

日本社会学会 RDF 知識統合システム (仮称) 創設の提案

田中 重人 (東北大学大学院文学研究科)

<http://www.sal.tohoku.ac.jp/~tsigeto/sociodb/>

これからの発展が見込まれる「セマンティック Web」技術に基づいて社会学関連の情報を収集・配信する「日本社会学会 RDF 知識統合システム」(仮称) を創設し、学会員の協力のもとに運営していくことを提案する。このシステムでは、RDF 記述によって情報を配信している既存の信頼できるサービス (たとえば国立国会図書館や国立情報学研究所) の文献情報を公開インターフェイス (API) によって収集する。また学会員、関連学会、各種機関から情報を収集する。これらの情報はすべて統合し、外部に配信する。これまで作成してきた「社会学文献情報データベース」(BJS) は、このシステムの基幹データとして統合する。学会員はこのシステムにアカウントを作成し、RDF 形式での情報の登録のほか、HTML その他の形式による文書を公開できる。将来的には、学会大会の報告申し込みや機関誌への投稿・査読のインターフェイスを兼ねることにより、学会において生産される知的資産を統合・管理・公開するシステムとして機能させることを目指す。

1 現状と問題点

「社会学文献情報データベース」(BJS) は、現在、つぎのような問題点をかかえている。

- 社会学文献調査への回答率が低い → 捕捉率が低い
- 入力されたデータのまちがいが多い
- 文献情報を入念にチェックする人手・時間の不足
- 外部機関との連携の遅れ

2 提案の目的

- 会員の入力に頼る割合をへらす／外部機関からとれる情報を利用する
- 文献は、そもそも外部機関に登録されるよう、図書館などに寄贈してもらう／目次などについて適切なコーディングのウェブページを公開してもらう
- 単なる書誌情報の入力ではなく、社会学者にしかできない情報の発信を考える
- 他の機関にはない書誌データを公開し、永続的な URI を提供する
- 単独のデータベースとしての検索だけでなく、さまざまなインターフェースで情報を提供する
- 研究者コミュニティの再創造

3 文献メタデータとセマンティック Web

BJS のデータ = 固定フィールドのテキストファイル

→ すべてのデータを自前で管理する必要がある。

RDF による記述 = 「ノード S」と「ノード O」の関係を「プロパティ V」であらわす → S V O .

```
# (本当は dc, dcterms, prism についても接頭辞宣言 (@prefix ...) が必要だが、省略)
@prefix bjs: <http://sociodb.jp/search/details.php?ID=> .
bjs:110101727 dc:creator <http://ci.nii.ac.jp/nrid/9000002337108> .
bjs:110101727 dcterms:isPartOf <urn:isbn:9784861631467> .
bjs:110101727 prism:pageRange "99-118" .
```

これらの記述はばらばらに書かれていてもかまわない。この形式のデータをほかのシステムからとってこれれば、それを自前のデータをあわせて解釈できる。複数の記述をまとめて解釈したとき、ノードをプロパティでつないだ有向グラフとして、データの全体像があらわれる (= semantic Web)。

4 日本国内の状況

公開 API による外部への RDF 提供と永続的 URI の提供。

国立情報学研究所: CiNii (Articles/Books): 2011-11 リニューアル <http://ci.nii.ac.jp/info/ja/api/api_outline.html>

国立国会図書館: NDL Search / Web NDL Authorities: 2012-02 正式公開 <<http://iss.ndl.go.jp>>

標準的なインターフェイスによって、正確な書誌情報が取得できるようになった。ISBN や著者 ID などの識別子がわかれば、対応する文献を自動的に取得して、情報をインポートできる。

→ 書誌情報を入力してもらうより、識別子を教えてもらうことが重要

5 日本社会学会 RDF 知識統合システム (仮称)

5.1 やりたいこと

- 外部機関からの情報自動収集システム
- 特定のクエリによる RSS フィード
- 学会員からの情報収集と公開
- 「社会学文献情報データベース」の既存データの活用
- 外部機関への情報配信用公開 API
- 文献情報以外の社会学的知識 (たとえば概念、命題、調査、データ、社会的事実) についてのリソースの公開
- 社会学的知識を記述するためのプロパティ語彙の整備
- 学会出版物の作成プロセスの統合

5.2 データ収集システムと公開 API

RDF を用いて情報を配信する仕組みとしては、ブログやニュースの記事配信で広く使われている RSS (RDF Site Summary)、学術レポジトリでメタデータを配信する仕組みとして普及している OAI-PMH (Open Archive Initiative Protocol for Metadata Harvesting) がある。また RDF で記述されたデータを対象に検索する仕組みとしては SPARQL (SPARQL Protocol and RDF Query Language) があり、国立国会図書館などでつかわれている。取得したい情報の URL がわかっているならば、こうした仕組みを使わなくても直接 HTTP でアクセスすればよい。

「日本社会学会 RDF 知識統合システム」においても、入力・出力の両方について、これらのインターフェイスに対応する。いずれも特にめずらしい技術ではなく、現在基盤 (B) 科研で構築をすすめている XOOPS で対応できるはずである。

5.3 社会学文献情報データベースの活用

「社会学文献情報データベース」(BJS) に収録されている 10 万件あまりのデータは、新システムであつかえる形式 (RDF) に変換して統合する。BJS で付与されている ID は、そのまま利用する。

※ RDF の記述を集積してつくられる知識構造 (プロパティ付き有向グラフ) をどういうデータ形式で格納するのかは、要調査

現在すすめている、BJS データ再構築の方向はこの RDF 化に向けたひとつのステップと位置付けられる。ただし、現行の文献情報のかたちは、新システムには完全には移せない。これは、広くつかわれているプロパティ語彙 (Dublin Core とそこから派生したもの) の規定が図書館で使われている目録規則よりゆるいためである。図書館規則よりのことを表現するためにいろいろ工夫することは可能だが、かなりの労力を要する上に、かえって流通性を下げおそれがあるため、ある程度のところでわりきるしかない。

5.4 学会員の参加と情報発信

日本社会学会員は、新システムにアカウントをつくり、RDF をデータベースに追加することができる。仕組みとしては、通常のプログサービスと同様のもので、ただしテキストや HTML だけでなく、RDF でかかれたファイルを投稿できる仕組みにしておけばよい。

RDF を記述するには複数の方式があるが、いずれも複雑であるため、入力補助のインターフェイスが必要である。具体的にとりうる手段としては、つぎのようなものがありうる：

- 対象となるふたつの URI とそれらを結ぶプロパティの 3 つの入力窓を用意する
- URI は検索結果からコピーする
- プロパティは、代表的なもののメニューを用意する
- 作成した RDF を日本語に翻訳して表示する
- 日本語の記述を RDF に変換する
- 現行の文献調査と同様の入力インターフェイスも準備する (既存データからわかるものについては、初期値をいれておく)

もちろん、自分で手書きするほか、アプリケーションで生成したものをコピーすることもできる。また、EndNote / CSV / 社会学評論スタイル文献記述などのデータからも RDF を作成できるようにする。

文法のチェックの仕組みをつくる。既存データとの整合性についても、投稿時にチェックする。(矛盾する記述については、書き換え希望を出せる)

内容についても、データベースに統合する前にチェックする仕組みが必要。特に、同一ユーザによる大規模なデータの書き換えなどは、一応 (データベース委員が?) チェックする。また、どの記述を誰が追加したかは記録しておく。

参加者は、各自のブログ記事を公開したり、ファイルをアップロードしたりできる。これらについてのメタデータは、データベースに自動的に追加される。この仕組みは、学会員が日記、書評、ワーキングペーパー、調査報告書などを公表するために使える。

また、一定のテンプレートを用意しておくことにより、学会報告要旨や機関誌への投稿に利用することが可能である。これらを支援する仕組みとして、文献情報を「社会学評論スタイルガイド」にしたがって整形する、リンクや文献参照やセクション構造について自動的に RDF を生成するといったサービスを提供する。

5.5 社会学的知識の集積

RDFは、対応するURIがあって必要なプロパティ語彙が定義されていれば、なんでも記述できる。文献情報にかぎらず、社会学の研究で使われるあらゆるものが言及対象になりうる。

セマンティックWebに社会学的知識を集積していくには、そのためのリソースを持続的なURIで公開しておく必要がある。

BJSにふくまれる内容分類やキーワードなどは社会学的知識の記述のためのベースとして活用できるだろう。また、各種の事典、教科書、会員の授業資料、調査の質問項目なども、重要なリソースになりうる。こうした知識を集積していくことを通じて、新システムが社会学研究者のコミュニティの再創造の核となることが期待できる。たとえば、ワーキングペーパーを公開した場合、そのペーパー内で参照している文献についての情報、主要な概念、使用したデータ、結論として主張したい命題などをシステムに登録できる(それらのデータから逆にたどることができる)。通常のブログやSNSではこうしたことをおこなうのはむずかしい。この点は、学会によるコミュニティをつくることの重要な利点といえる。

6 参考文献

- 神崎正英, 2005. 『RDF/OWL 入門: セマンティックウェブのための』 森北出版. <ISBN:9784627829312>

