

京都大学における Semantic Web 研究

石田 亨 荒井幸代

京都大学情報学研究科社会情報学専攻

京都大学 COE「知識社会基盤構築のための情報学拠点形成(代表 上林弥彦教授)」の一環として行う Semantic Web 研究について、現時点の計画を述べる。社会情報学専攻では、2003 年より Human Centered Semantic Web プロジェクトを、4 年間を目途に開始させている。その中心的課題は、グローバル社会の、あるいはコミュニティの共有情報に対するセマンティックスの支援である。

我々の Semantic Web 研究は、論理的推論のように形式的にセマンティックスを扱うものではない。共有情報の意味を捉えようとするユーザの要求を支援するために、セマンティックスの扱いは非形式的にならざるを得ないと覚悟している。研究のアプローチは、「社会的目標を掲げ、研究開発と実証実験を並行して進める」もので、社会情報学研究の方法論に基づいている。

1. グローバル社会におけるセマンティックスの支援

人種、学歴などによる情報アクセス機会の格差が、メディアリテラシーの格差を生み出していると言われていた。個人責任の自覚がないまま、金銭的な被害やトラブルに巻き込まれたり個人情報や外部に流されたりする事件も生じている。こうした問題に対して、「個人情報保護法案」などの枠組み作りが進んでいるが、我々が注目するのは、ユーザ個々人のメディアリテラシーの支援である。すなわち、情報の送り手を規制するのではなく、受け手側が「情報の内容を自ら判断する」ことに対する支援である。

具体的には、ユーザのメディアリテラシーを支援するために、Web により閲覧される情報の信頼性を評価すると共に、閲覧される情報に対する多様な見方を世界各国の Web から収集し併せて提示する。研究の進め方を以下に示す。膨大な情報を対象とするため、帰納的なアプローチを考えている。

- 1) 情報の信頼性を評価する。社会的フィルタリングが個人価値による選別であったのに対し、信頼性によるフィルタリングは共通価値による選別である。すなわち、信頼性の計算プロセスを、株式や為替と同様に、ユーザ個々人が推定する共通価値を統合するプロセスと捉える。共通価値の統合プロセスとして、信頼性の伝播モデルが幾つか提案されているが、単一のモデルですべての領域の信頼性伝播を説明することは困難である。たとえば、研究者サイトと企業サイトの信頼性伝播のモデルは異なると考えるのが自然である。そこで複数のモデルを重み付けして合成する手法を研究する。また、信頼性伝播

の計算を正確なものとするために、リンクや参照のセマンティクスをメタタグによって記述する。すなわち、**Semantic Web** を用いた **Web of Trust** を実現する。

- 2) 情報に対する多様な見方を収集するために、テキスト解析を行う。Web 情報のセマンティクスは、その情報への多様な見方との関係において初めて理解できるものとする。例えば、「イラク戦争に対するブッシュ大統領の発言」には、世界各国で賛否の声が沸きあがる。そうした見方を網羅し、また、各々の見方の総量を添えてユーザに提示する。この方法を、例えば、組織や個人を批判する Web 情報に適用すれば、どの程度の世論を背景にしたものであるかを知ることができる。

2. コミュニティにおけるセマンティクスの支援

我々は 2002 年に、日中韓馬の 5 大学 40 余名の教員・学生による、母国語によるオープンソースソフトウェア開発実験を行った。その際に、言語、文化の違いを超えるプロジェクトでは、セマンティクスの支援が必要であることを痛感した。ここで扱うのは予め確立した知識体系ではない。コミュニティのインタラクションの中で生まれ、育ち、共有されるプロジェクト固有の概念（プロジェクトオントロジー）を扱う。例えば「デジタル北京」という概念の言語を超えた共有化を考える。これは「Digital Beijing」とも「Digital Peking」とも訳すことができる。しかしながら、その選択は、プロジェクトが現代的北京に力点を置くか、歴史遺産に注目するかによって依存する。このように、プロジェクト固有の概念は、その議論の過程でセマンティクスが与えられていく。

また、異文化コラボレーションの失敗の多くは、翻訳の問題ではなくプロトコルの相違に基づくと考えている。例えば、国内においてさえ、企業と大学では仕事の進め方が異なる。まして、文化、言語を異にする参加者間で、仕事のプロトコルが異なるのは当然である。そこで、参加者のプロトコル（コラボレーションオントロジー）を記述しその共有を図る。研究の進め方を以下に示す。少量の情報を対象とするため、演繹的なアプローチとなる。

- 1) プロジェクトオントロジーの構築は、単に用語集や対訳表を作るのではなく、討論の e-mail や関連資料を語彙に関連つける。例えば、「デジタル北京」のセマンティクスは、経緯を議論した e-mail や資料によって与えられる。異文化コラボレーションにおける概念の共有は、参加者のインタラクションを支援し集積することによって可能となると考える。
- 2) コラボレーションオントロジーの構築も同様に、プロトコルの記述と関連資料の共有から始まる。プロトコルには表層と深層がある。表層のプロトコルは合理的な議論によって収束させることができるが、深層のプロトコルは個人差、文化差があり、容易に収束しない。例えば、スケジュールをどの程度遵守すべきかの感覚は、文化により大きく異なる。プロトコルの記述とその解釈実行は別物と考え、実行時の調整を試みる。