

## 集合知とウェブ

志村正道<sup>1</sup>

### 1. まえがき

約37億年前に誕生した生命がヒトに進化するまでには途方もなく長い年月が必要であった。進化を繰り返し、人間の祖先につながる原人が生まれたのは600万年前のことであった。さらに旧人を経て、現代人に繋がるが新人・ホモサピエンスが誕生したのは6万年前と言われる。生命誕生からの長い時間経過から見ればごく最近のことと言ってよい。旧人であるネアンデルタール人の脳の大きさはホモサピエンスのクロマニヨン人の脳とほぼ同じであったけれど、大きな相違は前頭前野の発達がないことであった。前頭前野の発達がないということは、ネアンデルタール人は発声機能が極めて貧弱で、言葉を発声し言葉によるコミュニケーションに欠けていたことを示している。一方クロマニヨン人は発声機能が発達しており言葉を生み出し、言葉により自分たちの持つ知識を空間的にも時間的にも蓄積し、共有し、利用することが出来た。この結果、ネアンデルタール人はおそらく自然と自然環境の変化のために対処できず、約2万年で絶滅してしまい、その後クロマニヨン人にとって替わられたという。しかもアフリカに偏在していた人類の祖先の中でもクロマニヨン人は急速にアジア、アメリカ大陸など世界中に広がっていった。

このように言葉と言葉、言語と言語によるコミュニケーションは人類の進化過程に大きな役割を果た

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すようとしている。すなわち個人を超えた集合の知は空間軸、時間軸に対して従来の知性を全く変えるものになろうとしている。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。本稿ではウェブの発展とこれに伴う集合知について考察する。

### 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学問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。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 のようにユーザが参加する仕組みである。しかしこの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 博士は集合知を次のように定義している。すなわち、「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。」

<sup>1</sup> 武蔵工業大学名誉教授

## 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、多数の人の平均的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より正確であるという事実に基づいている。

しかしながら、群衆の叡智が妥当であるのは次の3点の事項が成立している場合であるといわれる。

1. 認知力……市場判断はより早く、信頼性が高くかつ専門家の結果に比べてより客観的である。
2. 整合力……群衆が混雑や渋滞箇所における衝突回避の判断は極めて高度である。
3. 協調力……中央管理システムでなくても全体としてまさしく協調的に管理される。

上述の、認知とは問題解の候補の中から正しい解を見出すことであり、整合とは自律的なエージェントの中での行動計画を調整することであり、協調とは他者との共同的な作業による目的達成の行為を意味している。

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。しかしこれが正しく活用されるためには条件がある。群衆の叡智の対角にある群衆の愚行・衆愚 (The Madness of Crowds)、衆愚政治、あるいは共有地の悲劇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

集団による意志決定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恐れがある。
- (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から導かれた知性である。「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「烏合の衆」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に、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。よく知られたウイキペディア (Wikipedia) などで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。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監視され、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。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。また、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て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、候補者を包括するという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ーの製品のみ評価が集まってしまう、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採ることに由来する。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昆虫学者の W. M. Wheeler によって1911年に提案されたもので、このような群れの生命体は超個体と呼ばれる。

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。

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者に限定し、さらに監視人を設けたオンライン事典に Citizendium 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トラネット「Intelink」上にあり、その情報量は2万8000ページ以上、登録ユーザ数は3600人、16の政府機関で利用されているという。

### 3. 集合知の応用とビジネス

集合知の概念はインターネットの発展にしたがい多くの面で応用されるとともに従来存在しなかったような種々のビジネスの進出を促してきた。多数の人がインターネットを通じて各個人の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てきた。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ったのである。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

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付けられたタグ（名札）によって容易な検索を可能にした分類法を意味する。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。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。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ォ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500万行を超えたとされている。

#### (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 は葉の意味であるが、ブラウザに任意の URL を登録しておく機能のことである。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ーシャルブックマーク (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、ウェブを利用する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4万6000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) 価格比較

商品を購入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 coneco.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とその理論的背景をおいている。Amazon の推奨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、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する。

**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メ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急速にビジネスへの結びつきを強めてきた。ここでは、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 WorldWideWeb 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、WorldWideWeb 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 上の文書記述用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に関連づけて記述できるようにしたのが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。このときその付加情報を入れる文字列などのことをタグという。上述の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れてきた。このような Web を Web 1.0 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同時に、Web2.0 により実現されたと見なしてよい。

## 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. オライリによって言い出されたとしてされている。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っていった。この結果ウェブはむしろ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。フォークソノミ、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・ネットワーク・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、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し、ウェブに表示されている情報を閲覧することが可能であった。しかしこの場合、欲しい情報あるいは欲

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事が出来る。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無駄な検索

### 4.3 Web3.0

Web2.0は集合知を利用した画期的なビジネスを生み出してきたが、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。

Web1.0はポータルサイトを読むだけの Read-Only 型、Web2.0は双方向の流れが可能な 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ことが可能で、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ないし、商品のメーカー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、種々のビジネスが生まれたものの発展の余地は大きい。

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「ウェブ3.0は、レ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。個人化されたレ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と述べている。

### 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しての情報であるメタデータを付加した 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。その結果、データの意味が形式化され、人間の介在がなくても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使われた言葉である。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。なお S-web の検索エンジンには Maryland 大学の Swoogle などが知られている。

次に Semantic Deep Web (SDweb) 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精細な検索でも、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、欲しい情報を探し出すことは極めて困難である。このような情報は「深いウェブ」(Deep Web) と呼ばれる。ディープ 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。実際に、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 SDWeb である。SDWeb 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。それにはセマンティックな知識と推論機能を用いながら 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 とから構成されるのが一方法である。

## 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出し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き、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。

## 注

W3C (World Wide Web Consortium) : WWW で利用される技術の標準化を目的として1994年に発足した団体。

PDF (Resource Description Framework) : XML 技術に基づくメタデータ記述のための枠組み。

OWL (Web Ontology Language) : オントロジを用いた言語。

オントロジ (Ontology) : 語彙と語彙の関係あるいは対象の性質と対象間の関係を表す構造のこと。

## 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, pp. 1214 - 1221 (2006)
- [4] J. Geller, S. Chun, Y. An: Toward the Semantic Deep Web, IEEE Computer, pp. 95 - 98 (2008)
- [5] T. Segaran: Programming Collective Intelligence: Building Smart Web 2.0 Applications, O'Reilly & Associates Inc. (2007)
- [6] [http://en.wikipedia.org/wiki/Main\\_Page](http://en.wikipedia.org/wiki/Main_Page)
- [7] <http://mindreading.jp>
- [8] <http://www.as-mode.com>
- [9] <http://techon.nikkeibp.co.jp>
- [10] <http://www.w3.org>
- [11] <http://net.intap.or.jp/INTAP/s-web/swc2008/>
- [12] <http://www.kanzaki.com/>
- [13] [www.technorati.jp/tag/Web2.0BOOK](http://www.technorati.jp/tag/Web2.0BOOK)
- [14] <http://www.impressjapan.jp/books/>
- [15] <http://www.web-30.biz/>
- [16] <http://cci.mit.edu>