

氏名(本籍)	大塚 淳史 (栃木県)				
学位の種類	博士(情報学)				
学位記番号	博甲第 9637 号				
学位授与年月日	令和 2 年 4 月 30 日				
学位授与の要件	学位規則第4条第1項該当				
審査研究科	図書館情報メディア研究科				
学位論文題目	問い合わせ拡張による 情報要求の言語化支援技術に関する研究				
主査	筑波大学	教授	博士(工学)	佐藤 哲司	
副査	筑波大学	教授	博士(教育学)	芳鐘 冬樹	
副査	筑波大学	教授	工学博士	松本 紳	
副査	筑波大学	准教授	博士(情報学)	加藤 誠	
副査	筑波大学	教授	博士(工学)	宇津呂武仁	

論文の要旨 (2,000字程度)

本学位論文は、曖昧な情報要求から具体的なクエリを生成する情報要求の言語化に着目した情報アクセスの支援技術に関する研究の成果である。画像や音声などの様々なメディアが電子化され膨大に蓄積された情報化社会においても、情報検索で用いられるクエリは依然として、検索キーワードや質問文などのテキストである。しかしながら、利用者が自らの情報要求を明確に言語化することは難しい場合が主流である。このため本論文では、利用者が想起した単純なキーワードや曖昧な質問文を入力として、膨大な情報源から十分に絞り込まれた所望の検索結果を得るための的確なクエリを生成する手法を提案している。利用者が最初に入力したクエリや質問文から、想定される情報要求の候補を利用者に複数提示し、利用者が選択することでインタラクティブに情報要求を詳細化することを基本的な枠組みとし、どのような情報要求の候補を提示すべきかを研究課題として設定している。

全6章で構成されている本論文は、第1章において上述した研究の背景、目的と研究課題の設定を述べている。特に、情報アクセスの手段やアクセス対象の違いに着目し、2つのアプローチを採ることを述べている。第1のアプローチは、日々情報が追加されているWeb情報空間を対象とした全文検索、すなわち、順位付けされた検索結果が返却される検索を対象とする。このような情報源を本研究ではオープンナレッジと称して、様々な観点から多様な情報にアクセスするための情報要求を提示することを目指している。第2のアプローチは、追加更新のない限定されたWeb情報空間や、特定の製品やサービスに対するマニュアルなどを対象とするクローズドナレッジに対する情報検索

を対象とする。ここでは、利用者が求める検索結果は1つ、あるいは、極めて少数であることから、探索空間を絞り込むための情報要求を提示するとしている。

第2章では、本論文に関連する先行研究を概観している。最初に、オープンナレッジにおける情報要求の言語化支援に関する関連研究として、情報検索における利用者支援技術であるクエリ拡張技術を紹介している。また、本論文が対象としている Web 検索やコミュニティ QA に関する関連研究も紹介している。

次に、クローズドナレッジの情報要求の言語化支援に関連して、近年、研究が盛んになっている質問応答や機会読解技術に関する先行研究を紹介している。そこでは、機械読解の基盤技術となっている深層学習による言語処理技術、言語生成技術についても関連研究を示している。

第3章は、本論文の第一の核となるオープンナレッジを対象に、より多様な情報にアクセスすることを目的とした Web 検索における言語化支援技術を提案している。利用者が入力したクエリを多様な情報要求に展開する情報源として、様々な疑問や意見などを質問回答(QA)形式で自由にコミュニケーションしているコミュニティ QA の質問記事を利用している。コミュニティ QA の質問記事を言語化された情報要求とみなし、旅行や恋愛などのカテゴリや、クリスマスや新学期などの季節性をコンテキストとして、コンテキスト切り替え型のクエリ拡張手法を明らかにしている。

第4章は、本論文の第2の核となるクローズドナレッジを対象に、機械読解による質問応答における改訂質問生成技術を提案している。機械読解は、自然言語で記述された質問文に対する回答を、対象となるドキュメント集合の中から発見する技術であるが、曖昧な質問文を入力すると回答箇所を特定できず、回答精度が低下するという課題が知られている。本論文で提案する改訂質問生成では、入力された曖昧な質問文から、機械読解で使用するドキュメント集合の内容に基づいて具体化した質問文を生成する。この具体化された質問文を機械読解に入力することで、機械読解の回答を更新し、より特定性の高い回答が得られるようになる。生成された複数の質問文を利用者に提示することで、特定性を高めるための言語化された情報要求を提示できるとしている。本論文では、言語生成モデルをベースに改訂質問生成のために考案した、機械読解とコピー機構を組み合わせた深層学習モデルを提案している。

第5章は、オープンナレッジとクローズドナレッジそれぞれに対して、言語化支援を行うことの効果と、提案手法の限界について考察している。Web 検索においては、特に同音異義語のようなクエリに対して、提案手法が実現している自然言語で記述された質問文を提示することの有効性を論じている。これは、クエリを作成する際に失われたコンテキストをシステムが補完できていることに起因していると結論づけている。機械読解においては、日本語において特に顕著に発生する省略された主語などを補完できることが、回答精度の向上に繋がったと結論づけている。

第6章は、本論文のまとめの章であり、コミュニティ QA や検索対象となるドキュメント集合を情報源としてクエリを拡張する情報要求の言語化支援の意義を明らかにしている。

審査の要旨 (2,000 字以上)

【批評】

インターネットの普及・発展と情報のデジタル化によって、あらゆる情報に簡便にアクセスできる情報化社会が到来している。一方で、様々な要因によって情報アクセスの恩恵を受けられない情報弱者の存在も否定できない。

本学位論文は、こうしたデジタルデバイドと呼ばれる社会的背景を受けて、曖昧な情報要求から具体的なクエリを生成する情報要求の言語化に着目した情報アクセスの支援技術に関する研究の成果をまとめたものである。膨大な情報源から所望の情報をアクセスするには、具体的かつ詳細なクエリを用意する必要があるが、情報源に対する事前知識を持たない一般の利用者が適切なクエリを想起するのは困難であることは容易に想像できる。したがって、本論文が解決しようとしている課題設定は妥当であり、学術面での貢献に加えて社会的な要請にも応える重要なテーマであると言える。以下、本論文の章構成に従って批評を述べる。

第 1 章は、上記の社会的背景から研究の目的を述べ、対象とする情報源の特性をオープンナレッジとクローズドナレッジに分類・定義し、それぞれに対する本論文でのアプローチを明確化している。オープンナレッジとは、Web 情報空間に代表される、日々情報が追加され続けている更新を前提とする情報源であり、ここでは、利用者も知らない最近の情報にアクセスできるようにする、多様な観点を提示するクエリ拡張が重要であるとしている。一方、情報の追加更新が殆ど無いクローズドナレッジに対しては、最も適している情報すなわち特定性の高い情報を厳選してアクセスできる手段を提供することが重要であるとしている。このように情報源の特性あるいは情報アクセスの特性から研究課題を分解し、それぞれに対して適切なアプローチを採ることは、研究方法として適切であると言える。

第 2 章は、関連研究の章である。クエリ拡張に関する先行研究を、コンテキストウェア（状況依存・状況理解）なクエリ拡張と、根拠を提示するクエリ拡張の 2 つに整理したことは、研究の位置づけを明確化することに貢献している。また、本研究で提案する手法で利用する具体的な情報源である「コミュニティQA」に関する先行研究も網羅されている。クローズドナレッジに関する先行研究は、深層学習の広がりから注目されている機械読解を中心に整理されている。本研究を特徴付ける側面に沿って先行研究が整理されており、また引用文献の数も十分であると言える。

第 3 章は、オープンナレッジに対する新たなクエリ拡張手法を提案している章である。提案手法の核は、コミュニティQAに投稿された質問記事を Web 検索のクエリと見なすことである。利用者が入力した曖昧あるいは不十分なクエリを手がかりとして質問記事を選択し、利用者に提示することでクエリを補完する。その際に、クエリが入力された時期（季節）と一致する質問記事を選択することでコンテキストウェア

アなクエリ拡張を実現している。利用者が入力するクエリに使用される語彙には、それぞれに固有の周期があるとの仮説に基づいて、動的なトピックモデル(DTM)を用いて話題変動をモデル化している。実際のコミュニティQAデータを用いてクエリ拡張型Web検索システムを実装し、提案手法の有用性を評価している。コミュニティQAのカテゴリと春夏秋冬の季節性を2次元のファセットとして提示し、利用者が選択することでコンテキストアウェアなクエリ拡張が実現できることを示したことは、実用面からも十分に評価できる。一方で、異なる周期を持つトピックが混在している環境での効用性など、提案手法の限界についての言及がもう少しあっても良かったと思われる。

第4章は、クローズドナレッジに対する機械読解の手法を論じた章である。機械読解は、自然言語による質問応答で高い精度を達成できる技術であるが、ここでは、回答が複数個考えられる曖昧な質問や、回答を絞るために必要な情報が欠落している質問を扱うために、改訂質問生成モデルを考案している。さらに、文圧縮によって擬似的に短い(情報が欠落した)質問を生成する手法も提案しており、機械学習における学習データの自動生成も行っている。公開されている日本語と英語のデータセットを使用して提案手法を評価し、従来手法よりも高い精度を達成したことが示されている。また、本論文で示された文圧縮による質問生成の手法は、欠落情報の量を制御した柔軟な評価方法へと発展していく可能性があり、機械読解の評価に新たな展開が期待できる。

第5章は、オープンナレッジとクローズドナレッジのそれぞれで提案した手法を考察する章であり、両者における情報要求の言語化の違いと今後の展望について述べている。本論文では、オープンとクローズドそれぞれに対する言語化支援の手法を提案しているが、将来的な展望として両者を融合したアプローチの必要性についても言及している。

第6章は本論文の結論として、利用者の情報要求すなわち「知りたいこと」を言語化支援することの意義と、本論文で提案した手法を整理して述べている。

【最終試験結果】

2020年3月31日、図書館情報メディア研究科学学位論文審査委員会において、審査委員全員出席のもと、本論文について著者に説明を求めた後、関連事項について質疑応答を行った。引き続き、「図書館情報メディア研究科博士後期課程の学位論文の審査に関する内規」第12項第2号に基づく最終試験を行い、審議の結果、審査委員全員一致で合格と判定された。

【結論】

よって、本学位論文の著者は博士(情報学)の学位を受けるに十分な資格を有するものと認められる。