, 然后计算子主题的加权平均值;

(2)  $\alpha$ -nDCG@k: 该指标是 Clark 等<sup>[42]</sup>提出的在传统 DCG 指标扩展得到的一种评价方法, 该方法奖励含有新发现的子主题  $\alpha$ -nDCG 值, 惩罚包含已有子主题文档的  $\alpha$ -nDCG 值。其中  $\alpha$  是一个惩罚子主题冗余的变量因子。本文选取的 k 值为 5,10,20;

(3) ERR-IA@k: 该指标是由 Chapelle 等<sup>[43]</sup>提出的一种全新评价指标。在评价结果列表中的某文档时, 根据排名在其之前的文档的相关性进行计算。与 MAP-IA 类似, 该指标也是单独对每个子主题进行计算, 然后进行加权平均。本文选取的 k 值为 5,10,20。

本文选择非语言模型 TF-IDF 和语言模型 (Language Model) 对查询词进行检索排序。其中为了避免词项在文档中出现的“零概率问题”, 在语言模型中分别使用狄利克雷 (Dirichlet) 平滑 (默认参数  $\mu=2500$ , 下简记为: Dir) 和线性插值 (Jelinek-Mercer) 平滑 (默认集合参数  $\lambda=0.4$ , 文档  $\lambda=0$ , 下简记为: J-M)。综合使用三种模型, 在默认参数设置下, 使用 TREC 官方评价工具 ndeval, 分别计算得到所选取 200 个查询在上述三个评价指标下的得分。

## 4 实验结果分析

为了分析多样化检索结果与查询专指度之间是否存在关联，笔者将查询计算所得的评分按照查询检索词专指度的强弱分为两组（记为1,2，其中1表示强专指度，2表示弱专指度），作为下文分析计算的两组独立样本。

### 4.1 统计学指标分析

笔者首先计算了1,2两组数据分别在三种模型下各评价指标的平均值，结果如表3-表5所示。其中差值项为（样本组1-样本组2）所得差值，差值幅度项为（样本组1-样本组2）/样本组2\*100%计算所得。

表3 TF-IDF模型下各指标均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
强专指度	0.0742	0.2702	0.3072	0.3517	0.2203	0.2382	0.2509
弱专指度	0.0474	0.1455	0.1904	0.2323	0.1084	0.1269	0.1379
差值	0.0268	0.1247	0.1168	0.1194	0.1119	0.1113	0.1130
差值幅度	56.5%	85.7%	61.3%	51.4%	103.2%	87.7%	81.9%

表4 Dir模型下各指标均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
强专指度	0.0734	0.2018	0.2487	0.2947	0.1591	0.1804	0.1931
弱专指度	0.0615	0.1269	0.1789	0.2276	0.0975	0.1185	0.1312
差值	0.0119	0.0749	0.0698	0.0671	0.0616	0.0619	0.0619
差值幅度	19.3%	59.0%	39.0%	29.5%	63.2%	52.2%	47.2%

表5 J-M模型下各指标均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
强专指度	0.0488	0.1975	0.2253	0.2621	0.1583	0.1713	0.1817
弱专指度	0.0312	0.1070	0.1316	0.1658	0.0802	0.0911	0.1002
差值	0.0176	0.0905	0.0937	0.0963	0.0781	0.0802	0.0815
差值幅度	56.4%	84.6%	71.2%	58.1%	97.4%	88.0%	81.3%

由表3-表5可以观察到：

(1) 三种模型在强专指度查询下的各评价指标均要高于弱专指度查询下的各评价指标。虽然单从两类查询各指标平均值的差值来看并不明显，但是在差值幅度上，三种模型在强专指度查询下的各评价指标同比与弱专指度下的

各指标均有非常明显的提升，提升的幅度从19.3%~103.2%不等；

(2) 强专指度查询词与弱专指度查询词两组样本数据在ERR-IA@k的差值最为明显，差值幅度均在80%以上，而在MAP-IA的差值幅度相较于同模型下其他评价指标最则较小；

(3) 在 ERR-IA@k 和  $\alpha$ -nDCG@k 两类评价指标下, 当 k 取值越小时, 强专指度与弱专指度的差值幅度越大, 说明对于检索所得到的结果列表中排名较高的文档, 对应查询词的专指度强弱所体现的差异最为明显, 而随着检索结果的增加, 这种差异性逐渐减小。

综上所述可以看出, 强专指度查询与弱专指度查询对应的各组评价指标存在着一定的关联, 检索词专指度越强, 各项评价指标的得分越高, 表明多样化检索的效果越好。

## 4.2 差异性检验

由于单从统计学指标中的均值对数据进行观察存在一定的局限性, 为了进一步验证强专指度查询和弱专指度查询对应的各项评价指标数据存在明显的差异, 本文对两组数据采用统计学中的非参数检验进行了显著性

差异检验。

两独立样本的非参数检验是在对总体分布不甚了解的情况下, 通过对两组独立样本分析来推断样本来自的两个总体的分布、均值等是否存在显著性差异的方法。两独立样本的非参数检验方法包括曼-惠特尼 U 检验、K-S 检验、W-W 游程检验、极端反应检验等。本文选用曼-惠特尼 U 检验作为实验方法。曼-惠特尼 U 检验(又称“曼-惠特尼秩和检验”), 主要用于检验两个样本来自的两个总体的均值是否具有显著的差别。其原假设是: 两组独立样本来自的两总体无显著差异。

笔者将每个模型下同一评价指标在强专指度查询和弱专指度查询的数值作为两组独立样本, 在设定的显著水平为 0.05 条件下使用数据分析软件 SPSS 分别进行曼-惠特尼 U 检验, 得到检验结果如表 6 所示。

表 6 数据显著性差异检验结果统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
TF-IDF	0.018387	0.000210	0.000320	0.000214	0.000104	0.000093	0.000048
Dir	0.437669	0.004778	0.019015	0.019719	0.003411	0.006627	0.003715
J-M	0.001859	0.000989	0.000659	0.000647	0.000595	0.000404	0.000260

在针对三种模型下七个评价指标的曼-惠特尼检验所得的结果中, 除语言模型 Dirichlet 平滑下的 MAP-IA 这一指标大于 0.05, 其余各指标的检验结果均远小于 0.05, 其余各指标的检验结果均远小于 0.05。结合表 4 中的数据可知, Dirichlet 平滑的语言模型下, MAP-IA 这一指标在强专指度查询和弱专指度查询下的平均值差值最小, 其相差幅度也是所有指标中差异最小的, 仅为 19.3%, 可见两组查询在这一指标下

体现的差异性最不明显, 这与本节的非参数检验所得结果基本相符。

因此, 综合考虑其他评价指标的检验结果, 在设定的显著水平为 0.05 条件下, 研究认为两组独立样本来自的两总体存在显著差异, 即强专指度查询词和弱专指度查询词在多样化检索结果的各评价指标下存在着较为明显的差异, 进一步的可认为查询检索词的专指程度与多样化的检索结果存在一定的关联性。

## 5 结果讨论

综合上述研究，查询检索词的专指程度与检索结果多样化存在关联，且强专指度对应的查询词的多样化检索结果要优于弱专指度查询词的多样化检索结果。为深入分析造成这一现象的原因，笔者认为应从强专指度查询较弱专指度查询所表现出的特征入手。由表 2 可知，强专指度查询在平均查询子主题个数上要少于弱专指度查询的平均子主题个数。Drosou<sup>[32]</sup>等在通过覆盖度对检索结果多样化的概念进行界定中提到，即结果列表好的文档对查询的不同主题覆盖的越高，那么该结果列表的多样化程度就越高，从而最大可能的

满足不同用户的信息需求。基于这一观点笔者认为，子主题个数是造成强专指度查询和弱专指度查询在多样化检索中所反应的检索结果差异性的重要因素之一。

为了验证这一假设，笔者在强专指度查询样本组中选择对应子主题个数为研究特征，进一步研究其内部因子主题个数造成的差异。由表 2 可知，强专指度下平均子主题个数为 3.796，笔者以此作为分界，将强专指度查询按子主题个数进一步分为两类，即子主题个数 <4（向上取整）的查询词归为样本组 1，子主题个数 ≥4 为归为样本组 2，计算两组样本在统计学指标下的平均值如表 7- 表 9 所示。

表 7 TF-IDF 模型强专指度查询子主题分组均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
子主题 <4	0.0924	0.3240	0.3643	0.4086	0.2631	0.2822	0.2951
子主题 ≥4	0.0606	0.2300	0.2645	0.3093	0.1884	0.2053	0.2178
差值	0.0318	0.0940	0.0998	0.0993	0.0747	0.0769	0.0773
差值幅度	52.4%	40.9%	37.7%	32.1%	39.6%	37.4%	35.5%

表 8 Dir 模型强专指度查询子主题分组均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
子主题 <4	0.0960	0.2658	0.3192	0.3790	0.2196	0.2442	0.2611
子主题 ≥4	0.0567	0.1540	0.1961	0.2318	0.1139	0.1328	0.1424
差值	0.0393	0.1118	0.1231	0.1472	0.1057	0.1114	0.1187
差值幅度	69.3%	72.6%	62.8%	63.5%	92.8%	83.9%	83.4%

表 9 J-M 模型强专指度查询子主题分组均值统计表

	MAP-IA	$\alpha$ -nDCG@5	$\alpha$ -nDCG@10	$\alpha$ -nDCG@20	ERR-IA@5	ERR-IA@10	ERR-IA@20
子主题 <4	0.0574	0.2290	0.2669	0.3123	0.1902	0.2069	0.2199
子主题 ≥4	0.0424	0.1739	0.1942	0.2248	0.1346	0.1447	0.1532
差值	0.0150	0.0551	0.0727	0.0875	0.0556	0.0622	0.0667
差值幅度	35.4%	31.7%	37.4%	38.9%	41.3%	43.0%	43.5%

从表 7- 表 9 可以看出:

(1) 三种模型中, 子主题个数 <4 的强专指度查询相较于子主题个数  $\geq 4$  的强专指度查询各项评价指标均有着显著的提高, 提高幅度从 31.7%~92.8% 不等, 由于该分组是在强专指度查询词下的进一步细分, 因此两组样本的差值的幅度相较于强专指度与弱专指度的大类样本组下的差值变化幅度较为稳定, 集中在 35%~45% 的区间内;

(2) 与表 3- 表 5 中所体现的现象类似, 当选取较少的检索结果时, 两组数据体现的差异性较为明显, 随着检索列表数目的增加, 数据之间的差异性逐渐减小。

综合上述实验结果, 结合多样化检索的相关理论, 笔者认为子主题个数上的差别是专指度强弱程度上反映的特征, 而这一差异性的特征是导致多样化检索效果差异性的重要因素之一。强专指度查询由于所表达的用户的信息需求较为明确, 因此对应的子主题数目较少, 在有限的结果列表中数目较少的子主题更容易被全部覆盖, 从而提升整个检索的多样化效果。相反, 弱专指度查询由于用户的查询意图较为模糊, 往往在有限的结果中无法充分完全满足用户的所有需求。

## 6 结论

本文选取 TREC 提供的查询词, 通过进行不同查询专指度的分类, 在 TF-IDF、语言模型狄利克雷平滑和语言模型线性插值平滑三种模型基础上, 选取多样化检索中的常用评价指标, 对检索结果进行了统计学指标的分析 and 显著性

差异检验。实验结果表明: 查询词的专指度特征与多样化检索的效果间存在一定的关联。强专指度查询的多样化检索效果要明显好于弱专指度查询, 特别在靠前的检索结果列表中, 这种差异性最大。进一步的, 本文认为查询词专指度在子主题个数上的差距, 是造成最终检索结果差异性的重要因素之一。因此, 在今后的搜索引擎以及检索算法的改进中, 可以将查询词的专指度作为特征之一, 实现检索过程的进一步优化, 从而更好的满足用户的信息需求。

同时, 本文研究也存在着一定的局限性, 包括:

- ①仅选用 TREC 查询检索词, 数据集较为固定和单一且数据量有限, 所得结论也主要基于该测试数据集;
- ②仅选用了三种检索模型进行实验。今后还需在其他数据集上选用更多类型的检索模型进行更深入的检验, 实现对查询词专指度与多样化检索二者之间更全面的分析。

---

## 参考文献

---

- [1] 张晓娟, 陆伟. 利用查询重构识别查询意图[J]. 现代图书情报技术, 2013, 29(1):8-14.
- [2] Nguyen H. Capturing User Intent for Information Retrieval[C]. Nineteenth National Conference on Artificial Intelligence, Sixteenth Conference on Innovative Applications of Artificial Intelligence, July 25-29, 2004, San Jose, California, USA. 2004:371-375.
- [3] 陆伟, 周红霞, 张晓娟. 查询意图研究综述[J]. 中国图书馆学报, 2013, 39(1):100-111.
- [4] Carbonell J, Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries[C]. International ACM SIGIR Conference on Research and Development in Information Retrieval. 1998:335-336.

- [5] 宋巍. 基于主题的查询意图识别研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [6] Broder A. A taxonomy of web search[J]. ACM Sigir Forum. ACM, 2002, 36(2): 3-10.
- [7] Rose D E, Levinson D. Understanding user goals in web search[C]. Proceedings of the 13th international conference on World Wide Web. ACM, 2004: 13-19.
- [8] Marchionini G. Exploratory search: from finding to understanding[J]. Communications of the Acm, 2006, 49(4):41-46.
- [9] Lee U, Liu Z, Cho J. Automatic identification of user goals in Web search[C]. International Conference on World Wide Web. ACM, 2005:391-400.
- [10] Mendoza M, Baezayates R. A Web Search Analysis Considering the Intention behind Queries[C]. Latin American Web Conference. IEEE Computer Society, 2008:66-74.
- [11] 任珂, 陆伟, 丁恒. 查询专指度对检索效果的影响研究[J]. 现代图书情报技术, 2016, 32(11):34-43.
- [12] Baeza-Yates R. A multi-faceted approach to query intent classification[C]. International Conference on String Processing and Information Retrieval. Springer-Verlag, 2011:368-379.
- [13] 唐祥彬, 陆伟, 张晓娟, 等. 查询专指度特征分析与自动识别[J]. 现代图书情报技术, 2015, 31(2):15-23.
- [14] 章成志, 徐小琴. 信息检索系统的相关词提示技术与评测[J]. 情报理论与实践, 2007, 30(1):100-104.
- [15] Ingwersen P, Jarvelin K. The Turn [M]. Berlin: Springer, 2005.
- [16] Vries A P D. Relevant contextual features in XML retrieval[C]. International Conference on Information Interaction in Context. ACM, 2006:56-65.
- [17] Hafernik C T. The Relationship Between Query Length, Parts of Speech Usage and Web Search Query Specificity[D]. Pennsylvania: The Pennsylvania State University, 2013.
- [18] Calderón-benavides L, González-carro C, Baezayates R. Towards a deeper understanding of the user's query intent[C]. SIGIR 2010 Workshop on Query Representation and Understanding. 2010: 21-24.
- [19] Donato D, Donmez P, Noronha S. Toward a deeper understanding of user intent and query expressiveness[C]. ACM SIGIR, query representation and understanding workshop. 2011.
- [20] Phan N, Bailey P, Wilkinson R. Understanding the relationship of information need specificity to search query length[C]. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 709-710.
- [21] Najork M A, Zaragoza H, Taylor M J. HITS on the Web: How does it Compare?[C]. Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 471-478.
- [22] Arampatzis A, Kamps J. An empirical study of query specificity[C]. European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 2010: 594-597.
- [23] Yu B, Cai G. A query-aware document ranking method for geographic information retrieval[C]. Proceedings of the 4th ACM workshop on Geographical information retrieval. ACM, 2007: 49-54.
- [24] Tamine L, Chouquet C, Palmer T. Analysis of biomedical and health queries: Lessons learned from TREC and CLEF evaluation benchmarks[J]. Journal of the Association for Information Science and Technology, 2015, 66(12): 2626-2642.
- [25] Krynski T R, Jain D, Teodorescu D C. Determining search query specificity: U.S. Patent 8,370,319[P]. 2013-2-5.
- [26] Mu X, Lu K. Improving UMLS Metathesaurus Query Expansion Based on the Query Specificity and Length[C]. Proceedings of the ACM SIGKDD Workshop on Health Informatics. 2012.
- [27] Haiduc S, Bavota G, Oliveto R, et al. Evaluating the specificity of text retrieval queries to support

- software engineering tasks[C]. Software Engineering (ICSE), 2012 34th International Conference on. IEEE, 2012: 1273-1276.
- [28] Haiduc S, Bavota G, Oliveto R, et al. Automatic query performance assessment during the retrieval of software artifacts[C]. Proceedings of the 27th IEEE/ACM international conference on Automated Software Engineering. ACM, 2012: 90-99.
- [29] Hauff C, Hiemstra D, De J F. A survey of pre-retrieval query performance predictors[C]. Proceedings of the 17th ACM conference on Information and knowledge management. ACM, 2008: 1419-1420.
- [30] He X, Jhala P D. System for classifying a search query: U.S. Patent 7,603,348[P]. 2009-10-13.
- [31] Clarke C L, Craswell N, Soboroff I. Overview of the trec 2009 web track[R]. WATERLOO UNIV (ONTARIO), 2009.
- [32] Drosou M, Pitoura E. Search result diversification[J]. ACM SIGMOD Record, 2010, 39(1): 41-47.
- [33] Radlinski F, Bennett P N, Carterette B, et al. Redundancy, diversity and interdependent document relevance[C]. ACM SIGIR Forum. ACM, 2009, 43(2): 46-52.
- [34] Croft W B, Cronen-townsend S, Lavrenko V. Relevance Feedback and Personalization: A Language Modeling Perspective[C]. DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries. 2001, 3: 13.
- [35] Santos R L T, Peng J, Macdonald C, et al. Explicit search result diversification through sub-queries[C]. European Conference on Information Retrieval. Springer, Berlin, Heidelberg, 2010: 87-99.
- [36] Zhai C, Lafferty J. A study of smoothing methods for language models applied to ad hoc information retrieval[C]. ACM SIGIR Forum. ACM, 2017, 51(2): 268-276.
- [37] Zhu X, Goldberg A, Van G J, et al. Improving diversity in ranking using absorbing random walks[C]. Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference. 2007: 97-104.
- [38] 张语晨, 杨沐昀. 信息检索结果多样化探索[J]. 智能计算机与应用, 2012, 2(1):29-30.
- [39] Dang V, Croft W B. Diversity by proportionality: an election-based approach to search result diversification[C]. Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval. ACM, 2012: 65-74.
- [40] Turpin A, Scholer F. User performance versus precision measures for simple search tasks[C]. Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2006: 11-18.
- [41] Agrawal R, Gollapudi S, Halverson A, et al. Diversifying search results[C]. Proceedings of the second ACM international conference on web search and data mining. ACM, 2009: 5-14.
- [42] Clarke C L A, Kolla M, Cormack G V, et al. Novelty and diversity in information retrieval evaluation[C]. Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2008: 659-666.
- [43] Chapelle O, Ji S, Liao C, et al. Intent-based diversification of web search results: metrics and algorithms[J]. Information Retrieval, 2011, 14(6):572-592.