

特集 「編集委員今年の抱負 2009：経糸から横糸まで」

## 共有知と集合知の活用

馬場 雪乃 東京大学大学院情報理工学系研究科



### 1. 共有知を活用したアプリケーション

The International Semantic Web Conference (ISWC 2008) への参加中に、この記事執筆している。昨年度の会議では Wikipedia からの情報抽出に関する研究が数件見られたのに対して本年度はほとんど見受けられないというのが、昨年からの大きな変化だといえる。では、セマンティック Web 研究者が Wikipedia に対する興味を失ったのかといえば、けっしてそうではないようだ。

昨年度の ISWC では DBpedia<sup>\*1</sup>[Auer 07] という、Wikipedia 上の情報を構造化し RDF の形式で提供するサービスが発表された。また本年度の基調講演では、Wikipedia を含めた複数の Web 上のソースからの情報を、やはり構造化し、さらには Web 上でユーザが自由にそのデータを編集できるようにした Freebase<sup>\*2</sup> というデータベースが紹介された。Wikipedia では人々がもつ知識が共有されており、そのような共有知の提供媒体として Wikipedia は依然強い関心がもたれている。しかし Wikipedia 上のデータを構造化して提供するこれらのサービスが現れたため、セマンティック Web に関わる人達は Wikipedia を直接扱うのではなく、これらの構造化データを活用すればよい、というのがセマンティック Web 界限での風潮のように感じた。実際、DBpedia を用いてタグの表記揺れを解決したりタグにひもづけられた情報を提示したりするソーシャルブックマークサービス Faviki<sup>\*3</sup> や、自然言語での検索文を Freebase 用のクエリに変換することで自然言語検索を可能にした Powerset<sup>\*4</sup> のように、Wikipedia を構造化したデータを用い、Semantics を考慮したアプリケーションが提供され始めている。

セマンティック Web を実現するためには、マシンに推論などの処理を行わせるための大規模な構造化データがまずは必要である。すでにそれらのデータが、人々の共有知として集約され、提供されるようになってきた。さらに高度なセマンティック Web 技術を実現するためにも、また、人々に知の共有に対するさらなるモチベーションを与えるためにも、これらの構造化データを活用したアプリケーションをつくり出していくことが今後ますます重要視されるだろうと考える。折しも今回の ISWC から、Billion Triples Challenge という、DBpedia や Freebase、英単語オントロジー WordNet<sup>\*5</sup>

のデータセット (Billion Triples) を活用したアプリケーションを作成するコンテストが開始された。私自身もこのような流れに沿って、Triples を用いることで何が実現できるのかをじっくり考えていきたいというのが、新年の一つ目の抱負である。

### 2. 集合知からの潜在的知識の抽出

Wikipedia では、確かに共有知が形成されている。しかし Wikipedia 上に表れる知識は、あくまで人々が明確に認識している知識のみである。ある人が Wikipedia 上に記事を投稿するとき、または記事を修正するとき、その人の中で整理された知識が文章となって書き込まれる。だが一方、人間が頭の中で想起するものの中には、明確に整理されない、漠然とした意識が存在すると予想される。このような意識は、知識を整理し Wikipedia に投稿するといった行動には表れなくても、ほかの Web 上での行動には現れる場合もある。

例えば、delicious のようなソーシャルブックマークサービスでは、ユーザが気に入った Web ページをブックマークし、さらにページを分類するためにタグ付けを行う。これらの「ブックマークする」、「タグ付けをする」といった行動は、皆で知識を共有したいという願望からというよりも自身の利便性のための行動である。そこには「こういうページだからブックマークしたいと思った」、「こういう内容のページだから xxx というタグを付与する」といったユーザの意識が反映されている。同様に、Web ページの検索時にはユーザがクエリを入力し、検索結果として出力されたページリストのうちから目的の情報が含まれていそうなページへのリンクをクリックする。このような行動にも、「こういう情報がほしいから、xxx というようなクエリを入力する」といった意識が表れている。

特に delicious や、ほかの Web 2.0 サービスでは、ユーザの行動が反映された情報が大量に集まり、容易に入手可能となっている。このような、個々のユーザの意識が集まった集合知から、人間が明確に認知しているわけではない潜在的な知識を抽出し提示する手法を研究していきたいというのが、新年の二つ目の抱負である。

### ◇ 参考文献 ◇

[Auer 07] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. and Ives, Z.: DBpedia: A nucleus for a web of open data, 6th Int. Semantic Web Conf., Busan, Korea (Nov. 2007)

\*1 <http://dbpedia.org/>

\*2 <http://www.freebase.com/>

\*3 <http://www.faviki.com/>

\*4 <http://www.powerset.com/>

\*5 <http://wordnet.princeton.edu/>