arXiv:2302.06355v1 [cs.IR] 13 Feb 2023

电子商务自然语言查询数据集

Andrea Papenmeier、Dagmar Kern、Daniel Hienert firstname.lastname@gesis.org GESIS – 莱布尼茨社会科学研究所 德国科隆

Alfred Sliwa、Ahmet Aker、Norbert Fuhr firstname.lastname@uni-due.de 杜伊斯堡-埃森大学 德国杜伊斯堡

摘要

在线购物在日常生活中日益普遍。对于电子商务搜索系统而言，理解通过语音助手、聊天机器人或对话式搜索传达的自然语言，是了解用户真正需求的一项关键能力。然而，目前尚无可用于研究、涵盖产品搜索者自然且详尽需求的评估数据集。出于隐私问题和竞争方面的影响，公开可用的、包含日志中真实用户搜索查询的数据集寥寥无几。本文介绍了一个数据集，其中包含3,540条自然语言查询，涉及两个领域，描述用户在搜索自己想要的笔记本电脑或夹克时的需求。该数据集还包含对1,754条笔记本电脑查询中的模糊术语和关键信息的标注。这个数据集为产品搜索领域的自然语言处理和（交互式）信息检索研究开辟了多种机遇。

CCS概念

• **以人为中心的计算** → 人机交互（*HCI*）；**自然语言接口**；• **信息系统** →**查询意图**；• **计算方法** → 话语、对话与语用学。

关键词

数据集；自然语言查询；用户意图；电子商务。

1 引言

在线搜索产品是数百万用户的日常活动，电子商务的市场份额也在持续增长 [13]。深入理解自然语言，是电子商务和整个万维网搜索领域即将迎来的关键技术：语音助手依赖于对口语自然语言的处理，聊天机器人需要从书面自然语言中提取用户的信息需求，而对话式搜索这一研究领域则探索通过对话来支持用户找到合适信息的方法。针对儿童查询构建方式的研究表明，自然语言是与搜索引擎交互的直观方式 [9]。然而，关键词搜索是搜索引擎无法处理开放词汇、也无法从自然语言文本中提取关键信息的产物。关于查询和信息需求构建的研究，主要基于日志数据[19, 20, 32] 或专有数据 [8]。日志数据并不适合用于研究自然信息需求，因为系统会影响用户；也就是说，如果用户认为系统只能处理关键词，他们就会据此构建查询 [10]。目前已有一些规模较小的自然语言信息需求数据集，例如 Kato 等人收集的数据 [10]。在图书领域，CLEF 社交图书搜索数据集 [11] 提供了 120 条自然语言信息需求。这些数据来自一个在线讨论论坛，属于非受控的数据收集场景。在此前针对一个包含 132 条笔记本电脑自然语言查询的小型数据集开展的研究中，我们表明，与当前搜索引擎上的查询相比，自然语言查询有望揭示更多有关用户目标产品的信息 [14]。然而，现有数据集规模不足以训练自动化系统来处理自然语言信息需求。

在本项工作中，我们收集并整理了一个大型数据集，其中包含针对两个产品领域（笔记本电脑和夹克）的3,540条自然语言查询。与现有数据集不同，我们的产品查询是在受控实验中收集的，参与者具有广泛的领域知识。对于笔记本电脑领域以及夹克领域的一个子集，我们提供了描述中提及的关键产品事实，以及文本中包含的模糊词语的人工标注。借助该数据集，我们为自然语言处理领域以及产品搜索语境下的交互式信息检索贡献了一项宝贵资源。

2 相关研究

2.1 信息需求数据集

理解用户需求最常用的方法之一是从查询日志中提取信息。例如，Broder [4] 对网页搜索查询的用户意图进行了粗略区分，将其分为信息型、导航型或交易型。多项研究将日志中的搜索查询 [19, 20, 32] 或语音日志转录文本 [8] 聚类为不同的用户意图。用户自然语言表达的信息需求与查询日志中的简短类关键词查询之间显然存在很大差异。*TREC* [29] 主题以及基于这些主题开展的研究（例如 [2, 16]）对信息需求作出了更详细的描述。在这些研究中，搜索意图的情境和背景得到了更细致的阐述，并被组织成背景故事或搜索任务。CLEF 社交图书搜索数据集 [11]收录了从在线讨论论坛中提取的 120 条图书自然语言信息需求。Bogers等人 [3]同样聚焦于论坛，以提取自然语言信息需求。他们根据“相关性方面”（即对搜索目标的要求），对 1041 条信息需求（503 条属于图书领域，538 条属于电影领域）进行了标注。

另一个方向是对话式搜索，这种对话方式会进一步了解用户的实际信息需求。评估数据集基于论坛或对话支持系统中已有的问答交互。例如，MS Dialog [17] 包含 2,400 段关于微软产品的标注对话；Penha 等人 [15] 则提供了一个包含 8 万段对话的数据集，涵盖 14 个领域，这些对话提取自 Stack Exchange。

现实环境中的人与人之间的对话，以及在由人类模拟系统的绿野仙踪法情境中的对话，是自然表述的信息需求的另一来源。例如，CCPE-M 数据集 [18] 关注电影偏好，而 Frames 数据集 [5] 关注旅行领域，其对话是在使用 SLACK 进行的人际对话中收集的。对话式搜索的另一种方法是提出澄清问题。Qulac 数据集 [1] 通过一次众包实验，为 198 个 TREC 主题收集了 1 万个澄清问题及其答案。

关于口语信息寻求对话的数据集提供音频、转录文本和其他标注。口语对话与书面对话存在差异，可用于评估 Siri、谷歌、Google Now 或 Cortana 等软件代理。MISC 数据集 [25] 包含音频、视频、转录文本、情感和生理信号、计算机使用情况，以及基于先前研究主题的五种不同搜索任务的调查数据。SCSdata [26] 包含 101 段带有标注和视频的转录对话，用于根据九种搜索任务和背景故事解决信息需求。

2.2 电子商务中的产品搜索

产品搜索和电子商务是学术研究中相对较新的领域，在信息检索方面面临着专门的挑战：文档、查询、相关性、排序、推荐系统和用户交互都不同于网页搜索等知名研究领域 [28]。在用户意图层面，Su 等人 [24] 区分了三种不同的用户目标：目标查找、决策和探索，而这些目标在查询构建、浏览和点击方面具有不同的行为模式。Sondi 等人 [22]报告了另一种通过对日志中的查询进行聚类而生成的查询分类体系：浅层探索、大件商品购物、目标购买、小件商品购物和困难抉择型购物。对话式电子商务是一个新兴领域，用户通过语音或聊天与对话系统进行对话，以寻找合适的产品或获取帮助。对话需要自然，使顾客感到投入其中。因此，理解用户请求的整体意图至关重要 [28]。

SIGIR 论坛 [28] 最近一篇关于电子商务研究的论文列出了 28 个用于电子商务搜索和推荐的数据集。然而，大多数数据集侧重于产品目录和分类体系，以及评论和推荐。只有两个数据集包含带有用户查询的搜索日志。电子商务和全球参与者产品搜索领域的大多数研究工作都使用自己的查询日志来改进自身的搜索系统（例如 eBay [27],、亚马逊 [23] 或阿里巴巴 [31]），但会对日志内容保密。然而，即使这些数据集公开可用，它们也只包含关键词查询，而不包含用户在将其输入搜索栏之前所思考的自然信息需求。

尽管产品搜索是电子商务中的重要应用场景，但关于产品买家的真实信息需求，我们掌握的数据却很少。为填补缺乏公开且受控采集的研究数据集这一空白，我们在本研究中推出了一组包含3,540条自然语言查询的产品搜索数据集。

	笔记本电脑	夹克
参与者（总数）	1818	1818
有效查询（总数）	1754	1786
年龄（平均值、标准差）	36 (13)	36 (13)
性别分布（m、f、d）	700, 1040, 12	718, 1054, 12
领域知识（均值、标准差）	4.5 (1.5)	4.6 (1.3)

表1：笔记本电脑和夹克查询数据集的参与者统计信息。

3 数据集生成

3.1 数据收集

我们在众包平台 Pro- lific[1] 上招募了1,818名参与者参加实验。为避免个人语言水平对措辞的影响，参与者必须是英语母语者。此外，为避免小屏幕和小尺寸键盘造成的影响，参与者不得使用手机填写调查问卷。

参与者知情同意后，我们请他们描述一台笔记本电脑（设想他们当前使用的笔记本电脑最近坏了）或一件夹克（设想他们丢失了夹克）。我们将这些产品描述定义为自然语言查询。所有参与者都完成了这两项任务，但顺序随机。任务说明和完整问卷可在线查阅[2]。最后，参与者回答了有关其人口统计背景的问题：（1）年龄，（2）性别，以及（3）对两个领域的自我评估领域知识（从 1 = “不了解”到 7 = “知识丰富/专家水平”）。表1列出了参与者在两个领域的人口统计特征。

我们对所有描述进行了人工筛选，剔除了因形式（例如空字符串）或内容（例如文本描述的是另一种产品、文本没有描述任何产品，或文本是参与者对任务的评论）而无效的查询。在 1,818 名参与者提供的内容中，我们整理出一个数据集，其中包含 1754 条有效的笔记本电脑查询和 1786 条有效的夹克查询。

3.2 数据标注

收集自然语言查询后，我们招募了 20 名正在本机构参加研讨会的标注人员。每条笔记本电脑查询由三名标注人员进行标注，标注内容包括关键信息和模糊词语。关键信息是描述产品要求的词或短语，而模糊词语则是关键信息中含义模糊、需要根据具体理解来解释的词语。到目前为止，我们已从夹克语料库中抽取 363 条查询进行标注，以用于跨领域评估，标注内容为其中的关键信息。

在开始标注任务之前，标注人员收到了关键信息和模糊词语的定义、示例，以及如何处理否定表达和模糊性边界案例的指导[3]。标注人员还在全体会议上讨论了相关指南。

[1]https://www.prolific.co [2]https://git.gesis.org/papenmaa/chiir21_naturallanguagequeries/tree/master/Questionnaire [3]标注材料可在 https://git.gesis.org/papenmaa/chiir21_naturallanguagequer 获取

	笔记本电脑
查询（总计）	1754
每条查询的词数（均值、标准差）	35 (20)
关键信息	
已标注查询（总计）	1752
每条查询中已标注的词数（均值、标准差）	10.6（6.5）
标注者间一致性	.697
模糊词语	
已标注查询（总数）	1686
每条查询中已标注的词数（均值、标准差）	3.2 (2.5)
标注者间一致性	.653

表 2：笔记本电脑语料库的数据集统计。

标注过程是在 Doccano[4] 中以序列标注任务的形式进行的。标注人员标注了关键信息（由一个或多个词组成），例如（以下以粗体显示）：

> 我会买一台基础款笔记本电脑，品牌不限，而且评论不错。

此外，标注人员还标注了模糊词语，例如（此处以粗体显示）：

> 我会买一台基础款笔记本电脑，品牌不限，**且不错的**评论。

笔记本电脑查询的标注者间一致性（Krippendorff α 系数，词级）中，关键信息标注的一致性为 .697，模糊词语标注的一致性为 .653。夹克关键信息标注的平均标注者间一致性为 .697。我们随数据集[5]一并提供一致性度量的实现。表 2 展示了笔记本电脑领域的详细标注统计信息。

列表1展示了最终数据集中笔记本电脑领域的一个数据点，其中包含唯一标识符、领域、原始文本（未经处理）、关于撰写该文本的用户的信息，以及关键信息和模糊词语的标注。每个数据点都包含单独标注，以及一份合并标注，其中标出了至少两名标注人员标注的词语。标注词语通过词语本身及其在文本中的位置（字符级偏移量）来识别。关键信息的标注词语还以文本片段的形式提供，并纳入了各个单独标签。

3.3 数据集可用性

该数据集可公开获取[6]，采用 CC BY-NC-SA 3.0 许可证。该数据集由 GESIS – 莱布尼茨社会科学研究所托管，该机构是知名的社会科学数据提供方。我们以 JSONL 和 CSV 格式提供该数据集，并附有变量说明和标注指南。此外，我们还提供一个 Jupyter Notebook，其中包含用于将数据集导入 Python、执行基本统计分析、计算标注者间一致性以及访问单个数据点的代码。

列表1：数据集中单个数据点的结构。

```
1 {" 标识符 ": 1 88 7 , 2 " 领域 ": ' 笔记本电脑 ' 3 " 文本 ": ' 我想要一
台
用
模
注
 5 9
 7 2
  [
8 "
1 ' 长
1 ',
1 _ 2
1 8 5
 ": (
1
1
1
1
2
2
2
2
2
```

4 使用案例

所提出的自然语言查询数据集可用于自然语言处理（NLP）和信息检索（IR）领域的多种任务。前者可以利用该数据集理解自然语言查询中的语义意图，而后者则可以通过构建针对模糊条件的（特定领域）检索模型而受益。我们将在下一节讨论潜在的使用案例。

4.1 自然语言处理

4.1.1 拼写纠正。这些自然语言查询由用户在没有任何正式限制的情况下撰写，因此包含许多拼写和语法错误。我们的数据集中的以下示例展示了这一问题：“我想买一台联想电脑，因为它们也可以当平板电脑用，非常方便”。该示例包含拼写错误的词语，例如“*rhem*”，以及词语连写错误，例如“*isvery*”。此外，实体特定错误（如拼错的品牌名称，例如“*Lesovo*”）以及领域特定的俚语表达（例如“至少配有 *8 GB* 内存”）在该数据集中也屡见不鲜。由于包含不同类型的拼写错误和口语表达，

口语表达方面，该数据集需要借助纠错模型，才能开展产品搜索中的命名实体识别或信息检索等任务（参见 [30]）。其开发

[4]https://github.com/doccano/doccano [5]Jupyter Notebook 可在 https://git.gesis.org/papenmaa/chiir21_naturallanguagequeries/blob/master/IAA.ipynb 获取 [6]https://git.gesis.org/papenmaa/chiir21_naturallanguagequeries

可以通过研究该数据集，探索针对原始自然语言查询的预处理技术。

4.1.2 模糊性。 信息检索中的一个常见问题是用户的查询语言与系统的索引语言之间存在词汇不匹配 [6]。这是由于用户方面的信息需求较为模糊，无法明确划分不同概念的边界，例如：

> 我想要一款电池续航良好、内存较大且启动相对迅速的产品

领域知识越缺乏，模糊性问题就可能越严重 [14]。因此，需要开发能够识别产品搜索中模糊短语的自动化模型。

该数据集针对模糊表达进行了词级标注。对这类模糊词语进行分类，有助于区分具体条件和模糊条件。对于具体条件，检索模型可以尝试将其与零售商生成的产品信息进行匹配，以筛选结果。然而，对于模糊的查询条件，直接应用此类筛选并不简单。因此，产品检索系统可以利用产品的用户生成内容（例如用户评论），筛选出零售商生成的字段中未包含、但符合用户要求的信息，例如产品质量 [12]。在此前的工作中，我们证明了用户评论与自然语言查询具有很高的相关性

s

根据词汇匹配度量 [14]。

4.2 属性映射

自然语言查询与关键词查询在长度（即查询词项的数量）以及特定对象需要满足的条件数量方面有所不同。命名实体识别（NER）旨在识别自然语言文本中的不同类别，可应用于搜索查询 [7]。由于该数据集是在关键事实级别进行标注的，即标注产品需要满足的要求，因此可用于研究产品搜索中非结构化信息到结构化信息的自动匹配（参见 [28]）。识别这些关键信息的属性领域对产品检索系统很有用，因为可以利用产品的技术字段进行匹配。

与关键词查询相比，自然语言查询具有更复杂的结构，因此另一项值得关注的任务是解析其语义结构。理解并表征其含义，比使用 BM25 等词汇匹配方法更有益。

4.3 产品查询分类

该数据集针对笔记本电脑和夹克两个领域进行了标注。产品检索系统在应用匹配模型之前，需要识别搜索查询所属的类别。误解查询所属的领域会导致令人不满意的结果。针对关键词搜索查询数据 [21, 33]，产品查询分类已得到研究。然而，对自然语言查询进行产品查询分类，可能会促使人们研究更复杂的算法。

5 结论与未来工作

自然语言是产品搜索中一项颇具挑战性的内容。目前，只有少数研究致力于收集搜索引擎用户不带偏见的自然信息需求。我们提供了一个包含3,540条自然语言查询的数据集，涉及笔记本电脑和夹克。我们对其中1,754条关于笔记本电脑的查询进行了标注，标出了其中包含的关键信息和模糊词语；此外，还标注了363条关于夹克的查询。

由于该数据集是一个持续进行的研究项目的一部分，我们计划进一步丰富数据集。首先，我们打算为剩余的夹克查询标注模糊词语和关键信息，以便在第二个产品领域中构建一个完全可比的数据集。其次，我们计划添加经过清理的查询版本，修正其中的拼写错误和标点符号。第三，为了支持交互式信息检索和用户体验设计方面的研究，应将关键信息与结构化产品属性相匹配。在 [14], 中，我们对一个规模较小的数据集（N = 132）进行了初步研究，将关键信息映射到现有产品搜索引擎的分面，并对未匹配的关键信息进行聚类，以确定新的分面。要训练分类器，将关键信息匹配到正确的分面，需要真实标注，我们希望未来将其添加到数据集中。最后，对于部分产品查询，我们计划从产品池中添加相关产品，以便开展检索实验。

然而，对于深度学习方法而言，所提出的数据集样本量相当小，仍可进一步扩大。与以往的数据集类似，所提出的数据集仅涵盖两个产品领域。为深入了解基于该数据集的模型的泛化能力，应纳入更多产品领域。

致谢

本研究部分经费由 DFG（资助编号：388815326）、GESIS 和杜伊斯堡-埃森大学的 VACOS项目提供。

参考文献

[1] Mohammad Aliannejadi、Hamed Zamani、Fabio Crestani 和 W. Bruce Croft。2019。开放领域信息寻求对话中的澄清式提问。载于第*42*届 *ACM SIGIR* 信息检索研究与发展会议论文集（法国巴黎）（*SIGIR '19*）。计算机协会，纽约，美国纽约州，475–484。https://doi.org/ 10 1145 3331184 3331265 。/。[2] Peter Bailey、Alistair Moffat、Falk Scholer 和 Paul Thomas。2015。用户差异性与信息检索系统评估。载于第*38*届 *ACM SIGIR* 信息检索研究与发展国际会议论文集（智利圣地亚哥）（*SIGIR '15*）。计算机协会，纽约，美国纽约州， 625 634 10 1145 2766462 2767728 –。https://doi.org/。/。[3] Toine Bogers、Maria Gäde、Marijn Koolen、Vivien Petras 和 Mette Skov。2018。“这部电影讲的是那个女孩吗？”。载于转变数字世界，Gobinda Chowdhury、Julie McLeod、Val Gillet 和 Peter Willett（编）。施普林格国际出版公司，尚，323–334。[4] Andrei Broder。2002。网络搜索分类体系。*SIGIR*论坛36，2（2002年9月），3–10。https://doi.org/10.1145/792550.792552[5] Layla El Asri、Hannes Schulz、Shikhar Kr Sarma、Jeremie Zumer、Justin Harris、Emery Fine、Rahul Mehrotra 和 Kaheer Suleman。2017。Frames：为目标导向对话系统添加记忆的语料库。载于第*18*届*SIGdial*话语与对话年度会议论文集。207–219。[6] George W. Furnas、Thomas K. Landauer、Louis M. Gomez 和 Susan T. Dumais。1987。人机通信中的词汇问题。《*ACM* 通讯》30，11（1987），964–971。[7] Jiafeng Guo、Gu Xu、Xueqi Cheng 和 Hang Li。2009。查询中的命名实体识别。载于第*32*届*ACM SIGIR*信息检索研究与发展国际会议论文集（美国马萨诸塞州波士顿）（*SIGIR '09*）。计算机协会，纽约，美国纽约州，267–274。https://doi.org/10.1145/1571941.1571989

[8] Ido Guy。2016。通过对话进行搜索：移动网页搜索中的语音查询分析。载于第*39*届 *ACM SIGIR* 信息检索研究与发展会议论文集（意大利比萨）（*SIGIR '16*）。计算机协会，纽约，美国纽约州，35–44。https://doi.org/10 1145 2911451 2911525 。/。Yvonne Kammerer 和 Maja Bohnacker。2012。儿童使用谷歌进行网页搜索：自然语言查询的有效性。载于第*11*届国际儿童交互设计会议论文集（德国不来梅）（*IDC '12*）。计算机协会，纽约，美国纽约州，184 187 10 1145 2307096 2307121 –。https://doi.org/。/。[9] Makoto P. Kato、Takehiro Yamamoto、Hiroaki Ohshima 和 Katsumi Tanaka。2014。查询背后隐藏的认知搜索意图：一项关于查询构建的用户研究。载于第*23*届国际万维网会议论文集（韩国首尔）（*WWW '14 Companion*）。计算机协会，纽约，美国纽约州，313–314。https://doi.org/10.1145/2567948.2577279[10] Marijn Koolen、Toine Bogers、Maria Gäde、Mark Hall、Iris Hendrickx、Hugo Huurdeman、Jaap Kamps、Mette Skov、Suzan Verberne 和 David Walsh。2016。CLEF 2016 社交图书搜索实验室概述。载于欧洲语言跨语言评测论坛国际会议。施普林格，351–370。[11] Felipe Moraes、Jie Yang、Rongting Zhang 和 Vanessa Murdock。2020。属性在产品质量比较中的作用。载于*2020*年人类信息交互与检索会议论文集（加拿大温哥华）（*CHIIR '20*）。计算机协会，纽约，美国纽约州，253–262。https://doi.org/10 1145 3343413 3377956 。/。美国商务部。2020。零售电子商务季度销售额。技术报告。美国商务部。[12] https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf 访问日期：2020年10月22日。[13] Andrea Papenmeier、Alfred Sliwa、Dagmar Kern、Daniel Hienert、Ahmet Aker 和 Norbert Fuhr。2020。“一台现代、最新款的笔记本电脑”——用于产品搜索的自然语言查询中的模糊性。载于*2020*年 *ACM* 交互系统设计会议论文集。2077–2089。[14] Gustavo Penha、Alexandru Balan 和 Claudia Hauff。2019。推出 MANTIS：一个新颖的多领域信息寻求对话数据集。*arXiv* 预印本 *arXiv:1912.04639*（2019）。[15] Martin Potthast、Matthias Hagen、Michael Völske 和 Benno Stein。[无日期]。TREC 主题的探索性搜索任务。（[无日期]）。[16] Chen Qu、Liu Yang、W. Bruce Croft、Johanne R. Trippas、Yongfeng Zhang 和 Minghui Qiu。2018。在信息寻求对话中分析和刻画用户意图（*SIGIR '18*）。计算机协会，纽约，美国纽约州，989 992 10 1145 3209978 3210124 –。https://doi.org/。/。[17] Filip Radlinski、Krisztian Balog、Bill Byrne 和 Karthik Krishnamoorthi。2019。辅导式对话式偏好 elicitation：理解电影偏好的案例研究。载于*SIGdial* 年度话语与对话会议论文集。[18] Eldar Sadikov、Jayant Madhavan、Lu Wang 和 Alon Halevy。2010。按用户意图对查询改写进行聚类。载于第*19*届国际万维网会议论文集（美国北卡罗来纳州罗利）（*WWW '10*）。计算机协会，纽约，美国纽约州，841–850。https://doi.org/10 1145 1772690 1772776 。/。Yelong Shen、Jun Yan、Shuicheng Yan、Lei Ji、Ning Liu 和 Zheng Chen。2011。用于理解用户意图的稀疏隐藏动态条件随机场。载于第*20*届国际万维网会议论文集（印度海得拉巴）（*WWW '11*）。计算机协会，纽约，美国纽约州，7 16 10 1145 1963405 1963411 – 。https://doi.org/ 。/。[19] Michael Skinner 和 Surya Kallumadi。2019。。使用产品分类体系映射进行电子商务查询分类：一种迁移学习方法。。载于 *eCOM@ SIGIR*。[20] Parikshit Sondhi、Mohit Sharma、Pranam Kolari 和 ChengXiang Zhai。2018。《电子商务搜索查询分类体系》。载于 第*41*届*ACM SIGIR*信息检索研究与发展国际会议（美国密歇根州安娜堡）（*SIGIR '18*）。计算机协会，纽约，美国纽约州，1245 1248 10 1145 3209978 3210152 – 。https://doi.org/ 。/。[21] Daria Sorokina 和 Erick Cantu-Paz。2016。。亚马逊搜索：产品排序的乐趣。载于 第*39*届*ACM*信息检索研究与发展国际会议论文集。459–460。[22] Ning Su、Jiyin He、Yiqun Liu、Min Zhang 和 Shaoping Ma。2018。《产品搜索中的用户意图、行为与感知满意度》。载于 第十一届*ACM*网络搜索与数据挖掘国际会议论文集（美国加利福尼亚州玛丽安德尔湾）（*WSDM '18*）。计算机协会，纽约，美国纽约州，547 555 10 1145 3159652 3159714 – 。https://doi.org/ 。/。[23] Paul Thomas、Daniel McDuff、Mary Czerwinski 和 Nick Craswell。2017。《MISC：信息寻求对话数据集》。载于 *SIGIR*第一届信息检索对话式方法国际研讨会（*CAIR'17*），第5卷。[24] Johanne R. Trippas、Damiano Spina、Lawrence Cavedon 和 Mark Sanderson。2017。《人们如何参与纯语音对话搜索任务：初步分析》。载于 *2017*年人类信息交互与检索会议论文集（挪威奥斯陆）（*CHIIR '17*）。ACM，纽约，美国纽约州，325 328 10 1145 3020165 3022144 – 。https://doi.org/ 。/。[25] Andrew Trotman、Jon Degenhardt 和 Surya Kallumadi。2017。。eBay搜索架构。载于与第*40*届*ACM SIGIR*信息检索研究与发展国际会议同期举办的*SIGIR 2017*电子商务研讨会论文集，*eCOM@SIGIR 2017*，日本东京，*2017*年*8*月*11*日（*CEUR*研讨会论文集，第*2311*卷），Jon Degenhardt、Surya Kallumadi、Maarten de Rijke、Luo Si、Andrew Trotman 和 Yinghui Xu（编）。CEUR-WS.org。http://ceur-ws.org/Vol-2311/paper_14.pdf[26] Manos Tsagkias、Tracy Holloway King、Surya Kallumadi、Vanessa Murdock 和 Maarten de Rijke。2020。《电子商务搜索与推荐中的挑战与研究机遇》。载于 *SIGIR*论坛，第54卷。[27] Ellen M. Voorhees 和 Donna K. Harman。2005。*TREC*：信息检索中的实验与评估（数字图书馆与电子出版）。麻省理工学院出版社。[28] Chao Wang 和 Rongkai Zhao。2019。《基于多目标回退优化的拼写纠正算法》。。载于 与第*42*届*ACM SIGIR*信息检索研究与发展国际会议（*SIGIR 2019*）同期举办的*SIGIR 2019*电子商务研讨会论文集，法国巴黎，*2019*年*7*月*25*日。（*CEUR*研讨会论文集），Jon Degenhardt、Surya Kallumadi、Utkarsh Porwal 和 Andrew Trotman（编）。[29] Jizhe Wang、Pipei Huang、Huan Zhao、Zhibo Zhang、Binqiang Zhao 和 Dik Lun Lee。2018。《用于阿里巴巴电子商务推荐的十亿级商品嵌入》。载于 第*24*届*ACM SIGKDD*知识发现与数据挖掘国际会议论文集。839–848。[30] Zhao Yan、Nan Duan、Peng Chen、Ming Zhou、Jianshe Zhou 和 Zhoujun Li。2017。《构建面向网上购物的任务型对话系统》。载于第三十一届*AAAI*人工智能会议论文集（美国加利福尼亚州旧金山）（*AAAI'17*）。AAAI出版社，4618–4625。[31] Hang Yu 和 Lester Litchfield。2020。《采用多目标回退优化的查询分类》。载于 第*43*届*ACM SIGIR*信息检索研究与发展国际会议论文集。1925–1928。