

オントロジーを集めよう

— World Modeling Project —

Let's Collect Ontologies !

— World Modeling Project —

武田 英明
Hideaki Takeda

奈良先端科学技術大学院大学 情報科学研究科
Graduate School of Information Science, Nara Institute of Science and Technology

Abstract: In this paper, we propose “World Modeling Project” in which existing ontologies in our world should be collected, integrated, and served. Our life is increasingly “digitized”, i.e., more and more things around us and our activities are represented and stored in computers. Such digitization is mainly motivated by commercial advantages so that it tends to create a unique and strict symbolization. But there needs more flexible and rich symbolization to represent our real life. In the project, we will collect a wide variety of ontologies as rich symbolization of our life and offer an information infrastructure of modeling our life.

1 はじめに

いまや、われわれの周りのほとんどのものが電子化されてコンピュータに取り込まれている。ほとんどの人工物は電子的な方法を通じて設計製造され、流通され、販売されている(図1)。ほとんどの情報も電子的な方法で資料収集され、編集され、放送あるいは販売されている。そして、人間の活動そのものも、出生から学び遊び、働く活動も電子的な方法で記録されている。つまり、おおかれすくなかれわれわれの生活はすでにコンピュータにはいつてしまっている(図2)。現在のところ、それほど連携してはなく、孤立したものが多いが、この5年ぐらいにおいてこれらの電子化された情報は急激に関係し合あうことと思われる。すでに多くの情報の標準化が進められている。しかし、のこりの部分は混沌としている。

これは人工知能にとって格好の材料であると思われる。人工知能は主に記号の世界における処理を得意としてきた。しかし、いままでそのような記号の世界はトイワールドのような人工的な設定以外ではみつけるのが困難であった。なぜならば、そのような記号の世界は研究者が自ら構築するしかなく、手法の有効性以前に結果的に小さな世界しか対象にしえなかった。しかし、いま多

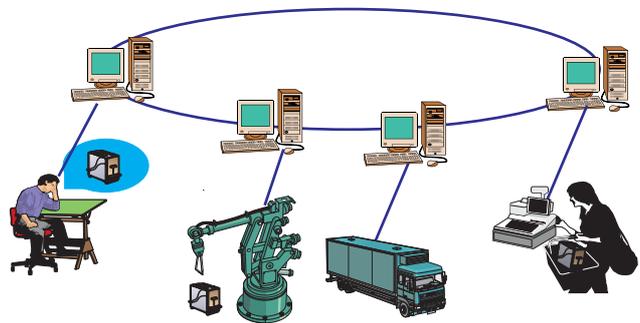


図 1: Artifacts in Computers

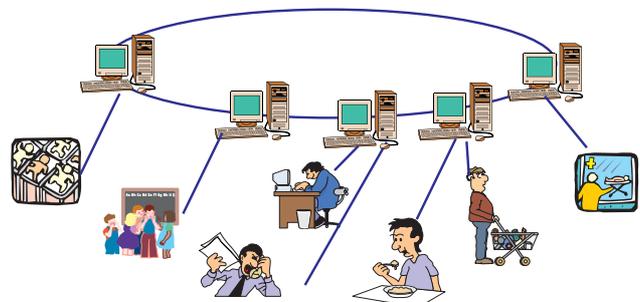


図 2: Life in Computers

数の実環境を反映した記号の世界が構築されようとしており、まさによい実験環境を提供するはずである。

また、別の意味でも人工知能にとって重要な材料である。それは記号化の行為そのものが人間の知的行為のひとつの重要な側面である。こうした電子化された世界は暗黙のうちに様々な分野、視点での記号化のプロセスを含んでいる。これを研究することはまさにわれわれの記号化能力の研究となりうるであろう。

本稿で主張するのは実はこのような研究をするというのではなく、このような研究を進めるための材料作りをしようという提案である。すなわち、そのような記号化された世界をオントロジーと捉え、さまざまな分野、視点でのオントロジーを収集しようという提案である。このプロジェクトの一義的な意味はさまざまなオントロジーのリポジトリをつくり、利用したい人に提供することである。その上で形式化や推論の提案が自由になされるようにするべきである。

具体的なメリットとは例えばつぎのようなものである。例えば、今自分の周囲をモデルが使用する時、我々の周囲には多数の物品が存在する。これらは今まであれば個別にモデル化されている。例えば、目の前のペンを記述したいとおもったときどうしたらよいであろうか。このペンはきっと製造元にはCADデータがあるであろう、流通経路においては文具の何らかの分類に含まれているだろうし、会社の経理にはなんらかの品目として登録されているだろう。あるいはこのペンの機能は（存在すれば）文具の百科辞典に他との差異とともに記述されているだろう。たったひとつの物品でもこのような状態である。ましてや部屋全体ではさらに混沌としているだろう（図3）。

このプロジェクトが狙うところはこのような状況に対して糊のような情報、すなわち関連づけの情報を提供することである。ただし、それは標準化といった一つの視点からものではなく、必要とされる視点に対応するような関連づけであり、一意的なものではない。

この提案の人工知能からみた理由は冒頭で述べたが、社会に対してはべつの貢献があると思われる。上記のような知能の働きからの記号化の研究は、単に効率性から定められがちな情報の標準化に新しい視点、人間からの視点による情報の構造化の方法論を提供すると期待できる。これは人間を中心とした情報化においては重要な役割を果たすと思われる。

2 なにをするのか

本プロジェクトの目的は2つある。一つは広く社会や研究コミュニティに世界の記号化を共有させることである。局所的に利用されている世界の記号化の情報を収集して、他の分野や視点の記号化と関連づけることによ

り、多様な利用を可能にする。二つめはオントロジーの研究者のための研究資源として提供することである。これは一つめの目的を実現するための機構を開発することでもある。

したがって、このプロジェクトが行なうべきことは

1. オントロジーの収集
2. オントロジーの整理
3. オントロジーの供給

である。このうち収集のフェーズにおいては、その形式や内容に拘らず、広く収集を行なう。整理のフェーズでは収集したオントロジーを相互に関連づける。このフェーズにおいては、研究開発と並行して、実施する。特に形式性の度合に則したオントロジー相互関係がこのプロジェクト自身を遂行する上で必要とされる研究である。そのような関連づけされたオントロジーを再び外部へ提供する。

3 オントロジー収集の方法論

オントロジーとは”概念化の仕様”であるといわれている[1]。ここでいう概念化とはどのように対象世界を記号的に表すかという方法である。もう少し詳細に言えば、オントロジーは語彙とオントロジーのための理論(ontological theory)からなるといえる[2]。語彙はどのような記号を用いて表現するかに対応し、オントロジーのための理論とはこれらの記号の間の関係であり、対象世界に存在するそれぞれの記号に対応する存在物間の関係を規定する。このオントロジーのための理論をどの程度の詳細度を考えるかによって、オントロジーのレベルが分けることが可能である[7]。例えば[6]では

- レベル1：概念の切り出しと階層的記述
- レベル2：意味定義の明示化・形式化
- レベル3：構築されたモデル実行に関わる計算論

と分けている。

このようなオントロジーは実はすでに多く存在する。実際、なんらかの人間の活動が存在する分野において、以下のいずれかの意味でのオントロジーが存在しない分野はありえないと思われる。

1. 網羅的でない分類：カテゴリーとその関係が示されるが、網羅的でもなくまた重複などもありうる分類。例)料理の本の料理一覧、Yahooのディレクトリ。
2. 網羅的な分類：その分野を網羅するカテゴリー。例)職業別電話帳の職業分類、商品分類

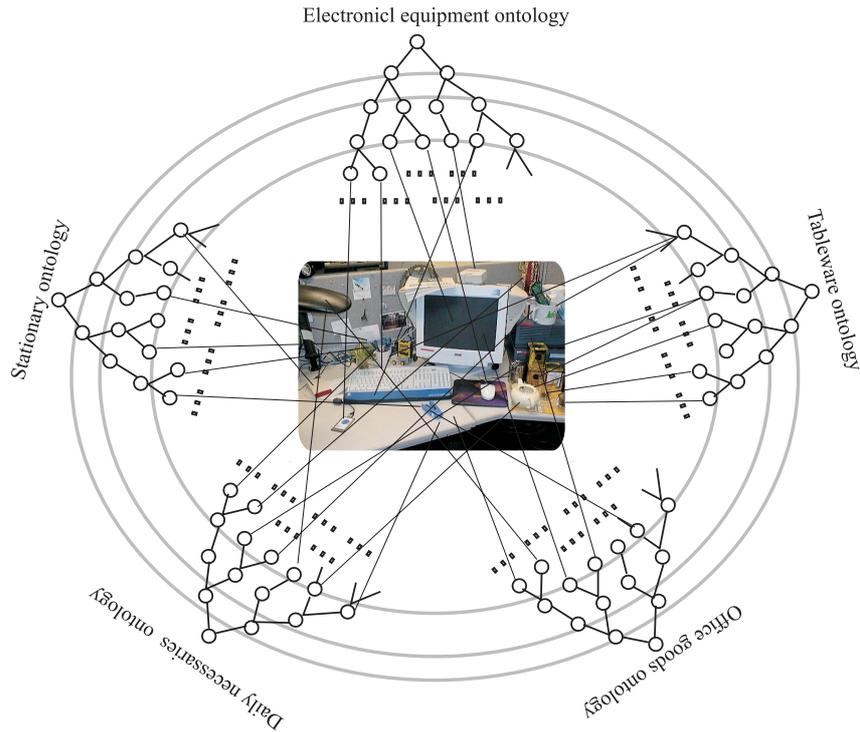


図 3: Ontologies for Your Desk

3. 定義のある分類：さらになんらかの形で各カテゴリーに定義がなされているもの。例) 工業規格 (JIS など)
4. 相互の関連のある記述：各カテゴリーあるいは概念の振舞いがそれらの相互関連として記述されているもの。例) 教科書や教習本の内容

1-3 は分類であり、4 は関係の記述が主であるので、上記の溝口分類には必ずしも対応しないが、1と2はレベル1であり、3はレベル3、3と4を組み合わせたものがレベル4に相当すると思われる。

これは当然、ある分野において唯一存在するものではなく、利用の目的からなどから多数存在すると思われる。例えば、1章で例題とした日常の物品であれば、JISにおける分類、製造会社における分類、流通経路における分類など、類似しているが観点の異なる分類が存在する。これらはどれが正しいというのではなく、どれも有用なものである。

対象分野としてはさまざまなものが考えられるが、複数の視点が存在するものが興味深い。例えば先に挙げた日常の物品のほかにも、産業・職業（各種産業統計、電話帳など）、人間の活動（法律、企業モデリング、自然言語など）など、すでに一定のオントロジー/分類をもつ分野は多数あると思われる。

収集にあたっては、形式化の度合いなどによって排除したり変形するのではなく、形式化や粒度が異なるオントロジーを広く受け入れるべきである。そのような非均質性がオントロジーの多様性のひとつの現われだからである。

4 オントロジー整理の方法論

先に述べたようにひとつの対象に対して多くの関連するオントロジーが存在する。これらの関係を発見して、利用可能にする統合化機構の開発が必要となる。先に述べたように、ここで目指しているのは標準化のような一意的な構造を作ることではない。むしろ多様な視点に対して、その視点に則した関係を提示できるような機構が必要である。この部分は研究トピックスであり、今後の進展が期待される。例えば、これまで [5] では多重オントロジーの研究があり、[4] では異なるオントロジーを併用してオントロジーを開発する仕組みが研究されている。

ただし、先にも述べたように形式化のためにオントロジーの多様性を阻害してはならない。すなわち、この統合化機構においては、例えば形式性が少なければ弱く疎

Cyc (例えば [3] の問題点はまさにこの立場が逆転している点である。あるものがあるがままにみるのではなく、自らの形式化の都合を優先させている。

な関係を形式性が強ければより強固で密な関係を見出すような、形式化の度合いに応じて異なる対応ができるようにしなければならない。

5 実施体制

このプロジェクトは、その性格上大学等の研究者の個人的研究や企業における一研究ではなく、産官学による公共的プロジェクトとして実施する必要がある。また、情報研究者に限らず多様な分野の人間が参加することが必須であるので、活動の多くは研究者の自発的参加に追うところが大きいと思われるが、bootstrap としての中核部分は一定の人月をかけて構築すべきである。成果は当然、広く公開すべきものである。

6 期待される研究成果

第一に、世界の記号化のインフラストラクチャとしての役割が期待される。具体的な成果としては収集されたオントロジーそのものとその組み合わせを可能にする統合化機構である。個々のオントロジーは特定の領域の特定の視点による概念化に過ぎないが、統合化機構により横断的、目的指向的に構成し直すことができるようになる。このため例えばモデル構築あるいは新しい標準化のデータなどの多様な利用が可能になる。例えばサーチなどにおいて位置指向サーチ [8] からさらに進んでもの指向サーチなどが可能になるであろう。

また研究的側面としては、対象世界と思考の関係の基礎データとして利用が期待される。例えば、人間の記号化能力の実際的な研究などがこのデータを利用して可能だと思われる。あるいは知識処理分野においては、知識発見の基礎データなどに用いることが可能であると思われる。

7 おわりに

我々の住む世界そのものを計算機の中に実現することは人工知能研究者の夢の一つであった。それはいま現実化しつつある。しかし、それは商業的利用の側面のみが協調され、歪んだ形での実現になりつつある。いまこそ、人工知能研究者が人間の立場、すなわち人間の思考や行動の立場からコミットするべきであろう。これは人工知能の研究を社会と結び付ける重要なチャンスであると思われる。

謝辞

本稿作成にあたっては、(株)けいはんな主催の I3 研究会での議論がよい刺激になりました。とくに、NTT

流通基盤プラットフォーム研究所の高橋克巳氏との議論は大変参考になりました。また、本学の伊藤英毅氏にも議論に参加してもらいました。あわせて感謝いたします。

参考文献

- [1] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University, August 1993.
- [2] N. Guarino. Formal ontology and information systems. In N. Guarino, editor, *Formal ontology in Information Systems. Proceedings of FOIS'98*. IOS Press, 1998.
- [3] R.V. Guha and D.B. Lenat. Cyc: a midterm report. *AI Magazine*, Vol. 11, No. 3, pp. 32-59, 1990.
- [4] R. Sekiuchi, C. Aoki, M. Kurematsu, and T. Yamaguchi. Doodle: A domain ontology rapid development environment. In *Proceedings of PRICAI'98*, 1998.
- [5] Hideaki Takeda, Kenji Iino, and Toyoaki Nishida. Agent organization and communication with multiple ontologies. *International Journal of Cooperative Information Systems*, Vol. 4, No. 4, pp. 321-337, December 1995.
- [6] 溝口理一郎. オントロジー概論. 日本ソフトウェア科学会チュートリアル「オントロジー」, number 22 in 講演会資料シリーズ. 日本ソフトウェア科学会, 1999.
- [7] 溝口理一郎, 池田満. オントロジー工学序説 — 内容指向研究の基盤技術と理論の確立を目指して —. *人工知能学会誌*, Vol. 12, No. 4, pp. 559-569, 1997.
- [8] 高橋克巳, 三浦信幸, 横路誠司, 島健一. Mobile info search: Information integration for location aware computing. 情報処理学会 第6回モバイルコンピューティング研究会, 電子情報通信学会 インターネット研究会, 1998.