

## 論文の内容の要旨

論文題目 知識共有システムにおける知識の獲得・加工・管理に関する研究

氏名 友部 博教

コミュニティにおいて、我々は多くの知識を持っている。自分の持っている知識を互いに共有することによって、より有益な知識を獲得することができる。この知識の共有という機能をより発展させるためには、決められた表現で記述された知識だけでなく、コミュニティにおける背景知識のような明確でない知識も共有できるような効果的な知識共有システムを構築する技術が必要である。Web 上の情報もこういった知識共有システム構築のための重要な情報基盤となっている。本論文では Web 環境における知識共有システムのための基礎技術に関する研究について述べる。

本論文は 2 部で構成されている。第 1 部では、自然言語で記述された膨大な知識を活用するためのアプローチとして、KRNL と呼ばれる知識表現について記述している。次に第 2 部では、コミュニティにおける背景知識として人間関係を Web 上から抽出する手法と知識共有システムにおける利用法について述べる。

第 1 部では、KRNL について述べる。これは、テキストで記述された知識の高次利用ができる新しい知識表現法である。

近年の計算機技術の発展によって、我々は Web 上から膨大な情報を得ることのできるようになった。その中でも特に多いのが自然言語で記された情報である。我々は一つの文書から望む情報や知識を得ることができない場合には、検索エンジンを用いて関連するいくつかのページを集め、それを組み合わせて知識を獲得する。しかし、検索結果で提示される文書が多数であるため、同じような情報を何度も読まなければならないなどの不都合、非効率性が生じる。このような場合、冗長な情報を削除し、ユーザの要求に関連する情報を統合したものを提示できれば便利である。

Web 上の情報は、個人の記述したものから、組織により記されたものまで、非常に雑多な自然言語文書が混じっている。これらを統合するには、何らかの枠組み（知識表現）を導入し、それに従った規則で知識を記述する必要がある。まだ同様に、複数ユーザからの知識入力を想定した知識共有システムも、同じことが言える。知識を入力するユーザは各々意図を持っているわけだが、この意図に従って自分勝手に知識を入力するのではなく、システムの仕様に基づいた知識入力を行うことで有効な知識共有が行われるのだ。つまり、複数ユーザによる知識共有システムを実現することが、文書統合への必要条件となる。

知識共有システムでは、複数のユーザから大量の知識が入力されることになる。その中

からユーザの要求に合致する知識を探すことになるのだが、既に入力された知識だけではユーザの要求に答えることができない可能性もある。しかし、さらに複数の知識を組み合わせることによって発見できる場合がある。よってこのような知識共有システムでは、ユーザの要求を満たす知識を発見できるように、持っている知識を複数組み合わせることが必要となる。

述語論理は式で表現された知識を組み合わせることによって推論する枠組みである。堅固な基礎に立脚する論理は、重要な知識表現の一つである。

また、論理は数学的基礎に立脚する広い範囲をカバーできる汎用的知識表現であるという特徴を持っている。論理は、述語や引数を自由に設定することができるということである。そのため述語論理は、自然言語からの知識を表現する重要な枠組みとなっている。

しかし、自然言語文を論理に変換する場合、表現が一意に定まらないという問題がある。述語を何に設定するか、また引数の数をいくつにするか、その引数の並べ方をどうするかなどに任意性が生じる。同じ知識であってもこの任意性によって表現が変わってしまい、知識の組み合わせによる推論が正しく行うことができない。

特に複数ユーザによる知識共有においては、表現が統一されないことが多い。同じ知識であっても、ユーザによって表現が異なる場合がある。また同一ユーザが変換する時でさえ、表現が一つでないことがある。

なぜこのような任意性が起るのだろうか。これは自然言語文で記された知識に対し、どの観点を軸にして述語表現に変換するか、ユーザによって異なるためである。例えば、「柴犬が通行人にかみついた」という自然言語文があるとしよう。あるひとは、「かみつく」という行動に観点を置き「かみつく(柴犬,通行人)」と書き、また別の人は「柴犬」という動作の主体に観点を置き「柴犬(かみつく,通行人)」と記述するかもしれない。すると、この自然言語で記された知識を他の知識と組み合わせるには、述語「かみつく」に関するルールと、述語「柴犬」に関するルールを完備する必要がでてくる。冗長なルールが増加することは推論処理にかかる時間が増大し、またルールを記述するコスト（時間や労力といったユーザの負担）も増してしまう。よって知識共有のためには、自然言語文で記された知識を述語表現に変換する際に観点について考慮しなければならないのである。

ここで導入する知識表現 KRNL(Knowledge Representation for Natural Language)は、自然言語文で記された知識を「二つのフレーズと、それを結び付ける関係」という観点で述語を数種類に絞って表現したものである。そしてその述語に従って、引数の数も限定している。論理との関係で言えば自然言語との対応の良い形で論理のサブセットを切り出したものであり、論理と同様な推論を行わせることができる。

述語と引数を絞ることは、述語論理における任意性を解消することにつながり、複数ユーザによる知識入力における表現を統一することができる。表現を統一することで、それらの知識を組み合わせる推論を行うことができる。しかし、述語と引数を絞ることによって、推論処理において本来は個別の知識であるはずのものが統一されてしまうという欠点

がある。また、述語と引数を絞り込むことである特定のドメイン特有の述語表現になってしまう。この場合には表現できない知識が増えるため、他のドメインでは利用できないという汎用でないものになる。また、ユーザは知識のドメインが変わるごとに、そのドメインに応じた知識の表現法を学ばなければならない。

本論文で提案する KRNL は、人間が自然言語で思考する時に用いる基盤的背景知識に注目して述語と引数を絞り込むことによって、自然言語文で記された知識ならばどのようなドメインでも利用することを可能にした汎用の知識表現である。また、自然言語との対応の良い形で表現されているので、ユーザは一定の規則（表現法）を学べば専門家でなくとも自然言語から KRNL への変換が任意性少なく行うことができる。逆に KRNL の表現は直接的に自然言語として解釈することができ、変換は容易に行うことができる。

第1部では、KRNL を用いることで新しい知識共有の道が開けることを示す。

第2部では、コミュニティにおける背景知識として人間関係を Web 上から抽出する手法と知識共有システムにおける利用法について述べる。

我々は現実世界においても、または Web 上のオンラインの世界においても多種多様なコミュニティに存在している。そのコミュニティは、大学の一研究室の様に数名から十数名ほどで成り立つ小さいものから、何千人という会員を抱える学会などの学術組織まで、様々である。そのコミュニティにおいて、人間関係は所属するメンバーを特徴づけるのに大きな役割を持っている。たとえば、コミュニティの中心人物ならば彼の発言は他のメンバーに大きな影響を与えるものになるだろう。またコミュニティがいくつかのクラスタに分かれていれば、それが同じ興味を持つ集まりであることを知ることができる。

人間関係を知っていればコミュニティにおけるコミュニケーションも活発にすることができる。たとえば、コミュニティにおける友人の重要度について知ることができる。また、あまり面識のない人物と出会ったとき、自分とはどのような人間関係でつながっているのか知ることができる。

このようなコミュニティの中で、学会は次のような特徴を持っている。

- Web 上で獲得できる情報が多い（特に、情報科学系の学会）
- 研究者は個人個人の線引きが明確である
- 研究活動を効果的に進めるにおいて人間関係が重要である

そこで、我々は最初の対象として学会組織を選んだ。特に、学会における有益な出会いを対象としている。

第2部では、学会における人間関係を自動的に抽出する手法について提案する。この方法は、従来の社会学において多くの質問から人間関係を抽出するのではなく、Web 上の情報を探索することによって人間関係を抽出する。人間関係を考慮することによって、我々は人間関係ネットワークを構築する。この人間関係ネットワークにおいて、ノードはコミュニティ内のメンバーを表し、エッジは二人の人間の人間関係を表すことになる。また、

エッジに付加されたラベルは、人間関係の種類を表している。このネットワークによって、我々は二人の間関係の情報を知ることができる。

このシステムは実際に 400 人規模の学会の全国大会においてサービス運用された。