

**■S3 群 (脳・知能・人間) - 3 編 (人工知能と学習)****7 章 Web インテリジェンス**

(章主任：高間康史) [2018 年 12 月 受領]

**■概要■**

Web インテリジェンスとは、一言で言ってしまうと Web を活用するための知的情報処理技術、ということになる。狭義には、HTTP プロトコルに従うインターネット上のアプリケーションが Web であるが、広義にその存在を考えれば、我々にとって最も身近かつ大規模な情報源、コミュニケーションの場であり、実社会で欠かすことのできない大きな存在である。また、Web が日々変化・成長を続ける存在である一方、インテリジェンスの定義も研究の進展に伴い変化し続けている。したがって、Web インテリジェンスに関して固定的な定義を与えることは現状では難しい。

日々変貌を遂げる Web インテリジェンスに関してハンドブックとしてまとめるにあたり、本章では個々の要素技術についてまとめるのではなく、より大きな視点で、人と情報環境との相互作用の観点から取り上げる項目を定めることとした。具体的には、Web パーソナライゼーション (7-1 節)、Web 情報アクセス (7-2 節)、Web とトラスト (7-3 節)、Web 知識創造 (7-4 節)、セマンティック Web (7-5 節) に関してその概要をまとめた。

本構成を定めたのは 2007 年頃であり、記事の多くもその頃に執筆されたものである。現在では Twitter や Facebook といったソーシャルメディアの普及、Instagram や YouTube などによるメディアの多様化、スマートフォンからのアクセスの一般化、フェイクニュース問題の深刻化など、当時とは状況が大きく異なっており、それに伴い Web インテリジェンスが対象とする研究も変化している。しかし、要素技術や対象に変化はあっても本章で取り上げた項目は普遍的なものであり、また当時を知る資料としても価値あるものと考えている。

**【本章の構成】**

本章では以下について解説する。

- 7-1 Web パーソナライゼーション
- 7-2 Web 情報アクセス
- 7-3 Web とトラスト
- 7-4 Web 知識創造
- 7-5 セマンティック Web

## ■S3 群-3 編-7 章

### 7-1 Web パーソナライゼーション

(執筆者：土方嘉徳) [2008年12月 受領]

Web が誕生した後、そのビジネス的な利用が始まるまでには、それほど時間がかからなかった。1990年代後半には、推薦システムに関する特集号が ACM の学会誌で生まれ、この頃には Web 視聴率調査のサービスが多数出現していた。企業の側から見ると、インターネットは消費者に直接アクセスすることのできるパスを提供しており、そのうえで実世界の店舗と同様の顧客に応じた販売活動を行うことは必須となっていた。このように、ユーザの特性（嗜好、興味、知識の程度など）に応じたアイテム（商品やサービス、コンテンツなど）の推薦、加工、提示の総称を、パーソナライゼーションと呼ぶ。

パーソナライゼーションの中心的な機能は情報推薦であるが、本章ではもう少し高い視点から、Web におけるユーザ適応について解説する。具体的には、パーソナライゼーションの歴史、Web 上でパーソナライゼーションを実現するための一般的なフレームワーク、パーソナライゼーションの機能、最後にパーソナライゼーションの今後の展開について説明する。

#### 7-1-1 パーソナライゼーションとは

インターネットビジネスにおける主要な機能としてパーソナライゼーション (Personalization)<sup>1)</sup> がある。この定義は、「ユーザに適した情報をユーザに適した形式で提示する技術やサービス」を言う。インターネットビジネスの世界でよく使われる情報推薦 (レコメンデーション) は、パーソナライゼーションの一種と考えられる。情報推薦の紹介は、7群6編6章6-3節にある。情報推薦の定義は、「ユーザの興味や嗜好に応じた情報を提示する技術やサービス」である。パーソナライゼーションは、ユーザの知識の程度や、利用時の習慣など、すべてのユーザの情報を含んでいる点で異なる。また、単に情報を提示するだけでなく、その見せ方までユーザ適応させる点で異なる。現在のところ、情報推薦には標準となる手法がある程度確立されているが、パーソナライゼーションで標準と呼べる手法はまだ存在していない。パーソナライゼーションの定義は、情報推薦の定義よりも広義であるため、汎用的な手法の開発に至っていないことが原因と思われる。また、評価用のデータセットの提供が困難であるため、情報推薦の研究の方が盛んであることも、理由の一つと思われる。パーソナライゼーションと似た意味の言葉としてカスタマイゼーション (Customization) がある。パーソナライゼーションは、システム側がユーザの情報を暗黙的、あるいは明示的に獲得し、その情報に合わせて情報やその見せ方を適応するのに対し、カスタマイゼーションは、ユーザがシステムを自分好みに改変する能力を意味する。

#### 7-1-2 パーソナライゼーションの歴史

Web はインターネットプロトコル上で動作する分散型ハイパーテキストシステムと言える。ハイパーテキストの目的は、情報と情報を人間が結び付けておくことで、ユーザに関連する情報を次々と閲覧していくこと (情報探索) を可能とすることである。インターネットの普及と同時に、ハイパーテキスト上で、ユーザのパーソナライゼーションを行う研究が行われ始めた。その起源は、1994年に行われた Workshop on Adaptive Hypertext and Hypermedia にさかのぼると

言える。このワークショップでは、ハイパーテキスト上の様々なユーザ適応と、そのためのユーザプロファイリング技術について議論がなされてきており、今では国際会議 International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, 及び、国際会議 International Conference on User Modeling, Adaptation, and Personalization に発展している。

また、1986年にMITのマローン (Malone) らが開発した Information Lens<sup>2)</sup> に始まり、従来の情報検索及び CSCW の分野で発展した情報推薦 (フィルタリング) の研究も、貢献を果たしてきている。情報推薦の研究は、評価用のデータセットを構築しやすかったこともあり、機械学習、データマイニング、データ工学などの各分野で、主要な研究テーマの一つにまで発展している。

2000年には米国 ACM の学会誌にて、パーソナライゼーションの特集号が組まれている。この特集号では、1990年代に行われた様々なパーソナライゼーションの研究事例、サービス事例、そしてビジネス的な効果 (影響) についても報告されている。

### 7-1-3 Web パーソナライゼーションのための基本フレームワーク

Web パーソナライゼーションを実現するための基本概念として媒介サーバ (Intermediary)<sup>3)</sup> がある。媒介サーバとは、サーバとクライアントの間に位置し、それらのメッセージのやり取りに介入し、ユーザに適したメッセージに改変するサーバである。媒介サーバには、以下のような実装形態が考えられる\*1。

#### (1) クライアントサイド

Web ブラウザへの Web ページのダウンロード中に、クライアントサイドでコンテンツをユーザに適応させる。クライアント上のスタンドアローンのアプリケーションや、Web ブラウザへのプラグインとして実装される (図 7・1(a) 参照)。任意の Web サイト中の Web ページも適応可能な点、ユーザの詳細なマウス操作などを検出できる点が利点である。ただし、インストールに関わるユーザの負担がある。

#### (2) プロキシサイド

サーバとクライアントの間に位置し、任意のクライアントからきたリクエストを加工して、任意の Web サーバに渡し、帰ってきた Web ページを加工して、クライアントに返す。一般的には、サーバともクライアントとも異なるサーバ (プロキシサーバ) 内で、独立のアプリケーションとして実装される (図 7・1(b) 参照)。単一のクライアントだけでなく、複数のクライアントで利用可能なこと、そして任意の Web サイトに対する適応を実現可能なことが利点である。ユーザの情報を常にトラッキングするためには、HTTP リクエストを必ずプロキシサーバを経由させる必要がある。そのため、通過する HTML 中の <A> タグ中の href 属性は、プロキシサーバの URL に書き換え、その URL 中にパラメータとしてオリジナルの URL を保存する必要がある。

#### (3) サーバサイド

サーバがクライアントからのリクエストに対応する Web ページを送信する際に、サーバサイドでコンテンツをユーザに適応させる。一般的には、Web アプリケーションサーバにおいて、

\*1 図 7・1 中の HTML の図において、黒字はオリジナルの部分、青字は <A> タグのみ変更した部分、赤字はユーザ適応した部分を表す。

EJB (Enterprise JavaBeans) などの形式で実装される (図 7・1(c) 参照). 動的にコンテンツを作成する Web サーバにおいては, コンテンツをリレーショナルデータベース (RDB) 内に保管していることが多く, RDB と親和性の高い Web アプリケーションサーバ上で実装するのが効率が良い. ユーザにインストール作業を強いることなく, パーソナライゼーションが実現できる.

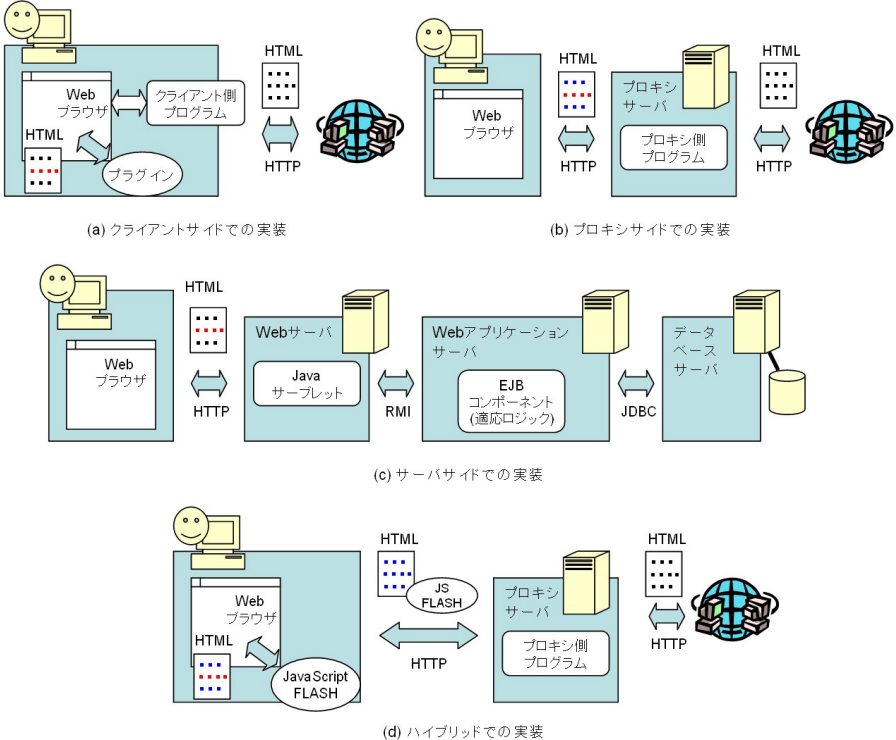


図 1・1 パーソナライゼーションの実装方式

#### (4) ハイブリッド

上記(1)~(3)の形態を組み合わせた実装形態である. パーソナライゼーションにおいては, ユーザに近いところで, ユーザの情報を獲得し, その情報に基づく適応結果を返す方が, よりきめ細かなサービスを行うことができる. また, 複数のユーザの情報を利用の方が, 協調フィルタリングのアルゴリズムを適用することができ, より高度なサービスを行うことができる. この両者を実現するには, (1)と(2)または(1)と(3)を組み合わせる必要がある. 任意の Web サイトに対する適応を実現するには, (1)と(2)を組み合わせる必要がある. また, ユーザへのインストール作業を回避するためにクライアントサイドのプログラムとして, JavaScript や FLASH のプログラムを, コンテンツの HTML に埋め込むこともある (図 1・1(d) 参照).

### 7-1-4 Web パーソナライゼーションの機能

Web パーソナライゼーション（特に、ハイパーテキスト上でのユーザ適応）の機能には、content-level adaptation (CLA) と link-level adaptation (LLA) がある<sup>4)</sup>。前者は、ハイパーテキストのリンク以外の部分のユーザ適応であり、後者はリンク部分のユーザ適応である。CLA は、更にテキストの適応、マルチメディアデータの適応、表現メディアの適応に分けられる。テキストの適応は数多くの研究がなされている。ある条件に適合したときのみテキスト部分を表示する表示する“conditional text”や、ユーザのリクエストにより更に詳細を表示する“stretchtext”などがある。LLA は更に、下記のように分類できる。

#### (1) Direct Guidance

次にたどるべきリンクに説明をつけたり、次にたどることを推薦する付加的なリンクを動的につけることにより、ユーザにどのリンクをたどればよいかという示唆を与える方式である。

#### (2) Ordering

ユーザに適している順にリンクをソートすることにより、ユーザが情報を自分の要求に合っていそうな順に見ることができる方式である。

#### (3) Hiding

リンクを見せなくすることにより、探索できるハイパー空間を制限する方式である。

#### (4) Annotation

リンクに付加的な装飾をつけることにより、ユーザがリンク先のページを既に訪れたかどうかや、リンク先のページがユーザにお勧めかどうかといった、リンク先についての簡単な説明を視覚的に見せる方式である。

#### (5) Generation

類似検索などの技術により現在閲覧中のページに関連のあるページを発見し、そのページへのリンクを付加する方式である。

また、Maglio らは Intermediary の役割として、customize, filter, annotate, aggregate, transcode, cache があると述べている<sup>3)</sup>。customize は、どちらかというパーソナライゼーションのことで、ユーザの情報によりページの適応を指している。filter は、コンテンツやリンクの選択的表示、aggregate は複数の情報源からの情報の結合、transcode は、情報の形式の変換（PC 用コンテンツから PDA 用コンテンツへの変換など）、cache は、後の利用時に高速にアクセス可能とするためのキャッシュである。

### 7-1-5 パーソナライゼーションの未来

パーソナライゼーションの研究の歴史は浅くはないが、研究によって得られている知見はまだ多くなく、多くの課題も残されている。特に、情報推薦の研究に見られる評価用データセットが確立されておらず、統一的な指標のもとで研究することができない。また、パーソナライゼーションはサービスと直接的に結び付くため、弊害として研究が個別のアプリケーションの提案の形態をとっていることも問題である。多くのサービスで用いられるような基盤技術が何かを模索する段階にきていると言える。

■参考文献

- 1) D. Riecken(ed.) : “Personalized Views of Personalization,” Commun ACM, vol.43, no.8, pp.26-158, 2000.
- 2) T.W. Malone, et al. : “Semi-structured Messages are Surprisingly Useful for Computer-Supported Coordination,” Proc. of CSCW'86, pp.102-114, 1986.
- 3) P. Maglio and R. Barrett : “Intermediaries Personalize Information Streams,” Commun ACM, vol.43, no.8, pp.96-101, 2000.
- 4) P. Brusilovsky : “Adaptive Hypermedia,” User Modeling and User-Adaptation Interaction, vol.11, pp.87-110, 2001

## ■S3 群-3 編-7 章

### 7-2 Web 情報アクセス

(執筆: 高間康史) [2008 年 11 月 受領]

Web の基本的特徴は、ページ (Web ページ) を基本単位とし、それらがハイパーリンクでつながれたハイパーテキスト構造を持っていることである。この基本的特徴は Web の登場時から今まで変化していないが、そのシンプルさの故にイノベーションを生み出す潜在能力を持っていた。その結果、動的なページ生成などの技術的進歩、学術利用からビジネス応用へといった用途の多様化などにより、Web 2.0 と呼ばれるほどの大きな変貌を遂げている。これに伴い、Web 上の情報へアクセスする手法、及びこれを支援する知的技術も時代によりその主流が変化してきている。

Web が普及した当初 (1999 年代半ば～後半) は、ハイパーリンクをたどりながら関連する Web ページを次々に閲覧する、ブラウジング (Browsing) が主流であった。その後、Web サイト、Web ページ数の爆発的増加に伴い、より効率良く情報を収集する手段としてサーチエンジン (検索エンジン) が登場したことにより、キーワードを用いて関連する Web ページを幾つか見つけた後、それを起点として近隣のページをブラウジングするアクセス手法が一般的となった。更に近年では、ブログ (Blog) やソーシャルタギングなどに見られるコンテンツの多様化により、サーチエンジンのような汎用的情報アクセスでは対処が困難となってきた。また、情報推薦などの個人適応や、ユーザ体験を重視したインタフェース改善なども、Web 情報アクセスに多大な影響を与えつつある。以下では、ブラウジング、サーチエンジン、新しい潮流の 3 段階に分け、それぞれ関連する人工知能技術について説明する。

#### 7-2-1 ブラウジングによる情報アクセス

ブラウジングによる情報アクセスの問題点は、リンク先に何があるかはたどってみるまで分からず、また全体的な空間構造を把握することができないため、リンクをたどっているうちに迷子になりやすい点である。前者の問題点を解決するために、たどるべきリンクをユーザに提案することによるブラウジング支援が提案されている。代表的システムである Web Watcher<sup>1)</sup> では、同一目的のユーザがあるページにおいてリンクを選択する傾向を学習し、リンク推薦を行う。

後者の迷子問題に対する解決策として、現在閲覧中のページを中心として、周囲のページとのリンク関係を視覚的に提示することによりブラウジング支援を行うシステムが提案されている。代表的なシステムに Web Cutter<sup>2)</sup> があり、探索対象サイト、マップサイズ、検索文字列などを指定すると、指定サイトからページを収集し、検索文字列を用いて関連ページをフィルタリングするなどの処理を行ってサイトマップを作成する。サイトマップは情報可視化技術を用いて提示され、インタラクティブに閲覧ページを選択することができる。

#### 7-2-2 サーチエンジンを用いた情報アクセス

サーチエンジンを用いた情報アクセスでは、クエリ (Query) の質を向上させることが重要であり、代表的な技術にクエリ拡張 (Query Expansion)、適合性フィードバック (Relevance Feedback) がある。クエリ拡張は、ユーザが入力した初期クエリに関連キーワードを追加することにより、

検索精度の向上を図る技術である。シソーラスなどを用いて、初期クエリ中のキーワードに関するキーワードを追加するアプローチや、検索結果を分析し、頻出キーワードなどを関連キーワードとして抽出するアプローチなどがある。関連キーワードが検索結果を絞り込む効果を情報可視化 (Information Visualization) 技術を用いてユーザに提示し、クエリに用いるものを選択させる半自動アプローチもある。VQuery<sup>3)</sup> では、関連キーワードの組合せによる検索結果の絞り込み効果をベン図を用いて可視化・提示し、インタラクティブなクエリ指定を可能としている。また、サーチエンジンに入力されたクエリログを分析し、共起するキーワードを関連キーワードとする社会的アプローチもあり、商用サーチエンジンにおけるキーワード提案機能などで利用されている。

適合性フィードバックは、ベクトル空間モデル (Vector Space Model : VSM) を用いた情報検索システムで用いられることが多い技術であり、検索結果の各文書 (Web ページ) に対してユーザが下した適合・不適合の判断に基づきクエリベクトルを更新する。このほか、対話的進化計算がクエリ更新に用いられることもある。適合性フィードバックでは、検索結果の評価や、探索を繰り返す必要があるなどのユーザ負荷が問題となる。また、適合・不適合の判断を下すことは必ずしも容易な作業ではない。ユーザ負荷を軽減するために、多段階での評価や、明示的な評価を必要としない暗黙的フィードバックなどの改善案が提案されている。

## 7-2-3 情報アクセスにおける新しい潮流

### (1) 主観的情報へのアクセス

かつて、情報は客観的であることが重視され、個人的意見や感想などの主観的意見はノイズの一種とみなされていた時代もあった。しかし現在では、個人にとっては商品購入などの意思決定や流行を知るために、企業にとっては消費者の意見などを知るために欠かせないコンテンツの一つとして重要視されている。商品に対する好評・不評などの意見は評判情報 (Sentiment Information) と呼ばれ、ある対象に対する評価記述の抽出や、対象に対する文書の肯定・否定の判別などが研究されている。また、ブログにおける単語の出現頻度の変化などから、注目されているトピックやその変遷などを分析する、動向情報 (Trend Information) の処理に関する研究も行われている。

### (2) プッシュ型情報アクセス

ユーザがサーチエンジンなどを用いて能動的に情報アクセスを行う形態 (プル (Pull) 型) に対し、ユーザが興味・関心を持つ情報をシステム側から提供する形態をプッシュ (Push) 型と呼ぶ。オンラインショッピングサイトにおける商品の推薦や、メールによる関連情報の配信などがある。個人の興味・関心に合わせた情報提供を行うために、Web パーソナライゼーション技術が用いられる (本章 7-2 節参照)。プッシュ型は 1990 年代後半に話題となったものの、当時はインターネットへの常時接続が一般的でなかったなどの理由で普及しなかった。しかし、携帯電話など常時接続が一般的となった現代では主要な情報アクセス技術の一つとなっている。

### (3) 情報可視化技術の応用

現在主流のサーチエンジンでは、検索結果を線形リスト形式で提示しているが、検索結果の閲覧にかかるユーザ負荷を軽減するために、情報可視化技術を用いてユーザに分かりやすく提示するシステムの研究開発が行われている。代表的なシステムに、検索結果をサブトピックご



とに階層的にクラスタリング (Clustering) して提示するシステムや、Web ページや商品などの検索アイテム間の関係をグラフ構造として提示するシステムなどがある。

#### (4) Web ページ単位の情報アクセスからの脱却

従来の Web 情報アクセスは Web ページ単位での情報収集・閲覧であったが、必要な情報断片の単位で情報へアクセスできることが、Web の更なる活用の観点から望ましい。このため、質問応答型 (Question Answering) の情報アクセスや、テキスト要約技術 (Text Summarization) の応用などが研究されている。また、セマンティック Web で用いられるメタデータ (Meta Data) などを活用した情報アクセスは、情報の編集・再構築を可能にするアプローチである (本章 7-6 節参照)。メタデータは自動的に付与される場合もあるが、フォークソノミー (Folksonomy) やソーシャルタギング (Social Tagging) など、人手によるメタデータを Web 上で共有することも一般的となっている。このようなユーザ参加型のアプローチはスケーラビリティに優れるだけでなく、新たな知識や流行を生み出す集合知 (Collective Intelligence) としても期待されている。

#### ■参考文献

- 1) R. Armstrong, D. Freitag, T. Joachims, and T. Mitchell : “WebWatcher: A Learning Apprentice for the World Wide Web,” AAAI Spring Symposium on Information Gathering from Heterogeneous, Distributed Environments, 1995.
- 2) Y.S. Maarek et al. : “WebCutter, A System for Dynamic and Tailorable Site Mapping,” 6th WWW Conference, 1997.
- 3) S. Jones : “VQuery: a Graphical User Interface for Boolean Query Specification and Dynamic Result Preview,” Working Paper 98/3 March 1998, Department of Computer Science, University of Waikato, New Zealand, 1998.

## ■S3 群-3 編-7 章

### 7-3 Web とトラスト

(執筆: 藤本和則) [2008年11月 受領]

トラストは「人間が相手の行為 (Action, Behave), 能力 (Competence, Capability), 意志 (Willingness) に対してもつ信念 (Belief) の一つ」とされ, 古くから経済学, 社会学, 心理学, 工学などの領域で幅広い視点から研究されてきた. コンピュータサイエンスの領域では, 近年, Web が知識社会のインフラとしての地位を固めるに伴い, Web 上のエージェントやコンテンツに対してのトラストの研究が進められている. Web におけるトラスト研究では, Web でつながった面識のない相手への信頼, Web 上の不特定多数のエージェントや多種多様なコンテンツへの信頼など, これまで扱われなかったあるいは中心とならなかった信頼が主要な研究対象となる. こうしたトラスト研究は, 従来のトラストの概念を Web 上での新しいトラストへと拡張し発展させるものである. 以下では, トラストの概念を整理したうえで, Web 上の情報を使ったトラストの計算モデルについて解説する.

#### 7-3-1 Web におけるトラスト

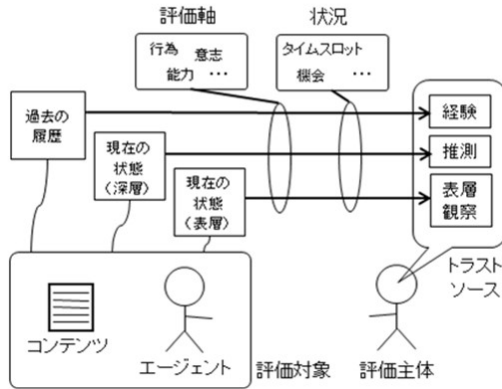
トラスト (Trust) は「信頼」と訳されることが多く, この点でリライアビリティ (Reliability) やクレディビリティ (Credibility) といった用語と混同されやすい. トラストは従来から戦略的な事柄を含む概念<sup>1)</sup>とされ, この見方が多様な信頼性の概念整理に役立つ. ここで, 戦略的とは, 信頼するか否かは「信念の度合い」のみではなく, 「効用やリスク」も合わせて決められるとする立場である<sup>2)</sup>. Web において, 特にトラストという言葉が使われる場合には, 適切なサービスの選択など暗に損得勘定が前提となる場合や, 複数エージェントにおけるゲーム理論のような視点など, 背後に戦略的な事柄があると考えてよい. 一方, 特にクレディビリティという場合には, Web 上のコンテンツや発信者に関する信念の度合いが研究対象とされ, その外側の戦略的な事柄とは切り離されることが多い.

Web におけるトラスト研究では, エージェントのみではなく, コンテンツに対してもトラストという言葉が用いられる<sup>3)</sup>. コンテンツにも積極的にトラストという言葉が使われるようになった背景には, Web の信頼性とその重要さを増したことに加え, 戦略的なレベルではコンテンツへの信頼はエージェントへの信頼と同じ構造を持つこと, また自動で更新される Web 上のコンテンツは擬人化されやすいことがあると考えられる. セマンティック Web の技術階層の最上位には, エージェントの処理結果の信頼性を保証するためのトラスト層が設けられる. ここで言う信頼性には, 理想的には, 本節に述べるトラストはもちろんのこと, 電子署名や暗号化などのセキュリティ技術も含まれる<sup>4)</sup>. トラスト層の研究はまだ緒についたばかり<sup>5)</sup>であるが, 多様化するトラストの研究を整理し, 具体的な規格として標準化する研究活動の意義は大きい. トラスト計算についてはその計算過程の透明性, すなわち, どの情報をどのように使ってトラストを導いたかを開示する仕組みの重要性が指摘されており, セマンティック Web の証明層に関連して多くの研究が進められている<sup>6)</sup>.

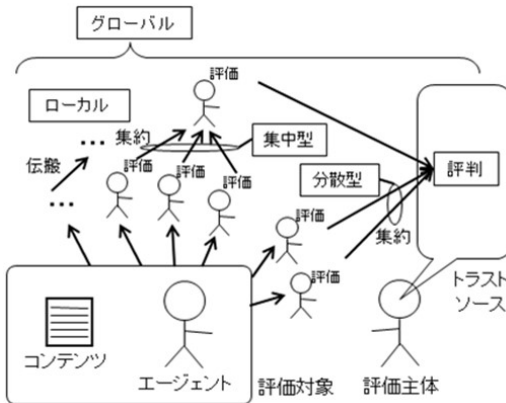
#### 7-3-2 トラストの計算モデル

トラストの計算モデルには個人的視点と社会的視点がある. 前者ではエージェントの内部で

形成されるトラストのモデル化に興味があり、トラストの認知モデルとの関連も大きい。一方、後者では社会ネットワークのなかで形成されるトラスト、特に評判 (Reputation) のモデル化に興味がある。以下では、それぞれについて解説する。



(a) 個人的視点



(b) 社会的視点

図 3・1 トラスト計算モデルの基本構成

### (1) 個人的視点

個人的視点の計算モデルは、経験、推測、表層観察などのトラストソースからトラストが生成されるという基本構成をとる (図 3・1(a))。トラストソースへの入力には、過去の履歴、現在の状態 (表層、深層) があり、それぞれ行為、能力、意志などの評価軸で評価される。評価にあたってはタイムスロットや機会などの状況が考慮される。現在の状態の「表層」と「深層」は、それぞれ対象の表面的な側面、意味的な側面を指す。表層については Web サイトのどのよ

うな要因が信頼性に影響を与えるかの心理的な検証<sup>7)</sup>など、深層については複数の情報間の矛盾に基づいたトラストの評価<sup>8)</sup>などが研究される。初期の計算モデルには、時間変化する重要度、効用、トラストの数値を用いて協調するか否かを定める計算式がある<sup>9)</sup>。その後、ファジィ認知マップを使ったモデル<sup>10)</sup>や、Web ページのクレディビリティを扱うモデル<sup>11)</sup>などが提案されている。

## (2) 社会的視点

社会的視点の計算モデルは、複数のエージェントがそれぞれ相手を評価するというネットワークから評判が生成されるという基本構成をとる (図 3・1 (b))。計算モデルの分類には、計算される評判の客観性からのものと、計算アーキテクチャからのものがある。前者の分類は、グローバル (各エージェントはネットワークにおいてユニークな評判を持つ) と、ローカル (各エージェントの評判は評価する主体によって異なる) から成る<sup>12)</sup>。グローバルについては、分散環境を前提に個々のエージェントの評価からグローバルな評判を計算するアルゴリズム<sup>13)</sup>などが研究される。また、ローカルについては、統計的な伝搬とファジィ的な集約を合わせた計算モデル<sup>14)</sup>など個人的視点にも通じる研究がされる。一方、後者の分類は、集中型 (ある機関が代表して計算する)、分散型 (各エージェントが個別に計算する) から成る<sup>15)</sup>。集中型の研究はオークションサイトや商品レビューサイトに適用され<sup>16)</sup>、分散型の研究は P2P ネットワークの分野への適用が多い<sup>17)</sup>。

## ■参考文献

- 1) E.M. Uslander : “Trust in the knowledge society, paper prepared for the Conference on Social Capital,” Cabinet of the Government of Japan, Mar. 24-25, 2003. (邦訳: “知識社会における信頼,” 宮川公男・大守 隆 (編) : “ソーシャル・キャピタル: 現代経済社会のガバナンスの基礎,” pp.123-154, 東洋経済新報社, 2004)
- 2) R. Falcone and C. Castelfranchi : “Social trust: a cognitive approach,” in C. Castelfranchi and Y-H. Tan (Eds.) : “Trust and deception in virtual societies,” pp.55-90, Kluwer Academic Publishers, 2001.
- 3) Y. Gil and D. Artz : “Towards content trust of web resources,” in Proceedings of the 15th International World Wide Web Conference, pp.565-574, 2006.
- 4) D. Artz and Y. Gil : “A survey of trust in computer science and the Semantic Web,” Web Semantics: Science, Services and Agents on the World Wide Web archive, vol.5, no.2, pp.58-71, 2007.
- 5) 武田英明: “信頼の Web,” コンピュータソフトウェア, vol.22, no.4, pp.19-25, 2005.
- 6) S. Sizov : “What makes you think that? The semantic web’s proof layer,” IEEE Intelligent systems, vol.22, pp.94-99, 2007.
- 7) B. Fogg et al. : “Stanford-Makovsky Web Credibility Study 2002: Investigating what makes Web sites credible today,” A Research Report by the Stanford Persuasive Technology Lab in collaboration with Makovsky & Company. Stanford University, 2002.
- 8) X. Yin, J. Han, P.S. Yu : “Truth Discovery with Multiple Conflicting Information Providers on the Web,” IEEE Transactions on Knowledge and Data Engineering, vol.20, no.6, pp.796-808, 2008.
- 9) S.P. Marsh : “Formalising Trust as a Computational Concept,” PhD thesis, Department of Computer Science and Mathematics, University of Sterling, 1994.
- 10) R. Falcone, G. Pezzulo, and C. Castelfranchi : “Quantifying belief credibility for trust-based decision,” in AAMAS-02 Workshop on Deception, Fraud and Trust in Agent Societies, pp.41-48, 2002.
- 11) B. Cugelman, M. Thelwall, and P. Dawes : “Website Credibility, Active Trust and Behavioural Intent,” in Proceedings of the Third International Conference of Persuasive Technology (PERSUASIVE 2008), pp.47-57, 2008.
- 12) C-N. Ziegler and G. Lausen : “Propagation Models for Trust and Distrust in Social Networks,” Information Systems Frontiers, vol.7, no.4-5, pp.337-358, 2005.

- 13) S.D. Kamvar, M.T. Schlosser, and H. Garcia-Molina : “The Eigentrust algorithm for reputation management in P2P networks,” in Proceedings of the 12th International World Wide Web Conference (WWW2003), pp.640-651, 2003.
- 14) S. Shekarpour and S. Katebi : “A Trust Model Based on Statistical Propagation and Fuzzy Aggregation for Semantic Web,” in: Second UKSIM European Symposium on Computer Modeling and Simulation (EMS2008), pp.459-464, 2008.
- 15) A. Josang, R. Ismail, and C. Boyd : “A Survey of Trust and Reputation Systems for Online Service Provision,” Decision Support Systems, vol.43, no.2, pp.618-644, 2007.
- 16) E. Chang, F. Hussain, and T. Dillon : “Trust and Reputation for Service-Oriented Environments: Technologies For Building Business Intelligence And Consumer Confidence,” Wiley, 2006.
- 17) 相良 毅, 井口 誠, 藤本和則 : “Web Trust 研究動向,” 人工知能学会誌, vol.21, no.4, pp.430-437, 2006.

## ■S3 群-3 編-7 章

## 7-4 Web 知識創造

(執筆者：庄司裕子) [2009年3月 受領]

## 7-4-1 知識創造のプロセス

ピーター・ドラッカーは、知識が重要な役割を果たす現代社会を「知識社会」と呼び、この中では知識が最も重要視される資産であると述べた<sup>1)</sup>。伝統的な社会が労働、資本、土地といった資産の価値を重視してきたのに対し、知識社会では知的資産 (Intellectual Capital) が絶対視される。したがって、企業や国が競争力を高めるためには、知識が如何に創られるかの仕組みやプロセスを理解し、知識創造を促して知の生産性を高めることが必要となってきた。このような背景から、1990年代以降、知識創造のプロセスに関する研究が注目を集めた。なかでも、組織の知識創造プロセスに関するモデルとして広く知られているのが、野中らによる SECI モデルである<sup>2)</sup>。

SECI モデルでは知識を暗黙知 (Tacit Knowledge) と形式知 (Explicit Knowledge) の2つに分類し、その相互変換に着目する。形式知とは言葉で表現できる知識であり、分析的に生み出せる知識、共有しやすい形態の知識である。一方、暗黙知は言葉で表現することが困難な知識であり、経験の集積のなかで獲得されるノウハウなど、一人一人異なる個人的な知識である。形式知と暗黙知が相互に変換しながら知識が創造されると仮定する SECI モデルでは、以下の4つの知識変換モード (すなわち知の創り方) が考えられる。4つの変換モードは単独に存在しても意味がなく、4つをスパイラル状に回しながら個人の知と組織の知が増幅し、知識創造がなされるという考え方である。

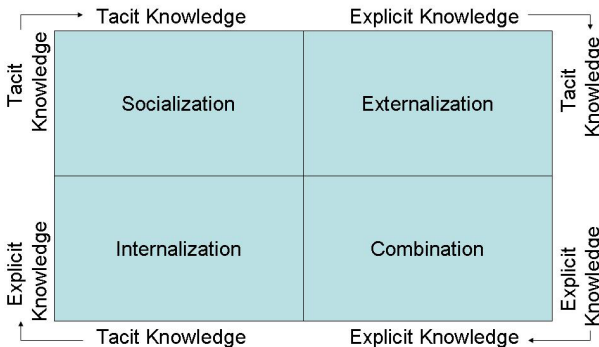


図 4・1 SECI モデルの4つの知識変換モード

- (1) 共同化 (Socialization) : 個人の暗黙知からグループの暗黙知を創造するプロセス。経験を共有することによって他者の持つ暗黙知を伝搬し獲得する。
- (2) 表出化 (Externalization) : 暗黙知から形式知を創造するプロセス。思いやノウハウを明示的な概念、すなわちコンセプトにする。
- (3) 連結化 (Combination) : 個々の形式知を組み合わせることで体系的な形式知を創造するプロセス。

ス。言葉やコンセプトを組み合わせる作業であり、情報技術の寄与が大きい。

- (4) **内面化 (Internalization)** : 形式知から暗黙知を創造するプロセス。形式知を実践することによって、頭で分かったことを自分のものとして体得する。

#### 7-4-2 Web と知識創造

近年、インターネットに代表される情報技術の進展によって、形式知を上手に蓄積・共有・流通できるようになってきた。Web 上にも膨大な知識が蓄積されているが、検索エンジンやテキストマイニングの技術の貢献により利用者は必要な知識に到達し利用できるようになっている。Web が知識社会のインフラとなりつつある現在、情報源としての Web は知識創造に不可欠な存在であると言っても過言ではない。

一方、暗黙知については、言葉で表現できないという特性上、Web 上で直接ハンドリングすることは難しい。暗黙知は言葉で伝えることが難しいため、対話形式でのコミュニケーションを通して共有・流通されるのが一般的である。少人数での対話形式を前提とするため、知識の伝搬速度は遅い。専門家どうしの組織内での技能継承や、地域コミュニティでの慣習の共有などはその一例である。このような従来のコミュニティに加え、Web 上に従来とは異質のコミュニティが形成されてきている。Web 上のコミュニティでは、大多数のメンバーが誰でも容易に知識を発信することが可能であり、メンバー間の知識共有も瞬時に行われる。従来のコミュニティは比較的少数の固定メンバーによる閉鎖的な集団である場合が多かったが、大多数のメンバーによる大量の知識共有や創造が可能になったことで、コミュニティのあり方が大きく変化した。Web によって知識創造プロセスのオープン化が促されたと言える。Web は知識を利用するためのツールにとどまらず、知識創造の場としての機能を果たすようになってきている。

Web 上での知識創造スパイラルを上手に回すためには、まず、質の高いメンバーの参加を促進して質の高い知識がコミュニティに蓄積されることが必要である。そして、蓄積された知識を上手く活用して新たな知識が創造されコミュニティに還元されれば、結果として更に多くのメンバーの参加につながる。企業のような固定的な組織と異なり、コミュニティというオープンな場においては、良質なメンバーを維持しながら SECI モデルのスパイラルを回すことが重要である。成功例として注目されているものとして、オープンソース・コミュニティがある。Linux に代表されるようなオープンソース・ソフトウェアは、当初、無償だが信頼性は低いと思われていたが、Linux の品質は高く評価されている。

暗黙知の形成や流通を支援する情報システムに関しては、これまで主に CSCW や発想支援の分野で研究が進められ、様々なグループウェアが開発されてきた<sup>3)4)</sup>。これらの研究では固定メンバーによる閉じた組織での知識創造活動を対象とする。は Web 上のコミュニティ内での知識創造を支援するためには、閉じた組織を対象としてきた従来のグループウェアとは異なり、参加メンバーが流動的であり、時間的にも空間的にも分散した環境で行われることを考慮する必要がある。また、従来のグループウェアが問題解決を支援するものであったのに対し、コミュニティではあらかじめ設定された問題ではなく気づきや出会いといった問題発見を促すことが求められる<sup>5)</sup>。コミュニティでの知識創造を支援するための研究も試みられており、支援システムの事例としては角らによるエージェントサロンなどが挙げられる<sup>6)</sup>。

## ■参考文献

- 1) P.F. Drucker : “Post-Capitalist Society,” HarperBusiness, 1993. (上田惇生, 佐々木実智男, 田代正美(訳) : “ポスト資本主義,” ダイヤモンド社, 1993.)
- 2) I. Nonaka and H. Takeuchi : “The Knowledge-Creating Company,” Oxford University Press, 1995. (梅本勝博(訳) : “知識創造企業,” 東洋経済新報社, 1996.)
- 3) 國藤 進 : “創造的思考のための知識獲得支援グループウェア GRAPE,” 日本創造学会編, 創造性研究 10, 共立出版, 1994.
- 4) 角 康之, 西本一志, 間瀬健二 : “協同発想と情報共有を促進する対話支援環境における情報の個人化,” 電子情報通信学会誌, vol.J80-D-1, no.7, pp.542-550, 1997.
- 5) 大澤幸生(編著) : “IT Text 知識マネジメント,” オーム社, 2003.
- 6) 角 康之, 間瀬健二 : “エージェントサロン : パーソナルエージェント同士のおしゃべりを利用した出会いと対話の促進,” 電子情報通信学会誌, vol.J84-D-1, no.8, pp.1231-1243, 2001.



## ■S3 群-3 編-7 章

### 7-5 セマンティック Web

(執筆: 大向一輝) [2018 年 12 月 受領]

#### 7-5-1 セマンティック Web とは

セマンティック Web は Web 上の情報の意味をソフトウェアに理解させ、知的処理を行うことを目的とした一連の技術ならびに研究分野である<sup>1)</sup>。Web の発明者である Berners-Lee によって提案された。セマンティック Web は 1960 年代から研究が進められてきた意味ネットワーク (Semantic Network) の一種であるが、成功を収めた Web の要素技術に即した形で知識を表現することで、大規模化と実用を目指す点が特徴的である。

意味ネットワークでは、あらゆる知識を概念 (ノード) とその関係性 (リンク) として定義する<sup>2)</sup>。セマンティック Web では、機械可読なネットワークの記述方法として RDF (Resource Description Framework) を採用している。RDF では、知識の最小単位は主語・述語・目的語からなる 3 つ組 (トリプル) である。主語と目的語はノードに相当し、リソースと呼ばれる。また、述語は主語となるリソースから目的語となるリソースへの有向リンクであり、プロパティと呼ばれる。あるトリプルの目的語が別のトリプルの主語となることが許容されており、トリプルを多数接続することで大規模なネットワークを構成することができる。すべてのリソースとプロパティには識別子としてグローバルに一意的な URI (Uniform Resource Identifier) が与えられ、これによって誰もが他者の知識に言及し、拡張することが可能である。

URI による識別とリンクは、HTML 文書を対象としてきた従来の Web と同様の方法論である。RDF はより粒度の小さい知識やデータをグローバルに共有するための Web 技術の拡張であると言える。セマンティック Web では、これに加えてクラス・インスタンスの定義やクラス間の上位・下位関係あるいは部分・全体関係といった知識体系をトリプルを用いて定義するため、RDFS (RDF Schema) ならびに OWL (Web Ontology Language) と呼ばれるオントロジー言語を提供している。

オントロジー言語では、クラス間の階層関係や排他的関係の明示化のほか、特定のプロパティが適用可能な主語・目的語のクラスを限定するなどの制約を定めることができる。オントロジー言語を解するプログラムは、これらの知識体系と RDF 形式のデータから、推論や矛盾の発見が可能になる。更には、ルールに基づく処理を定義する SWRL (Semantic Web Rule Language) や、クエリ言語 SPARQL、証明の機構、情報の真正性を保証するための電子署名などの技術を組み合わせることで、信頼性の高い知的処理を行うことがセマンティック Web の最終的な目標である。

#### 7-5-2 セマンティック Web と Linked Data

セマンティック Web の各要素技術は研究開発が進められている段階であり、その目標の達成には至っていない。また、各分野の専門家が RDF やオントロジー言語を用いて自身の持つ知識やデータを表現するにあたり、記述コストが過大であるという批判もある。一方、近年のウェブ上でのデータベースの普及や、データ公開の取り組みが増加したことでデータフォーマットが乱立し、統合的な処理が困難な状況にある。これらの課題に対して、汎用的なデータ表

現や公開のための手法が求められており，セマンティック Web 技術に基づく Linked Data が注目されている．Linked Data は RDF に基づく情報の構造化，Web 標準に適合したデータ公開，コミュニティでの語彙の共有を含む方法論を指す<sup>3)</sup>．

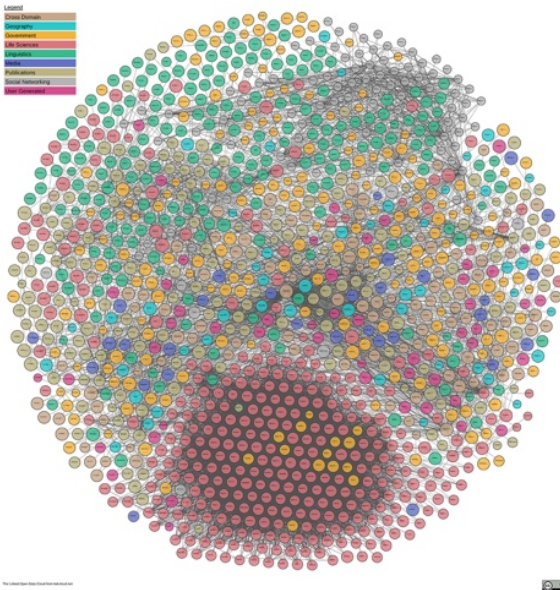
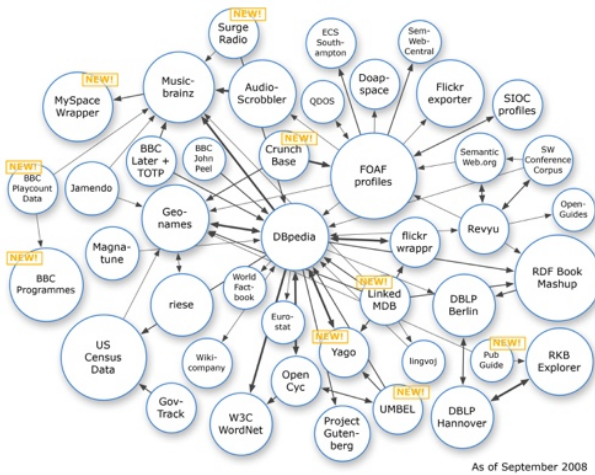


図 5・1 Linked Data 対応サイトの推移 (上) 2008 年 (下) 2018 年  
(CC BY The Linked Open Data Cloud from lod-cloud.net)

Linked Data の要件として，概念や事物に URI を与えることのほか，HTTP で URI にアクセスすると詳細な情報が RDF で入手可能であること，またそのデータに他サイトの URI が含まれていることが求められる．Linked Data の条件を満たすサイトの一覧は有志によって継続的に調査・収集されており，サイトのジャンルやリンク関係が可視化されている (図 5・1)．なお，

公開されたデータについて自由な再利用を許諾するライセンスが付与されている場合には LOD (Linked Open Data) と呼ばれることがある。

異なるサイト間でのデータの共通化のためには、クラスやプロパティの定義を揃えることが望ましい。セマンティック Web ではすべてのクラスやプロパティに一意な URI が与えられており、同一の URI を用いて表現されたデータは同様に解釈できることが担保される。この特性を生かして、コミュニティ内で必要なクラスやプロパティを定義し、その URI を公開・共有している例がある。これを語彙 (Vocabulary) と呼ぶ。

主要な語彙としては、書籍のタイトル、人名、日時の表記などに用いられる DCMI Metadata Terms (<http://dublincore.org/documents/dcmi-terms/>)、ソーシャルメディアにおけるプロフィールや知人関係を記述するための FOAF (Friend of a Friend) (<http://xmlns.com/foaf/spec/>)、緯度・経度を表すための Basic Geo Vocabulary (<https://www.w3.org/2003/01/geo/>) などが広く利用されている。また、大手検索エンジン提供者を中心として、店舗情報やレビューなど多様な Web サイトの情報に対応した語彙を定義する Schema.org (<https://schema.org>) が注目されている。政府レベルでは各省庁の文書の構造を定型化し、行政の効率性を高めることを目的として語彙の統一が図られている。アメリカ政府の NIEM (National Information Exchange Model)、EU の Joinup といった先行事例を参照し、日本でも共通語彙基盤の整備が進められている<sup>4)</sup>。

新たにデータ公開を行う場合には、これらの既存の語彙を組み合わせて記述し、不足している語彙を新たに定義することで、相互運用性を確保しつつ柔軟な知識表現が可能になる。

#### ■参考文献

- 1) T. Berners-Lee, J. Hendler, and O. Lassila: "The Semantic Web," Scientific American, pp.29-37, 2001.
- 2) F. Lehmann and E. Rodin: "Semantic Networks in Artificial Intelligence," Pergamon Press, 1992.
- 3) T. Heath and C. Bizer: "Linked Data: Evolving the Web into a Global Data Space," Morgan & Claypool, 2011.
- 4) 加藤文彦, 武田英明, 田代秀一, 平本健二, 松澤有三: "IMI 共通語彙基盤," デジタルプラクティス, vol.9, no.1, pp.33-53, 2018.