クエリ形式とランキング手法が検索結果のスタンス分布に与える影響の分析

池元　太陽†　　山本　岳洋††

† 兵庫県立大学 社会情報科学部　〒 651–2197 兵庫県神戸市西区学園西町 8–2–1
†† 兵庫県立大学 大学院情報科学研究科　〒 651–2197 兵庫県神戸市西区学園西町 8–2–1
E-mail: †fa22t006@guh.u-hyogo.ac.jp, ††t.yamamoto@sis.u-hyogo.ac.jp

あらまし　本研究では，賛否両論あるトピックに関するウェブ検索において，クエリ形式の違いとランキング手法が検索結果の偏りに与える影響を分析した．具体的には，「学校に制服は必要か」といった多様な意見が存在するトピックを使用し，質問クエリとキーワードクエリに分けられるクエリ形式と，特性の異なる 3 つのランキング手法（BM25，e5，PRP）が，検索結果の偏りを示すスタンス分布にどのような影響を及ぼすかについて分析した．分析の結果，クエリ形式の違いによって検索結果におけるスタンス分布が変化することを示したが，その傾向はランキング手法によって異なることが確認された．また，検索有効性が高いランキング手法は，クエリ形式の違いによる影響を受けにくい一方で，検索結果の偏りを強める可能性があることを示した．本研究で得られた知見は，検索システムの評価において，検索有効性のみならず，クエリの表現の違いやランキング手法の選択によって生じる検索結果の偏りを考慮することの重要性を提示するものである．

キーワード　情報検索，スタンス分布，検索モデル，クエリ形式

1　はじめに

賛否両論あるトピックにおいて，ウェブ検索は，ユーザが多角的な視点から情報を収集し，責任ある意見形成を行うための重要な手段として機能している．例えば，「学校に制服は必要か」，「死刑制度は廃止すべきか」といったトピックには，1 つの明確な正解が存在せず，多様な価値観や意見が対立し，継続的に議論されている．このようなトピックに対して，ユーザはウェブ検索を通じて，様々な主張や根拠に触れ，それらを批判的に評価することで自身の意見を形成していくことが望まれる [19].

一方で，検索結果の偏りがユーザの意見形成や態度変容に強い影響を与えることが，これまでの研究により指摘されている [8,9,17]．検索結果の上位に提示された文書の賛否のスタンスに偏りがある場合，それを受け取ったユーザは偏った意見を形成する可能性がある．本研究では，このような検索結果の偏りを捉える指標として，スタンス分布に着目する．スタンス分布とは，検索結果として提示された個々の文書が，そのトピックに対してどのような立場をとっているか（賛成・中立・反対など）の割合を示すものである．この分布が特定のスタンスに偏っている場合，ユーザが接触する視点が限定され，多角的な情報収集が阻害される可能性がある．したがって，スタンス分布は，検索結果がユーザの意見形成に与える影響を定量的に評価するための有効な指標であると考えられる．

検索結果のスタンス分布は，ユーザの検索行動によって異なる可能性がある．既存研究では，同じ情報要求を持つユーザ間でも，入力するクエリの表現は異なること [1,2,22,27]，及びその表現の違いが検索結果やランキングに大きな変化をもたらすことが示されている [1,2,12,27]．また，ユーザが利用する検索エンジンやシステムによって検索結果を提示するためのランキング手法も異なることが想定され，それらが検索性能および検索結果を変化させることが示されている [23,28]．しかし，クエリの表現の違いやランキング手法の選択が検索結果のスタンス分布に与える影響については検証されていない．

そこで，本研究では，クエリの表現方法の違いの 1 つであるクエリ形式と，アドホック検索において代表的な 3 つのランキング手法に注目し，それらが検索結果のスタンス分布に与える影響を調査する．これにより，賛否両論あるトピックに関する情報収集におけるバイアスへの理解を深め，ユーザの責任ある意見形成を支援するための，偏りの少ない検索結果の提示に寄与することを目指す．

上記の目的を達成するため，以下の 2 つのリサーチクエスチョン（RQ）を設定する：

- **RQ1**：同じ情報要求でも，クエリ形式が違えばスタンス分布は異なるのか？
- **RQ2**：同じクエリでも，ランキング手法が違えばスタンス分布は異なるのか？

これらのリサーチクエスチョンに答えるため，本研究では以下の実験を実施した．まず，賛否両論あるトピックを扱った既存のデータセットを用いて，質問形式のクエリとキーワード形式のクエリを，賛成・中立・反対のスタンス別に大規模言語モデル（LLM）を用いて生成した．次に，生成したクエリを用いて，3 つのランキング手法（BM25，e5，LLM を用いたペアワイズランキングプロンプティング（PRP））それぞれで検索結果を求めた．そして，検索結果からスタンス分布を求め，クエリ形式とランキング手法が検索結果のスタンス分布に与える影

響を定量的に評価した．

実験の結果，以下の知見が得られた：

1. 同じ情報要求でも，クエリ形式の違いによってスタンス分布は異なることを示した．また，ランキング手法によって，クエリ形式の違いによるスタンス分布の差異の大きさと影響の性質は異なることを示した．
2. 同じクエリでも，ランキング手法の違いによって検索結果のスタンス分布は異なることを示した．特に検索有効性の高いランキング手法で，クエリのスタンスが検索結果に強く反映される傾向が見られた．
3. クエリ形式やランキング手法に関わらず，文書コーパスの偏りは検索結果のスタンス分布に影響を与えることを示した．特に BM25 は，文書コーパスのスタンス分布に近似しやすく，コーパスの偏りの影響を受けやすいランキング手法であることを示した．

本研究の主な貢献は，同一の情報要求であっても，クエリ形式の違いとランキング手法の選択によって，検索結果のスタンス分布が変化することを定量的に明らかにしたことである．

2 関 連 研 究

2.1 検索結果とユーザの意見形成

検索結果はユーザの意見形成および態度変容に強い影響力を持つことが多くの研究で示されている [8,9,17]．Epstein ら [9] は，ウェブ検索を用いた有権者の投票に関する意思決定において，偏った検索結果がユーザの態度や信念を変化させる検索エンジン操作効果（Search Engine Manipulation Effect，SEME）と呼ばれる現象が確認され，検索結果の偏りがユーザの意見形成に影響を及ぼすことを示した．そして，検索エンジン操作効果の要因として，Epstein らは，上位にランキングされているページの意見をより重視する順序効果が影響していると分析したが，後の Draws ら [6] の研究で，特定の視点（一方のスタンス）を支持する文書を閲覧するほど，その視点を受け入れやすくなるという接触効果に起因する可能性が高いことが示唆されている．

2.2 賛否両論あるトピックにおける検索行動と検索結果のバイアス

賛否両論あるトピックに関する検索結果において，検索行動および検索結果の提示過程に内在する複数のバイアスが，間接的にユーザの意見形成に影響を及ぼすことが示されている．具体的には，ユーザが自身の事前信念に合致する情報を優先的に選択する確証バイアスに加え，ユーザの満足度最大化を目的として設計された検索エンジンのランキングアルゴリズムによるランキングバイアス，さらにユーザが検索結果の上位項目を優先的に選択するポジションバイアスが相互に作用することで，検索エンジン操作効果を引き起こし，ユーザの多角的な視点に基づいた意見形成を阻害する可能性が指摘されている [19]．

特に，ユーザは自身の政治的信条などの事前信念に合致する検索クエリを選択する傾向があり，そのようなクエリを検索エンジンに入力することで，事前信念に偏った検索結果が提示されやすくなる．この現象は，ユーザの事前信念をさらに強化する方向に作用することが報告されている [7,22]．同様に，ユーザの事前信念は入力するクエリの文言に反映されることが知られており，クエリ自体に事前信念との明確な意味的関連性が見られない場合であっても，検索結果としてはユーザの信念や態度に合致した情報が提示されることが明らかにされている．その結果，対立する態度を持つ個人間で受け取る情報が大きく異なることが報告されている [11]．

さらに，賛否両論あるトピックを対象とした議論検索においては，クエリへの適合性を重視した高い検索有効性を持つランキング手法が，情報の公平性を犠牲にし，検索結果におけるスタンス分布の不均衡を助長する可能性があることが示されている [16]．このようなランキングの偏りと確証バイアスが結びつくことで，ユーザは検索結果を十分に精査しなくなり [14]，事前信念と一致する情報のみを選択的に閲覧するため，信念が補強される傾向が強まることが示唆されている [25]．

2.3 クエリ形式が検索有効性に与える影響

同じ情報要求であっても様々なクエリの表現方法があり，ユーザによって入力されるクエリは異なるが [1,2,22,27]，本研究では，異なるクエリの表現としてクエリ形式に注目する．クエリ形式は質問クエリとキーワードクエリの 2 種類に大別される [26]. キーワードクエリは，従来から主要なクエリ形式として広く用いられてきたが [26]，近年では質問クエリの利用が増加傾向にある [5]．既存研究では，一般的な検索タスクにおいて，質問クエリとキーワードクエリの検索有効性の差異に統計的な有意差は確認されなかったと報告している [26]．

しかし，特定のタスクにおいて，クエリ形式が検索有効性に影響を与えることが示されている．松田ら [28] は，多言語検索において，クエリ形式の違いが検索有効性に与える影響について分析した．語彙ベースの検索モデルである BM25 では，その影響は限定的であったが，DPR [13] などの意味ベースの検索モデルにおいて，質問クエリの方がキーワードクエリよりも高い検索有効性を示す傾向にあることを明らかにした．また，Wang ら [24] は複数の情報を統合する必要がある複雑な検索タスクにおいて，キーワードクエリの方が質問クエリよりも検索有効性が高く，単一の情報を要求する単純な検索タスクにおいては，質問クエリの方がキーワードクエリよりも検索有効性が高い傾向にあることを示し，タスクの複雑性がクエリ設計の有効性に影響を及ぼすことを報告した．

2.4 ランキング手法が検索有効性に与える影響

既存研究では，語彙ベースの検索から意味ベースの検索，さらに LLM を用いたリランキングへと至るランキング手法の進展が，検索有効性に与える影響について多角的に検討されてきた．BM25 は，語彙の一致に基づくランキング手法として，高い効率性と解釈性を備え [20]，多様な検索タスクにおけるベースラインとして広く利用されている．一方で，意味ベースの密検索モデルは，Transformer に基づく埋め込み表現を用いるこ

とで，語彙の不一致により関連文書が検索されない問題を克服し，BM25 を上回る検索性能を発揮する可能性があることが示唆されている [23].

さらに，BM25 や密検索モデルによって取得された文書候補に対し，LLM を直接リランカーとして適用する手法は，クロスエンコーダやハイブリッド検索などの高性能な手法を上回る検索有効性を達成しており，LLM の高度な文脈理解能力がリランキングにおいて有効であることが示唆されている [4].

また，デジタル化されていない歴史的文書のように，ノイズや複雑な文書構造を含む多数の文書を対象とした検索タスクにおいては，意味ベースの検索モデルよりも BM25 が高い性能を示す傾向が確認されており，検索の対象となるデータの特性に応じて適切なランキング手法を選択する重要性が指摘されている [10].

そこで，本研究では，語彙ベースの手法として BM25，意味ベースの手法として e5，および LLM を直接ランキングモデルに適用する手法としてペアワイズランキングプロンプティング（PRP） [18] の 3 つのランキング手法を用いる．

2.5 本研究の位置付け

このように，クエリ形式やランキング手法が検索有効性にどう影響するかについては様々なタスクで検証されているが，それらが賛否両論あるトピックにおける検索結果のスタンス分布に与える影響については検証されていない．そこで，本研究では質問クエリとキーワードクエリという 2 つのクエリ形式がスタンス分布に与える影響と，検索モデルとして一般的に用いられ，高い検索有効性を持つ 3 つのランキング手法がスタンス分布に与える影響を定量的に評価して分析する．

3 実験方法

3.1 実験の概要

リサーチクエスチョンを明らかにするため，本研究では以下の手順で実験を行った．まず，賛否両論あるトピックを扱った既存のデータセットを用いて，質問クエリとキーワードクエリからなるクエリペアを，賛成・中立・反対のスタンス別に 30 件ずつ LLM を用いて生成した．次に，生成したクエリペアを用いて，データセットの文書コーパスを対象に，3 つの手法で文書をランキングした．そして，得られたランキング上位 k 件のスタンス分布を求め，クエリ形式とランキング手法による影響を複数の指標で評価した．

3.2 データセット

本研究の実験には，Draws ら [6] の研究において，検索結果のスタンスの偏りがユーザの態度にもたらす影響を分析するために構築された，賛否両論あるトピックに関するデータセットを用いた．表 1 にデータセットの統計情報を示す．このデータセットでは，ユーザの意見が特定の立場に偏っていないことと，確信度の低い意見が多数を占め，論争の余地があることをトピックの選定基準として，“Is obesity a disease?” といった健康課題に関するトピックや “Should bottled water be banned?” といった環境問題に関するトピックなど，学術的かつ賛否両論あるトピックが 5 つ選定されている．

表 1 データセットの全トピックと各トピックにおけるスタンス別文書数（表 2 に示すスタンス評価値が+1 から+3 の文書を賛成文書，0 の文書を中立文書，-1 から-3 の文書を反対文書として集計）．

トピック	文書数		
	賛成	中立	反対
Is obesity a disease?	24	13	19
Is cell phone radiation safe?	12	11	33
Should bottled water be banned?	26	9	21
Should zoos exist?	22	6	28
Are social networking sites good for our society?	20	14	22

表 2 7 段階のリッカート尺度によるスタンスの評価値とラベル，及びラベルを具体的に表す例文（文献 [6] の表 1 を引用）．

Viewpoint Label		Example(Topic:“Should Zoos Exist?”)
+3	strongly supporting	“There is nothing wrong with zoos!”
+2	supporting	“I‘m in favor of zoos, let’s keep them.”
+1	somewhat supporting	“Zoos are not great, but they benefit society.”
0	neutral	“We present arguments for and against zoos.”
-1	somewhat opposing	“Despite some benefits, I’m against zoos.”
-2	opposing	“We should strive towards closing all zoos.”
-3	strongly opposing	“Horrible places! All zoos should be closed.”

各トピックに 56 件の文書からなるコーパスがあり，各文書にはクラウドワーカによって二値の適合性と，トピックに対する見解（スタンス）が「強く支持」から「強く反対」までの 7 段階のリッカート尺度で与えられている．文献 [6] に記載されているスタンスの評価基準とその例を表 2 に示す．また，不適合文書におけるスタンスの評価値は 0 である．しかし，不適合文書は各トピックのコーパスで 0 から 2 件と非常に少ないため，本研究の実験において不適合文書が検索結果に与える影響は限定的であり，主に適合文書間の順序関係に着目したリランキングの問題に近い性質を持つ．

表 1 から，大半のトピックのコーパスが，約 20 件の賛成文書と反対文書，及び約 10 件の中立文書で構成されていることが分かる．しかし，“Is Cell Phone Radiation Safe?” というトピックのみ，反対文書数が賛成文書数より多く，他のトピックと相対的に比較しても，偏りのあるコーパスであることが分かる．このような文書コーパスの偏りが検索結果のスタンス分布に及ぼす影響について，4 節の実験結果で議論する．

3.3 クエリの生成

本研究では，データセットのトピックから，質問クエリとキーワードクエリで構成されるクエリペアを LLM で生成した．表 3 に生成したクエリペアの一例を示す．例は使用するデータセットにおけるトピック “Should bottled water be banned?” に対する生成クエリペアである．

既存研究において，ユーザは事前信念に沿った検索クエリを選択する傾向にあり，選択したクエリを検索エンジンで使用することで，事前信念を補強するような検索結果が表示される傾向にあることが確認されているため [7, 11, 22]，ユーザの検索行動を考慮して，クエリペアを賛成・中立・反対のスタンス別

表 3　GPT-4o を用いて生成したクエリペアの例（トピック：“Should bottled water be banned?”）.

	質問クエリ	キーワードクエリ
賛成クエリ	What are the environmental benefits of banning bottled water? How does banning bottled water reduce plastic waste?	Environmental benefits of banning bottled water Banning bottled water and plastic waste reduction
中立クエリ	What are the environmental impacts of bottled water? What are the health benefits and risks of bottled water?	Environmental impacts of bottled water Health benefits and risks of bottled water
反対クエリ	Why shouldn't bottled water be banned? What are the benefits of bottled water?	Reasons to keep bottled water Advantages of bottled water

に複数生成した．本節では，クエリペアの生成方法と生成されたクエリペアの品質について説明する．

3.3.1　クエリペアの生成

クエリペアの生成には，OpenAI の GPT-4o (2024-11-20)[1] を用いた．生成する際の temperature は 0 に設定した．そして，データセットの各トピックについて，賛成・中立・反対の 3 つのスタンス別に，質問クエリとキーワードクエリからなるクエリペアを 30 件ずつ生成した．

例として，クエリペアを生成する際に与えるプロンプトを以下に示す．プロンプトは，クエリの例を数件提示する Few-shot プロンプトであり，与えるクエリの例は生成するクエリのスタンスによって異なる．また，赤字の stance には，賛成クエリを生成する際は supportive，中立クエリを生成する際は neutral，反対クエリを生成する際は opposing を指定する．

> You are an AI assistant that generates diverse search queries based on users' information needs. For the given query, please infer the user's information need and then generate 30 **pairs** of search queries that a user with a **stance** stance might formulate.
> Each pair must consist of one **Question Query** (e.g., "Should we adopt X?", "What is the best way to achieve Y?") and one related **Keyword Query** (e.g., "X benefits," "Y advantages," "reasons to support Z"). The two queries in each pair must seek **exactly the same information**.
>
> ---
>
> Query: {original_query}
>
> ---
>
> ### Output Format
> Please return the generated queries as a comma-separated list of 30 pairs. Each pair should be formatted as: **["Question Query", "Keyword Query"]**.
> Example:
> [["Question Query 1", "Keyword Query 1"], ["Question Query 2", "Keyword Query 2"], ..., ["Question Query 30", "Keyword Query 30"]]

Bailey らの研究 [2] では，クラウドワーカを利用した多様なクエリの収集の際に，事前に作成した情報要求を与え，その情報要求に応えるために検索エンジンで使用するクエリを入力することを求めた．本研究で使用するプロンプトではそれらのクエリ収集プロセスを参考にして，与えられたトピックに対する情報要求を考え，各スタンスを支持するユーザが作成，入力すると考えられる検索クエリを生成するように指示した．そして，生成するクエリペアが，それぞれのペアで同じ情報要求を持つことを強調した．

表 4　生成したクエリペアの意味的類似度の平均（標準偏差）.

	平均	最大	最小
賛成クエリペア	0.880(0.050)	0.972	0.743
中立クエリペア	0.914(0.041)	0.978	0.775
反対クエリペア	0.887(0.038)	0.967	0.782

3.3.2　クエリペアの意図保持性の評価

生成したクエリペアが同じ情報要求を満たすことを確認するため，英語テキストの高性能な埋め込みモデル `BAAI/bge-large-en-v1.5`[2]を用いて，質問クエリとキーワードクエリのコサイン類似度を算出した．既存研究では，形式の異なるクエリが同じ情報要求を満たしているかを測る意図保持性の評価に Universal Sentence Encoder [3] や OpenAI の text-embedding-3-large[3] などのモデルが用いられているが [12, 28]，本研究では検索タスク向けに最適化された埋め込みモデルであり，情報検索分野において高い性能が報告されている点[4]を考慮し，bge-large-en-v1.5 を採用した．表 4 に算出したコサイン類似度の値を示す．

生成したクエリペアは，どのスタンスにおいてもコサイン類似度の平均値が 0.8 を超えており，最小値は 0.7 以上であった．このコサイン類似度の目安として，Iovine らの研究 [12] で，意図保持性を満たすコサイン類似度の閾値が定義されている．Iovine らは，質問クエリとキーワードクエリの間で双方向の書き換えを可能にする教師なしモデルの提案において，モデルの学習に使用するクエリペアの類似度の閾値を 0.6 としている．これは，ペアとなる質問クエリとキーワードクエリが類似したドメインを共有していることを満たす基準として適用された．また，モデルを評価するためのテストデータとして使用するクエリペアの類似度の閾値を 0.8 としている．これは，同じ情報

1：https://platform.openai.com/docs/models/gpt-4o?snapshot=gpt-4o-2024-11-20

2：https://huggingface.co/BAAI/bge-large-en-v1.5

3：https://platform.openai.com/docs/models/text-embedding-3-large

4：https://huggingface.co/spaces/mteb/leaderboard(2026 年 2 月 2 日閲覧)

要求を満たし，意味的にほとんど等しいことを示す，より厳格な基準として適用された．本研究で生成したクエリペアは，コサイン類似度の最小値が 0.6 より大きく，平均値も 0.8 より大きいため，既存研究で定義されている意図保持性の基準を概ね満たしているといえる．

3.4 ランキング手法

本研究では，BM25，e5，PRP の 3 つのランキング手法を使用した．

3.4.1 BM25

BM25 [20] は，クエリと文書の単語の一致に基づくランキング手法であり，文書内の単語頻度（TF）と逆文書頻度（IDF）を用いて適合性を評価する．本研究では，語彙ベースの疎検索モデルとして BM25 を用いる．

3.4.2 e5

e5 [23] は，情報検索のタスクに特化して学習されたテキスト埋め込みモデルであり，入力されたクエリや文書のテキストを意味的な特徴を捉えた高次元のベクトルに変換する．本モデルは対照学習を用いて，クエリと適合する文書のベクトルは近くに，適合しない文書のベクトルは遠くなるような埋め込み空間を構築する．本研究では，語彙一致に依存しない意味ベースの密検索モデルとして `intfloat/e5-base-v2`[5]を用いる．

3.4.3 ペアワイズランキングプロンプティング（PRP）

PRP [18] は，LLM に 2 つの文書を提示し，どちらがクエリに対してより適合性が高いかを判断させる手法である．この手法では，LLM に与えるプロンプトを工夫することで，モデルが文書間の相対的な優劣を比較し，ランキングに必要なスコアを生成する．本研究では，LLM を直接的にランキングモデルとして利用するためのアプローチとして，PRP を用いる．以下に PRP で使用するペアワイズランキングのプロンプトを示す．プロンプトは既存研究 [18] を参考にし，GPT-4o-mini (2024-06-01)[6]を LLM として使用した．

> Given a query {query}, which of the following two passages is more relevant to the query?
> Passage A: {document1}
> Passage B: {document2}
> Output Passage A or Passage B:

LLM はプロンプト内のテキスト順序に敏感であるため [15]，PRP では，文書の順序を入れ替えて，2 回問い合わせを行うことで，入力順序に対するバイアスを低減させている．各クエリにおいて，コーパス内の全ての組み合わせで PRP を行ってスコアを割り当て，合計したスコアが高い順にソートしてランキングする．

3.5 評価指標

3.5.1 スタンス分布

本研究では，検索結果の偏りを定量的に評価するために，検索結果における賛成文書・中立文書・反対文書の割合を示すスタンス分布を求める．具体的には，ランキング上位 k 件において，スタンス評価値が $+1$ から $+3$ の文書を賛成文書，0 の文書を中立文書，-1 から -3 の文書を反対文書として，賛成・中立・反対のスタンス別文書数の割合を計算する．

3.5.2 Root Normalized Order-aware Divergence(RNOD)

本研究では，質問クエリによって得られたスタンス分布とキーワードクエリによって得られたスタンス分布の差異を評価する指標として RNOD を用いる．RNOD [21] は，カテゴリ間に定義された順序構造を考慮しながら 2 つの確率分布間の差異を測定する距離指標である．RNOD の値が小さいほど，2 つの分布がより類似していることを示す．

RNOD の特徴は，カテゴリ間の距離を考慮して分布間の差異を評価できる点にある．たとえば，上位 5 件が全て賛成文書である検索結果 A，上位 5 件が全て中立文書である検索結果 B，上位 5 件が全て反対文書である検索結果 C を比較する．このとき，賛成文書が全て中立文書に置き換わっている A と B よりも，賛成文書がすべて反対文書に置き換わっている A と C の方を，分布間の差異として大きいものとして評価できる．検索結果間のスタンス分布の差異を測定するため，本研究では，賛成，中立，反対を区別して評価可能な RNOD を採用した．

本研究では，賛成，中立，反対という 3 つの要素と順序からなるカテゴリ集合 $C = \{\text{賛成}, \text{中立}, \text{反対}\}$ を考える．質問クエリによって得られた上位 k 件のスタンス分布を $P = (p_{\text{賛成}}, p_{\text{中立}}, p_{\text{反対}})$ と表す．例えば，質問クエリによる上位 5 件の検索結果において，賛成文書が 3 件，中立文書が 1 件，反対文書が 1 件出現していれば，$P = (p_{\text{賛成}} = 0.6,\ p_{\text{中立}} = 0.2,\ p_{\text{反対}} = 0.2)$ となる．同様に，キーワードクエリによって得られた上位 k 件のスタンス分布を $P^* = (p^*_{\text{賛成}}, p^*_{\text{中立}}, p^*_{\text{反対}})$ と表す．このとき，2 つのスタンス分布 P と P^* の RNOD は以下の式で計算される：

$$\mathrm{RNOD}(P \parallel P^*) = \sqrt{\frac{\mathrm{OD}(P \parallel P^*)}{|C| - 1}} \tag{1}$$

ここで，$\mathrm{OD}(P \parallel P^*)$ は以下の式で定義される：

$$\mathrm{OD}(P \parallel P^*) = \frac{1}{|C|} \sum_{i \in C} \mathrm{DW}_i \tag{2}$$

$$\mathrm{DW}_i = \sum_{j \in C} \delta_{ij} (p_j - p^*_j)^2, \qquad \delta_{ij} = |i - j| \tag{3}$$

ただし，δ_{ij} はカテゴリ間の距離を表し，本研究では 賛成$= 0$, 中立$= 1$, 反対$= 2$ とした．例えば，$\delta_{\text{賛成反対}} = |0 - 2| = 2$ となる．

4 実験結果

本節では，本研究の実験結果として，まず，RQ1 に関する結果を示し，次に，RQ2 に関する結果を示す．

4.1 RQ1：同じ情報要求でも，クエリ形式が違えばスタンス分布は異なるのか？

本節では，クエリ形式の違いがスタンス分布に与える影響を

5：https://huggingface.co/intfloat/e5-base-v2
6：https://platform.openai.com/docs/models/gpt-4o-mini

表 5　ランキング上位 5 件のスタンス分布（%）.

		BM25			e5			PRP		
		賛成文書	中立文書	反対文書	賛成文書	中立文書	反対文書	賛成文書	中立文書	反対文書
質問クエリ	賛成クエリ	37.9	18.3	43.9	39.2	25.6	35.2	63.9	16.0	20.1
	中立クエリ	38.7	16.1	45.2	31.9	23.6	44.5	45.7	18.3	36.0
	反対クエリ	36.9	13.5	49.6	22.8	23.2	54.0	15.7	13.1	71.2
キーワードクエリ	賛成クエリ	39.2	20.7	40.1	42.8	24.4	32.8	63.3	16.1	20.5
	中立クエリ	34.4	24.3	41.3	33.5	21.3	45.2	43.9	19.2	36.9
	反対クエリ	35.5	20.9	43.6	24.5	17.3	58.1	16.7	13.9	69.5

表 6　クエリペアにおける質問クエリ上位 5 件のスタンス分布とキーワードクエリ上位 5 件のスタンス分布の RNOD の平均（標準偏差）.

	BM25	e5	PRP
賛成クエリ	0.236(0.162)	0.213(0.149)	0.096(0.115)
中立クエリ	0.214(0.138)	0.159(0.130)	0.109(0.124)
反対クエリ	0.242(0.134)	0.174(0.138)	0.124(0.172)
平均	0.231(0.145)	0.182(0.139)	0.110(0.137)

ランキング手法別に分析する．表 5 に各ランキング手法における上位 5 件のスタンス分布を示す．そして，表 6 に各ランキング手法の，質問クエリにおける上位 5 件のスタンス分布とキーワードクエリにおける上位 5 件のスタンス分布の RNOD の平均値を示す．また，実験はランキング結果上位 3 件，上位 5 件，上位 10 件でスタンス分布を求めて評価したが，どの件数においても同様の傾向が得られた．

4.1.1　BM25 におけるクエリ形式の影響

表 5 に示す BM25 の結果において，質問クエリでは中立文書の割合が低く (賛成クエリ：18.3%，中立クエリ：16.1%，反対クエリ：13.5%)，キーワードクエリで高い（賛成クエリ：20.7%，中立クエリ：24.3%，反対クエリ：20.9%) 傾向が観察された．この結果は，BM25 が語彙の一致に基づいたランキング手法であることに起因すると考えられる．質問クエリには機能語や文の構造を含む表現が多く含まれるため，それらと語彙的に強く対応する文書が上位にランク付けされやすくなる．その結果，賛成・反対といった立場を示す語彙を多く含む文書が上位にランキングされたと考えられる．一方，キーワードクエリでは話題や対象を表す語が中心となるため，特定の立場を示さずに事実や背景を説明する文書とも一致しやすく，中立的な文書が相対的に多く検索されたと考えられる．

また，表 6 の RNOD の結果から，BM25 は本研究で用いた 3 つのランキング手法の中で RNOD の平均値が最も大きく（平均：0.231），クエリ形式の違いによるスタンス分布の差異が大きいことが確認された．これは，語彙ベースのランキング手法がクエリ表現の表層的な違いに敏感であるという既存研究の知見 [1,2,28] とも一致しており，BM25 においては同一の情報要求であっても，クエリ形式の違いが検索結果のスタンス分布に影響を及ぼす可能性が高いことを示している．

4.1.2　e5 におけるクエリ形式の影響

表 5 に示す e5 の結果において，質問クエリでは中立文書の割合が高く（賛成クエリ：25.6%，中立クエリ：23.6%，反対クエリ：23.2%），キーワードクエリでは中立文書の割合が低い（賛成クエリ：24.4%，中立クエリ：21.3%，反対クエリ：17.3%）傾向が観察された．この傾向は BM25 における影響とは異なるものである．また，特徴的な傾向として，賛成クエリにおいては賛成文書の割合が高く（質問クエリ：39.2%，キーワードクエリ：42.8%），反対クエリにおいては反対文書の割合が高い（質問クエリ：54.0%，キーワードクエリ：58.1%）というクエリのスタンスに応じた偏りが，キーワードクエリで強まる傾向が観察された．この結果は，e5 ではキーワードクエリが，クエリの潜在的なスタンスをより強調し，スタンスの偏りを増幅させる可能性があることを示唆している．

また，表 6 に示す RNOD の結果から，e5 の RNOD の平均値は BM25 よりも小さい値を示し（平均：0.182），クエリ形式の違いによるスタンス分布の差異が BM25 よりも小さいことが確認された．これは e5 が語彙の一致に依存せず意味的類似性に基づいて検索を行うため，クエリの表現の違いによる影響を緩和していることを示す．

4.1.3　PRP におけるクエリ形式の影響

表 5 に示す PRP の結果において，質問クエリによるスタンス分布とキーワードクエリによるスタンス分布で同様の傾向を示した．また，表 6 に示す RNOD の結果から，PRP は本研究で用いた 3 つのランキング手法の中で RNOD の平均値が最も小さく（平均：0.110），クエリ形式の違いによるスタンス分布の差異が最も小さい手法であることが示された．これらの結果は，PRP がクエリと文書を同時に入力するペアワイズ比較に基づく手法であり，クエリ形式の差異よりも，クエリが示す本質的な立場や意図を優先して評価していることを示唆する．

4.2　RQ2：同じクエリでも，ランキング手法が違えばスタンス分布は異なるのか？

本節では，表 5 の結果から，ランキング手法の違いがスタンス分布に与える影響を分析する．まず，BM25 においては，e5 や PRP と比べて，スタンス分布に大きな偏りは見られなかった．質問形式の反対クエリにおいては反対文書の割合がやや高い傾向にあったが（質問・反対クエリ：0.496），全体として，賛成文書と反対文書が比較的均等に分布している．これは，BM25

が単語の一致に基づく手法であり，文脈やスタンスといった意味的特徴を十分に捉えることができないため，クエリのスタンスに対する感度が低いことに起因すると考えられる．この特性は，BM25 が多様な視点を含む検索結果を比較的維持しやすい一方で，特定の立場に基づく情報探索には適さない可能性を示している．

一方で，e5 ではクエリのスタンスに応じた分布の偏りが観察された．具体的には，賛成クエリでは賛成文書の割合が高く（質問クエリ：0.392，キーワードクエリ：0.428），反対クエリでは反対文書の割合が高まる傾向を示した（質問クエリ：0.540，キーワードクエリ：0.581）．e5 は，高次元の埋め込み空間を用いて，クエリと文書の意味的特徴を捉えることが可能なため，クエリが持つ潜在的なスタンスに感応し，類似したスタンスを持つ文書を上位にランク付けしたと考えられる．この結果は，意味ベースの検索モデルが検索有効性を向上させる一方で，ユーザの立場を強化する方向に検索結果のスタンスが偏る可能性を持つことを示唆している．

さらに PRP では，その傾向が顕著であり，賛成クエリでは賛成文書の割合が 6 割以上を占め（質問クエリ：0.639，キーワードクエリ：0.633），反対クエリでは反対文書の割合が 7 割程を占めた（質問クエリ：0.712，キーワードクエリ：0.695）．これは，LLM が単なる意味的類似性ではなく，クエリと文書の間のスタンスの整合性や論調の一致を重視してランキングを行っている可能性を示している．既存研究 [18] が指摘するように，LLM は高度な文脈理解能力を持つ一方で，クエリに含まれる前提や価値観を強く反映した検索結果を生成・選択する傾向があり，本研究の結果はその特性が検索結果のスタンス分布にも表れていることを示している．

4.3 文書コーパスの偏りがスタンス分布に与える影響

本節では，3.3 節で言及した文書コーパスの偏りが検索結果のスタンス分布に与える影響について検討する．表 7 にランキング上位 5 件におけるトピック別のスタンス分布を示す．また，表 8 に各トピックにおけるデータセット全体の文書コーパスのスタンス分布とランキング上位 5 件のスタンス分布の RNOD の平均を示す．RNOD の値が小さいほど，ランキング結果から得られたスタンス分布がコーパスのスタンス分布に類似していることを示す．まず，表 7 の結果において，コーパスにおける反対文書の割合が高い “Is cell phone radiation safe?” というトピックのみ，クエリのスタンスやランキング手法に関わらず，検索結果の反対文書の割合が高い傾向が確認された．これは，文書コーパスの偏りがクエリの表現やランキング手法に関わらず，検索結果の偏りを増強することを示唆する．

また，文書コーパスのスタンス分布と各ランキング手法で上位 5 件のスタンス分布の差異を定量的に評価した表 8 の結果において，3 つのランキング手法のうち，BM25 における RNOD の平均が最も低い（BM25：0.218，e5：0.246，PRP：0.308）ことが確認された．これは BM25 を用いてランキングしたスタンス分布は，コーパスのスタンス分布に近似することを示し，BM25 がコーパスの偏りの影響を受けやすい手法であることを示唆する．

5 議　　論

本研究は，クエリ形式およびランキング手法の違いが検索結果のスタンス分布に与える影響を明らかにすることを目的として実施した．実験結果から，クエリ形式とランキング手法の双方が検索結果のスタンス分布を変化させることを示した．これにより，同一の情報要求であっても，クエリの形式やランキング手法によって，意図せず異なる検索結果が提示される可能性があることを明らかにした．ユーザが入力するクエリの表現や検索システムが採用するランキング手法によって，検索結果に偏りが生じる可能性を考慮する必要性があることを指摘した点は，本研究の重要な意義である．

RQ1 に関して，同一の情報要求であっても，質問クエリとキーワードクエリというクエリ形式の違いが，検索結果のスタンス分布に影響を与えることを明らかにした．BM25 はクエリの表層的な違いに敏感に反応し，クエリ形式の違いによるスタンス分布の差異が大きいことが示された．一方で，クエリと文書の意味的類似性を考慮できる e5 と PRP では，スタンス分布の差異が小さく，クエリの表現の揺らぎに対して頑健な手法であることが示された．また，クエリ形式の違いがスタンス分布に与える具体的な影響については，ランキング手法によって異なることが確認された．これらの結果から，ユーザの些細なクエリ表現の違いが，意図しない検索結果の変化を生む可能性があるといえる．

次に，RQ2 に関して，ランキング手法の違いがスタンス分布に与える影響は明確であった．クエリと文書の意味的類似性を考慮できる e5 や PRP でランキングしたスタンス分布は，クエリのスタンスに応じた検索結果の偏りが強まる傾向が確認された．これは，意味ベースや LLM ベースのランキング手法は，検索有効性は高い一方で，クエリに内在するスタンスを強く反映し，検索結果の偏りを強める可能性があるため，ユーザが受け取る情報の多様性を損なう危険性があることを示す．

さらに，本研究では，文書コーパスの偏りが検索結果のスタンス分布に与える影響について検証した．分析の結果から，検索結果のスタンス分布が，コーパスに内在する偏りの影響を強く受けることが確認された．実際に，コーパスにおける反対文書の割合が高いトピックでは，クエリ形式やランキング手法に関わらず，検索結果のスタンス分布における反対文書の割合が高まる傾向が確認された．また，実験で使用した 3 つのランキング手法のうち，BM25 を用いてランキングしたスタンス分布が，コーパスのスタンス分布に最も近似していたことから，BM25 のような単語の一致に基づく語彙ベースの検索手法は，コーパスの偏りの影響を受けやすい手法である可能性が示唆された．

以上の結果から，本研究は，検索結果のスタンス分布に生じる偏りが，クエリ形式，ランキング手法，及び文書コーパスの偏りという複数の要因の相互作用によって形成されることを示したといえる．

表 7　ランキング上位 5 件のトピック別スタンス分布（%）.

			BM25			e5			PRP		
			賛成文書	中立文書	反対文書	賛成文書	中立文書	反対文書	賛成文書	中立文書	反対文書
質問クエリ	賛成クエリ	Is obesity a disease?	48.0	24.0	28.0	46.0	16.7	37.3	78.0	21.3	0.7
		Is cell phone radiation safe?	8.7	11.3	80.0	8.7	25.3	66.0	17.3	16.7	66.0
		Should bottled water be benned?	26.0	19.3	54.7	26.0	34.0	40.0	52.7	34.0	13.3
		Should zooz exist?	56.7	14.0	29.3	74.0	8.7	17.3	94.7	5.3	0.0
		Are social networking sites good for our society?	50.0	22.7	27.3	41.3	43.3	15.3	76.7	2.7	20.7
	中立クエリ	Is obesity a disease?	57.3	15.3	27.3	38.7	26.7	34.7	78.0	16.0	6.0
		Is cell phone radiation safe?	8.0	8.7	83.3	12.7	25.3	62.0	11.3	16.0	72.7
		Should bottled water be benned?	37.3	18.0	44.7	32.7	30.7	36.7	43.3	34.7	22.0
		Should zooz exist?	47.3	17.3	35.3	53.3	6.7	40.0	59.3	10.7	30.0
		Are social networking sites good for our society?	43.3	21.3	35.3	22.0	28.7	49.3	36.7	14.0	49.3
	反対クエリ	Is obesity a disease?	53.3	17.3	29.3	28.0	18.7	53.3	44.0	11.3	44.7
		Is cell phone radiation safe?	10.0	10.0	80.0	10.0	22.7	67.3	0.7	9.3	90.0
		Should bottled water be benned?	40.0	14.0	46.0	32.0	32.7	35.3	26.0	32.0	42.0
		Should zooz exist?	42.0	7.3	50.7	30.0	6.0	64.0	2.7	0.0	97.3
		Are social networking sites good for our society?	39.3	18.7	42.0	14.0	36.0	50.0	5.3	12.7	82.0
キーワードクエリ	賛成クエリ	Is obesity a disease?	39.3	42.7	18.0	35.3	39.3	25.3	82.7	17.3	0.0
		Is cell phone radiation safe?	7.3	13.3	79.3	5.3	22.0	72.7	10.0	14.0	76.0
		Should bottled water be benned?	44.0	12.0	44.0	50.0	13.3	36.7	54.0	39.3	6.7
		Should zooz exist?	55.3	12.0	32.7	77.3	6.7	16.0	97.3	2.0	0.7
		Are social networking sites good for our society?	50.0	23.3	26.7	46.0	40.7	13.3	72.7	8.0	19.3
	中立クエリ	Is obesity a disease?	36.7	40.0	23.3	31.3	32.7	36.0	76.0	20.7	3.3
		Is cell phone radiation safe?	5.3	16.0	78.7	5.3	21.3	73.3	4.7	15.3	80.0
		Should bottled water be benned?	44.0	27.3	28.7	45.3	17.3	37.3	44.7	36.0	19.3
		Should zooz exist?	50.0	13.3	36.7	60.0	6.0	34.0	61.3	5.3	33.3
		Are social networking sites good for our society?	36.0	24.7	39.3	25.3	29.3	45.3	32.7	18.7	48.7
	反対クエリ	Is obesity a disease?	42.7	29.3	28.0	26.7	11.3	62.0	39.3	9.3	51.3
		Is cell phone radiation safe?	6.0	14.0	80.0	6.0	16.0	78.0	0.0	6.0	94.0
		Should bottled water be benned?	50.0	30.0	20.0	50.0	14.7	35.3	25.3	38.7	36.0
		Should zooz exist?	50.0	9.3	40.7	29.3	5.3	65.3	15.3	0.0	84.7
		Are social networking sites good for our society?	28.7	22.0	49.3	10.7	39.3	50.0	3.3	15.3	81.3

表 8　各トピックにおけるデータセットのコーパスのスタンス分布とランキング上位 5 件のスタンス分布の RNOD の平均（標準偏差）.

		BM25（標準偏差）	e5（標準偏差）	PRP（標準偏差）
質問クエリ	賛成クエリ	0.226（0.117）	0.264（0.097）	0.329（0.069）
	中立クエリ	0.216（0.102）	0.238（0.104）	0.279（0.093）
	反対クエリ	0.194（0.101）	0.231（0.098）	0.284（0.074）
キーワードクエリ	賛成クエリ	0.230（0.112）	0.268（0.082）	0.345（0.054）
	中立クエリ	0.216（0.100）	0.237（0.129）	0.298（0.088）
	反対クエリ	0.225（0.093）	0.235（0.098）	0.314（0.068）
平均		0.218（0.104）	0.246（0.101）	0.308（0.074）

しかし，本研究は，英語のトピックに限定された小規模なデータセットを対象としている点に限界がある．加えて，検索結果の評価は上位件数に基づいており，実際のユーザの検索行動や意見形成への影響までは直接的に検証していない．今後は，多言語および多文化環境における分析や，大規模なデータセットを対象とした検索実験，さらにはユーザ実験との統合を行っていく必要がある．

6 ま と め

本研究では，クエリ形式およびランキング手法の違いが，検索結果の偏りを示すスタンス分布に与える影響を調査した．実験の結果，検索有効性が高い意味ベースの検索モデルや LLM ベースのランキング手法で，クエリ形式の違いによる影響を抑制する傾向が確認されたが，クエリのスタンスに応じて検索結果の偏りが強まる傾向が観察された．また，文書コーパスの偏りが，クエリ形式やランキング手法に関わらず，検索結果に影響を与えることが確認され，特に BM25 などの語彙ベースの検索手法がその影響を受けやすい手法である可能性を示した．本研究は，検索有効性のみならず，スタンス分布を用いた検索結果の偏りの評価の重要性を示した．

謝　　辞

本研究は，JSPS 科研費 JP24K03228，JP25K03229 の助成を受けたものです．ここに記して謝意を表します．

文　　献

[1] Marwah Alaofi, Luke Gallagher, Dana Mckay, Lauren L. Saling, Mark Sanderson, Falk Scholer, Damiano Spina, and Ryen W. White. Where do queries come from? In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 2850–2862, 2022.

[2] Peter Bailey, Alistair Moffat, Falk Scholer, and Paul Thomas. UQV100: A test collection with query variability. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 725–728, 2016.

[3] Daniel Cer, Yinfei Yang, Sheng yi Kong, Nan Hua, Nicole Limtiaco, Rhomni St. John, Noah Constant, Mario Guajardo-Cespedes, Steve Yuan, Chris Tar, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Universal sentence encoder. *arXiv preprint arXiv:1803.11175*, 2018.

[4] Murilo Cunha, Marilia Silveira, Brenda Santana, Larissa Freitas, and Ulisses Corrêa. Optimizing and evaluating a retrieval-augmented generation system for normative document retrieval in hospital settings. In *Proceedings of the 31st Brazilian Symposium on Multimedia and the Web*, pp. 385–393, 2025.

[5] Brian Dean. We analyzed 306m keywords. here's what we learned about google searches. `https://backlinko.com/go`

ogle-keyword-study.

[6] Tim Draws, Nava Tintarev, Ujwal Gadiraju, Alessandro Bozzon, and Benjamin Timmermans. This is not what we ordered: Exploring why biased search result rankings affect user attitudes on debated topics. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 295–305, 2021.

[7] Axel G Ekström, Guy Madison, Erik J Olsson, and Melina Tsapos. The search query filter bubble: Effect of user ideology on political leaning of search results through query selection. *Information Communication and Society*, Vol. 27, No. 5, pp. 878–894, 2024.

[8] Robert Epstein and Ji Li. Can biased search results change people's opinions about anything at all? A close replication of the search engine manipulation effect (SEME). *Plos one*, Vol. 19, No. 3, e0300727, 2024.

[9] Robert Epstein and Ronald E. Robertson. The search engine manipulation effect (SEME) and its possible impact on the outcomes of elections. *Proceedings of the National Academy of Sciences*, Vol. 112, No. 33, pp. E4512–E4521, 2015.

[10] Haruki Fujimaki and Makoto P. Kato. KASYS at the NTCIR-18 SUSHI task. In *Proceedings of the 18th NTCIR Conference on Evaluation of Information Access Technologies*, pp. 422–455, 2025.

[11] Hussam Habib, Ryan Stoldt, Andrew High, Brian Ekdale, Ashley Peterson, Katy Biddle, Javie Ssozi, and Rishab Nithyanand. Algorithmic amplification of biases on google search. *arXiv preprint arXiv:2401.09044*, 2024.

[12] Andrea Iovine, Anjie Fang, Besnik Fetahu, Jie Zhao, Oleg Rokhlenko, and Shervin Malmasi. CycleKQR: Unsupervised bidirectional keyword-question rewriting. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 11875–11886, 2022.

[13] Vladimir Karpukhin, Barlas Oguz, Sewon Min, Patrick Lewis, Ledell Wu, Sergey Edunov, Danqi Chen, and Wen-tau Yih. Dense passage retrieval for open-domain question answering. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 6769–6781, 2020.

[14] Suzuki Masaki and Yusuke Yamamoto. Characterizing the influence of confirmation bias on web search behavior. *Frontiers in Psychology*, Vol. 12, No. 771948, pp. 1–11, 2021.

[15] Anna Neumann, Elisabeth Kirsten, Muhammad Bilal Zafar, and Jatinder Singh. Position is power: System prompts as a mechanism of bias in large language models (LLMs). In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, pp. 573–598, 2025.

[16] Sachin Pathiyan Cherumanal, Damiano Spina, Falk Scholer, and W. Bruce Croft. Evaluating fairness in argument retrieval. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pp. 3363–3367, 2021.

[17] Frances A. Pogacar, Amira Ghenai, Mark D. Smucker, and Charles L.A. Clarke. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *Proceedings of the ACM SIGIR International Conference on Theory of Information Retrieval*, pp. 209–216, 2017.

[18] Zhen Qin, Rolf Jagerman, Kai Hui, Honglei Zhuang, Junru Wu, Le Yan, Jiaming Shen, Tianqi Liu, Jialu Liu, Donald Metzler, Xuanhui Wang, and Michael Bendersky. Large language models are effective text rankers with pairwise ranking prompting. In *Findings of the Association for Computational Linguistics: NAACL 2024*, pp. 1504–1518, 2024.

[19] Alisa Rieger, Tim Draws, Nicolas Mattis, David Maxwell, David Elsweiler, Ujwal Gadiraju, Dana McKay, Alessandro Bozzon, and Maria Soledad Pera. Responsible opinion formation on debated topics in web search. In *Advances in Information Retrieval*, pp. 437–465, 2024.

[20] Stephen Robertson and Hugo Zaragoza. The probabilistic relevance framework: BM25 and beyond. *Foundations and Trends in Information Retrieval*, Vol. 3, No. 4, pp. 333–389, 2009.

[21] Tetsuya Sakai. On variants of root normalised order-aware divergence and a divergence based on Kendall's Tau. *arXiv preprint arXiv:2204.07304*, 2022.

[22] Marieke van Hoof, Corine S Meppelink, Judith Moeller, and Damian Trilling. Searching differently? How political attitudes impact search queries about political issues. *New Media & Society*, Vol. 26, No. 7, pp. 3728–3750, 2024.

[23] Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*, 2024.

[24] QianYing Wang, Clifford Nass, and Jiang Hu. Natural language query vs. keyword search: Effects of task complexity on search performance, participant perceptions, and preferences. In *Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction*, pp. 106–116, 2005.

[25] Ryen W. White and Horvitz Eric. Belief dynamics and biases in web search. *ACM Transactions on Information Systems*, Vol. 33, No. 4, pp. 1–46, 2015.

[26] Ryen W. White, Matthew Richardson, and Wen-tau Yih. Questions vs. queries in informational search tasks. In *Proceedings of the 24th International Conference on World Wide Web*, pp. 135–136, 2015.

[27] Guido Zuccon, Joao Palotti, and Allan Hanbury. Query variations and their effect on comparing information retrieval systems. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, pp. 691–700, 2016.

[28] 松田明梨, 加藤誠. 多言語検索における質問クエリとキーワードクエリの性能評価. 第 17 回データ工学と情報マネジメントに関するフォーラム, 6L–04, 2025.