

# 人工知能と集合知



---

志村 正道 Masamichi Shimura

東京工業大学・武藏工業大学 名誉教授

---

## 1. まえがき

人工知能の研究はよく知られたダートマス会議から数えても既に50年以上の歴史を持っている。人間の持つ知能をコンピュータシステムに組込むことにより、知的な振舞いをするようなシステムの構築を目的としてきた。その成果として多くのシステムが実用化されてきたが、これからも総合的なものとして進展して行くであろう。近年情報システムにおける大きな転換がインターネットの展開とその利用環境の開発によってもたらされてきた。インターネットはコミュニケーションという従来作り上げてきた人間社会の大きな文明に影響を与えてきた。すなわち情報の発信・受信・処理の一連の流れの中に多くの人工知能の技術が取り入れられるようになったのである。それは知識処理、マイニング、オントロジーなどの技術である。これらの技術の統合の結果生まれたのが集合知の概念であり、集合知を利用したネット上のビジネスの爆発的な広がりであり、全く新しい社会が出現してきたといってよいほどの変革をもたらした。このような集合知の背景にあるのがウェブの発展であり、インターネットの普及である。

人類の進化過程において、言葉と言葉あるいは言語と言語によるコミュニケーションは大きな役割を果たしてきた。同じようにインターネットによる情報の伝達と知識の蓄積はまさしくクロマニヨン人の急速な発展と対比して、大きな全世界的な変化を生み出そうとしている。すなわち個人を超えた集合の知は空間軸、時間軸に対し従来の知性を全く変えるものになろうとしているのである。現代社会においてインターネットの利用は極めて大きなインパクトをビジネスのみならず日常生活にも与えてきた。これらの技術が実社会にどのような影響を与え、どのように進展してきたか、またそれらを支える技術にはどのようなものが存在したかを考察する。また、今後のウェブの発展とこれに伴う集合知の形態について考究し、将来の展望について論じる。

## 2. 集合の知性

### 2.1 集合知

近年インターネットの世界で集合知(Collective Intelligence)とよばれる概念が社会的にも注目を浴びてきている。すなわち集合知とはインターネットを介した極めて多数の主張が作り出す知性である。これは情報の収集のみならず、得られた知識の評価や物の販売・購入などの行動に至るまで広範囲の連鎖的ともいえる結果を創出している。すなわち集合知とは多くの情報から般化によって知識が生成され、その妥当性の評価もまた多数の人々によって自然発生的になされる。大多数の考えは正しい、あるいは妥当であることを意味している。このことによって新たな仕組みが普及された。例えば、WikipediaやAmazonに見られるようなユーザが参加する仕組みである。しかしこのような仕組みすなわち集合知の利用に価値があるか否かを見極めることは重要なことであり、次のような仮説の上に成立していると考えられている。「多数決の論理であり統計的なデータの集合と考えれば、その成果はほぼ妥当である。」分かりやすく言えば、多数の人の知識が蓄積され、有用で利用の容易な形に体系付けられた知識ベース化されたものを集合知という。MITのThomas W. Malone博士はこの集合知を次のように定義している。すなわち、「Collective intelligence is groups of individuals doing collectively that seem intelligent：集合知とは、知的に見えるような行為を集合的に行う個人の集合である。」

### 2.2 群衆の叡智

集合知に対して、群衆の叡智／英知(Wisdom of Crowds)という言葉がある。群衆の叡智とは、2004年出版のJames Surowieckiによる書名で、その副題は「Why the Many Are Smarter than the Few and How Collective Wisdom Shapes Business, Economics, Societies, and Nations：何故多数は少数より妥当なのか、またどのように集合的英知は商業、経済、社会そして国家を形成していくのか」である。「群衆の叡智」という言葉はもともと1841年にCharles Mackayによって出版された著書“Extraordinary Popular Delusions and the Madness of Crowds”に由来するといわれる。群衆の叡智は、多数の人の平均評価は各個人の評価より正確であり、なおかつ専門家による評価値より正確であるという事実に基づいている。

群衆の叡智が実際に活用されているのは次の3つの要素に基づく事象と言われている。

1. 認知 (cognition) …問題解の候補の中から正しい解を見出すこと。
2. 調整 (coordination) …自立的なエージェント間での行動計画を調整すること。
3. 協調 (cooperation) …中央管理システムなしに全体として協調的に管理されること。

例えば、専門家より多数の判断の方が、客観的で素早い対応が可能な市場判断は、認知に相当するものであり、渋滞箇所や雑踏での衝突回避の例では調整が働くと考えられるし、統制管理のないフリーマーケットなどは協調行為に基づくものと考える。

これらのことが実現されるには、群衆の全てが賢人である必要はなく、極端に言えば暴徒の集団でもよいということである。しかし上述のように群衆の叡智が正しく活用されるための条件がある。群衆の叡智の対角にある群衆の愚行・衆愚(The Madness of Crowds)とかあるいは衆愚政治などの群衆の意志決定が極端な方向に偏る傾向がある。ガリレオ・ガリレイの地動説

が天動説に替わっていたかもしれないという話がある。

### 2.3 集団の知性の信頼性

集団による予測が誤ることはしばしば生じることであるが、それが妥当であるためには次のような基本的な性質が担保されていなければならない。

#### (1) 多様性(diversity of opinion)

各個人の背景や観点がそれぞれに異なっているならば、結果的に全体としては多くの解候補をもつ可能性が高い。また解候補の多様性が低い場合には、適切な解が存在しない怖れがある。

#### (2) 独立性(independence of members from one another)

各個人の意見や提案が他者の影響を受けないよう、各個人の独立性が確保されている必要がある。とくに小集団の場合には、多様性が低いために偏った結論に集約される危険性がある。

#### (3) 分散性(decentralization)

問題を抽象化せず、各個人が直接得られる情報に基づいて判断する必要がある。各個人ごとに得られる情報の種類は異なることが多いと考えられるが、多様性を保つためにも、各個人共通の尺度で判断すべきでない。

#### (4) 集約性(a good method for aggregating opinions)

上記3点の特性を生かして得られた知識を全体で共有し、その中から比較検討して最終的な結論を導く仕組みが必要である。

群衆の叡智とは多数の個人やグループが集団として意志決定を行う場合には、意見の多様性、各個人の独立性など上に述べた条件が満足されていれば妥当で適切な結論が導かれるという現象のことである。「みんなの意見は案外正しい」ということである。しかし常に正しいというわけではなく、前述の「鳥合の衆」とか「衆愚政治」などとよばれるような群衆の叡智に相反する事態も生じる。

個々の情報が集約される際、情報の正しさはそのまま残され、情報の誤りは他の個人によって削除あるいは修正されるという巧妙な仕組みが作られている。集合知の信頼性はこのようにして担保されている。よく知られたウィキペディア(Wikipedia)などはまさしくこのような仕組みによって広くユーザの信頼を得ているのである。すなわちユーザによって作成されたコンテンツは他の多くのユーザによって修正され、その結果、1768年に英国で生まれ、アメリカで出版されるようになったブリタニカ百科事典(Encyclopedia Britannica)と同程度の正確さを有しているという報告もある。すなわち個の意見が偏らない限り、その平均値あるいは集約された知識は共通で正しいものののみが残されるということである。個々のユーザあるいは参加者の質は問われない。これに対して群衆の叡智とは多くの個を集めれば、与えられた問題に対しの専門家がその集団の中に存在するようになるということである。したがって群衆の叡智とは参加者の平均的意見を採用することではなく、解候補を包括する意味での群衆である。

集団の極端な偏りは一部のグループによって全体が扇動されたり、愉快犯的な誤った行動や

天の邪鬼的意見が喧伝されるなどの弊害が生じがちである。このことは意見集約時においても発生する可能性がある。極端な場合には評価結果による相互修正が無限ループに落ち込む危険性も否定できない。さらに集合知は集団参加者の平均的な集約結果が貴重な情報を隠してしまう怖れもある。例えば、家電に対する評価では有名メーカの製品のみに評価が集まってしまい、少数の人が持っている特別な評価結果が埋没してしまう危険性がある。

集合知の考え方は民主主義の根底にあるもので、議会のみならず多くの会議などでも多数決の原則に従っている。集合知は個の協調や競争によって生まれる知性であり、単なる個では知性を持たないような蟻や蜂などの昆虫が集団では高い知能を持っているが如き行動を探ることに由来する。昆虫のみならずバクテリアや動物の集団についても同じである。集団的知性の概念は蟻の観察から1911年に昆虫学者のWilliam M. Wheelerによって提案され、このような群れの生命体を超個体と呼んでいる。

アメリカのMITでは、集合知に関する研究センター(MIT Center for Collective Intelligence)を2006年10月に創設し、多面的にかつ集中的に研究を展開している。このセンターでは、新しいコミュニケーション技術が人間の作業を如何に変化させてきているかについて、MIT全学で研究を進めていて、その目的は深いレベルでの集合知を理解することであり、また新しい可能性を創発しようとしている。

前述のWikipediaは集合知を利用した百科事典であるが匿名性がある。これに対し著者を許可された学識あるものに限定し、さらに監視人をもうけた事典としてCitizendumがある。また2006年に運用開始された「Intellipedia」は合衆国政府の機密情報を扱うインターネット「Intelink」上にあり、その情報量は2万8,000ページ以上、登録ユーザ数は3,600人、16の政府機関で利用されているという。これらのシステムは集合知の代表的なシステムであるWikipediaがもつ欠点を補おうとして考えられたものであり、集合知の利用がすべての領域について有効ではないことを意味している。ただし、集合知とは正しい結果を生み出すとは限らないし、またWikipediaについて言えば、次のような観点に立っている

- (1)すべての記事は偏ることなく、中立的な観点にしたがって書かれていかなければならない。
- (2)信頼性の高い情報に限定し、公表、出版の事実、視点、理論及び議論だけを掲載しなければならない。
- (3)独自の分析結果・合成結果を掲載してはならない。

したがってWikipediaでの基準は「真実・真理」ではなく「検証可能性」なのである。あくまでも「現時点で認識されている知識の集大成」しか期待していないと言うことである。

### 3. 集合知の応用とビジネス

前節までに述べた集合知の概念はインターネットの発展とともに応用されると同時に従来存在しなかったような新しい種々のビジネスの進出に繋がる。多数の人がインターネットを通じて各個人の持つ情報や知識を集約することによって新たな付加価値を持つ情報や知識が生まれる。この付加価値とも言える効果が新しいビジネスを生み出すことになるのである。その分野は予想・予測、分類仕分け、知識作成、ソフト開発、ウェブサイト登録システム、協調型プロ

ジェクト、価格比較、推奨など広い分野に利用されている。以下これらについて説明を加える。

(a)予想・予測

未来の予測を行うサイトでスポーツの勝ち負けや株式市場の予想を行うものであり、「あした新聞」や「FTPredict」がよく知られている。あした新聞Predictionは谷川正剛氏のプレディクションが最近リニューアルされたシステムで、ユーザの予想により記事を作成し、株式情報などをユーザに提供する予測市場サイトである。海外では、Financial Timesの「FTPredict」、映画関連の「Hollywood Stock Exchange」やスポーツ・サイトの「Trade Sports」等が知られている。

(b)分類（フォークソノミ）

大衆(folk)の分類法(sonomy)という造語である。ウェブサイト上の情報をタグ(名札)を付けることにより検索可能な分類を行うことを意味する。図書館などでは定められた語彙に基づいて分類が行われるが、フォークソノミではユーザが作成したキーワードが用いられ、探索がより容易になる。実際には写真共有サイトのやソーシャルブックマークのDel.icio.usなどがある。また、はてなブックマーク、動画共有サイトのニコニコ動画やWikipediaもその分類はフォークソノミに基づいている。

Flickrでは撮った写真をウェブ上で整理・分類・展示したり、あるいは共有することによって画像掲示板やソーシャル・ネットワーキング・サービスのようなコミュニティとしてのサイトともなっている。

(c)知識生成

多人数からの提案を集約し、妥当な知識を生成しようとするもので、ウィキペディア(Wikipedia)などはその代表的な例である。Wikipediaはよく知られているようにWiki Wiki WebとよばれるソフトとEncyclopedia(百科事典)からの造語であり、現在では、世界中で164の言語により、記事項目は1千万以上、日本語版でも53万項目の巨大なウェブ上の事典となっている。またウィキペディアには次のような姉妹プロジェクトもある。すなわち、引用の百科事典であるウィキクオートや生物種データベースウィキスピーシーズなどである。

(d)ソフト開発

オープンソースのソフトウェアをウェブ上で多くの人々の協力によって開発していくもので、よく知られている例にOSのLinuxがある。Linuxは1991年に当時ヘルシンキ大学院生であったLinus Torvaldsによって開発されたOSのカーネルで、Unixに類似したオープンソースのソフトである。このLinuxはソースコードがインターネットに公開された結果、世界中のプログラマの手によって優れたOSに進化してきた。ちなみに開発当初のソースコードは約1万行であったが、翌年にはユーザ数が1000人、ソースコードも4万行、1997年にはユーザ数は約350万人となり、ソースコードは約80万行となった。現在ではカーネル部のソースコードは500万行を超えたと言われている。

(e)ウェブサイト登録システム

いわゆるブックマーク(Bookmark)のことであり、ブラウザのNetscape Navigatorにおいて最初に使用された言葉である。マイクロソフトのInternet Explorerでは、「お気に入り(Favorite)」とよばれる。Bookmarkは栄の意味であるが、ブラウザに任意のURLを登録しておく機能のことである。

ある。とくに不特定多数のユーザが利用できるようなブックマークのことをソシアルブックマーク(Social Bookmark)という。実際には、はてなブックマーク、Googleブックマーク、Firefoxブックマーク、Yahooブックマークなどがある。ちなみに、はてなブックマークは「はてな」の伊藤直也氏により開発されたシステムで、ユーザは15万6000人、2480万のブックマークを持っている。

#### (f) 協調型プロジェクト

ウェブによって多くの人々があるテーマに沿って議論の場を構成することが出来る。多くの人々の意見を聞く試みは特に新しいものではないが、ウェブを利用してことで電話や通信などに比較して極めて多数の人々の意見を収集し、双方の議論を通して集約することが出来る。例えば、2006年に開催されたIBMのInnovationJamはオンラインのブレインストーミングであり、IBM社員とその家族、顧客、大学など104カ国から15万人がイノベーションに関して議論し、結果として4万6,000件以上のアイデアが生まれたと言われている。実際にこの中から10個のプロジェクトが抽出され、これ自体大変興味ある試みであった。また、2007年にはSNS(Social Networking Service)のmixi上でエースコックが新商品開発のプロジェクトを立ち上げた。mixiのコミュニティでは700以上の案が出て、2種類のカップ麺の商品化が行われたという。

#### (g) 價格比較

商品を購買しようとする消費者を対象に、商品情報や販売価格あるいは既購買者のいわゆるくちコミ情報を提供しているサイトである。商品の価格を調べて提示するだけのサイトではなく、多数の販売業者が他の販売業者の付けた価格を参考にしながら自社の価格を決めていくと同時に、購買者からの商品の評価を記載することで、購買者が参加するようなシステム作りを可能にしている。代表的サイトには価格.comやcone.co.netが知られている。この他多くの類似のサイトを運営している通販業者も多い。

#### (h) 推奨 (Recommendation : リコメンデーション)

ユーザの趣向や購買履歴などのデータに基づいて、商品やサービスあるいは情報等を提示するサービスのことである。その手法としては協調フィルタリング、コンテンツベース・フィルタリング、ルールベース・フィルタリングなどが知られている。

協調フィルタリングとは、趣向や購買履歴の類似したユーザに同じ購買商品などを提示する方法で、同じような人は同じようなものに興味を示すという事実にその理論的背景をおいている。Amazonの推薦機能に採用されているし、はてなアンテナやlivedoorグルメにも協調フィルタリングの技術が実装されている。なお、代表的な協調フィルタリングシステムとしてはインターネット上のニュースに対するフィルタリングを目的としたアメリカのGroupLensや音楽を対象としたRingoなどがある。

コンテンツベース・フィルタリングとは過去に購買・チェックした商品リストを分析し、同種の特徴を持つ商品を抽出して提示する方法である。

ルールベース・フィルタリングとはデータマイニングと同様なルールを作成し、そのルールに従って商品を推奨する方法である。

### (i) クラウドソーシング

集合知の応用ビジネスとしてクラウドソーシング(Crowdsourcing)とよばれるサイトがある。アウトソーシング(Outsourcing)は業務を外注することで、社外の職業人を雇用することであるが、クラウドソーシング は一般にインターネットなどを通じて必ずしもプロでない不特定多数の人々に対してアウトソーシングを行うことである。クラウドソーシングという言葉は、新しいライフスタイルを提案しているアメリカの雑誌WIREDに掲載された記事「The Rise of Crowdsourcing」(2006年6月)に見られる。一般にプロフェッショナルや専門家を社員にするとか、社外の専門家集団に依頼するような従来のシステムでは、必ずしも当該プロジェクトに適切な人材を確保することが容易ではなく、また雇用効率がよくない場合が業種によってはしばしばある。その点クラウドソーシングでは、適切な人的ネットワークに仕事を発注することが可能となり、人材の安価な調達法として今後増加していくと思われる。以下いくつかの事例を述べておく。

A Swarm of Angels : 映画製作に必要な資金集め、その制作、完成した映画の流通に至るまでの全行程を実行しようとするプロジェクトのサイトである。

CrowdSpirit : 電化製品のクラウドソーシングで、電化製品のデザインや開発、また制作の資金集めを行い、製品の販売利益に応じてユーザに還元される。

CafePress : 自分でデザインしたTシャツやマグカップなどの商品を販売できるサイトで、サイト上に自分の店舗を持ち、販売するサイトである。

news-press.com : Gannett社のサイトで読者参加型の新聞を制作するものである。すなわち読者に情報提供や調査を呼びかけることで、低く抑える仕組みになっている。実際にフロリダ州フォートマイヤーズで成果を挙げたといわれる。

## 4. ウェブと集合知

### 4.1 Web1.0

以上述べてきたように、集合知はインターネットの発展により爆発的にビジネスへの結びつきを強めてきた。本稿においては、ビジネスについての記述を避け、技術的な面についてのみ述べることにする。

最初のブラウザであるWorld Wide Webは1990年に欧州原子核研究機構(CERN)のティム・バーナーズ=リーによって、NEXTコンピュータ社のOSである NeXTSTEP上で開発された。文書にアクセスするグローバルハイパーテキストプロジェクトから生まれたもので、後にWorld Wide Webと誤用されないようにNexusとよばれるようになったが、World Wide WebはHTTPを使った最初のプログラムであった。World Wide Webはインターネット上で提供されるハイパーテキストシステムであり、単にウェブ(Web)と呼ばれることが多い。このウェブはハイパーテキストの特長を生かして、文字や画像をクリックすることにより他のWebページに繋がりリンクを張ることが出来る。

インターネットのサービスとしては、ニュース、メール、Gopher、Telnet、WWWなどが提案され、使用されてきた。WWWはWeb1.0の世代からWeb2.0へ進展し、さらにWeb3.0が見え隠れするようになってきた。

Web1.0とは、2000年頃までのディレクトリ型の検索エンジンが主であった、一世代前のWebであり、スタティックなHTML(Hyper Text Markup Language)で作られた。HTMLとはWeb上の文章を記述するためのマークアップ言語でW3Cによって標準化されている。HTMLは文書やデータの表示属性を表しているだけであるが、ユーザが独自のタグを用いることにより、データの属性と内容を関連づけて記述できるようにしたのがHTMLの後継言語と言われるXML(Extensible Markup Language)である。HTMLやXMLでは、文書中のタグや属性について定義化されている。文書構造や書式、文字飾りなどを指示したり、リンクを持たせたりするには、定義によって定められた「<」と「>」で囲まれた標識により指示されるが、そのような付加情報を入る文字列などのことをタグという。上に述べたような定義を共有しておけば、他のアプリケーションでの再利用が可能になり、XMLで記述したデータは業務システムに直接取り込んで処理できるなどの利点をもっている。XML文書はタグを含むテキストであるため、データは人間が読めるし、タグと構造化によりコンピュータでも理解や処理が可能となっている。

さて、2004年頃から新しいアイデアに基づいたWebサイト、Web関連のサービスおよびその技術が世の中に現ってきた。このようなWebをWeb1.0に続いて、Web2.0という。Web2.0はWeb1.0の延長上にあると言うより、かなり質的な相違があると言われている。本稿の集合知はこのWeb2.0と深く関わりがあるとともに、Web2.0により実現されたと考えてよい。

## 4.2 Web2.0

Web2.0の由来は1997、1998年に出版された著書名に見られるが、現在広く使われているWeb2.0なる言葉は、2004年にT.オライリーによって言い出されたとされている。また、2005年にはサンフランシスコで同氏によって主催された「Web2.0会議」によって認知度が上がった。

Web1.0と呼ばれていた時代のウェブはHTMLによって文書が記述され、ウェブサーバに登録してブラウザによってその文書を閲覧する仕組みであった。ウェブはインターネットにおけるサービスの一つであって、利用可能なサービスはその他にもeメール、ネットニュース、FTPなどがあった。しかしインターネットが急速に普及してくると、メールとウェブが主役となるとともにブラウザが高機能となり、マルチメディアのコンテンツが表示可能となつたのである。ウェブはむしろ新しいプラットフォーム化するようになったと考えてよい。すなわち、ブラウザはコンテンツビューワからOSやコンピュータによらないコンテンツ利用のためのシステムとなって行ったのである。

Web2.0は次のような7項目の機能を要件として持っている。フォーカソノミ、リッチインターフェイス、ロングテール、ユーザ協調、ユーザ参加、オープンソースソフトウェア、分散指向である。とくにWeb2.0は様々な新しいビジネスを立ち上げ、従来からの発想にない形態のウェブが生まれることになった。ブログやソーシャルネットワーキングサービス(SNS)のようにユーザである参加者が自分自身によってコンテンツの作成や公開を行うことがインターネットの世界では普通のこととなり、事実コンテンツは膨大な量となり、質も高くなっている。

Web1.0はユーザがウェブに自由にアクセス出来、そのウェブに表示されている情報を閲覧す

ることが出来た。しかしこの場合、欲しい情報をあるいは欲しい情報が表示されているサイトをユーザが探索しなければならない。Web2.0ではユーザ個人がウェブにアクセスするだけでなく情報発信を可能にしている。すなわち、Web1.0ではユーザは情報の受け手であったが、Web2.0では情報の送り手となることが出来る。情報を発信できるとともに、ユーザ同士のつながりが形成されるようになった。このことにより多数の人の集約した意見は専門家よりも正しいという集合知の実現となるのである。上に述べたようにWeb1.0に比べてWeb2.0は大きな進歩であり、ユーザにとって極めて有用なものとなったと言ってよい。また新しいビジネスを生み出す原動力となったことも明白である。しかし、Web2.0の陰の部分も否定できない。以下にWeb2.0の弱点とも言うべき現象をまとめておく。

- ・批判的・反対的意見が殺到するブログ炎上
- ・サイトの運営を妨害するいわゆる荒らし
- ・誤りの集合知
- ・扇動やマインドコントロールによる真実の曲解
- ・ウェブの肥大化による情報取り出しの困難
- ・一様で大量な情報からの必要情報の体系化困難
- ・不適切な検索語や検索エンジンによる冗長で無駄な検索

#### 4.3 Web3.0

Web2.0は集合知についての節で述べたように画期的なビジネスを生み出したが、このWeb2.0に続くWebがいわゆるWeb3.0と期待されている。以下このWeb3.0について述べることにする。ただし断っておくがWeb3.0の確たる定義はないし、それほど実態があるものでもないというのが現在での一般的な考え方である。

Web1.0はポータルサイトを読むだけのRead-Only型であり、Web2.0は双方向の流れを可能とするRead-Write型であった。Web3.0はユーザ同士が相互作用できる情報互換型と言ってよいであろう。すなわち、アプリケーションとコンテンツが相互にしかもシームレスに連携する事が可能で、インターネットTV・携帯ユビキタス通信・モバイルブロードバンドなどがキーワードとなり得る。例えば、GoogleやAmazonを通して商品購入などを行う場合、他人が何を購入したいかは分からぬし、商品のメーカも消費者であるユーザの購入形態を直接には把握できない。全ての情報は流通経路のプラットフォーム企業が握ることになり、大きなビジネスが生まれたわけである。

このようなWeb2.0に対して、Web3.0ではこの過程を可視化しようとするものであり、ソーシャルメディアとパーソナライゼーション(個人化)という言葉が生まれてきた。また2008年2月には、イギリスのGuardianウェブ版の中で「Web3.0は、リコメンデーションとパーソナライゼーションである」との記事がある。個人化されたリコメンデーションサービスが新しい音楽や製品、レストランなどの情報を人々に届けるようなウェブであることを述べている。

#### 4.4 セマンティックウェブ

Web3.0はセマンティックウェブ(Semantic Web : S-web)であるとも言われている。すなわち、Web3.0は未定義であるが、S-webは1998年にW3Cのティム・バーナーズ=リーが提案して以来多面的に研究が行われてきた。すなわちWebの利用をより効率的にするためにデータに意味を付して、ウェブページの閲覧に高度な機能を持たせることがS-webの目的であった。膨大な情報をもつWebに一層利便性を持たせるための技術で、コンテンツの内容を理解するのではなく、その構造や属性に関する情報であるメタデータを附加したWebである。このためS-webは、XMLによって記述された文書にRDFやOWLを用いてタグを付ける。データの意味を形式化することによって、人間の介在をなくし、自動的に情報収集や分析などが可能となる。

以上より、S-webとWeb3.0は明確に同じものではないということが知れる。Web2.0で用いられているHTMLがもはや限界となり、その限界を越える一つの手法としてS-webの提案があったわけである。一方、Web3.0はWeb2.0の画期的な技術の開発とその成果に対して、必然的とも思われる次世代WebあるいはWeb2.0の進化したものとして現れた言葉である。勿論、Web3.0とS-webは部分的に共通する利便性や機能を持っていることは否定できないが、いずれにしろ今後の問題であろう。

セマンティックWebは、膨大なWebページの中から、ユーザが求めている情報を探し出しやすくするために、検索性能の向上や利便性を高める次世代Web技術と言える。Webサイトを記述しているHTMLは、コンテンツ(情報内容)を表示するための言語であるが、その意味を解釈するための「メタデータ(意味付けされた情報)」を付けることで、セマンティックWebは困難な情報検索の効率を大幅に高めることが可能なのである。そのために、「メタデータ」と「オントロジー」という2つの技術を取り入れているのであり、ここでメタデータとは、ページの内容を説明する附加情報を、オントロジーとはメタデータを記述する用語の定義を意味している。

ところで知能システムについて、最も引用件数の高い文献の1、2位を参考文献の[8]および[9]に示しておいた。すなわちその重要なキーワードは「セマンティックWeb」と「オントロジー」と考えてよい。

オントロジーはシステムの内側と外側で共通した情報の理解を持たせることができ、システム間での相互操作性を保たせることができる。1995年からのLenatのCycや上位オントロジーのためのIEEEのSUMO(Suggested Upper Merged Ontology)は時間をかけても少人数グループによるオントロジーであり、グループ間の互換性に乏しい。もともと人間社会はフラクタル構造であって、Webシステムも同様にフラクタルであると考えてよい。このためフラクタルの階層毎のコミュニティに固有のオントロジーを、共通したグローバルなオントロジーとして用いることが要求される。このような構成のオントロジーを実現するために有効なWebがセマンティックWebなのである。

上位オントロジーの標準化の候補となるべく、IEEE(米国電気電子工業会)のStandard Upper Ontology (SUO) Working Group で議論されているのがSUMO(Suggested Upper Merged Ontology)である。このSUMOは独自の視点から体系的なオントロジーを作るのではなく、既存のよく使われているオントロジーを適切に結合させることで、包括的かつ一貫的なオントロジーを構成し

ようとしている。すでにオントロジーを取込んだセマンティック検索エンジンとして Swoogle が知られている。このシステムはMaryland大学のプロジェクトで1,000のオントロジーがインデックスされているといわれる。

次にSemantic Deep Web(SD-web)について触れておこう。WWWで公開されている情報の中でも入手が困難な情報が多くある。人間であれば不可能でないような検索でも、知的でなく機械的にしか検索できないような検索エンジンではウェブ全体を検索して、ユーザが欲しい情報を探し出すことは極めて難しい。このような情報は「深いウェブ」(DeepWeb)と呼ばれる。したがって Deep Webの検索には対象とするデータベースなどの内容を予め検索エンジンに登録しておくことが要求される。このような隠れた見えない情報すなわちコンテンツは、見えるものの1万倍にもなるという膨大な量である。

このような状況で提案されたのが上述のSD-webである。SD-webはe-ビジネスやサービスの自動化に有用であるといわれる。そのためにはセマンティックな知識と推論機能を用いてWebを歩き回る手法が重要となる。データ源にアクセス可能なDeep Webとオントロジ操作のソフトが利用できるS-webとから構成されることが実現への道である。なお、Deep Webは中京大学で田村により IntraSiteとしてシステムが開発され運用されている。

## 5. あとがき

WebとWeb関連の技術の発展は極めて著しいものがある。インターネットによるサービスはメールやニュースを主体として現在では非常に広い分野でのビジネスを創生してきたし、人間の知に関する考え方さえ大きく変化させてきた。

Webは文書にアクセスするためのハイパーテキストプロジェクトが生まれてから20年足らずのうちに、Web1.0、Web2.0と呼ばれる時代を経て、新しいWebの胎動が見られるに至るまでの技術の進展とそれに伴う集合体の知性についての概念を多くの文献に沿って述べてきた。このような現象はただ単にインターネット上の一つのサービスという枠を越え、ビジネスを含む新しい一つの分野を作りあげてきたと言っても過言ではあるまい。さらにWebとWebに関する技術は今後も急速に進化していく、インターネットに繋がる多くのユーザに有用でかつ利便性の高いシステムが提供されていくものと容易に予測できる。

Web2.0と呼ばれるシステムの普及にしたがって集合知を利用したアプリケーションが多種多様な形で実用システムに組込まれ、新しい世界を作ってきた。勿論集合知の例として政党が引合いに出されるように、いろいろな形態で古くから活用され、有用な方策を与えてきたこともあり、その背景には集合知の是非が問われるような課題も存在する。本文でも示唆してきたように、集合知は常に正しいとは限らないわけで、その時点において妥当とされる結果しか生み出さないことも認識しておかなければならない。一方インターネットの発達とともに急速に展開してきたことから、この集合知はWebとの相性が極めて良いということも利用する側からは重要な観点となる。現在では正しさを追求されない主観的な問題に対してと、ビジネス面での適用が大部分である。しかしながら将来的には、ビジネスのみならずサイエンスの面でも集合知によるアプローチが大きな成果を上げてくることを期待したいし、社会面での貢献も忘れてはならない。

## 参考文献

- [1] NHK取材班：グーグル革命の衝撃，NHK 出版、2007.
- [2] 佐々木俊尚：ウェブ国産力、2008.
- [3] 大向一輝：Web2.0と集合知、情報処理、Vol. 47. No. 11, 2006, pp. 1214–1221.
- [4] 古崎晃司、来村徳信、溝口理一郎、“Web2.0時代のオントロジー利用雑感”、人工知能学会研究会資料、SIG-SWO-A602-06, 2008.
- [5] 田村浩一郎、“IntraSite2について”、中京大学人工知能高等研究所ニュース、No. 23, 2008, pp. 9–12.
- [6] 丸山不二夫、浦本直彦、石田愛、“Cloud Computingの世界I：Googleの分散処理技術”、電子情報通信学会誌、Vol. 91, No. 6, 2008, pp. 496–502.
- [7] 丸山不二夫、浦本直彦、石田愛、“Cloud Computingの世界II”、電子情報通信学会誌、Vol. 91, No. 7, 2008, pp. 604–609.
- [8] S. A. McIlraith, T. C. Son, and H. L. Zeng, “Semantic Web Services”, IEEE Intelligent Systems & Their Applications, Vol. 16, No. 2, 2001, pp. 46–53.
- [9] B. Chandrasekaran, B. Josephson. Jr., and V. R. Benjamins, “What are Ontologies, and Why Do We need Them?”, IEEE Intelligent Systems & Their Applications, Vol. 14, No. 1, 1999, pp. 20–26.
- [10] James Geller, Soon Ae Chun, and Yoo Jung An, “Toward the Semantic Deep Web”, Computer, Vol. 41, No. 9, Sep 2008, pp. 95–99.
- [11] Ed. H. Chi, “The Social Web : Research and Opportunities”, Computer, Vol. 41, No. 9, Sep 2008, pp. 88–91.
- [12] Tim Burners-Lee and Lalana Kagal, “The Fractal Nature of the Semantic Web”, AI Magazine, Vol. 29, No. 3, 2008, pp. 29–34.
- [13] Bernardo A. Huberman, “Crowdsourcing and Attention”, IEEE Computer, Vol. 41, No. 11, 2008, pp. 103–105.
- [14] Valentina Tamma, Terry R. Payne, “Is a Semantic Web Agent a Knowledge-Savvy Agent?”, IEEE Intelligent Systems & Their Applications, Vol. 23, No. 4, 2008, pp. 82–85.
- [15] J. Geller, S. Chun, Y. An, “Toward the Semantic Deep Web”, IEEE Computer, 2008, pp. 95–98.
- [16] T. Segaran, “Programming Collective Intelligence : Building Smart Web2.0 Applications”, O'Reilly & Associates Inc, 2007.
- [17] [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)
- [18] <http://mindreading.jp>
- [19] <http://www.as-mode.com>
- [20] <http://techon.nikkeibp.co.jp>
- [21] <http://www.w3.org>
- [22] <http://net.intap.or.jp/INTAP/s-web/swc2008/>
- [23] <http://www.kanzaki.com/>
- [24] [www.technorati.jp/tag/Web2.0BOOK](http://www.technorati.jp/tag/Web2.0BOOK)
- [25] <http://www.impressjapan.jp/books/>
- [26] <http://www.web-30.biz/>
- [27] <http://cci.mit.edu>
- [28] <http://www.ontologyportal.org/>