

DODDLE: 計算機可読型辞書を利用した 領域オントロジー構築支援環境 (2)

3 A F - 7

- 概念定義構築支援 -

関内 律恵子* 小森 聡* 樽松 理樹** 青木 千鶴* 山口 高平*

* 静岡大学 ** 岩手県総務部県立大学整備室

1 はじめに

文献[1]においては、領域オントロジーの概念階層を構築支援するために、WordNetを利用したがWordNet(を含めてほとんどのMRD)は、概念定義情報をあまり保持しないため、概念定義構築支援には利用できない。本稿では、概念定義情報を保持する数少ないMRDとして、EDR電子化辞書(日本電子化辞書研究所)を利用し、概念定義構築支援の可能性について検討する。

2 EDRの利用とその限界

EDR電子化辞書[2](以下EDRと略記)は、約40万個の概念の定義を与えており、カバーする概念数は多いといえる。しかしながら、その定義は、言語処理の観点から与えられており、8個の比較的一般的な概念関係子(agent, object, a-object, goal, implement, cause, place, sence)のみが与えられているので、領域固有の概念関係子に関する情報は保持されていない。また、保持されている概念関係子の値についても、その値は例示レベルの値が直接的に数多く記述され、集約作業は何ら行われていないので、抽象化等の作業が必要である。以下の節においては、EDRで提供されている概念関係子の値をどのように処理すれば、所与の問題領域における適切な値を設定する上で有用な示唆となるかという点に焦点を当て、検討を進める。

3 概念定義構築支援

文献[1]と同様に、ベストマッチの選択を行い、そのベストマッチ固有の概念関係子ごとに、各概念関係子に割り当てられるすべての値を、再びEDRの概念階層上にマップして、階層木を得る。この階層木を昇降することによって有用な示唆を与えてくれる概念を探索することが課題であるが、本節では、概念階層木の上昇だけに焦点を当てて検討する。

DODDLE: A Domain Ontology Rapid Development Environment Using Machine-Readable Dictionaries(2)
-Conceptual Definition Support-
Rieko Sekiuchi
Graduate School of Science and Engineering, Shizuoka University

3.1 支援方法

最初、概念関係子の値を全て含むように概念階層木の一般化を行ったが、根節近くまで一般化が進んでしまい、有益な情報は得られなかった。

そこで、経験的に概念関係子のすべての値に対して30%以上包含する上位ノードを優先的に探索することにした。また、それらの上位ノードで全ての値がカバーされない場合、30%未満の他の上位ノードも利用することにした。例えば図1においては、一般化により7・8の上位ノードを得て、その後7・8だけではカバーしきれないので、9・10の上位ノードも追加される。

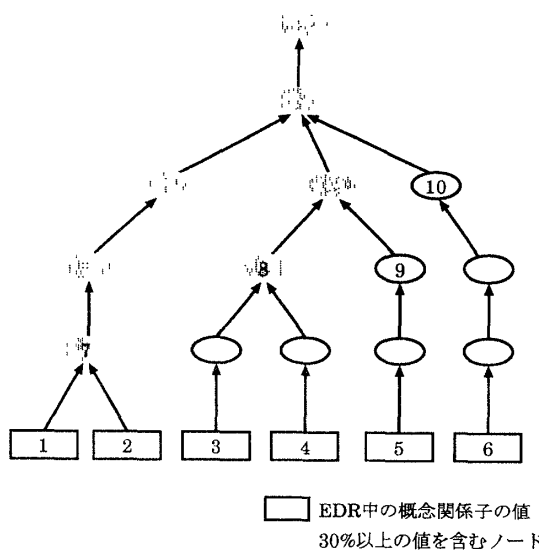


図1: 概念関係子の値による階層木

さらに、ベストマッチにおいて目的とする概念関係子の記述が欠落していたり、あるいは概念関係子の値の総数が少ない場合は、ベストマッチの上位概念の定義を継承したうえで、再度上記の方法により一般化を行う。このようにして、与えられた概念の概念関係子の正しい値を示唆するような上位ノードを複数個得ることになる。

3.2 実験と結果

3.1の概念定義構築支援法を国際統一売買法第2部[3]に含まれる46個の法的概念に適用する実験を行った。

これらの語彙に対し、ベストマッチが存在したものは42概念であった。以下、これらの42概念に対して、ユーザが概念定義（概念関係子とその値）を最初に与え、ユーザの各概念関係子の値と3.1の方法により得られた上位ノードの構造的関係のパターンを調べることによって、概念定義構築支援の可能性について検討した。その調査結果を図2に示すが、支援可能なパターンとしてP1～P3、支援不可能なパターンとしてN1とN2があることが判った。

	P1	P2	P3
支援可能			
(1)	12	5	6
(2)	34	5	10
	N1	N2	
支援不可能			
(1)	7	2	
(2)	5	1	

R:根節点 G:上位ノード U:ユーザ値

図2: 構造的関係のパターン

P1については、(上位ノード)Gと(ユーザ値)Uの距離に依存して支援可能性は決定されるが、距離が大きくても、下降方法を検討することにより距離が縮まる可能性があるため、支援可能とみなした。P2については、根節点からGまでのパス中にUが含まれるので、そのパスを提示することにより支援可能である。P3については、Gの直接の上位ノードから下降方法を検討することによりUまでの距離を縮めることが可能となるので、支援可能とみなした。

一方、N1,N2については、GとUは大きく離れており、Gから上昇した後、下降を開始するノードが不明であったり、下降するための探索空間が大きくなりすぎて、支援は困難といえる。

実験においては、ベストマッチのもつ概念定義を、ベストマッチに固有の定義と、上位概念から継承された定義に分け、まず、前者について実験を行い(図2の(1)の段)、その後、後者について追加実験を行った(図2の(2)の段)。図2の(1)の段を合計すると23ケースが支援可能といえ、ユーザが定義した概念関係子の値の総数(79個)から考えると、約3割が支援可能といえる。一方、図2の(2)の段を合計すると、49ケースが支援可能といえ、支援率は約6割となる。

また、図3に構造的関係のパターンの具体的な実験結果を示す。P1～P3については、GとUの意味は比較的近く、GがUを示唆するために有用であることが判るが、N1とN2については、GとUの意味はかなり離れており、有用ではないことが判る。

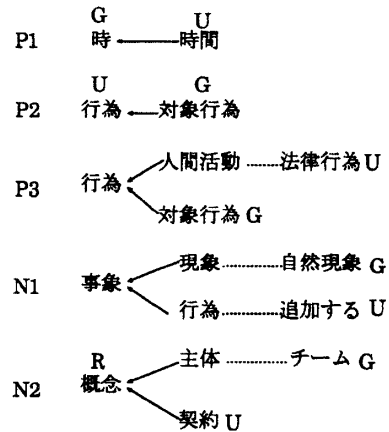


図3: 構造的関係のパターンの例

3.3 評価・考察

3.2の概念定義構築支援実験において、支援率が約6割になりうる事から、本支援方法の適用可能性はある程度は示せたといえる。しかしながら、P1とP3においては、GとUの距離が大きい場合、ユーザに大きな負担を強いることなく、Uに近づいていくように階層木を下降する必要がある。下降時に分岐が多く発生する場合は、自動的に下降する事は困難が伴うことが推測され、領域理論的な知識の活用法等について、今後検討していかねばならない。

また、今回は、上位ノードを獲得するために、全値の30%という閾値を実験的に定めたが、閾値の設定法などについても検討していく必要がある。

4 おわりに

本稿では、MRDを利用した領域オントロジー構築支援環境DODDLEの概念定義構築支援部について述べた。概念定義の構築支援は、利用可能なMRDが限定され、また、本来、概念定義はユーザ(専門家)に大きく委ねられる作業であり、支援には限界があると考えている。しかしながら、概念の意味や構造が問題領域に依存して変化するConcept Drift問題は、AIにおいて、理論的にも工学的にも興味あるかつ有用な技術的課題であり、今後、両面から追求できればと考えている。

参考文献

- [1] 小森聡 他: “DODDLE: 計算機可読型辞書を利用した領域オントロジー構築支援環境(1) - 概念階層構築支援 -”, 第55回情報処理学会全国大会3AF-8, (1997)
- [2] EDR 電子化辞書使用説明書, 日本電子化辞書研究所, (1993)
- [3] 曾野和明, 山手正史: “国際売買法”, 青林書院, (1993)