

問題解決のための知識マネジメントシステムにおける オントロジーの取り扱い

吉岡 真治*1

Masaharu YOSHIOKA

*1北海道大学大学院工学研究科

Graduate School of Engineering, Hokkaido University

高度な問題解決のためには、教科書やマニュアルといった体系的に整理した知識だけでは不十分であり、ノウハウと呼ばれるような十分に整理されていない情報が役に立つことが知られている。本研究では、体系的に整理可能な形式的な知識・情報と文書情報をうまく組み合わせることにより、問題解決のために役立てる知識マネジメントシステムの研究を行っている。本稿では、この知識マネジメントシステムにおいてオントロジーを導入するための方法論とその問題点について述べる。

1. はじめに

問題解決の過程において必要とされる知識は、問題の性質に応じて異なる。単純な作業にたいしては、教科書やマニュアルなどにより整理された知識が有効であり、高度な問題においては、以前の作業記録の参照や、経験に基づいた判断が必要となる。また、後者のような高度な問題においても、問題が具体化した段階では、整理した知識が利用可能な状況になる場合もある。

本研究では、この理解に基づき、作業記録の文書のように形式化が不十分な情報と、論理式で記述できるような形式化可能な情報や知識をうまく組み合わせ、問題解決に役立つ知識マネジメントを実現するシステムを提案している [1, 2]。このシステムでは、問題解決の過程について、問題のレベルを区別することなく、既存の文書を参照しながら新しい文書を作成する過程として支援を行う。このシステムで利用する知識・情報については、定常的な作業については問題解決の過程を記述した文書をテンプレート化することにより形式的に記述し再利用性を高める。また、非定常的な作業については、差分を文書情報として蓄積し、形式的な知識と関係付けて整理する事を行っている。

本システムでは、各々の文書知識について、どのような問題に適用可能であるかという情報と、どのような状況の時に問題を適用可能なかという知識選択のための文脈の情報を一階述語論理により形式的に記述することにより、類似性の判定を行っている。現在のシステムでは、類似性の判定において、一階述語論理の述語部分の名前の一致により類似性の判断を行っている。

本稿では、このシステムにオントロジーを導入するための方法論を述べ、類似性の判定においてオントロジーによる概念階層構造を用いた場合の問題点などについて考察を行う。

2. 問題解決のための知識マネジメントシステム

ここでは簡単に、本研究で提案している問題解決のための知識マネジメントシステム (Documentation Management system for Problem Solving (DMaPS)) について概観する [1, 2]。

2.1 文書知識による形式的知識と文書情報の統合

DMaPS では、中心的な知識ベースを前提とするのではなく、個別のサブ問題ごとに文書情報と問題の対象を表現するための形式的モデル、形式的モデルに適用可能な形式的知識を組み合わせた文書知識を利用する。

本研究では、中心的な知識ベースを前提とするのではなく、個別のサブ問題ごとに形式的知識と文書情報を統合した知識 (以下では、文書知識と呼ぶ) を蓄積する。本研究では、形式的知識を表現するにあたり一階述語論理を採用する。

この文書知識を有効に活用するためには、各知識の適用対象や条件を記述する必要がある。一番簡単な知識検索のためのインデックスは、文書情報を自然言語処理による情報検索のインデックスを作成することである。しかし、文書情報によるインデックスでは曖昧性が高い。例えば、同じ建築構造物の基礎構造を設計する場合でも、その建物が原子力発電所なのか、マンションなのかによって、設計の検討する項目が大きく異なると考えられる。また、建築構造物、基礎構造、評価というキーワードがあっても、基礎構造の評価なのか、建築構造物の評価なのかで、文書の中身も大きく異なる。

よって、DMaPS では、文書知識が持つ形式的知識を利用したインデックスの利用も行う。具体的には、問題の対象を表す簡単なモデルを一階述語論理で記述する。この時、先に述べたような同じようなモデルを用いても、問題が対象としている部分が異なる文書の違いを表現するため、サブ問題の対象を表現した論理式と、問題の文脈を規定する問題に区別してモデルを表現する。

さらに、同じ建築構造物を問題の対象としていても、評価のための知識なのか、設計のための知識なのかによって、文書の中身が大きく異なる。よって、タスクについても問題対象を表すモデルと同様に、形式的な表現を行う。さらに、タスクに応

連絡先: 吉岡真治, 北海道大学大学院工学研究科, 〒 060-8628
北海道札幌市北区北 13 条西 8 丁目, Tel/FAX: 011-706-7107, yoshioka@db-ei.eng.hokudai.ac.jp

じた検索を支援するために、関連するサブタスクがある場合にはその情報についても記述可能にする。

また、今回提案する知識マネジメントシステムは、暗黙知や形式的な情報からより有用な新しい知識を創造するという考え方に基づいたシステムである。しかし、本当に有用な知識を作成するためには、いろいろな事例に適用した結果を踏まえ、知識に対して何度も修正や変更を行い洗練していく必要がある。また、既にマニュアル化されている知識でも、法規制のように、時代の流れに応じて変化していく知識も存在する。

よって、本研究で利用する知識においては、バージョン管理が重要であると同時に、これらの知識を使って行った作業記録に対しても、どのような知識を利用したかを明記することが必要である。

このバージョン情報を用いることにより、古い作業記録について、見直しをする必要がある部分が明確になったり、その知識を使った結果の作業がうまく行ったのかどうかを判断して、知識の良し悪しを検討することも可能になる。

DMaPS では、文書知識を XML を用いて記述する。この文書知識のための DTD を図 1 に示す。文書知識に記述される情報を分類し対応する要素を示すと以下ようになる。

文書情報：name と document サブ問題について文書で記述された情報

形式的知識：rules 各文書知識に対応する形式的知識

問題の記述：model 各文書知識が対象としている問題を表現する情報を一階述語論理のモデルにより表現

タスクの記述：task 各文書知識が対象としている問題のタスクを表現する情報と関連するサブタスクを一階述語論理のモデルにより表現

バージョン情報：version と origin 文書知識のバージョンや参考にした知識に関する情報

この DTD に基づき地盤沈下の安全性評価を例にした文書知識の実例を図 2 に示す。

```

<!ELEMENT template (name, version, origin*, document, model, rules)
<!ELEMENT name (#PCDATA)
<!ELEMENT version EMPTY
<ATTLIST version major CDATA #REQUIRED
<ATTLIST version minor CDATA #REQUIRED
<!ELEMENT origin (name, version)
<!ELEMENT document (#PCDATA)
<!ELEMENT model (target*, context*)
<!ELEMENT target (#PCDATA)
<!ELEMENT context (#PCDATA)
<!ELEMENT task (targettask, subtask*)
<!ELEMENT targettask (#PCDATA)
<!ELEMENT subtask (#PCDATA)
<!ELEMENT rules (rule*)
<!ELEMENT rule (conditions, consequences)
<!ELEMENT conditions (#PCDATA)
<!ELEMENT consequences (#PCDATA)

```

図 1: 文書知識の DTD (template.dtd)

2.2 システムの動作

また、DMaPS では、これらの文書知識を組み合わせながら必要に応じて編集するプロセスになる。ただし、適切な文書知識がない場合はユーザが補完する事となる。この組み合わせた結果は、全体として文書の記録となると同時に、各文書知識のモデルを組み合わせ、問題全体を表現する一階述語論理のモデルが作成される。

ユーザが新たな問題に適用可能な文書知識を検索する場合には、キーワードによる検索に加え、問題全体を表現するモデルと文書知識に表現されているモデルとの類似性の比較を行うことにより、多様な観点からよりユーザの持っている問題に適切な知識をランキングすることが可能になっている。

DMaPS の動作は以下ようになる。

1. 文書知識の検索

DMaPS における類似性の判定は図 3 に示す 3 つの項目を用いて比較し関連する文書知識を検索する。ユーザが持っている問題が一階述語論理の形で表現されている場合には、モデルの類似度を使った検索を行う。自然言語による問い合わせも並行して行うことが可能である。各検索結果は、類似度に応じてランキングが付けられる。

2. 文書知識の選択

検索された文書知識から、適切な文書知識を選択する。文書知識のモデルを組み合わせ、全体の問題を表すモデルを作成する。

3. 形式的知識の適用

全体の問題を表すモデルに対し形式的知識を適用して演繹を行い、新たな検討項目などを導出する。矛盾が発生した場合には、矛盾の解消を行う。また、この時点で全ての問題解決が終わっていれば、作業を終了し、そうでない場合には、1 に戻って、別のサブ問題について適用可能な知識の検索を行う。

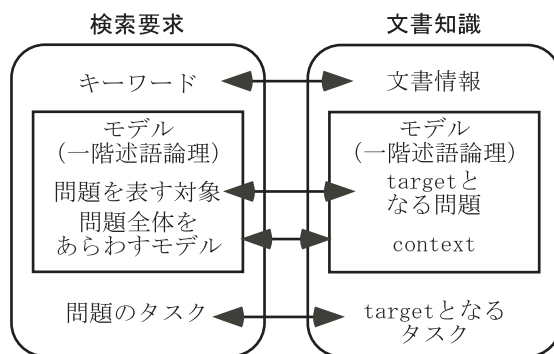


図 3: 文書知識との比較

ここで、キーワードに関する類似性については、文書検索などで用いられている TF・IDF などの指標を用いる。

モデルの類似性については、現在のモデルと文書知識が仮定しているモデルとの類似度を比較する操作である。この類似度の判定では、次に述べるターゲットと文脈による 2 つの比較を行う。

ターゲット ターゲットとなるモデルの要素の述語部分を比較する。述語部分が一致しているとターゲットが一致、そうでない場合はターゲットが不一致と判定する。

文脈 問題領域を表わす全体のモデルと、各文書知識の文脈情報を比較する。文脈情報の比較の目的は、問題領域を表

```

<?xml version="1.0" encoding="EUC-JP"?>
<!DOCTYPE template SYSTEM "template.dtd">
<template>
<name> 建築基礎設計における地盤沈下の安全性評価 </name>
<version major="1" minor="0" />
<origin>
<name> 地盤沈下の安全性評価 </name>
<version major="1" minor="0" />
</origin>
<document>
建物の基礎を設計するにあたり、地盤沈下による安全性の評価は重要である。地盤沈下には、以下の3種類がある。


- 圧密沈下
- 即時沈下
- 地震による沈下


安全性の評価として、各沈下が不同沈下として起こった場合に、基盤底面の破壊が起こるかどうかを判断する必要がある。
</document>
<model>
<target> (基礎構造 基礎構造) </target>
<context> (建物 建物) </context>
<context> (地盤 地盤) </context>
<context> (上 建物 基礎構造) </context>
<context> (接続 建物 基礎構造) </context>
<context> (上 基礎構造 地盤) </context>
<context> (埋設 基礎構造 地盤) </context>
</model>
<task>
<targettask> (安全性評価 基礎構造) </target>
<subtask> (評価 圧密沈下) </subtask> <subtask> (評価 即時沈下) </subtask> <subtask> (評価 地震沈下) </subtask> </task>
<rules>
<rule> <conditions> (地盤 *地盤)(上 *構造 *地盤)(基礎構造 *構造) </conditions>
<consequences> (圧密沈下 圧密沈下)(発生 圧密沈下 *地盤) </consequences> </rule>
<rule> <conditions> (地盤 *地盤)(発生 *地震 *地盤)(地震 *地震) </conditions>
<consequences> (地震沈下 地震沈下)(発生 地震沈下 *地盤) </consequences> </rule>
<rule> <conditions> (地盤 *地盤)(上 *構造 *地盤)(基礎構造 *構造) </conditions>
<consequences> (即時沈下 即時沈下)(発生 即時沈下 *地盤) </consequences> </rule>
</rules>
</template>

```

図 2: 文書知識の実例

わず言葉が違って、そこに関係する状況が等しい場合には、役に立つ知識である可能性があるという仮定に基づくものである。よって、この比較は、単純に同じような論理式が幾つ存在するかというのを数えるのではなく、文書知識中の項とモデルの項のマッピングを行うことにより一致する論理式がどれだけ多くなるかによって比較を行う。

この時、ターゲットとして考えられている論理式の項の部分と文書知識中のターゲットの項は必ず対応付けることとする。また、対応付けを行う項については、かならず、ターゲットの項を含む述語の項である、もしくは、すでに対応づけられた項を含む述語の項であることを制約条件とする。

図4を例にとり比較の説明を行う。まず最初にターゲットとなるモデルの要素の比較を行う。この場合は、ユーザが「(基礎構造 基礎構造)」を問題として選んでいる場合を想定しており、文書知識Bと文書知識Cでは、ターゲットが一致しており、文書知識Aでは、対象となるターゲットが異なるという評価がなされる。

次に、文脈を比較するにあたっては、ターゲットの情報により、文書知識Aにおいては、「基礎構造=工場」、文書知識B,Cに対しては、「基礎構造=基礎構造」という項の対応付けが行われる。その結果、文書知識Aでは、「基礎構造=地面」という項の対応関係付けが行われ、下線の要素の対応関係が付けられる。また、文書知識Bでは、「マンション=アパート」などの対応関係がなされ、文書知識Cでは、「粘土地層=砂地層」などの対応関係がなされて、各々下線の要素の対応関係が付けられることになる。

この比較により、文書知識B,Cは文書知識Aよりもより適切であると考えられる。また、文書知識BとCについては、類似部分と非類似部分を示すことにより、ユーザの知識選択に役立つ情報を提供できると考えられる。

さらに、タスクについても同じように類似性を判断することにより、全ての項目が類似している適切な知識が存在する場合は、その知識を正確に選択することができる。また、全ての項目から類似しているとは言えない知識に対して、タスクが似ているのか、問題の対象が似ているのかといった、類似性に関する説明が可能となる。

3. オントロジーの導入

まず、DMaPSにオントロジーを導入する検討を行うにあたり、グラフのマッチングをクラス・インスタンスという観点から分析を行う。

DMaPS行っているモデルのマッチングでは、項の対応関係について自由度を与えて述語によるマッチングを行っていた。ここで、項を一つだけ持つ論理式の述語をクラス、その項をクラスに属するインスタンスと考え、二項以上の項を持つ論理式によって表現される論理式をインスタンスの関係表現しているものとする。

この時、DMaPSで行っているマッチングは、クラス名が同一のインスタンスあるいは、対応付けがなされたインスタンスに対して、二項以上の述語を持つ論理式によって関係付けられた特定の関係にあるインスタンスを見つける行為に相当する。そのため、本マッチングにおける操作は、クラスの名前がマッチしているインスタンスを中心として、そのインスタンスに対して同等の役割を持つものを発見する操作に相当する。

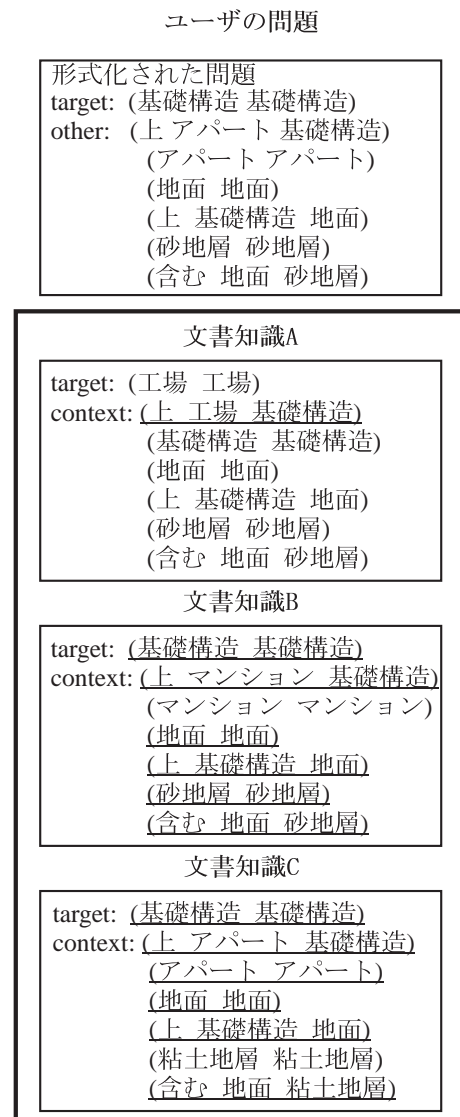


図4: 問題対象を表わすモデルと文書知識のモデルとの比較

一方、図4に示した例のように、マンションとアパートが似ているとか、砂地層と粘土地層が似ているといったクラスの類似性に関する類似性を扱うことができていない。

よって、この比較を行うためにオントロジーによる概念階層表現の導入を検討している。また、このようなオントロジーの記述形式としては、その流通の容易性などを考慮してセマンティックウェブで用いられている技術の利用を考えている。

本稿では、DMaPSにおいて利用する記述レベルとその記述レベルで出せることについて考察を行う。本マッチングの拡張において、概念階層構造の表現が最低限必要な表現レベルと考えている。

この要請を満たすためには、RDF-Schema[3]の利用が考えられる。概念階層構造を利用することにより、クラスに相当する述語の名前が異っている場合においても、述語間の類似度を定義することが可能になり、柔軟な比較が可能になると考えられる。しかし、同じような概念であっても、概念の細分類が違おうと適用可能な知識が大きく異なる場合がある。例えば、図4の例において、砂地層と粘土地層と言うのは、同じ地層概念の細分類と考えることができるが、砂地層の上に建物を建てる場合と粘土地層の上に建物を建てる場合では、適用可能な知識が大きく異なる場合もある。よって、類似性の尺度の計算を行う際に、単純な概念間の類似度だけでは不十分な場合が考えられる。

概念階層構造に加え、より詳細なオントロジー定義を行うことを考えると、OWL[4]の利用が考えられる。OWLを用いて、詳細な概念定義を行うことにより、属性やその他の情報を用いたより詳細な概念間の類似度が定義可能になると考えられ、より適切な類似度の計算が可能になると考えられる。しかし、一方で、オントロジー定義にかかるコストや計算コストが大きくなり、コストに見合った支援が行えるのかどうかについて適切に検討を行う必要がある。

一方、オントロジー構築の立場から、DMaPSで行っているマッチングを考える。先ほど述べたように、DMaPSでは、クラスの名前がマッチしているインスタンスを中心として、そのインスタンスに対して同等の役割を持つものを発見する操作を行っている。これは、特定の目的において、同じ役割を果たすものを見つける操作であり、ある観点に基づくオントロジー構築に対して役立つ情報が見出されている可能性がある。この点についても、さらに検討を行っていきたいと考えている。

4. まとめ

本報では、問題解決のための知識マネジメントシステム(Documentation Management system for Problem Solving (DMaPS))について簡単に説明し、この類似性の判断にたいして、オントロジーを導入することについての考察を行った。

今後の展望としては、システムを改良し、様々な事例に応用することにより、今回検討した内容についての確認と更なる改良を目指していきたいと考えている。

参考文献

- [1] 吉岡真治, 五十嵐優, 社本康広. 文書情報と形式化された知識を融合する知識マネジメントの研究(第1報)- 文書情報と形式的知識の融合 -. 2002年度人工知能学会全国大会講演論文集. 人工知能学会, 2002. 2F2-07 (CD-ROM).
- [2] 吉岡真治, 五十嵐優, 社本康広. 文書情報と形式化された知識を融合する知識マネジメントの研究(第2報)- タスク

の種類に応じた知識の記述 -. 2003年度人工知能学会全国大会講演論文集. 人工知能学会, 2003. 1G2-08 (CD-ROM) (to appear).

- [3] Rdf vocabulary description language 1.0: Rdf schema. Technical report, W3C, 2003. <http://www.w3.org/TR/rdf-schema/>.
- [4] Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, and Lynn Andrea Stein. Owl web ontology language reference. Technical report, W3C, 2003. <http://www.w3.org/TR/owl-ref/>.