通过递归算法审问审计搜索查询建议偏差

Fabian Haak
fabian.haak@th-koeln.de 德国科隆应用技术大学

Philipp Schaer
philipp.schaer@th-koeln.de 德国科隆应用技术大学

arXiv:2601.02962v1 [cs.IR] 6 Jan 2026

摘要

尽管搜索查询建议在在线信息搜索中发挥着重要作用，但与搜索引擎的其他大多数方面相比，它们尚未得到充分研究。造成这一现象的原因多种多样，其中最突出的问题是上下文稀少，以及每个搜索查询最多只有十条建议所带来的数据基础有限，这使得识别搜索查询建议中的偏差颇具挑战。迄今为止，减少数据稀疏性并提高搜索查询建议偏差识别有效性的最成熟方法，是考察同一查询在一段时间内后续搜索中出现的建议。本文提出一种新的替代方法，用于识别搜索查询偏差，并纳入较少高层级的建议，以充实偏差分析的数据基础。我们采用递归算法审查技术，构建建议树，从而能够获取更隐蔽的搜索查询建议。基于这些建议，我们研究政治领域人物相关搜索中的主题群体偏差。

CCS 概念

•**信息系统** → **查询建议**；信息检索；**网络搜索引擎**。

关键词

搜索查询建议、偏差、算法审计、网络搜索引擎

ACM 参考格式：
Fabian Haak 和 Philipp Schaer。2022年。通过递归算法审查审计搜索查询建议偏差。载于第*14*届*ACM*网络科学会议（*2022*）（*WebSci '22*），*2022*年*6*月*26*日至*29*日，西班牙巴塞罗那。ACM，纽约州纽约，美国，9页。

1 引言

搜索引擎是查找政治议题相关信息的一种广受欢迎且值得信赖的方式，而不仅仅是寻找被视为 [3, 4, 25]的在线信息来源。用户通过查询 [2],传达其信息需求，而搜索引擎针对这些查询提供的搜索查询建议往往会引导搜索方向。尽管查询建议存在不同的定义和类型，例如查询扩展、自动补全、查询预测或查询细化，但我们使用“查询建议”这一术语来描述基于当前输入到搜索引擎搜索框中的内容而提供的搜索查询建议。（搜索）查询建议会影响用户的搜索方式，并生成不同的搜索引擎结果；众所周知，这些结果会影响意见形成，进而影响总体公众意见 [12, 32]。在理想且中立的环境中，查询建议对信息搜索过程的影响不会造成任何危险。然而，尤其是在具有明确立场的领域中，例如政治，查询建议可能带有倾向性、不具代表性，或偏向某些立场、观点或主题。

关于有偏查询建议的研究，旨在寻找识别有偏查询建议的方法，尽管查询建议存在稀疏上下文、非短语性以及领域／主题依赖性等不利条件。由于有害性、粗俗性或其他形式的负面归因而存在问题的建议，大多数时候会被搜索引擎拦截或被用户识别出来 [14]。更棘手的是由主题的有偏分布或其他潜意识层面的信息曝光偏移所引发的偏差类型。它们无法从单个查询建议列表中识别出来，而是需要更大的数据集，而这类数据集十分罕见。

分析查询建议偏差的另一个问题在于，信息搜索通常包括反复改述和多次搜索，直到结果符合预期，或信息需求得到明确、用户感到满意为止 [9]。由于搜索不会在较长时间内持续进行，因此分析一组从未在同一时间出现过的建议并不具有代表性。此外，查询建议会根据用户输入动态生成。

本研究旨在采用递归算法审问（RAI）技术 [26], ，该技术此前曾用于比较谷歌和必应生成的建议。在本文提出的方法中，我们将所谓的建议树与系统化的查询改写相结合，以获取尽可能完整的查询建议集合。

本研究运用这一方法，尝试揭示人物相关搜索的查询建议中是否存在主题群体偏差，其中以政治人物姓名作为根查询输入。我们旨在回答两个问题：

（1）政治领域的人物相关搜索查询建议是否存在偏向特定群体的主题偏差？（2）与纵向“顶层建议分析”层级的分析 [6] 或类似的感知导向方法 [15] 相比，深度优先建议网络分析能否揭示更多或不同的偏差？

2 相关工作

搜索引擎对塑造政治观点有着显著影响：Epstein 和 Robertson [12] 创造了“搜索引擎操纵效应（SEME）”这一术语，并估计搜索查询建议对用户搜索内容的影响可能改变全国选举的结果。尽管有人质疑 SEME 的影响程度 [32], ，但它凸显了搜索引擎偏差问题的紧迫性和严重性。带有偏差的搜索结果以及由此产生的

搜索建议可能会造成问题，因为用户对搜索引擎抱有高度信任 [4]，而这种信任可能被利用。此外，研究表明，查询建议可以被操纵，而且无法仅凭建议本身判断它们是否经过操纵 [30]。这进一步揭示了查询建议影响力所带来的危险。

偏差可能源于查询建议所依据的数据和算法 [20],，也可能源于用户认知方面的影响 [1]。搜索引擎展示数据中固有的信息；因此，搜索结果中的偏差与人们的思想和自然语言文本中的偏差一样无处不在 [5, 10, 24]。谷歌也承认这一点，并表示他们知道其搜索引擎呈现的内容有时带有误导性和偏差 [14]。由于查询建议是基于以往的搜索生成的，包括用户输入以及与查询建议的交互，因此它们极易形成“破坏性反馈循环[s]” [16],。贝蒂娜·武尔夫（前德国总统克里斯蒂安·武尔夫的妻子）就曾遭遇过这种情况。2012年，武尔夫对谷歌采取了法律行动，因为搜索她的姓名时，会直接出现与“陪侍”或“卖淫”相关的查询建议。由于媒体关注度上升，这些搜索及其诽谤性建议被更多人搜索和点击，形成了典型的马太效应。谷歌和其他搜索引擎提供商表示，他们确实会过滤建议以避免负面影响；不过，这仅适用于单条有害建议 [31]。

算法审计是一种描述并调查算法行为及结果的方法 [29]。审计方法已被用于大规模调查搜索引擎，研究其在党派倾向方面的政治偏见 [18],，以及其在政治搜索情境中的个性化效应 [21, 27]。

有偏差的搜索结果一直是讨论和研究的对象。Kulshrestha 等人研究并比较了谷歌搜索结果和 Twitter 搜索中的偏见 [19]。他们的偏差检测方法依赖作者、发布与传播平台等外围信息，以及从这些来源获得的信息。尽管已有许多系统被提出，以实现搜索引擎结果的算法公平排序，其中包括一项仍在进行的 TREC赛道 [11],，但现代网络搜索引擎中的搜索建议组件或自动补全功能，在当前有关公平性和偏差的研究中仍未得到充分关注。

迄今为止，针对搜索查询建议中的偏差检测，相关尝试寥寥无几。最初的偏差研究聚焦于人物相关搜索，尤其关注政治人物 [7]。后来，Mertens等人 [22] 和Bonart等人 [6] 发现，在政治人物搜索中存在显著的性别方面偏差。前一项研究使用了Twitter搜索数据集，后一项则分析了针对政治人物的搜索查询建议中的主题偏差，以及建议随时间的变化。他们通过纵向方式收集建议来创建数据集，即在一段时间内定期获取查询建议，并分析某个查询的全部建议 [6]。然而，他们每次搜索仅分析至多十条不重复的单词建议，且不考虑列表位置和频率，因此在研究中未纳入查询建议的许多特性。因此，其他方法尝试将这些特性，以及人们对查询建议的感知和交互方式纳入考量。这些方法旨在检测事实可感知偏差，并引入新的考虑列表顺序的指标和随时间变化的频数统计 [15]。这些研究采用的偏差检测流程经过扩展，纳入了搜索建议在自动完成列表中的排名，并结合一种感知意识模型 [15],，将用户如何感知和使用搜索查询建议的研究发现 [17] 转化为偏差检测的方法和指标。使用同一数据集，他们提出了新的指标，将建议的频率和列表位置纳入考量，从而提高分析结果在推断事实可感知偏差方面的有效性。尽管这些研究在研究领域和偏差检测方法上与本文提出的方法有诸多相似之处，但所采用的数据集存在显著差异。此前两项研究都在较长时间内收集建议，形成纵向数据；而我们分析的是横截面建议数据。

研究者还提出了搜索引擎偏差的其他特征，例如用户输入查询时出现的动态建议、查询改写策略，以及信息搜索过程中产生的认知效应 [1]，但其中只有少数得到了系统性考察。

3 审计搜索查询偏差

本文提出的方法旨在通过扩展分析所依据的数据，分析搜索查询建议中的主题群体偏差。以往用于搜索查询建议偏差识别的方法主要存在两个问题：

（1）公开可用的数据集十分稀缺，利用以往发表的方法创建搜索查询建议数据集耗时较长。（2）仅对基础层级建议（针对根查询的建议）进行偏差分析，会忽略许多用户在搜索过程中改写查询并反复搜索不同搜索查询时可能接触到的建议。

本文旨在克服这些问题。为此，我们将 RAI方法 [26]与类似于 [6, 15] 的基于回归的主题搜索查询建议流程结合到一个偏差检测框架中。

该方法的组成部分和步骤如图1所示。首先，选择根术语。接下来，使用 RAI 为选定的根查询创建一个包含搜索查询建议的数据集。随后对所得数据进行预处理，将处理后的建议向量化并按主题聚类。最后，通过执行回归分析来分析群体偏差。

3.1 使用递归算法审查提取查询建议

以下章节概述了使用 RAI 创建建议树的方法。

首先，汇总一份根查询列表。对于每个根查询，要获取以该词为根的建议树，首先需要检索该根查询的查询建议（通常最多十条），并将其添加为叶节点，原始词则作为其父节点。然后，将这些建议分别作为新的输入查询，抓取相应的建议，并将其作为上一条建议的子节点添加到树中。如此递归重复，直到生成深度为 n 的树，或者当前层级的任何查询都不再有建议为止。

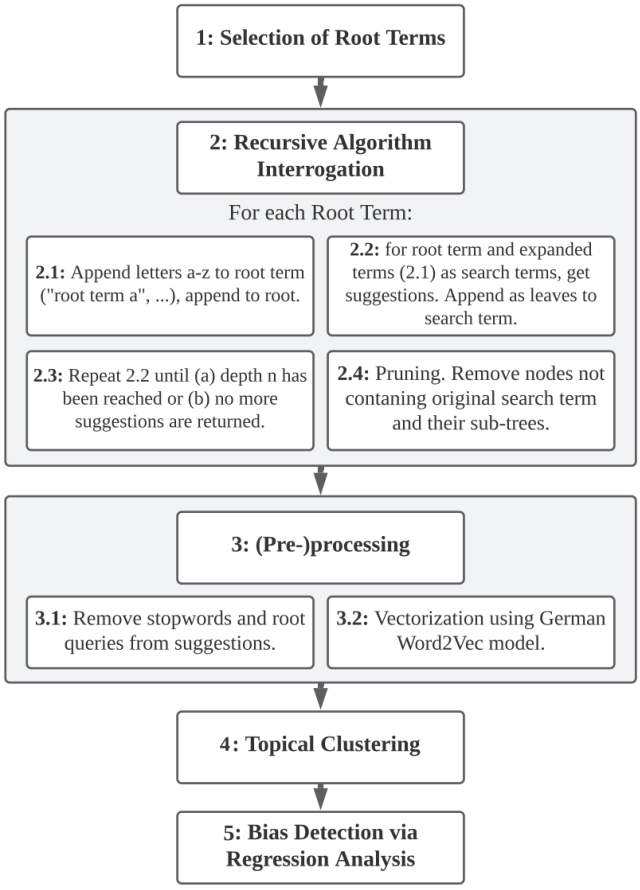


图1：[26]提出的使用建议树进行基于 RAI 的偏差检测方法示意图。

这会生成由建议构成的广度优先网络，并以树状结构（建议树）汇编。除了最初提出的 RAI 方法外，在我们的方法中，还会通过在每个搜索词末尾依次附加 a 到 z 的每个字母来扩展查询。此过程拓宽了搜索树的第一层，仅第一层就最多可从 270 条建议开始（26*10 加上根查询的 10 条建议），而单独针对根查询最多只能生成十条建议。我们的目标是找出用户输入查询时可能接触到的建议。大量建议凸显了以往检测搜索查询建议偏差的方法存在的一个重大问题：每个搜索词最多分析十条建议，忽略了在线信息搜索的迭代特性，以及用户输入过程中建议不断出现的情况。下一步是对建议树进行剪枝。删除与根查询在语义上无关的叶节点及其整个子树。例如，对于输入查询“volker wissing”（现任联邦数字事务和交通部长），其中一条建议是“wissing”。它不仅是政治人物的姓名，也是一个城市的名称。将“wissing”作为新的输入查询所得到的建议，与沃尔克·维辛这个人无关，而是与这座城市有关。我们假设用户在尝试查找沃尔克·维辛的相关信息时，不太可能会点击这一特定建议。因此，这两条建议很可能都不相关。图2展示了另一个剪枝示例，以红色突出显示。此处，“电动汽车税”作为一条建议被返回，随后该建议及其覆盖的子树均被移除。剪去这些子树，可以防止这些建议干扰偏差检测结果。

建议树最初用于比较谷歌和必应这两种搜索引擎生成的建议，并描述两者在更深层建议层级上的行为 [26]。研究发现表明，谷歌生成的查询数量更多，且多样性也高于必应。尽管偏见分析并非他们的研究目标，RAI 相较于以往创建查询建议数据集的方法仍具有优势。以往的方法会反复收集查询词的基础层级建议，且收集时间往往较长 [6]。这样做的优点是能够考察建议如何随时间变化。然而，除了十条顶层建议之外，其他所有建议都被忽略了；用户在针对根词以外的任何查询搜索信息时可能接触到、也确实会接触到的大多数建议，例如通过点击某条建议接触到的建议，都未被纳入考量。因此，这种做法会损害以此方式开展的偏差分析结果的有效性。此外，如果根查询来自非常同质化的选择，并具有许多显著的共同特征，那么它们的基础层级建议往往彼此相似。例如，抓取德国政客的建议时，许多建议都很相似，因为它们内容十分笼统，或与当事人担任的职务有关，而非涉及个人特征（如“年龄”“推特”“丈夫/配偶”“选举”或“收入”）。使用 RAI 的另一项优势是，无须再耗费大量时间抓取许多查询的建议来构建数据集。总体而言，通过 RAI 和建议树解析，纳入给定根查询的全方位查询建议，是解决这一问题的可行方案。最后，数据收集方法是我们的方法与这些研究之间唯一决定性的差异。因此，将本研究的发现与这些研究的发现进行比较，我们便能直接考察数据收集方式对查询建议偏差发现的影响。此外，由于我们试图反映用户在迭代搜索会话中接触建议的平均情况，因此可以判断这种迭代式接触中的偏差，是否不同于用户仅针对根查询接触建议时的偏差。

3.2 方法论方法

方法论方法中各组件的结构和顺序如图1所示。

3.2.1 数据集创建与预处理。

根术语的选择与建议抓取。由于针对（群体）个人的偏差被认为尤为棘手，我们选择姓名作为根查询。为使结果能够与前述研究的结果进行比较，我们聚焦于政治领域，并选取部分德国政治人物的姓名作为根查询。我们选取了德国联邦部长及其前任、现任德国各州州长，以及在2021年联邦选举中得票最多的七个德国政党的领导人及其前任，

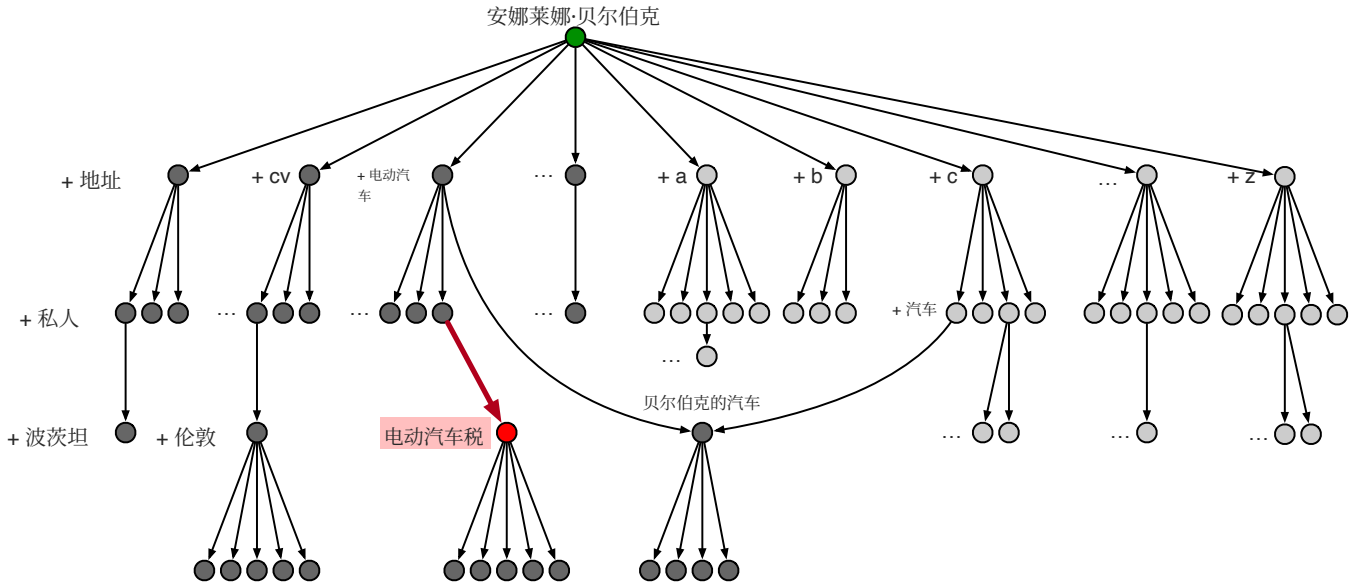


图2：基于根词“安娜莱娜·贝尔伯克”（德国联邦外交部长、绿党前联合主席）构建的建议树示例，其中该根词以绿色标记。添加到原查询字符串中的建议词以 + 标记；否则，将替换后的查询字符串完整写出。为保持图示清晰，省略了部分标签。深灰色节点表示基于原查询的建议词。浅灰色节点表示基于首字母建议或其后续建议的建议词。与初始根查询无关的建议节点以红色标记。这些节点将被剪枝，包括其后续子树。

共计55个根术语。对于所有这些人物，我们从德国联邦选举联邦选举主任网站获取了其性别、党派归属、政治职务和出生年份等元属性的人口统计信息 [8]。虽然根术语的数量相对较少，但这些元属性的分布反映了德国高级政治人物的整体情况。德国联邦议院议员中约有69%为男性 [8], 35 ，而在作为根查询的55名政治人物中，约有64%为男性。类似地，他们的党派归属分布与联邦议院的情况非常相似。随后，我们按照第3.1小节所述的RAI程序创建了深度为8的建议树。为收集德语查询建议并支持变音符号，我们对该方法进行了调整。我们在2022年2月的两天内从谷歌抓取了建议。我们选择将研究聚焦于谷歌，以确保研究结果能够与同样使用谷歌建议作为数据集的既往研究进行比较。抓取工作共获得21,407条建议。

预处理：清理数据。 接下来，预处理会从数据集中移除不必要且存在问题的数据。首先，除根查询外，所有查询中的根术语（政治人物的姓名）都会被移除。这样可以统一查询建议，并实现彼此之间的树形结构比较。在后续层级的建议中被重新表述的查询部分（例如“地址”

预处理步骤	唯一建议
RAI 抓取	8,899
剪枝	8,689
移除有歧义的内容根词	8,079
移除建议不使用向量表示	7,919

表1：预处理进度各步骤中的数据集唯一建议数量。

以及子树“(annalena baerbock) address”中的“private”→“(an- nalena baerbock) address private”→“(annalena baerbock) address private potsdam”（参见图2）不会从查询中删除，因为我们假设，更深层的子树反映了用户对这一主题的兴趣，而改写后的词语对于查询的含义至关重要。删除根查询后，还剩下8,899条唯一建议；每个根术语平均有307条唯一建议，最少为48条（“andreas bovenschulte”），最多为800条（“olaf scholz”）。表1展示了数据集在每个预处理步骤之后的建议数量。接下来，按照第3.1小节所述执行剪枝。剪枝会删除所有根不包含原始根术语的子树。为处理拼写错误和变音符号变体，我们为每个根术语手动创建了一个不会被剪掉的变体列表。对所有建议树进行剪枝后，共从谷歌数据集中删除了210条建议。

表2：谷歌建议数据的回归分析结果。仅纳入存在显著差异的元属性，并突出显示显著效应（$p \leq \alpha = 0.05$）。

	聚类1：政治		聚类2：地点		簇3：个人		数量建议	
	R2	p	R2	p	R2	p	R2	p
性别：女性	**0.074**	**0.045**	**-0.096**	**0.022**	0.003	0.702	-0.026	0.235
部长	**0.106**	**0.015**	0.001	0.8	**-0.092**	**0.025**	0.01	0.473
前部长	-0.043	0.129	-0.027	0.234	**0.103**	**0.017**	0.04	0.141
CSU	**-0.071**	**0.049**	0.003	0.705	0.031	0.201	0.056	0.082

表3：K均值聚类词项示例，已翻译为英语。

聚类1：政治	聚类2：地点	簇3：个人
劳动部	柏林	美发师
联邦部长	多特蒙德地址	离婚妻子
北约与俄罗斯	亚琛大学	电视露面
...	...	...

除根术语外，还删除了停用词，并将不含停用词的结果建议单独保存。审查生成的建议树后，我们发现其中一个根术语存在问题。“Gerd Müller”是一名前联邦部长，但有一位同名的知名度更高的足球运动员。数据表明，大多数建议与这名运动员有关，而不是与这位政治人物有关；鉴于大多数建议数量稀少且主题存在歧义，我们决定删除该树及其所有相关建议。这使数据集中的建议数量减少到8,079条（参见表1）。

预处理：向量化。为了对建议进行语义表示，我们使用一个在 CoNLL 2017 语料库上训练、由 [13] 提供的大型 Word2Vec 模型对这些建议进行向量化 [23] 。我们将多种词汇覆盖率、复杂度和数据基础各不相同的向量模型与不同的聚类方法进行了比较。随后，我们手动评估最终聚类内部主题的一致性。根据这些评估，Word2Vec 生成的聚类质量最高：每个聚类都能轻松识别出一个描述性特征，而且生成的聚类在主题上彼此区分明显。对于包含多个词项的建议，我们将各词项向量的均值用作句子表示。尽管这种方法可能会导致对并非句子、而是关键词集合的查询建议得出次优结果，但实际表现良好。我们决定舍弃含有向量模型无法表示的词元的建议。不过，这种情况只出现在很小比例的建议中，其中大多数包含与冠状病毒相关的词项，例如“coronaidiot”或“corona-test”，或拼写错误的词项。完成这最后一步预处理后，我们将所有树中的 7,919 条唯一建议汇总到用于偏见分析的数据集中。

3.3 语义聚类

由于我们的目标是比较每位政治人物的查询建议中的主题，以识别与元属性群体（性别、政党、年龄）相关的偏见主题关联，因此需要将查询建议划分到主题聚类中。我们通过执行 k-means 聚类来实现这一点。第 3.2.1 节对向量化和聚类方法的评估表明，k-means 生成的聚类优于潜在狄利克雷分配、基于密度的空间聚类以及基于简单相似度度量的聚类方法。我们通过计算一至十个聚类时各聚类内部平方距离之和，确定理想的聚类数为三个。平均轮廓分析也支持这一结论，并显示三个聚类时轮廓得分最高 [28]。手动比较三个、四个和五个聚类的内容还发现，后两种情况下各有一到两对聚类包含的建议类型基本相同，这进一步表明，对给定的查询建议集执行 k-means 聚类，可以找出数据中的三个主要类别。此外，采用三个聚类有助于将我们的结果与前文提及的、研究政治领域中与人物相关的搜索查询建议偏差的研究进行比较；这些研究使用了非常相似的聚类 [6, 15]。

最终聚类。通过研究分配给各聚类的术语，为最终聚类取了描述性名称。表3展示了各聚类中的一些示例术语。第一个聚类可称为政治术语。其范围从直接包含政治术语的建议（“最高法院辩论”“部长电子邮件”“2022年选举”“政府发言人”），到带有经济或职业背景的建议（“汽车行业”“警察”“律师”）。分配到第二个聚类的建议有一个共同点：都包含地点名称，例如城市或国家（“莱比锡教授解剖学”“波鸿建筑师”“亚琛大学”）。尽管有这一共同特征，第二个聚类中的建议除此之外并无其他共性，也可能适合归入另外两个聚类之一（“伦内施塔特去世”“柏林选举”）。第三个聚类由可描述为个人相关的术语组成。许多建议包含与家庭或关系相关的词语，如“妻子”或“子女”；另一些则与外貌有关（“头发”“美丽”“有一只玻璃眼”），或者不是德语（未翻译：“自由日”）。对于其他任何领域而言，该聚类内部主题的多样性都可能造成问题。由于政治人物通常表现得非常专业且深思熟虑（例如，他们很少在公开场合谈论非政治相关话题），我们认为区分政治与非政治内容就足够了。第二个聚类中的建议往往可以归入另外两个聚类中的任意一个。

由于地点建议数量众多且具有导航性质，单独划分出地点聚类可以减少噪声，从而提高比较另外两个聚类的有效性。

3.4 主题偏见分析

从中立的角度来看，对于一位重要政治人物而言，其性别、年龄，以及在一定程度上的政治派别，不应影响人们对该政治人物哪些方面感兴趣。因此，我们预计不同群体政治人物的查询建议不会有所差异；若存在显著差异，则表明有偏差。借助按主题聚类的建议，我们可以考察具有共同元属性特征的不同群体政治人物的查询建议，其主题分布是否不同，进而判断是否存在系统性主题群体偏差。我们关注的政治人物群体由性别、年龄、党派归属和政治职务等元属性界定。我们通过考察女性和男性政治人物的建议是否存在差异，来研究一种称为性别偏差的群体偏差。另一个属性是政治人物的年龄，以出生年份为依据。此外，偏见分析还会考虑党派归属，具体取决于此人所属的政党。在这一类别中，会注明其所属的德国政党：德国社会民主党、德国基督教民主联盟、基社盟、自由民主党、德国选择党、左翼党或联盟90/绿党。最后，政治职务说明此人因何种群体归属而被列入根术语名单。该元属性的类别包括2017年和2021年立法期的联邦部长、各联邦州现任州长，以及上述德国政党的领导人。

目标是检测不同元属性群体之间主题曝光度的差异所体现的群体偏差。为揭示显著差异，我们进行了线性回归。元属性的二元虚拟变量作为自变量，而每位政治人物各个聚类的建议百分比则作为因变量。针对主题聚类 $c \in 1, ..., k$ 的回归模型可表示为

$$y_{i,c} = \beta_0 + \beta_1 x_{i,1} + ... + \beta_p x_{i,p} + \epsilon_i, \quad (1)$$

其中，ϵ_i 为独立误差项，$i = 1, ..., N$ 为观测索引。显著的系数 β_p 表明该聚类和属性 $x_{i,p}$ 存在主题偏差。为避免多重共线性，每个属性选取一个变量作为基准类别，并将其省略。

政治人物查询建议中的偏差发现。对 Google 查询建议进行回归分析所得显著发现的概览见表 2。尽管 RAI 检索到的查询建议完整度前所未有，但与以往研究相比，我们并未发现其对所关注的元属性群体存在更显著的偏差。

仅以性别作为解释变量进行线性回归分析，结果显示存在显著偏差。回归分析结果表明，三个聚类中有两个存在性别偏差。最值得注意的是，结果揭示了主题层面的性别偏差：根据作为根查询使用的姓名所对应人物的性别，与职业和政治主题相关的查询建议所占比例存在显著差异。有趣的是，在给定的数据集中，搜索女性政治人物时返回的建议中，政治或经济主题的建议多于搜索男性政治人物时返回的建议（$0.045 = p \leq \alpha = 0.05$）。R平方度量表明，政治人物的性别可以解释政治建议聚类中建议占比差异的7.4%。女性政治人物姓名对应的政治建议更多，这与以基准层面为中心的偏差分析结果相反；此类分析发现，针对女性的政治建议显著较少（参见 [6, 15]）。

包含地点名称的聚类建议在男性和女性政治人物之间也存在不均等分布。针对女性政治人物的查询建议中，可归入该聚类的建议显著较少（$0.022 = p \leq \alpha = 0.05$）。性别差异解释了聚类2建议占比方差的9.6%。

除性别之外，政治职务也是影响查询建议主题分布的因素。德国现任部长获得的政治和经济类聚类1建议显著更多（$0.015 = p \leq \alpha = 0.05$），而被划分到包含个人查询建议的聚类中的建议则显著更少（$0.025 = p \leq \alpha = 0.05$）。回归结果表明，群体归属“部长”解释了聚类1数据方差的10.6%，以及聚类3数据方差的9.2%。

年龄和大多数政党的党派归属，并未显示为影响本研究所考察的三个主题聚类之主题分布的偏差因素。只有基社盟政治人物获得的政治类查询建议显著少于其他政党的政治人物（$0.049 = p \leq \alpha = 0.05$）。鉴于先前关于查询建议主题偏差的研究采用了非常相似的聚类，且聚类描述也具有可比性（ [6, 15], ），我们预期，在比较这些研究所使用的相同元属性群体的查询建议主题分布时，也会发现类似的偏差结果。总体而言，我们发现的偏差案例数量相近。与先前研究的发现形成鲜明对比的是，我们发现的性别偏差，在政治和经济相关建议方面与先前研究指出的偏差方向相反。对于所有元属性群体，建议总数似乎均不存在偏差。整体较小的系数可能是政治领域查询建议本身不存在偏差所致；其他研究也未发现更多显著偏差。

4 讨论

尽管在德国政治人物姓名的查询建议中发现的偏差很少，但在偏差检测的数据集创建中采用 RAI 的方法非常成功。只需付出很少的努力，这项技术便能生成大量数据，使研究人员能够分析用户搜索某个人时可能接触到的偏差“全貌”。尤其是通过添加字母来扩展根查询，事实证明，这是高效收集某个根查询的大部分甚至全部建议的有效方法。

我们的研究结果与先前关于政治领域人物相关搜索的搜索查询建议中群体偏差的研究发现部分相悖。

我们没有发现女性政治人物获得更多个人建议，而是发现她们获得了更多政治建议。与这些先前研究相比，研究方法中唯一显著不同的因素，是纳入了n元语法（ [6]和 [15] 都只使用单词建议来检测偏差），以及使用横截面数据而非纵向建议数据。纳入n元语法所带来的影响只能推测。仅使用单词建议的一个优势是，所采用的主题聚类方法能产生好得多的结果。相反，不纳入n元语法会忽略大多数建议，并严重损害任何有关偏差的研究发现的有效性。假设n元语法与主题无关，我们得出结论：该领域的主题群体偏差会因基础层级建议与建议层级较低处相对罕见的建议而异。

与其他类型的文档相比，查询建议的偏见分析面临着独特的挑战。其中一个主要挑战是上下文缺失，另一个重要方面则是人们如何感知查询建议。在查询建议的主题偏差分析中，引入感知意识因素（包括排名感知因素和频率感知因素），提高了结果的有效性 [15]。尽管本研究的目标同样是提高有效性，但我们既未直接纳入排名感知因素，也未直接纳入频率感知因素。由于数据集是在单次爬取中收集的，因此无法像以往的方法那样整合频率。不过，通过后续建议中重复出现的词语（“karl lauterbach girlfriend” ⟶ “karl lauterbach exgirl-friend”），我们也考虑了词语的频率。排名感知度量是区分查询建议列表中不同位置的建议可感知性的重要手段。尽管我们本可以采用排名感知度量，但我们决定使用基于集合的度量。由于没有统计证据表明，与建议树更深处的建议相比，根查询处或其附近深度的建议在可感知性上应受到不同权重，因此无法有效地纳入排名权重。

由于本研究未能支持以往关于政治领域查询建议中主题群体偏差的发现，因此需要进一步研究，以评估所提出的方法在搜索查询建议主题偏差检测方面的有效性。本研究与以往研究的结果相矛盾，可能有以下原因：当前的最佳实践方法——聚类和回归分析——可能并不十分适用于非短语式查询建议的稀疏特性；或者，基础层级建议中的偏差与根查询词相关的完整查询建议集合中的偏差存在显著差异。由于大多数关于搜索查询建议偏差的研究都在政治领域开展，研究结果有限也可能与该领域有关。

展望。 尽管仍存在一些不确定性，但本研究提出的方法前景十分可观，因此我们已经开始规划后续研究。后续研究可以采用更多技术来构建规模更大、能够反映“感知”的[15] 数据集，从而进一步发挥 RAI 的优势。例如，一种可行的方法是将基于智能体的算法审计或用户生成的查询日志纳入方法组合。我们认为，本方法的主要弱点在于所采用的向量化和主题聚类方法。尽管使用基于与搜索查询差异很大的文本语料库训练的词嵌入，有助于将本研究的结果与以往研究进行比较，但这种做法存在问题。在本研究中，聚类后的建议数量并不多，因此我们能够通过定性分析确保，尽管词嵌入并非最优，聚类仍取得了不错的结果。此外，与 K 均值相比，采用更复杂的聚类方法可以减少噪声，提高结果的整体有效性。分析中未纳入考虑排名的因素和指标，原因在于我们不了解如何为建议树不同深度上的建议赋予权重。后续可通过用户研究进一步探讨这一问题，并将相关因素纳入语料库构建，或对不同排名的建议分别赋予不同权重。将该技术应用于更多搜索引擎和其他领域，有助于考察这种方法是否适用于检测搜索查询建议中的偏差，并深入了解这一主题。此外，研究不同搜索引擎搜索建议中的偏见，也可成为后续研究的一个重要方向。

结论。 在本研究中，我们提出了一种新的搜索查询偏差识别方法。该方法结合了最先进的建议解析技术，以及经过验证的查询建议主题偏差检测方法。我们的目标是通过将查询建议更完整地表示为搜索词，提升查询建议偏差分析的有效性。

通过系统地抓取根查询的基础层级建议、扩展根查询，再递归抓取所有这些查询在一定深度范围内的建议，我们试图创建一个数据集，涵盖用户搜索根查询相关信息时可能遇到的几乎所有建议。按照查询建议偏差识别的最佳实践，我们随后采用聚类和回归分析，查找针对具有特定元属性（例如性别）的不同群体政治人物的查询建议中，主题分布是否存在偏差模式。

结果显示存在一些群体偏差。首先，在主题层面识别出的性别偏差表明，针对女性政治人物的建议中，被归类为政治类的建议显著更多。此外，政治人物的政治角色也被发现是一个造成偏差的元属性。德国政府现任部长获得的政治建议比例显著更高，个性化建议则更少。然而，搜索前部长时，系统提供的个性化建议显著更多。这些发现与其他关于政治领域中人物相关词语搜索查询建议偏差的研究存在显著差异。因此，我们推测，与仅采用每个根词的十条基础层级建议相比，纳入完整的查询建议集合导致了这种差异。

参考文献

[1] Leif Azzopardi. 2021. 搜索中的认知偏差：信息检索中认知偏差的回顾与反思。。发表于*CHIIR '21*：*ACM SIGIR* 人类信息交互与检索会议，澳大利亚堪培拉，*3*月 *14 19, 2021* 27 37 10 1145 3406522 3446023 -。–。https://doi.org/。/。[2] Nicholas J. Belkin.。1980 知识状态异常作为信息检索的基础。《加拿大信息科学杂志》5, 1 (1980), 133–143. [3] Edelman Berland. 2017. 2017年爱德曼信任度晴雨表。https://www.edelman. com/trust/2017-trust-barometer[4] Edelman Berland. 2022. 2022年爱德曼信任度晴雨表。https://www.edelman. com/trust/2022-trust-barometer

[5] Tolga Bolukbasi、Kai-Wei Chang、James Y. Zou、Venkatesh Saligrama 和 Adam Tauman Kalai. 2016. 男人之于计算机程序员，正如女人之于家庭主妇？去偏词嵌入。。发表于《神经信息处理系统进展》第29卷：2016年神经信息处理系统年会，2016年12月5日至10日，西班牙巴塞罗那。4349–4357. https://proceedings.neurips.cc/paper/2016/hash/a486cd07e4ac3d270571622f4f316ec5-Abstract.html [6] Malte Bonart、Anastasiia Samokhina、Gernot Heisenberg 和 Philipp Schaer. 2019. 网络搜索引擎查询建议中的偏差调查。Online Information Review 44, 2（2019年1月）, 365–381. https://doi.org/10.1108/OIR-11- 2018-0341 出版方：Emerald Publishing Limited. [7] Malte Bonart 和 Philipp Schaer. 2018. 政治相关查询的查询建议与搜索引擎结果之间的跨期联系。发表于 EuroCSS *2018* 1812 08585 数据集挑战。2022 [8] 科隆。https://arxiv.org/abs/。Bundeswahlleiter.。联邦选举官。https://www.bundeswahlleiter.de/en/index.html[9] Fei Cai 和 Maarten de Rijke. 2016. 信息检索中的查询自动补全综述。Foundations and Trends® in Information Retrieval 10, 4 (2016 273 363 10 1561 1500000055), –。https://doi.org/。/ Jacob [10] Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova. 2019. BERT：语言理解的深层双向变换器预训练。发表于《2019年北美计算语言学协会分会会议论文集：人类语言技术》，第1卷（长篇与短篇论文）。计算语言学协会，明尼阿波利斯，明尼苏达州， 4171 4186 10 18653 1 19 1423 –。https://doi.org/。/v /N - [11] Michael D. Ekstrand、Graham McDonald、Amifa Raj 和 Isaac Johnson. 2022. TREC 2021公平排序赛道概览。发表于《第三十届文本检索会议（TREC 2021）论文集》。[12] Robert Epstein 和 Ronald E. Robertson. 2015. 搜索引擎操纵效应（SEME）及其对选举结果可能产生的影响。《美国国家科学院院刊》112, 33（2015年8月）, E4512–E4521. https: //doi.org/10.1073/pnas.1419828112 出版方：美国国家科学院 栏目：PNAS Plus。[13] Murhaf Fares、Andrey Kutuzov、Stephan Oepen 和 Erik Velldal. 2017. 词向量、再利用与可复现性：迈向大型文本资源社区资源库。。发表于《第21届北欧计算语言学会议论文集》，NODALIDA 2017，瑞典哥德堡，2017年5月22日至24日。271–276. http://www.ep.liu.se/ecp/article.asp?issue=131&article=037&volume[14] 2019 谷歌。。谷歌如何打击虚假信息。https://www.blog.google/ documents/37/How_Google_Fights_Disinformation.pdf/ [15] Fabian Haak 和 Philipp Schaer. 2021. 面向查询建议的感知感知偏差检测。发表于《信息检索中的偏差与公平性进展》，Lu- dovico Boratto、Stefano Faralli、Mirko Marras 和 Giovanni Stilo（编）。Springer International Publishing，Cham，130–142. [16] Djoerd Hiemstra. 2020. 减少查询自动补全中的错误信息。CoRR abs/ 2007 02620 2020 2007 02620 .（）。[17] https://arxiv.org/abs/。Kajta Hofmann、Bhaskar Mitra、Filip Radlinski 和 Milad Shokouhi. 2014. 用户与查询自动补全交互的眼动追踪研究。发表于第23届 ACM 国际信息与知识管理会议论文集（CIKM '14）。计算机协会，纽约，纽约州，美国，549–558. https://doi.org/10.1145/2661829.2661922 活动地点：上海，中国。[18] Desheng Hu、Shan Jiang、Ronald E. Robertson 和 Christo Wilson. 2019. 审查谷歌搜索摘要中的党派倾向。发表于万维网会议（WWW '19）。ACM，纽约，纽约州，美国，693–704. https://doi.org/10.1145/3308558. 3313654[19] Juhi Kulshrestha、Motahhare Eslami、Johnnatan Messias、Muhammad Bilal Zafar、Saptarshi Ghosh、Krishna P. Gummadi 和 Karrie Karahalios. 2019. 搜索偏见量化：调查社交媒体和网络搜索中的政治偏见。《信息检索期刊》22, 1（2019年4月）, 188–227. https://doi.org/10.1007/ s10791-018-9341-2[20] L. Introna 和 H. Nissenbaum. 2000. 定义网络：搜索引擎政治。《计算机》33, 1（2000年1月）, 54–62. https://doi.org/10.1109/2.816269[21] Huyen Le、Raven Maragh、Brian Ekdale、Andrew High、Timothy Havens 和 Zubair Shafiq. 2019. 衡量 Google 新闻搜索的政治个性化。发表于万维网会议（WWW '19）。ACM，纽约，纽约州，美国， 2957 2963 10 1145 3308558 3313682 –。https://doi.org/。/。Armin [22] Mertens、Franziska Pradel、Ayjeren Rozyjumayeva 和 Jens Wäckerle. 2019. 推文如何，回复便如何？：政治人物数字通信中的性别偏差。。发表于第11届 ACM 网络科学会议论文集，WebSci 2019，美国马萨诸塞州波士顿，2019年6月30日至7月3日。193–201. https://doi.org/10.1145/ 3292522.3326013[23] Joakim Nivre 等. 2017. Universal Dependencies 2.0 – CoNLL 2017共享任务开发集与测试集。http://hdl.handle.net/11234/1-2184 LINDAT/CLARIAH-CZ 数字图书馆，查理大学形式与应用语言学研究所（ÚFAL）、数学与物理学院。[24] Safiya Umoja Noble. 2018. 《压迫算法：搜索引擎如何强化种族主义》（第1版）。纽约大学出版社，纽约。[25] Lily Ray。2020年。2020年谷歌搜索调查：用户对搜索结果的信任度有多高？https://moz.com/blog/2020-google-search-survey [26] Ronald E. Robertson、Shan Jiang、David Lazer 和 Christo Wilson。2019年。审计自动补全：建议网络与递归算法审查。载于第*10*届 *ACM* 网络科学会议论文集（*WebSci '19*）。ACM，纽约，美国， 235 244 10 1145 3292522 3326047 – 。https://doi.org/ 。/ [27] 2018 。Ronald E. Robertson 、 David Lazer 和 Christo Wilson。 。审计政治相关搜索引擎结果页面的个性化与构成。载于*2018*年万维网会议论文集（*WWW '18*）。国际万维网大会指导委员会，日内瓦共和国和州，CHE，955–965。https://doi.org/10.1145/3178876.3186143 活动地点：法国里昂。[28] Peter J. Rousseeuw。1987年。轮廓系数：一种用于解释和验证聚类分析的图形辅助工具。*J. Comput. Appl. Math.* 20（1987），53–65。https://doi.org/ 10 1016 0377 0427 87 90125 7 。/ - ( ) - Christian Sandvig、Kevin [29] Hamilton、Karrie Karahalios 和 Cedric Langbort。2014年。算法审计：用于检测互联网平台歧视的研究方法。数据与歧视：将关键关切转化为富有成效的探究 22（2014），1–23。[30] Peng Wang、Xianghang Mi、Xiaojing Liao、XiaoFeng Wang、Kan Yuan、Feng Qian 和 Raheem A Beyah。2018年。误导性建议之战：搜索自动补全操纵的语义分析。。载于*NDSS*。[31] Tamar Yehoshua。2016年。谷歌搜索自动补全。https://blog.google/ products/search/google-search-autocomplete/ [32] Katharina A. Zweig。2017年。观察观察者：Epstein 和 Robertson 的“搜索引擎操纵效应”。https://algorithmwatch.org/en/watching-the-watchers-epstein-and-robertsons-search-engine-manipulation-effect/ 报道方：AlgorithmWatch 图书馆目录：algorithmwatch.org 栏目：工作论文。

表4：作为根查询词的政治人物及其元数据。

Name	建议	聚类1	聚类2	簇3	性别	政党	出生年份	Role
亚历山大·高兰德	183	0.268	0.678	0.055	male	AfD	1941	前党魁
安德烈亚斯·博文舒尔特	48	0.188	0.708	0.104	male	SPD	1965	州长
安德烈亚斯·朔伊尔	270	0.27	0.637	0.093	male	CSU	1974	前部长
安格拉·默克尔	602	0.204	0.748	0.048	女性	CDU	1954	前部长
安雅·卡利切克	140	0.257	0.686	0.057	女性	CDU	1971	前部长
安娜莱娜·贝尔伯克	654	0.229	0.734	0.037	女性	绿党	1980	部长
安妮·施皮格尔	229	0.358	0.568	0.074	女性	绿党	1980	部长
Annegret Kramp-Karrenbauer	250	0.288	0.652	0.06	女性	CDU	1962	前部长
Armin Laschet	633	0.3	0.616	0.084	male	CDU	1961	前党魁
Bettina Stark-Watzinger	88	0.33	0.614	0.057	女性	FDP	1968	部长
博多·拉梅洛	259	0.29	0.602	0.108	male	左翼党	1956	州长
杰姆·厄兹代米尔	320	0.338	0.616	0.047	male	绿党	1965	部长
克里斯蒂安·林德纳	447	0.221	0.685	0.094	male	FDP	1979	部长
克里斯蒂娜·兰布雷希特	219	0.347	0.612	0.041	女性	SPD	1965	部长
Daniel Günther	325	0.243	0.566	0.191	male	CDU	1973	州长
Dietmar Woidke	73	0.342	0.616	0.041	male	SPD	1961	州长
Franziska Giffey	261	0.295	0.636	0.069	女性	SPD	1978	前部长
弗里德里希·默茨	458	0.365	0.563	0.072	male	CDU	1955	党领袖
格尔德·穆勒	742	0.102	0.756	0.142	male	CSU	1955	前部长
海科·马斯	439	0.339	0.629	0.032	male	SPD	1966	前部长
Helge Braun	310	0.332	0.603	0.065	male	CDU	1972	前部长
Hendrik Wüst	197	0.33	0.548	0.122	male	CDU	1975	州长
Horst Seehofer	268	0.25	0.683	0.067	male	CSU	1949	前部长
Hubertus Heil	286	0.357	0.549	0.094	male	SPD	1972	部长
Janine Wissler	231	0.364	0.528	0.108	女性	左翼党	1981	党领袖
Jens Spahn	613	0.219	0.688	0.093	male	CDU	1980	前部长
Julia Klöckner	280	0.304	0.632	0.064	女性	CDU	1972	前部长
卡尔·劳特巴赫	684	0.196	0.737	0.067	male	SPD	1963	部长
卡塔琳娜·巴利	152	0.289	0.671	0.039	女性	SPD	1968	前部长
克拉拉·盖维茨	89	0.371	0.596	0.034	女性	SPD	1976	部长
Lars Klingbeil	219	0.333	0.626	0.041	male	SPD	1978	党领袖
Malu Dreyer	345	0.214	0.675	0.11	女性	SPD	1961	州长
Manuela Schwesig	285	0.186	0.761	0.053	女性	SPD	1974	州长
马可·布施曼	185	0.303	0.638	0.059	male	FDP	1977	部长
马库斯·泽德	519	0.231	0.672	0.096	male	CSU	1967	州长
米夏埃尔·克雷奇默	236	0.233	0.568	0.199	male	CDU	1975	州长
南希·费瑟	201	0.388	0.587	0.025	女性	SPD	1970	部长
奥拉夫·朔尔茨	797	0.27	0.681	0.049	male	SPD	1958	部长
彼得·阿尔特迈尔	190	0.279	0.668	0.053	male	CDU	1958	前部长
彼得·琴彻	160	0.206	0.669	0.125	male	SPD	1966	州长
菲利普·罗斯勒	153	0.216	0.621	0.163	male	FDP	1973	前党魁
赖纳·哈泽洛夫	149	0.268	0.705	0.027	male	CDU	1954	州长
罗伯特·哈贝克	413	0.293	0.654	0.053	male	绿党	1969	部长
Saskia Esken	217	0.406	0.567	0.028	女性	SPD	1961	党领袖
Steffi Lemke	160	0.394	0.55	0.056	女性	绿党	1968	部长
Stephan Weil	227	0.269	0.551	0.181	male	SPD	1958	州长
Susanne Hennig-Wellsow	81	0.333	0.63	0.037	女性	左翼党	1977	党领袖
Svenja Schulze	207	0.391	0.531	0.077	女性	SPD	1968	部长
Tino Chrupalla	160	0.275	0.656	0.069	male	AfD	1975	党领袖
托比亚斯·汉斯	223	0.22	0.682	0.099	male	CDU	1978	州长
乌尔苏拉·冯德莱恩	533	0.139	0.841	0.021	女性	CDU	1958	前部长
福尔克·布菲耶	234	0.192	0.701	0.107	male	CDU	1951	州长
沃尔克·维辛	212	0.358	0.533	0.108	male	FDP	1970	部长
温弗里德·克雷奇曼	269	0.204	0.736	0.059	male	绿党	1948	州长
沃尔夫冈·施密特	347	0.179	0.274	0.548	male	SPD	1970	部长