あらすじ記述による書籍検索のためのクエリ構成法

京塚　萌々† 　許　　揚† 　田島　敬史†

† 京都大学大学院情報学研究科　〒 606-8211 京都府京都市左京区吉田本町
E-mail: †{kyozuka,xuyang}@dl.soc.i.kyoto-u.ac.jp, ††tajima@i.kyoto-u.ac.jp

あらまし　本研究では，ユーザが記憶しているあらすじを利用した書籍検索を支援する手法を提案する．現時点であらゆる書籍の全文検索を行うことは不可能なので，書籍の内容の要約を収録したデータベースに対しキーワード検索を行うことになる．提案手法では，ユーザによるあらすじの記述をクエリとするが，記述内の単語がデータベース内の要約に含まれているとは限らず，意図した書籍を探すのが困難になる．そこで本論文では，ユーザが作成したクエリから単語を除去して生成したクエリの検索結果を，除去した単語に似た語が含まれている順にランキングすることで，ユーザの意図した書籍をランキング上位に提示する手法を提案する．

キーワード　質問緩和, 質問修正

1　はじめに

図書館にはレファレンスというサービスがあり，利用者からの質問や相談に応じて資料を探す手助けをしている．国立国会図書館は今後の図書館のレファレンスサービスや一般利用者の調査研究活動を支援する目的で全国の図書館のレファレンス業務で行われた質問回答サービスの記録をレファレンス協同データベース [1]に蓄積している．そこでは「昔読んだ本のあらすじは曖昧に覚えているが，タイトルを思い出せないので探してほしい.」といった質問を多く見つけることができる．また，Yahoo! 知恵袋 [2]などの質問サイトにも同様の質問が数多く寄せられている．

このような質問の答えを Web 上で検索エンジンを使用して見つけようとする場合，現時点であらゆる書籍の全文検索を行うことは不可能なので，書籍内容の要約を収録したデータベースに対し質問者の記憶に基づいてキーワード検索を行うことになる．しかし，質問者は記憶違いをしておりキーワードが間違っている可能性がある．また，質問者の記憶に残る単語がたとえ書籍の本文に含まれていたとしてもデータベース内の要約に含まれているとは限らない．このような問題が起こった場合，意図した書籍を見つけ出すのは困難になる．

そこで，本論文では，あらすじ記述による書籍検索を支援する手法を提案する．提案手法は 2 段階のランキングから成る．まず，質問者の記憶に残るあらすじに基づいて作られた元のクエリから単語を取り除いてできる全ての組み合わせを緩和クエリとして生成する．1 段階目のランキングは緩和クエリのランキングである．検索結果が信頼できると思われる順に緩和クエリをランキングする．2 段階目のランキングでは各緩和クエリの検索結果を質問者による元のクエリと似ていると思われる順にランキングする．各クエリの検索結果をクエリランキング順に連結して最終的なランキングとする．

2 段階のランキングのうち，1 段階目のクエリランキングは筆者らが以前 [1] で提案した手法を用いる．本研究では，2 段階目の各緩和クエリの検索結果を質問者による元のクエリと近いと思われる順にランキングするために誤認識度という尺度を導入し，それを用いたランキング手法を提案する．

提案手法の性能を評価するため，質問サイトからあらすじを記述して書籍を探している質問のうち正解が見つかっているものを収集した．提案手法を用いて国立国会図書館サーチ [3]による検索を行い，提案手法の有効性を評価した．

2　関連研究

本章では本研究と関連する研究について述べる．

クエリ推薦に関しては多くの研究がなされており，いくつかの形式に分類される．

クエリ拡張 (query expansion) は与えられたクエリに単語を追加して検索を支援するもので，多くの手法が提案されている [2]．特に検索エンジンが収集したログをもとにクエリを拡張する手法は多数存在する．Wang [3] らはユーザが入力したクエリのログを解析し，効果的にクエリに単語を追加したりクエリ中の単語を置換したりすることでクエリを改良する手法を提案した．このようにクエリログを利用してユーザの検索意図を推測する研究は多いが，ユーザの記憶違いが過去の検索履歴に反映されるとは考えにくいため本研究とは解決しようとしている問題が異なると考えられる．

クエリ緩和 (query relaxation) は，与えられたクエリから単語を取り除くまたは上位語と置換するなどしてクエリを生成するものである．Muslea [4] は検索結果が空集合であるクエリを機械学習の手法を用いて緩和することで検索結果が空でないクエリを生成する手法を提案している．クエリ緩和は検索結果の再現率を高めるのに効果的であると言える．本研究ではクエリ緩和のアプローチを用いてユーザの意図した書籍を検索結果上位に提示することを目指す．

1：http://crd.ndl.go.jp/reference/
2：https://chiebukuro.yahoo.co.jp
3：http://iss.ndl.go.jp

本研究ではユーザが記述した文章を元に検索を行う．このような単語数の多いクエリ (long query) による検索は単語数が少なく短いクエリに比べ適切な検索結果を実現できないという研究結果がある [5]．そこで，余分な単語を取り除き，より短いクエリを生成するクエリ緩和のアプローチによって適切な検索結果を実現する研究が行われている [6], [7], [8]．これらの研究の目的は本研究に近いと言えるが，本研究では全文検索が難しい書籍検索に特化する．

次に，ユーザ自身が検索意図を指定してクエリ緩和を行うというアプローチについて述べる．金子ら [9] は，検索クエリ内の各単語について「緩和度」という要素を導入した.「緩和度」はユーザの各単語に対する確信度を示すもので,「緩和度」を導入することでユーザがどの程度各単語を他のキーワードに置き換えていいと思っているかシステムに伝えることができる．この研究はユーザの知識の不完全さに対応してクエリ緩和を行うものなので，本研究と共通する部分があると言える．しかし，この研究はユーザが自覚しているクエリ語に対する確信度に基づいてクエリ緩和を行う手法を提案したものであり，本研究の目的のようにユーザが自覚していない記憶違いを意図していない．そのため，本研究とは方向性が異なると言える．

本研究では正確とは限らないクエリでの検索を行う．このようにあいまいなクエリで情報検索を行う研究として Ochiai ら [10] の研究や隅田ら [11] の研究がある．これらの研究はユーザの記憶があいまいである場合を想定している．Ochiai らの研究ではユーザのエピソード記憶 (過去の経験に関係する記憶) に基づく情報検索の傾向をユーザ実験で調べており，情報を見た一定期間後に記憶を頼りにユーザがその情報に関して入力するクエリは動詞を多く含むために検索性能が低くなると報告している．この研究はどのような単語が検索性能を低下させるか分析しており興味深いが，クエリ推薦は行っていないところが本研究と異なる．隅田らの研究は，ユーザの記憶があいまいなために抽象的なクエリが入力された場合でも対応できるような映画検索エンジンの構築を目的としており，本研究の目的と似通う部分がある．この研究は，ウェブ文書の情報を用いてクエリ中の抽象的な語を具体的な語に置き換えることでクエリ拡張を行う手法と，文間アライメント認識 [12] によってクエリ文とウェブ上の映画のあらすじ情報との類似性を調べることで検索する手法の 2 つを提案している．このように，ウェブ上の情報を用いてクエリ拡張を行うアプローチもあるが，本研究ではクエリ緩和によるアプローチを試みる．

ここまでは記憶違いによって与えられたクエリ語が検索対象のデータベースに含まれない場合を考えた．これに加えて，与えられたクエリ語が検索対象のデータベースに含まれない原因として，表記揺れやタイプミスも考えられる．Voorhees [13] の研究は同義語辞書を用いて表記揺れしたクエリを修正する．Cucerzan ら [14] の研究ではクエリログを用いてスペルミスの修正を行う．

本研究では書籍のあらすじ検索を扱う．あらすじに関する研究として，岩井ら [15] の研究が挙げられる．岩井らは，ショッピングサイトの商品レビュー文からあらすじが書かれている部分を機械学習によって抽出している．このように，文章中からあらすじと思われる文を抽出する研究は行われているが，本研究はあらすじ検索のクエリ語に着目しているので異なるといえる．

3 提 案 手 法

本研究で提案する手法について述べる．

この研究の目的は，過去に読んだ本のあらすじ記述からクエリを生成・ランキングし，検索結果の上位に正解書籍を提示することである．

手法全体の流れを説明する (図 1)．まずユーザのあらすじ記述から元クエリを生成し，元クエリに含まれる単語集合のすべての部分集合に対応するクエリを緩和クエリとして生成する（クエリ生成）．生成した緩和クエリとその検索結果に対して 2 段階のランキングを行って最終的な検索結果のランキングを得る．1 段階目のランキングでは，検索結果に正解書籍が含まれる可能性が高いと思われる順に緩和クエリをランキングする（クエリランキング）．2 段階目のランキングでは，各クエリの検索結果を元クエリと内容が似ていると考えられる順にランキングし直す（検索結果のランキング）．1 段階目でクエリをランキングした順に，2 段階目でリランキングした各クエリの検索結果を連結することで最終的な検索結果リストを得る．

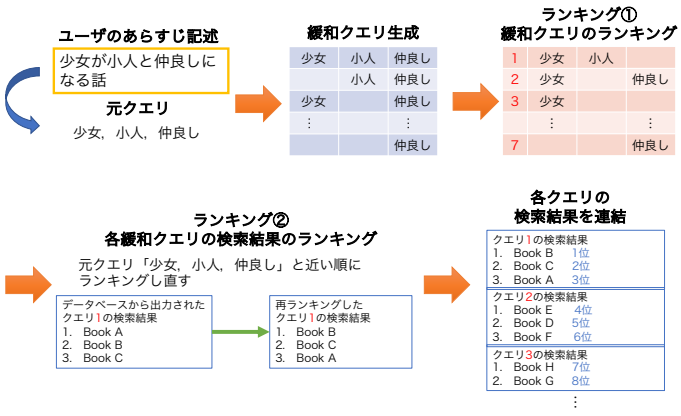

図 1　提案手法の流れ

3.1 クエリ生成

あらすじ記述から主要な内容であると思われる一文を抜き出し,「が」「を」などの助詞にあたる部分を取り除いて単語の集合の形にする．これを元クエリとする．

元クエリに含まれる単語集合のすべての部分集合に対応するクエリを生成する．これらを緩和クエリとする．

3.2 クエリランキング

この節では生成したクエリをランキングする手法について述べる．この手法は [1] で既発表である．

本研究では Probability Ranking Principle (PRP) [16] を用いて生成した緩和クエリのランキングを行う．まず生成したクエリ q ごとに検索結果に正解書籍が含まれる確率 $QP(q)$ を求

める．そのクエリによる検索結果数を $hit(q)$ とすると，そのクエリの各検索結果が正解書籍である平均確率 $p(q)$ は

$$p(q) = \frac{QP(q)}{hit(q)}$$

として求められる．$p(q)$ が大きい順にクエリをランキングすることで，正解書籍の順位の期待値を最小化でき [16]，検索結果の上位に正解書籍を提示できると考えられる．ここでは質問者の記述の文法構造を利用して $QP(q)$ を計算する．このクエリランキング手法は筆者らが過去に提案したものである [1]．

まず元クエリの単語を次の 4 種類に分類する．

- 主語 (「〜が」「〜は」となる部分)
- 述語 (「 する」「 である」となる部分)
- 目的語 (「 を」「 で」「 によって」となる部分)
- その他 (主語・述語・目的語に対する修飾語など)

どの役割の語が書籍内容の要約に含まれやすいかについては筆者ら [1] が国立国会図書館サーチで調査を行っており，述語の役割を果たす語の書籍内容の要約に含まれる確率が他の役割の語より低かったと結論づけている．この理由として，Ochiai ら [10] の研究で述べられた，あいまいな記憶に基づくクエリは動詞を多く含み不正確になるということと同時に，述語は言い換えや表記揺れが多くなるということが考えられる．したがって，この文法構造を利用してクエリをランキングすることで質問者の記憶違いや表記揺れをある程度考慮できると考えられる．

$QP(q)$ はクエリ q に含まれる各単語が属する文法的役割の語の組み合わせが書籍内容の要約に含まれる確率となる．簡単のため各役割の語が書籍内容の要約に出現する事象が独立であると仮定し，クエリに含まれる各単語の役割の語が書籍内容の要約に含まれる確率の積として $QP(q)$ を定義する．

3.3 検索結果のランキング

この節では，新たに各緩和クエリの検索結果をランキングする手法を提案する．元クエリから除去された単語と，緩和クエリの検索結果内の書籍の要約に含まれる単語が似ているか測る尺度として誤認識度を導入し，誤認識度を利用して検索結果をランキングする．

誤認識度は，他の語と取り違えやすい語には次の二つの特徴があるという仮説に基づき，人間がある 2 単語を取り違えやすいかどうかを推定する尺度である．

- あまり一般的でない語は他の語と取り違えやすい
- 同じぐらい一般的な語と取り違えやすい

本研究では，ある単語が一般的かどうかを，検索エンジンで単語を検索した際にヒットする件数で測る．

正解の単語 c を間違い語 w と取り違える際の誤認識度を次のように定義する．

$$M(c,w) = \log_2\left(\frac{1}{|H(w) - H(c)| \times H(c)}\right)$$

ここで $H(x)$ は検索エンジンで単語 x を検索したときの検索結果の件数である．

誤認識度は，正解の単語の検索件数が少なく，かつ間違い語と検索件数が近いとき大きな値をとり，記憶違いしやすいことになる．上で提案した誤認識度は 2 単語間に定義されるものであるが，本研究では緩和クエリと検索結果の要約文の間の誤認識度を定義し，検索結果をランキングする．元クエリから単語集合 W を取り除いて生成した緩和クエリの検索結果として出力された書籍内容の要約に含まれる単語の集合 A について次のように誤認識度 $M(A,W)$ を定義する．

$$M(A,W) = \frac{1}{|W|}\sum_{w\in W}\max_{a\in A}\log_2\left(\frac{1}{|H(w) - H(a)| \times H(a)}\right)$$

元クエリから取り除かれた各単語 w_i について，要約内の単語 A に含まれる単語のうち最も w との誤認識度が大きくなる単語 $\tilde{a}_i$ を選択する．次に W に含まれる全ての w_i について，w_i と $\tilde{a}_i$ の誤認識度の平均を計算し，これを W と A の間に定義される誤認識度としている．

各クエリの検索結果について，$M(A,W)$ が大きい順にランキングする．

4 実験計画

本研究に際して行う実験について述べる．

4.1 実験の手順

4.1.1 質問データ収集

まず Yahoo!知恵袋より書籍を探す質問と回答のデータを収集した．Yahoo!知恵袋で「うろ覚え 児童書」と「うろ覚え 絵本」を検索クエリとして収集した解決済みの質問のうち，質問者が書籍のあらすじを記述することで書籍を探しており，かつ回答された書籍に対し質問者から探していたものであったという旨のコメントがついているものを収集した．その内，正解書籍のデータを国立国会図書館サーチで検索し，書誌データに「要約・抄録」のデータが含まれるものを抽出した．

4.1.2 クエリ生成

3.1 で述べたように，収集した質問データからあらすじを説明していると思われる主要な部分を手作業で抜き出し，「が」「を」などの助詞にあたる部分を取り除いて単語の集合の形とし元クエリとした．ここで各単語と元の記述での文法的役割を関連づけた．元クエリのキーワードのすべての組み合わせを緩和クエリとした．

分類の例を表 1 に示す．

表 1　単語分類の例

正解書籍のタイトル	質問者によるあらすじ	(役割，単語) の組
木かげの家の小人たち	少女が小人と仲良し	(主語，少女) (目的語，小人) (述語，仲良し)
先生のつうしんぼ	先生は給食のカレーのにんじんが嫌い	(主語，先生) (その他，給食) (その他，カレー) (その他，にんじん) (述語，嫌い)

「木かげの家の小人たち」は 3 語からなるクエリなので緩和

クエリが $2^3 - 1 = 7$ 件,「先生のつうしんぼ」は 5 語からなるクエリなので緩和クエリが $2^5 - 1 = 31$ 件生成される.

4.1.3 クエリランキング

4.1.2 で元クエリそれぞれから生成された緩和クエリのランキングを行う.元クエリの各役割の語が書籍データベースのあらすじにも含まれる確率として,筆者ら [1] が国立国会図書館サーチのデータより得た数値を使用する.(主語:0.442, 述語:0.048, 目的語:0.545, その他:0.441)「木かげの家の小人たち」が正解となる質問について 3.2 の手法でランキングしたクエリ 7 件を表 2 に示す.

表 2　クエリランキング例 (上位 6 件)

述語	その他	目的語	主語
仲良し		小人	
		小人	少女
仲良し			少女
		小人	
仲良し			
			少女
仲良し		小人	少女

4.1.4 誤認識度によるランキング

4.1.2 で生成した緩和クエリを国立国会図書館サーチで検索する.質問データを収集する際に児童書を対象としているので,「児童書総合目録」というデータベースを対象に検索を行う.それぞれの検索結果を 3.3 で述べた誤認識度によってランキングする.あらすじ検索の結果を見るので検索結果のうち書誌データに「要約・抄録」のデータが含まれるものを用いる.また,本実験で利用する国立国会図書館サーチは,検索結果数が 500 件を超えるとき上位 500 件のみを検索結果リストとして表示するため,国立国会図書館サーチの検索結果の上位 500 件に含まれかつ書誌データに「要約・抄録」のデータを含むものを使用することになる.

誤認識度による検索結果ランキングはともに元クエリから除去された語と検索結果の要約に含まれる語の検索結果件数を用いる.本実験では形態素解析器 Janome[4] を用いて要約内の単語抽出を行い,国立国会図書館サーチの「児童書総合目録」[5] で検索して検索結果件数を取得した.

4.1.5 検索結果の出力

最終的な検索結果として,次の 3 つのランキング手法によるものを比較する.

- クエリランキングのみ (図 2)

国立国会図書館サーチで各緩和クエリによる検索を行い,4.1.3 で述べたクエリランキングの順に各クエリの検索結果を連結して最終的な検索結果とする.各クエリの検索結果は,国立国会図書館サーチで検索して出力された検索結果リストの順番をそのまま用いる.

- 誤認識度によるランキングのみ (図 3)

国立国会図書館サーチで各緩和クエリによる検索を行い,検索結果全体を誤認識度が大きい順にランキングして最終的な検索結果とする.

- クエリランキングと誤認識度によるランキングの併用 (図 4)

国立国会図書館サーチで各緩和クエリによる検索を行い,各クエリそれぞれについて誤認識度が大きい順に検索結果をランキングし直す.その後,4.1.3 で述べたクエリランキングの順に各クエリの検索結果を連結して最終的な検索結果とする.

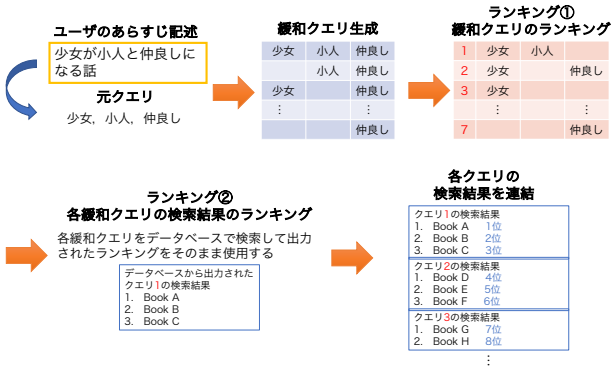

図 2　クエリランキングのみ

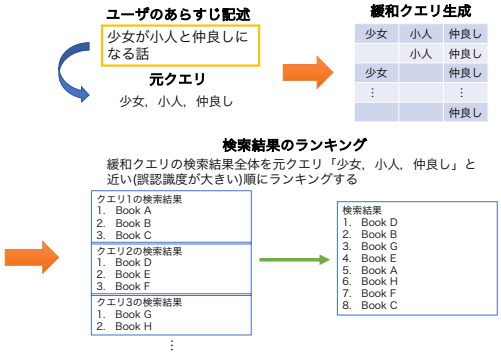

図 3　誤認識度ランキングのみ

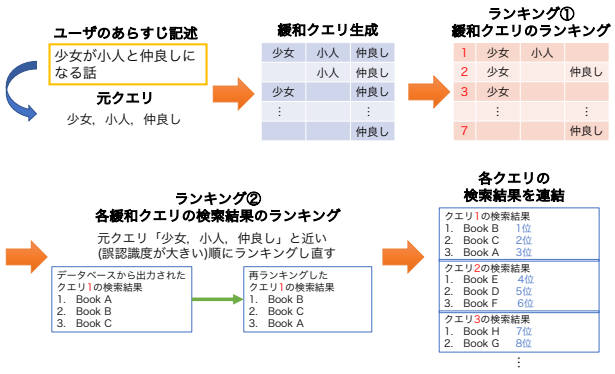

図 4　クエリランキング+誤認識度ランキング

4 : https://mocobeta.github.io/janome/

5 : http://iss.ndl.go.jp/information/target/

4.2 実験の評価

4.2.1 平均逆順位による評価

本研究の目的はユーザの意図した書籍を検索結果上位に表示することであるので，以下の式で定義される平均逆順位 (Mean Reciprocal Rank, MRR) を計算し比較する．

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$$

ここで $|Q|$ は対象とするクエリ件数，$rank_i$ は i 番目の質問に対して出力された最終的な検索結果における正解書籍の順位である．ただし，4.1.4 で述べたように，国立国会図書館サーチでは上位 500 件のみを検索結果リストとして表示するため，生成したどの緩和クエリの検索結果でも正解書籍が上位 500 件に含まれなかった場合，正解書籍がヒットしない．その場合，$\frac{1}{rank_i} = 0$ とする．

5 実験結果

本研究では，表 3 の 7 件のクエリについて実験を行った．

表 3　使用したクエリ

元クエリ	正解書籍名	正解がヒットする最長クエリ
少女 小人 仲良し	木かげの家の小人たち	少女 小人
姉弟 宝物 見せ合う 説明	たからものくらべ	宝物
どろぼう 捕まる 待遇 良い 刑務所 入る	ゆかいなどろぼうたち	どろぼう
先生 給食 カレー にんじん 嫌い	先生のつうしんぼ	先生 給食 にんじん
ミサ サミー 名乗る ゲーム 世界 閉じ込められる	恐怖のテレビゲーム：吸血鬼メビナの復讐	ミサ ゲーム
北極グマ かあさんグマ 子グマ アザラシ 仲良くなる	北極のムーシカミーシカ	北極グマ アザラシ
ゼプクククク 働く 嫌 失敗 反省	欲ばりのセブググググ：南アフリカ共和国	働く

各クエリについて，4.1.5 で述べた手法でランキングした際の正解書籍の順位を表 4 に示す．

それぞれの手法の MRR を表 5 に示す．

6 考　　察

6.1 誤認識度のみによるランキング

誤認識度によるランキングの MRR は本実験で比較した 3 種類のランキング手法のうち最も低かった．特に，「先生のつうしんぼ」「北極のムーシカミーシカ」が正解となるクエリに関して大幅に正解書籍の順位を下げた．その一方，「ゆかいなどろぼうたち」が正解となるクエリに関しては他の手法より正解書籍を上位にランキングすることができた．

表 4　正解書籍の順位

正解書籍名	誤認識度のみ	クエリランキングのみ	併用
木かげの家の小人たち	7	13	11
たからものくらべ	235	211	207
ゆかいなどろぼうたち	256	677	462
先生のつうしんぼ	428	2	2
恐怖のテレビゲーム：吸血鬼メビナの復讐	89	81	77
北極のムーシカミーシカ	269	2	3
欲ばりのセブググググ：南アフリカ共和国	125	120	76

表 5　MRR

誤認識度のみ	クエリランキングのみ	併用
0.025186941	0.157688362	0.136768965

誤認識度によるランキングで正解書籍の順位が下がっているクエリの特徴として，元クエリに含まれる単語のうち正解がヒットする最長クエリに含まれる単語が占める割合が高いことが挙げられる．逆に，正解書籍の順位が高くなったクエリは元クエリに含まれる単語のうち正解がヒットする最長クエリに含まれる単語が占める割合が少ない．このことから，誤認識度のみによるランキングは元クエリに含まれる単語のうち正解書籍の要約に含まれない語が多い場合に向いたランキング手法であると考えられる．

6.2 クエリランキングのみによるランキング

クエリランキングのみでランキングした際の MRR は本実験で比較した 3 種類のランキング手法で最も高くなった．しかし，「ゆかいなどろぼうたち」のように，他の 2 手法より大幅に正解書籍の順位を下げた場合もある．このランキング手法で正解書籍の順位が低くなっているクエリ (正解書籍が「ゆかいなどろぼうたち」「たからものくらべ」のクエリ) の特徴として，元クエリの単語のうち正解書籍がヒットするクエリ語の割合が少ないことが挙げられる．一方，クエリランキングによって正解書籍の順位が高くなっているクエリは元クエリの単語のうち正解書籍がヒットするクエリ語の割合が比較的高い．このことから，クエリランキングのみによるランキングは元クエリに含まれる単語のうち正解書籍の要約に含まれる語が多い場合，つまり比較的正解書籍の要約文に近い内容のクエリについて効果的に働くと考えられる．

6.3 誤認識度によるランキングとクエリランキングを併用したランキング

誤認識度によるランキングとクエリランキングを併用した際の MRR は本実験で比較した 3 手法のうち 2 番目に高かった．MRR ではクエリランキングのみの手法より低くなったが，クエリ 7 件のうちクエリランキングのみの手法より正解書籍の順位が低くなったのは 1 件のみで，残り 6 件はクエリランキングのみの手法と同じか，より高く正解書籍をランキングすること

ができた．元クエリのうち正解にヒットするクエリ語が少ないとき，誤認識度によるランキングを併用することで，元クエリに含まれているが検索クエリには含まれていない語に近い語を含む要約文を持つ書籍データをより上位に移動させることができるために，クエリランキングのみを用いた場合より正解書籍を上位にランキングできたと考えられる．

7 結　　論

本論文では，クエリ緩和により生成したクエリの検索結果を，元クエリから除去された単語に類似したものが含まれている順にランキングする手法を提案した．また，ユーザによる書籍のあらすじ記述で書籍検索を行うための手法として，本論文で提案した検索結果のランキング手法に，我々が過去の研究で提案したクエリ生成手法や生成したクエリのランキング手法を組み合わせた手法を提案した．次に，書籍のあらすじに関する記述によって書籍を探す質問データに対し，提案した手法を適用することで提案手法の有効性の確認を試みた．

本研究における今後の課題について述べる．

まず，本研究における実験ではクエリ 7 件を用いたが，提案手法の有効性を確かめるためにはより多くのクエリで実験することが必要である．

また，本研究では誤認識度を計算する際に国立国会図書館サーチの検索結果件数を使用した．しかし，人間が一般的に取り違えやすい単語かどうかを測るには google 検索 [6] などの web 検索による検索結果件数を用いるほうが自然であると考えられる．検索結果件数の取得に用いる検索エンジンを変えて比較することで，書籍検索特有の傾向があるかどうかを調査することもできるので，他の検索エンジンとの比較を行いたい．

本研究では元クエリから除去された単語と各書籍データの要約文に含まれる単語との類似を測るために誤認識度という尺度を提案し用いたが，単語間の類似を測る尺度としては他にも Bollegala ら [17] の WebPMI などが挙げられる．また，元クエリから除去された単語と各書籍データの要約文に含まれる単語の共起度を用いる手法も考えられる．他の手法についても検討を行い，よりあらすじ検索に効果的な手法を提案したい．

以上より，本研究全体における課題として

- より多数のデータで実験し提案手法の有効性を確認する
- 他の検索エンジンによる検索結果件数を用いた場合と比較する
- より効果的なランキング手法について検討する

ことを挙げる．

8 謝　　辞

本研究は，JST CREST (JPMJCR16E3)，JSPS 科研費 18H03245，JSPS 科研費 16K12430 の支援を受けたものである．

6 : https://www.google.co.jp

文　　献

[1] Momo Kyozuka and Keishi Tajima. Ranking methods for query relaxation in book search. In *Proc. of IEEE/WIC/ACM International Conference on Web Intelligence*, pp. 466–473, December 2018.

[2] Efthimis N Efthimiadis. Query expansion. *Annual review of information science and technology (ARIST)*, Vol. 31, pp. 121–87, 1996.

[3] Xuanhui Wang and ChengXiang Zhai. Mining term association patterns from search logs for effective query reformulation. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pp. 479–488. ACM, 2008.

[4] Ion Muslea. Machine learning for online query relaxation. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pp. 246–255, New York, NY, USA, 2004. ACM.

[5] Michael Bendersky and W. Bruce Croft. Analysis of long queries in a large scale search log. In *Proceedings of the 2009 Workshop on Web Search Click Data*, WSCD '09, pp. 8–14, New York, NY, USA, 2009. ACM.

[6] Yan Chen and Yan-Qing Zhang. A query substitution-search result refinement approach for long query web searches. In *Proceedings of the 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Volume 01*, WI-IAT '09, pp. 245–251, Washington, DC, USA, 2009. IEEE Computer Society.

[7] Giridhar Kumaran and Vitor R. Carvalho. Reducing long queries using query quality predictors. In *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '09, pp. 564–571, New York, NY, USA, 2009. ACM.

[8] Niranjan Balasubramanian, Giridhar Kumaran, and Vitor R. Carvalho. Exploring reductions for long web queries. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pp. 571–578, New York, NY, USA, 2010. ACM.

[9] Yasufumi Kaneko, Satoshi Nakamura, Hiroaki Ohshima, and Katsumi Tanaka. Query relaxation based on users' unconfidences on query terms and web knowledge extraction. In *Proceedings of the 11th International Conference on Asian Digital Libraries: Universal and Ubiquitous Access to Information*, ICADL 08, pp. 71–81, Berlin, Heidelberg, 2008. Springer-Verlag.

[10] Shuya Ochiai, Makoto P. Kato, and Katsumi Tanaka. Recall and re-cognition in episode re-retrieval: A user study on news re-finding a fortnight later. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, CIKM '14, pp. 579–588, New York, NY, USA, 2014. ACM.

[11] 隅田飛鳥, 池田和史, 服部元, 小野智弘. 9-10 あいまいな記憶下での抽象的クエリを許容する映画検索エンジン (第 9 部門ヒューマンインフォメーション). 映像情報メディア学会年次大会講演予稿集 2012, pp. 9–10. 一般社団法人映像情報メディア学会, 2012.

[12] 水野淳太, 渡邉陽太郎, エリックニコルズ, 村上浩司, 乾健太郎, 松本裕治. 文間関係認識に基づく賛成・反対意見の俯瞰. 情報処理学会論文誌, Vol. 52, No. 12, pp. 3408–3422, dec 2011.

[13] Ellen M. Voorhees. Query expansion using lexical-semantic relations. In *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '94, pp. 61–69, New York, NY, USA, 1994. Springer-Verlag New York, Inc.

[14] Silviu Cucerzan and Eric Brill. Spelling correction as an iterative process that exploits the collective knowledge of web users. In *EMNLP*, Vol. 4, pp. 293–300, 2004.

[15] 岩井秀成, 池田郁, 土方 嘉徳他. レビュー文を対象としたあらすじ分類手法の提案. 電子情報通信学会論文誌. D, 情報・システム / 電子情報通信学会 編, Vol. 96, No. 5, pp. 1222–1234, 2013.
[16] C. J. van Rijsbergen. *Information Retrieval*, pp. 113–114. Butterworths, 1979.
[17] Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka. Measuring semantic similarity between words using web search engines. pp. 757–766, 01 2007.