

研究論文

知的インフラストラクチャ構想

— 情報文化学的存在としてのインターネット —

Intelligence Infrastructure Initiatives

— The Internet : Its Information-Culturological Aspect —

天野 真家 Shin-ya AMANO

湘南工科大学 工学部 情報工学科

Department of Information Science, Shonan Institute of Technology

稲垣 耕作 Kosaku INAGAKI

京都大学大学院 情報学研究科

Graduate School of Informatics, Kyoto University

三浦 康之 Yasuyuki MIURA

湘南工科大学 工学部 情報工学科

Department of Information Science, Shonan Institute of Technology

要 旨

地球上にはいまやインターネットが張り巡らされ、世界のどこで何が起きているのかリアルタイムで看視することができる。マスコミはニュースを WWW サーバから報道し、種々のセンサーネットは局地的なデータを刻々と発信している。個人の情報発信は引きも切らない。今やインターネットは良きにつけ悪きにつけ人間の生活に深くかかわっている。さまざまな深刻な凶悪犯罪さえも、そこに予告され実行されている。しかしながら、そのような重要な情報があっても発見することは容易ではない。一つの原因はインターネットに知性が存在しないことに存する。もし、インターネットに知性を持たせることができれば、このような犯罪予告は早期に発見され実行を未然に防ぐことができたかもしれない。目を転じれば、環境破壊という問題がある。それは様々な思わぬ要因が絡んで起きる予測困難な複合汚染の側面をもっている。経済政策、社会政策も予測を超えた思わぬ影響を人々の生活に与えている。この種の人間には予測困難な影響もインターネット上の情報から推論により予測できる可能性がある。インターネットはすでに単なる技術として論じるだけのものではなく、人間社会に深く関与した存在として情報文化学的に論じられる必要が生じていると思われる。本論文はインターネットに知性を与え、近未来を予測する技術と構想の必要性を情報文化学的観点で論じたものである。

Abstract

The Internet prevails all over the Earth and provides people with means of watching news, events, weather and so on in real-time. Mass media distribute news, various sensors deliver local data every second and even individuals open their life records or opinions through the Internet. The Internet pervades every aspect of people's life both with merits and demerits. Even serious crimes were and maybe are preannounced on the Internet. If the Internet acquire intelligence and watch itself, it will prevent the crimes by finding such posts among piles of information trash. On the other hand, the multiple reasoning over the information on the Internet may produce very important information that humans can hardly discover. It will also foretell the serious influence of environmental destruction, propagating effects of economic changes and social policies in the near future. This paper proposes the Intelligence Infrastructure Initiatives that aim at giving the Internet the intelligence from the standpoint of the information-culturology.

1. まえがき

インターネットは WWW が出るに及んで研究者だけのツールから一般社会での地位を占めるようになった。そして、今、その機能は社会に大きな影響を与え始めている。誰もが簡単に情報を発信できる手段を提供したことによる。個人の情報発信は引きも切らず、今やインターネットは良きにつけ悪きにつけ人間の生活に深くかかわっている。膨大なインターネット上の情報から有益な情報を得るために検索エンジンがある。しかし、Google^[1] はキーワードによる統計型のエンジンであるために、検索結果は膨大になり、その結果を検索者が自ら再検索しなければならぬ。True Knowledge^[2] は未完成ながら、

一段進んだ検索エンジンであり、知識が自然言語で書かれた Wikipedia^[3] とは異なり「コンピュータが理解し処理可能な離散的形式で記述される」。日本でも次世代検索エンジンのための「情報大航海プロジェクト」^[4] が始まっている。

しかし、これらはなお検索エンジンの枠を超えていないと考えられる。本論文で提唱する構想はインターネット上の情報を動的な公理とみなして推論を行い人間が見過ごしてしまいがちな何段もの推論を必要とする帰結を得て近未来予測を行うものである。

現在では、さまざまな深刻な凶悪犯罪さえも、WWW 上に予告され実行されている。2008 年の「秋葉原連続殺人事件」は記憶に新しい。また、贖犯罪を予告するという形態で威力業

務妨害犯罪が新聞紙面を賑わしている。もし、インターネットに知性を持たせることができれば、このような犯罪予告を自動的に発見でき、実行を未然に防ぐことができるだろう。

また、一層積極的な有効利用という観点からインターネットを情報源とみれば、それは「顕在はしていないが極めて有用な情報」を内包している可能性がある。「風が吹けば桶屋が儲かる」のような何段もの推論で人間の推理能力の枠を超えて重要な情報が得られる可能性がある。ガソリン消費量を削減するためにバイオ燃料開発を推進することがマヨネーズや豆腐の価格を上げ、更には食糧危機を引き起こし、さらにそれがきっかけで重篤なインフレを引き起こすという予測が可能になるかもしれない。地価の高騰を抑えようと単純に税や公定歩合を上げたことが深刻なデフレを引き起こし、首都圏での不動産取得が多少困難にすぎないという地域的現象ですんだものが、全国的な不況となり、多くの失業者を生み出すという事態になることが、事前に推測できていたかもしれない。同様に、環境破壊も様々な思わぬ要因が絡んで起きる複合汚染である側面を持つが、人間の苦手な何段にも重なる推論が事前に可能になれば、環境アセスメントも容易になる可能性がある。かつて、ユナイテッド・カリフォルニア銀行の頭取 Paul Erdman^[5] は早魃を予想して世界のココアの半分の量を先物で買い、結局この投機に失敗したが、気候観測ネットワーク情報が当時、世界規模で存在していればもう少し精度の高い予測ができた可能性は高い。

以上の可能性には事例がある。第二次世界大戦時、ロンドン在住の民間人がドイツ軍の大規模な移動について記した書物を出版してヒトラーを慄然とさせた。この情報は諜報によって得られたのではなく、どの新聞にもある地域の名士の結婚記事、死亡記事のような公開情報からの推論で得ることができたものである。

「ハーゼ少将が第十七師団司令官でヌーレンベルクに駐在するというくだりは、ヌーレンベルクの新聞の死亡欄から得た情報です。この新聞の死亡記事欄に、最近移住した第十七師団の司令官としてヌーレンベルクに赴任して早々のハーゼ将軍が葬儀に参列したと出ていたのです。」

—— 参考文献^[6]より引用

一民間人がこのように新聞情報を丹念に集めることにより、ドイツ軍の戦闘序列を構築できたのである。さらにアメリカに8部しかない極秘文書である軍増強計画書でさえ新聞情報を収集、推論することにより再現することが出来たという。アメリカの秘密情報の95%は新聞などの定期刊行物に発表されているとトルーマン大統領は述べている^[6]。

インターネットはすでに単なる技術として論じるだけのものではなく、人間社会に深く関与し影響を与える存在として情報文化学的に論じられる必要がある。コンピュータには人工知能という知性が備わっているが、ネットワークにも知性を持たせることにより近未来を予測し、人間社会に深刻な影響を与える問題に対して何らかの回答を出す能力を与えることができるのではないか。そのような機能を実現する構想を、筆者らは本論文において Intelligence Infrastructure Initiatives (III: トリプル I) として提唱する。

2. Intelligence Infrastructure Initiatives の基盤としての ユビキタスコンピューティング

インフラとしての意味で、ユビキタス・コンピューティングという概念が使われ、コンピュータが「遍在する」環境が強調されている^{[7][8][9]}。このコンセプトは Xerox PARC の Mark Weiser が最初「The Computer for the 21st Century^[10]」の中で用いたコンセプトを拡張的に解釈したものである。M.Weiser のコンセプトの性質をより端的に示すことばを使えば、欧州連合が提唱している「disappearing computer」、あるいは、M.Weiser 自身の論文の中で繰り返し強調される概念「pushing computers into the background」であり、コンピュータ本体を人間の目に触れさせず、人間が本来行いたいことをコンピュータの操作という間接的手段によるのではなく、直接に行えるようにすることにより人間活動を支援するという意味であった。

人間が「行いたい」ことの対極として、人間が「知りたい」事を教示してくれるコンピューティングは可能であろうか。これが III の動機である。単に静的な知識を知りたいというだけなら、既に Wikipedia 運動がある。III では、現時点で明示的には存在していない近未来の知識を得る事を目標とする。それにはユビキタスコンピューティングが前提になる。

M.Weiser は、コンピュータの理想的な在り様としての「the computers themselves to vanish into the background」により、「(神は) 見えない」というコンセプトをこの ubiquitous という言葉に託したかもしれないが、一般には明示的にはそのような解釈よりも、「遍在」に重点が置かれ、「どこでもコンピュータ」と解釈されている。すなわち、ユビキタスと言うコンセプトで言われているのは、以下の2点である。

- 1) コンピュータは人間の目につかないように置かれ、使用されるべきである。
- 2) 様々な物の中で使われる。つまり、遍在する。

本論文で提唱するコンセプトと技術は、地球規模で張り巡らされたネットワークにこのような「ubiquitous」形態で接続されているコンピュータ環境を更に徹底的に、高度に利用しようとするものである。このような状況は単体の機械 -- 例えば自動車やテレビ -- が世界中に普及しているという状況とは全く異なる。地球を覆っている膨大な機器群が有機的に一台の機器として機能する可能性を秘めているのである。さらに、ネットワークの末端に接続されているものはいまや単なるコンピュータだけではない。たとえば、センサーがある。カメラ、電子百葉箱、RFID 用センサーなど多様である。いながらにして世界で起きている時事情報、イベント情報、天候情報、地域情報などが処理可能なデジタルデータとして入手できる。このような地球規模の情報を使い、演繹推論を施せば、今現在起きているイベントだけでなく近未来に起きる事象を予測できる可能性がある。「風が吹けば桶屋が儲かる」は演繹推論の極致であるが、III では、相当の信頼性を持ってこのような予測を立てることが、データの裏づけを伴い、かつ推論過程を伴い可能になることを目指す。分野により扱う情報の種類、専門とする

世界知識が異なるため、分野毎の多くの ontology が必要となる。それらを開発する機能を備えたネットワークコンセプトが Intelligence Infrastructure Initiatives である。図 1 は、その概念図を示す。

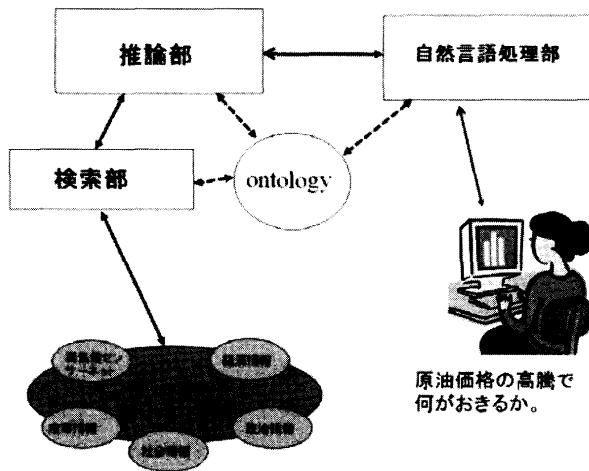


図 1 III の概念図

図 1 は、質問文に対して自然言語処理を行って形式的問い合わせ言語に変換し、ontology を用いて WWW 上の情報に対して演繹推論を行い、多くの情報を得るようである。それらはそのままでも情報として有用である可能性があるが、推論結果を ontology 上に置かれた評価用の知識を用いて評価し、有用性の判定を行うことが必要である。III では、このように ontology を言語理解、推論、結果評価などさまざまな知的活動のための知識ベースとして用いることを想定している。

3. 情報文化的存在としての「考えるネットワーク」

3.1 従来のネットワーク利用の限界

2 節で論じたネットワークの在り方をここで再び整理してみる。

電話	距離の克服
WWW	情報源 (距離の克服), 人間活動の代替
ubiquitous	人間活動の見えざる支援

通信はその動機において遠隔の二者があたかも近接しているようにすることを目的としている。これは情報伝達における「距離の克服」である。インターネットにおいても、この点は変わらなかった。インターネットの技術的前身である ARPANET においては、コンピュータの遠隔操作 (telnet) とメール、さらにファイル転送が主要アプリケーションであった。その後、WWW が出て、距離の克服だけでなく、サーバという形で人間活動の代替を行うことができるようになった。さらに、CGI のようなデータベースとのインタフェース、およびスクリプト言語処理系、JAVA アプレットに代表されるネットワークロードバランシング&エグゼキュータブルモジュールのような機能が開発されるにつれ、WWW サーバは単にファイルを転送し、それを閲覧するという機能から、窓口サービスを行うものに変質し、背後にある様々なサーバとの連携、利用者が意識しないブ

ログラムの実行という機能を持つようになり、人間活動の支援を行なうことができるようになった。この機能は、まさしく M.Weiser が言及していたユビキタスコンピューティングの機能の一つであるといえる。ショッピング、バンキング、あるいはブログのような簡易情報発信などであり、これらは、携帯電話、軽量小型の PC + 無線 LAN + ホットスポット、あるいは、日本においてはほとんどの家庭、企業にも備えられたユビキタスの存在である PC + インターネットによって利用できる。「遍在」あるいは「背後に隠れるコンピュータ」のどの意味においても、ユビキタスコンピューティングがまだ不器用な姿ながら実現しつつあると言える。

しかし、このようなネットワークあるいはコンピューティングは、利用者に無意味で困難なコンピュータの操作をさせないように背後にコンピュータを隠すという消極的なコンピュータ利用形態にすぎない。コンピュータやネットワークが表面上姿を消すだけではなく、知性を持ち、もっと積極的に我々の生活に有意義に関わってくることはできないものであろうか。それが III の動機である。

3.2 考えるネットワーク

インターネットは今や地球を覆っている。そこに上げられている情報は、単独ではほとんどの人には何ら関係がない「A 氏と B 嬢が結婚した」というような情報が圧倒的であろう。更に言えば、我々が現在 WWW で見ている情報はそのような断片的情報でしかない。しかし、トルーマン大統領がかつて報道機関に自粛を求めたように、それらのなんでもない情報群は推論により組み合わせられると、極秘情報に変質する可能性を孕んでいる。このことは単に諜報の世界に限定されるものではない。より重要なことは社会全体にかかわる有益な情報を得る情報源になる可能性があることである。

III において有益な情報とは近未来予測をさす。それは自然環境変化の予測であったり、経済予測であったり、犯罪抑制であったりと多面にわたる。ハブを退治するために導入したマンガースが薄明性のため、夜行性ハブの退治に役立たず、逆にヤンバルクイナを捕食するなど生態系に悪い影響を与えていると言われるが、そのようなことは、当時 WWW があれば、そして III があれば一層容易に予測できたのではないだろうか。

潜在的に有用な情報はインターネット上に溢れているであろう。そのような情報が、組み合わせにより「情報の synergy」を生み出すと考えるのは自然な発想である。

この観点から III は次の特徴を持つ。

1. 動的知識はネットワークから得る：ニュース、天候など時限的動的知識をネットワークからリアルタイムで取得する。
2. 推論を進めるメタ知識を ontology としてもつ：「風が吹く」から「桶屋が儲かる」までをつなぐ推論を進める知識である。

動的知識とは時限的に真である知識である。「東京で雨が降っている」、「原油価格が高騰している」など。メタ知識は推論過程を制御する知識である。「原油価格が高騰すると、x に対し

でYという事態が生起する」という形式の知識である。ここにYは述語、xは対象変数であり、Y(x)という形をとる。

このメタ知識をオープンソースとして開発することがIIIの目標である。

4. 必要な技術

本システムの実現に必要な技術を列挙すると、図1を詳細化した図2に示すようになる。これらの技術の現状とIIIを実現するために今後研究が必要な技術を検討する。

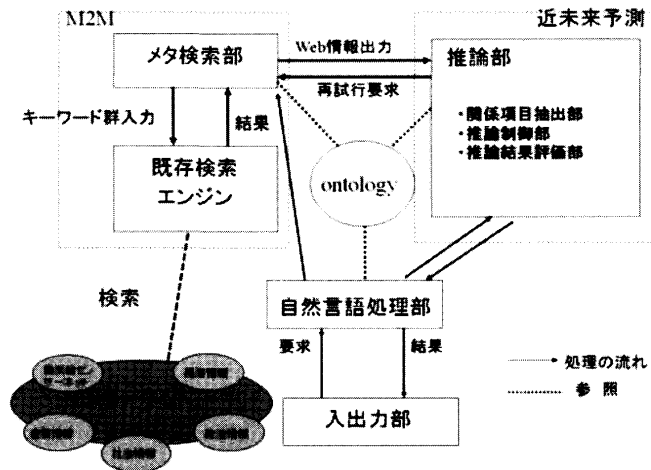


図2 IIIシステムの技術的概念図

4.1 自然言語処理部

利用者と末端のコンピュータとは自然言語で会話するので、自然言語理解の技術が必要になる。

利用者の要求文を入力する機能、要求文を構文・意味解析して「メタ検索部」で用いる「問い合わせ言語」に変換する機能、WWWサイトから検索で得られた文章を言語理解した上で意味構造に変換する機能を担当する。この意味構造は「問い合わせ言語」で表現される構造と同じ形態とすることにより、III内の言語的内部表現はすべて統一された構造をもつようにすることができる。

自然言語理解に必要な技術は下記のものである。

- 辞書引き・形態素解析
- 統語解析
- 意味解析
- 運用論的解析

4.1.1 辞書引き・形態素解析

形態素解析の技術は1970年代以来、特に仮名漢字変換、機械翻訳の研究を通じて蓄積されている。茶釜^[11]のようなフリーのものも存在している。しかし、構文解析、意味解析と辞書を統一することが望まれるので、新たな開発が望ましい。辞書引き・形態素解析の役割は2種あり、入力される質問文の仮名漢字変換と、WWWサイトから得られた漢字仮名交じり文の単語の認定と文節への切断である。構文解析以上の解析は文節単位で行われるからである。

仮名漢字変換のための形態素解析における同音語^[1]の問題は意味解析、運用論的解析を通してしか解決できないので、通常、形態素解析の結果は曖昧性を含むものとなる。選択制限規則を用いれば、「あつい」を「厚い(本)」,「熱い(湯)」などと原理的には分離できるが、「熱い(鉄板)」,「厚い(鉄板)」のようにどちらも可能な場合には選択制限では分離できない。この問題の解決には運用論的解析が必要となる。運用論の中でも特に文脈解析を用いてどちらかに決定することになる。

仮名漢字変換の過程はシステムと人間との対話の過程でもあり、しかもこのヒューマンインタフェースは日本語ワードプロセッサ以来非常にポピュラーなものになっているので人間側に抵抗がない。システム側には、この過程においてある程度の自然言語処理が対話的に行われる結果、辞書引き・形態素解析レベルの問題、更には上記の語彙の意味レベルの曖昧性は同音語の決定という過程を通してほぼ解消されるというメリットがある。しかも、この過程における誤りの修正法に関しては、仮名漢字変換として周知されているためにシステム側からプロンプトを出す必要がないというメリットがある。

一方、WWWサイトから得られた漢字仮名交じりのベタ文は形態素解析で文節に切り分けた形にして構文解析部に送る。このとき、自立語に関しては辞書引きで得られた品詞をはじめとする文法情報を用い、付属語部分の文法情報は形態素解析で得られた情報を用いる。

4.1.2 構文解析

構文解析は現在においても困難な問題である。現在の構文解析には、

- (1) 文法が不足していて、正しい文を受理できない
 - (2) 文法に誤りがあり
 - 1) 正しい文に対して誤った解釈を行う
 - 2) 誤った文(非文)を正しいと誤認する
 - (3) 曖昧性をもった解析をおこなう
- の3種の問題がある。

(1), (2)は文法開発の問題である。1980年代以来企業において機械翻訳システムの開発競争が盛んになり大規模な文法開発がおこなわれてきたが、現在に至るもこれらの問題を抱えたままであり、更なる研究の必要がある。

(1)と(2)の問題はトレードオフの関係にあり、(1)の問題を解決するために規則を追加すると、その副作用で(2)の問題が広がるという状況を呈する。その原因は品詞体系、各種の語彙的属性の体系、文法の適用条件などの設計にあるため、容易に解決できない。たとえば、ある問題を解決するために品詞細分を行い、品詞を1つ増やすということを行うと^[2]、それに伴って辞書の見直し、文法規則の見直しが起こり、大規模システムにおいてはそのため結果する事態を予測することも困難である。このため、トレードオフがある状態に達した時に、開発を停止させざるをえなくなり、その時点でその文法能力が決まるのである。

(3)の問題は理論的には意味解析、運用論的解析を用いなければ解決できない。

IIIにおいてこれらの問題をどのように解決するかを以下で論じる。

(1) 文法が不足していて、正しい文を受理できない場合（解析が中断する場合）

この場合、解析器は構文解析の失敗を認識している。構文解析が終了状態まで到達せずに中断するので、解析器は失敗を認識できる。この問題に対して、言語理解システムの人間へのフロントエンドである入出力部は「理解できません。別の表現で言い直してください」などと対応することができる。よりきめ細かい対応をするなら、失敗原因をいくつかの場合に分類して、それに応じて対応を変える。たとえば、あまりに長い文の場合、「もう少し短い文にしてください」とするなどで対応する。

なお、文法的に正しくない文（非文）の場合にも、この、終了状態に達する以前の解析中断という現象が起きる。非文に対しては当然、文法規則は用意されていないから「文法不足」という現象としては同じである。構文解析器はこれら2つの場合を判別できないので、「理解できません。別の表現で言い直してください」のような両方の場合を含めた応答にする必要がある。

(2) 文法に誤りがあり（解析が正常終了する場合）

1) 正しい文に対して誤った解釈を行う場合

2) あるいは、誤った文（非文）を正しいと誤認する場合

文法は解析器にとって公理なので、このいずれの場合も、構文解析器は解析の失敗を認識できない。これはシステムにとって構文解析は成功しているのであるが、人間が調べると誤っているということになる場合である。これによるシステムの人間に対する不正な応答は、専門家なら文法の誤りのために起きた現象であることに気付くであろうが、一般ユーザにとっては意味不明の応答になる。

この場合は、構文解析は失敗ではないので、処理は次の段階に進む。意味解析、あるいは運用論的解析で構文的誤りの存在を発見できれば、その時点でなんらかのメッセージを出力して対応することができるが、一般にシステムは下位ルーチンは成功していると仮定して設計されるので、実際には誤りを見つけることは困難である。人間の要求はこうして最後まで実行され推論結果が表示される。この時点で、利用者は要求したものとは異なる回答であると認識し、新たに要求文を入力しなおすことになる。システムにとっては正しく、人間にとっては正しくないという現象に対しては、人間の再試行しか解はない。

IIIにおいては、機械翻訳と異なり対話が主機能であるので、長く複雑に入り組んだ文は扱わないように対話そのものを制御できる。この特性により、質問文の入力においてはこのような状況を抑制することはかなりの程度可能であろう。しかし、WWWサイトの文章の解析においては、何らかのヒューリスティクスを用いることが必要となるだろう。部分的な失敗には成功部分だけから意味構造を作る試みをする、失敗した文は捨て、他の文に意義ある内容があることを期待する、などである。

4.1.3 意味解析と運用論的解析

(1) 語彙的意味の解析

語彙的な意味の曖昧性は同音異義語としてかな漢字変換過程でほぼ解決されると考えられる。ただし、平仮名のままの自立語、あるいは同音同字異義語⁽³⁾も存在するので細部では意味解析は必要である。しかし、意味解析を行うために辞書のすべての単語について最初から意味記述をすることは「世界」を完全に記述するということになり、理論的には可能であっても、工学としては開発を行う人的資源の問題で可能ではない。より充実した語彙的意味の記述をおこなうために、ontologyを公開にして、オープンソースのような方法で開発することが必要となろう。

(2) 質問文の意味の解析

自然言語理解を行うためには一般には「世界」の知識が必要になるが、IIIが想定する利用法は、対象とする世界が図2に示すように、インターネットに含まれる世界全体になり極めて広い。一方で、利用者の意図は基本的にはむしろ狭いと考えられる。「何かについての情報を知りたい」という一点に絞られるし、当面はそのように絞ってもよい。例えば、「Twenty In Ten 構想⁽⁴⁾で、どのような影響が考えられるか」、「今後10年でバイオ燃料を現在の7倍に増加させるとどのような影響がどの分野に生じるか」というような経済上の質問が想定される。この観点からすれば、むしろIIIで用いる質問文の自然言語理解は容易であると考えられる。また、意味論的問題よりも、現実世界との接点となる運用論的問題のほうにより重点が置かれる。上記の例では「Twenty In Ten 構想」が何であるかを知らなければならない。「影響」が具体的には何であって、どのようにして調査できるのかを知らなければならない、それに対する知識はあらかじめontologyに記載されている必要がある。

自然言語の質問文は形式言語である問い合わせ言語に変換する必要がある。問い合わせ言語の形式は、文の意味を示す意味表現形式であると同時にIIIにおけるontology表現形式でもあるように統一を図る。ontologyの項目はフレーム形式でXMLで表現して標準化を図る。

質問文はWWW検索と、ontologyの検索に用いられるが、単なるキーワード検索を行うのではない。質問文は述語を伴う形式になるので、述語の意味とその述語の格要素、それらへの修飾子の成す構造全体が質問文の意味を形成する。IIIにおける質問文は、その性質上、条件を含む複文であることが多いと考えられるので、単文の意味だけではなく、複文の構造も表現可能な形式とする。

次の質問文を例にとる。

「原油価格が高騰すると何が起きるか？」は次のように分析しておく（実際には上記したようにXMLを用いたフレーム形式で記述する計画である）。

条件： 原油価格が高騰する

質問： 何が起きるか

ontologyの表現形式と、質問文の表現形式を「意味構造」

として統一することにより, ontology に「原油価格が高騰すると, 電気料金が上がる」などの知識が存在すれば, 「電気料金」と「何」のマッチング操作だけで回答を得ることができるが, WWW 上の知識を導入する場合は, このような簡単な操作では回答は得られず, 推論が必要になる。これについては 4.3 以下で詳述する。

4.2 機械翻訳

人間の要求に対して WWW 上に存在するあらゆる関連情報を収集するために, WWW データの機械翻訳機能を必要とする。この機能は III の基幹機能に属すると言ってもよい。しかしながら, 機械翻訳システムの開発はそれだけでも III の半分程度の規模の開発になる。現在では多くの汎用検索エンジンは機械翻訳機能を備えている。この機械翻訳システムは 1980 年代から企業において研究開発されてきたものであり, 現在の汎用機械翻訳システムとしては最高の翻訳性能であると考えられる。汎用機械翻訳システムとは独立に III 専用の目的に特化された機械翻訳システムを新規開発することは III に独特のヒューリスティクスを得るという意味で大きな意義があるが, III の最初のバージョンでは汎用検索サイトの提供する翻訳機能を利用することが速やかに III システムを稼働に導く良策であると考えられる。

4.3 メタ検索と WWW 検索方式

ここは, メタ検索部が WWW サーバと自動で通信を行う Machine-to-Machine(M2M) の機能を受け持つ。WWW 検索には既存の検索エンジンを用いる。検索エンジンの言語処理機能は弱体であるから, メタ検索機能を開発し, 検索結果を絞る方式を研究する。メタ検索部の機能は 2 種類ある。

- (1) 質問文を意味表現した結果からキーワードを設定し, 汎用検索エンジンに引き渡す。
- (2) 得られた結果を推論部に送る。

III においては恒久的に真とは限らない知識を扱う。そのような知識は時事情報, 天候情報などとして WWW 上に存在する。近未来予測の前処理としての WWW 検索に用いる検索エンジンは, Google のような汎用検索エンジンを想定している。III の場合, 汎用検索エンジンで行っている一般のキーワード検索と異なる点は, メタ検索部のレベルで「何」という語でも検索が可能ならしにしなければならないことである。汎用検索エンジンでは, 単純に「何」をキーワードにすれば, 「何」という「語」を持つ文を含むサイトを検索結果とする。このような結果は III の検索に用いることができない。上記「原油価格が高騰すると何が起きるか?」の例の場合, メタ検索部は汎用検索エンジンに対するキーワードとしては「原油」, 「価格」, 「高騰」を設定する。この 3 個のキーワードを論理積の AND で満たすサイトをまず検索する。それらのキーワードを持つ文が得られたら, メタ検索部はその文を近未来予測部・推論部の関係項目抽出部に送る。関係項目抽出部はそのサイトの内容である自然言語文に対してスクリーニングとしての粗い言語処理を行う。この方式は未定であるが, まず上記のキーワードを含むよ

うな文を検索するのが妥当であろう。現在の汎用検索エンジンは人間にその機能を託して膨大ないわゆる「ごみ」を出力するが, III ではシステムがスクリーニングを行ってこれを防ぐ必要がある。このために, 汎用検索エンジンの検索特性を考慮したヒューリスティクスを構成することが必要となる。このレベルでは通常, 大量のサイトが見つかるので純粋な言語処理を行うと時間的に困難が生じるから, このヒューリスティクスによる高速なスクリーニング方式の創案が重要な課題である。

こうして, 「原油」, 「価格」, 「高騰」をキーワードにもつ文で, ヒューリスティクスを用いて有意義と考えられる文だけに対して深い自然言語理解を行い, 意味構造を作成する。この構造がメタ知識の

条件: 原油価格が高騰する

結果: a が起きる

に対応する形式になっていれば, 回答として「a が起きる」に対応する結果を返すことができる。

これは第一次の回答であり, 演繹推論の次段階として, 近未来予測部は, 得られた a のインスタンス A に対して,

条件: A が起きる

質問: 何が起きるか

という質問を自動生成して更なる検索を続ける。このような質問文の自動生成は, 推論を制御するメタ知識としてあらかじめ ontology に与えておく。上記のメタ知識は形式的には述語「起きる」を用いて,

$\exists x(\text{起きる}(A) \rightarrow \text{起きる}(x))$

を満たす x を得るものである。

一般には, 前件と後件の述語は同じ必要はなく,

$P(A) \rightarrow Y(x)$

の形式もあり得る。メタ知識としては,

「後期高齢者保険料 (A) が上がる (P) と, 何 (x) に, どのような変化がある (Y) か」という質問文を意味する知識になる。この知識を見つけた推論部は, 「後期高齢者保険料」と「上がる」をキーワードとして汎用検索エンジンで検索を行い。得られた結果の中で前件を満たすものから, 後件の Y と x を見出すことになる。このようにして, 「後期高齢者保険料が上がると, 病院に行く老人が減る。」を見出すだろう。こうして, 次々と生起する現象の連鎖が得られる。これは「風が吹けば桶屋が儲かる」式の推論になる。

考えられる最大の問題は自明の推論が膨大に得られる可能性があることである。メタ知識とそれから得られる結果を厳しく抑制すれば自明なもの数は減るが, 重要なものを落とす可能性が増すというデータベースの適合率と再現率のトレードオフ類似の問題はここでも何らかのヒューリスティクスで解決されなければならない。

4.4 ontology

III で行う推論は, 第一義的には演繹推論である。ユーザの質問を満足する回答を得るための推論規則であるメタ知識はここに記述する。これは推論過程を制御する知識であるので, III のように, 膨大な, しかし単独では大きな意味を持たない

知識を組織化して意味を見出すという目的において、重要な意義を持つ知識だけを選別するという意味で極めて重要な役割をもつ。III の ontology の重要性は主としてこのメタ知識の存在にある。この部分は 4. 3 での「後期高齢者保険料…」の例にみるように、分野に依存する知識となり、また、大量な知識となる可能性が高いので、一組織、一個人で開発することは極めて困難であると思われる。III はオープンソースのように開かれたシステムとして多くの専門家で作成できることを目指したい。

そのほか、「酸素は気体である」のような通常の恒久的に真であるとみなすことができる知識も可能な限りここに記述する。

4.5 近未来予測／推論部

推論部ではメタ検索部で得られた記事から自然言語理解を行って意味構造を作成する。その意味構造から関係項目を取り出す関係項目抽出部、それを用いて推論を行う推論制御部、推論結果を取りまとめる推論評価部から構成される。ontology はこの各過程で用いられる。

推論で用いる知識は、

- a) ontology 上の、基本的かつほぼ恒久的、静的な知識（「エチルアルコールは燃料となる」、「原油価格が上がれば、諸物価があがる⁶⁾」など）
- b) WWW 上に基本的には自然言語でおかれている大量かつ動的な知識（「原油価格が高騰」など）

である。推論部は、この 2 種の知識をシームレスにつないで用いることにより、非常な速度で増減、変化する WWW という地球規模の巨大知識ベースを扱うシステムを構成する。

ここで論じている演繹推論は、定理証明のように命題の真偽を定めるためのものではなく、定理発見と同様のプロセスになる。ただし、WWW 上の知識は恒久的に真ではなく、命題の真偽に変化がある。III では、ある変化から、どのような変化が次々と生じていくか、その波及効果、あるいは、波及の様相を調べるために演繹を行うのである。そこで用いられた知識は明日には存在なくなっている可能性があったり、今日とは逆に偽となっている可能性があるという意味で動的である。

III に対する質問の形態は次の 3 種が想定される。

(1) 命題論理に対応する質問

質問例：「Twenty In Ten 構想の実施で、原油価格は下落するか」

この質問に答えることは、「Twenty In Ten 構想の実施で、原油価格は下落する」という命題の真偽を決定することになる。このような質問には、様々なセンチメント（市場心理）による真と偽の見解が WWW 上でも公表されていると思われるので、第一段階では、それらを列挙することになる。なんらかの評価関数を用いて真偽の確率を計算できるようにすることも考えられるが、それは III の目指すものではない。

(2) 第一階述語論理に対応する質問

質問例：「Twenty In Ten 構想の実施で、価格が下落する

ものは何か」

条件： Twenty In Ten 構想の実施

質問： 価格下落(x)（ここに「価格下落」は述語である）を真にする x を WWW 情報から取得することに対応する。この質問には、最初に得られた x の価格が下落すると、その影響はどのように波及していくかを調べることもメタ知識により含まれていて、III の目指す検索となる。

(3) 第二階述語論理に対応する質問

質問例：「Twenty In Ten 構想で、どのような影響が考えられるか」

条件： Twenty In Ten 構想の実施

質問： Y(x)

この質問では、述語変数 Y と、対象変数 x が表れている。これは典型的に III の目指す近未来予測を行うことになる。

5. 予備的シミュレーション

質問文：「Twenty In Ten 構想が実施されると、どのような影響があるか」

第 1 段階：言語理解の結果、「Twenty In Ten」をキーワードにして汎用 WWW 検索を行う。検索で次の文を得る。

「例えば最近では、バイオエタノール等の代替燃料の利用などにより今後十年間でガソリンの消費量を 20% 削減する「Twenty in Ten」構想を打ち出しました。」

第 2 段階：上記の文の言語理解処理により、「バイオエタノール消費量が増加する」を帰結する。この帰結は上記 WWW から得られた文には明示的に示されていない。このような含意を明示的に文として示すための知識は ontology に存在する必要がある。

次に、「バイオエタノール消費量が増加すると、何が起きるか」という質問文を ontology に登録されている推論規則であるメタ知識により生起する。「バイオエタノール消費量 増加」で汎用 WWW 検索を行うことにより次の文を得る。

「バイオエタノールは主にトウモロコシとセルロース系原料から生産される・・・。」

第 3 段階：上記の文の含意である「トウモロコシの需要が増加する」が ontology から帰結される必要がある。以下、第 2 段階と同様にして、「トウモロコシの需要 増加」で WWW 検索すると、次の文を得る。

「この背景には、世界的な原油高、環境保護への取り組みでバイオエタノールの需要が高まり、工業用に消費されるトウモロコシが増加した結果、飼料高を招き豚肉など食料品価格が高騰していることがある。」

こうして、「Twenty In Ten 構想の実施により、豚肉など食料品価格が高騰する」が出力される。

このシミュレーションでは、実際に WWW を Google で検索している。キーワードで検索された文書から重要な文を抽出するためには、相当な言語理解能力が必要であることがわかる。高度な ontology の開発が極めて重要である。人間の書いた文章というのは含意が多く、目的とすることが明示的に書かれていないことが多い。おそらくこれは良いことではない。科学技術の論文ではもう少し明示的に目的とすることを書くであろうが、マスメディアの文章はそうではないので、ontology が推論だけではなく、言語理解のためにも重要な意味をもつと考えられる。

6. おわりに

インターネットが社会的存在になるにつれて、その上には社会の縮図として社会問題が投影されている。実社会はさまざまな慣習、法制、教育、文化により秩序が保たれているが、インターネット上は無法地帯の状況に近いと言ってよい。インターネットは情報科学が創り出した、「もうひとつの社会」であり、情報文化学の立場から論じられるべき存在となってきた。インターネットの社会にも外の社会から法制が定められ、あるいはネチケットという形で教育、慣習、文化が論じられているが、それにより秩序が作られているかといえば、むしろ普及速度の速さのために悪化している傾向がある。インターネットの自浄機能も必要ではないか。

III は、インターネットに知性を持たせ、近未来を予測することによってネット社会を正常な社会に導くと同時に、インターネットを人間にとってこれまで以上に有意義な存在にすることを目的としている。

注

- (1) 筆者らはこの用語を同音異義語だけでなく「横浜」と「横濱」のような同音同義異字も含む概念として用いる。
- (2) 副詞になりえる名詞（昨日）と純粋な名詞（機能）。形容動詞という品詞を作るか、名詞+助動詞とするかなどの問題がある。
- (3) 「上げる（利益を上げる、棚に上げる）」、「勉強する（物理を勉強する、値段を勉強する）」などの意味の違い。
- (4) 2007 年米国大統領一般教書にある構想。今後 10 年間でガソリン消費を 20% 削減するというもの。
- (5) この種の命題の寿命がどの程度であるかは定かではないが、WWW 上に存在するとは限らないこと、WWW から得ると検索のコストが高つくことから、ontology に収容するほうが得策であろうと考えられる。

参考文献

- [1] Google: <http://www.google.co.jp/>.
- [2] True Knowledge: <http://www.trueknowledge.com/>.
- [3] Wikipedia: <http://wikipedia.org/>.
- [4] 情報大航海プロジェクト: <http://www.igvpj.jp/>.
- [5] Paul Erdman: <http://www.timesonline.co.uk/tol/comment/obituaries/article1722875.ece>.
- [6] 加藤千幸: 「国際情報の読み方」講談社, pp12-17 (1986)
- [7] 青山友紀編: ユビキタスコンピューティング世界を実現する革新的ネットワーク技術, 情報処理 Vol.43, No.6, pp611-652 (2002).
- [8] 天野真家, 中島秀之編: ユビキタスコンピューティングとネットワーク社会の到来に向けて, 情報処理 Vol.45, No.9, pp897-938 (2004).
- [9] 稲垣耕作, 天野真家: ユビキタス: メディア化と環境化, 情報文化学会誌, Vol.13 No.1, pp3-10 (2006).
- [10] Mark Weiser: The Computer for the 21st Century, <http://www.ubiq.com/hypertext/Mark>Weiser/SciAmDraft3.html>.
- [11] 茶筌: <http://chasen.naist.jp/hiki/ChaSen/>.