

報告番号	※甲	第	号
------	----	---	---

主論文の要旨

論文題目 大規模テキストからの手続き型知識の自動獲得に関する研究

氏名 小澤 俊介

論文内容の要旨

メールやブログなどのインターネットを介した情報発信、並びに、新聞や書籍、論文などの紙媒体が盛んに電子化されることにより、膨大な量の電子化テキストが蓄積されている。こうした電子化テキストの中には人間が知りたい知識が含まれているため、これらを参照することにより、知識を得る機会が増加している。しかし、十分に整理されていないため、求める知識を得るまでに多くの時間を要したり、求める知識が得られないこともある。近年では、WikipediaやKnolなどの知識共有サイトが数多く作られ、参加者によって知識が整理、及び、提供されている。人間が知りたい知識を整理することは多くの人の役に立つが、あらゆる知識に対して、人手により収集し、整理することは困難であるため、テキストから知識を自動獲得することが望まれる。

これまで、意味解析や談話解析などの高度な自然言語処理技術への応用を目的として、固有表現や上位下位関係などの知識を自動獲得する研究が盛んに行われてきたが、これらは人間が参照しても、直接的には役に立たないことが多い。一方、評判情報やレシピ情報など、人間に対して役立つ知識の獲得はまだあまり行われていない。しかし、これらを獲得して整理することは重要であり、人間の知識探索を容易にすることができる。

近年では、人間が何か知りたいことがある場合、QAサイトで質問するという手段がよく用いられる。つまり、QAサイトに投稿された質問は人間が知りたい知識を表している。QAサイトに投稿される質問は大きく分けて、回答が人名や地名などの名詞で表される質問（factoid型）と回答が意見や理由などの文章で表される質問（non-factoid型）の2種類があり、このうち、より多く尋ねられるのはnon-factoid型の質問である。さらに、non-factoid型の質問は手続き型、叙述型、意見型、理由型、定義型の5種類に分類でき、これらのうち、QAサイトで最もよく尋ねられるのは手続き型の質問である。すなわち、人間が強く知りたい知識は手続き型知識である。

本論文では、人間に対して有用な知識であり、かつ、人間が強く知りたい知識である手続き型知識を獲得する手法を提案する。手続き型知識を体系的に獲得するため、知識をモノとコトを用いて5種類（モノに関する知識、モノとモノの関係知識、コトに関する知識、モノとコトの関係知識、コトとコトの関係知識）に分類する。これらのうち、手続き型知識が該当するコトの知識に関わる分類に基づき、コトに関する知識、モノとコトの関係知識、コトとコトの関係知識の各分類に対して、ノウハウ、言語資源の用途、外出行動に対するアドバイスをそれぞれ獲得する手法を述べる。

本論文は全5章から構成される。第1章は本論文の序論であり、まず、人間が知りたい知識をQAサイトと関連付けることにより、手続き型知識は人間が強く知りたい知識であることを述べる。次に、手続き型知識を体系的に獲得するため、モノとコトを用いて知識を形式的に分類する。続いて、コトの知識獲得に関する研究動向を示す。最後に、本論文の位置づけとアプローチについて述べる。

第2章では、コトに関する知識に対する手続き型知識として、ノウハウを獲得する手法を提案する。本手法では、ノウハウ獲得においてモノとその使われ方の情報が重要な役割を果たすという仮説に基づき、ノウハウを獲得する。まず、対象のモノを含むパッセージを獲得し、それらから5つの手法を用いることによりノウハウ候補を抽出する。次に、パッセージに含まれる手がかり表現、及び、モノとその使われ方として用途表現を利用することにより、ノウハウ候補がノウハウであるか否かを判定した。評価実験により、ノウハウ獲得においてモノとその使われ方が重要な役割を果たすことを定量的に明らかにした。

第3章では、モノとコトの関係知識に対する手続き型知識として、言語資源の用途情報を獲得する手法を提案する。言語資源の用途情報について説明し、言語資源検索における用途情報の有用性について述べる。言語資源の用途情報には、作成者、及び、利用者の視点からの用途情報がある。作成者の視点からの用途情報には、言語資源データベースSHACHIに格納されているメタデータを利用する。一方、利用者の視点からの用途情報は構文構造を考慮して作成した抽出ルールを用いて学术论文から獲得する。獲得した用途情報に対して、その精度を判定することにより、学术论文から言語資源の用途情報を精度よく獲得できることを示す。また、自動獲得した用途情報をSHACHIに格納することにより、言語資源検索システムを構築する。自動獲得した用途情報の格納前と格納後の言語資源検索の性能を評価した結果、自動獲得した用途情報が言語資源検索に有効に働くことを確認した。

第4章では、コトとコトの関係知識に対する手続き型知識として、外出行動前に入手すると役立つアドバイスをWebから獲得する手法を提案する。本手法では、外出行動前に入手すると役立つアドバイスを獲得するために、アドバイス文の特徴分析に基づいて生成した素性を用いて、アドバイス文か否か、及び、アドバイス文が外出行動前に参照しておきたいものか否かを機械学習を用いて判定した。設定した各素性の有効性を評価するため、提案手法と提案手法から各素性を除去した手法とを比較した結果、全ての素性が性能の向上に寄与することを示し、行動前アドバイス文か否かの判定について精度75%、再現率81%を達成し、本手法の有効性を確認した。また、提案手法を用いた事前情報提供アプリについて説明する。

最後に、第5章で本論文をまとめ、今後の研究課題、及び、将来の展望について述べる。