

実用材料開発のためのデータベース「PoLyInfo RDF」の構築： セマンティックウェブ技術による境界無きデータ駆動

Construction of “PoLyInfo RDF” database for practical material development:
Crossover data-driven by using semantic web technology

物材機構¹ ○石井真史¹, 竹村太郎¹, 谷藤幹子¹

NIMS¹, ○Masashi Ishii¹, Taro Takemura¹, Mikiko Tanifuji¹

E-mail: ISHII.Masashi@nims.go.jp

【序】実際の材料開発においては、純粋な物理・化学的な興味以外の要素も考慮する必要がある。例えば、採算性、安全性、環境問題などが学術的興味以上に重要になり、またそれが逆に新しい材料開発における研究のモチベーションとなる。セマンティックウェブ技術は、こうした境界無き情報の流通を促す国際的取り組みであり、W3C (World Wide Web Consortium) [1]の指針のもとで、次第に利用者が増えている。

この技術は、あらゆる情報を主語・述語・目的語の三種 (Triple という) で記述する RDF (Resource Description Framework) と呼ばれるプロトコルにより実現される。現在、世界の巨大データベース (DB) の多くが RDF を取り入れており、化学分野では日化辞 (日本化学物質辞書) RDF や PubChem RDF などよく知られている。本研究では、NIMS の高分子 DB、PoLyInfo の重合情報を RDF 化し、原料モノマーを介して、既に RDF 化している巨大 DB との統合を図った。

【実験方法】今回設計製作した PoLyInfo RDF の重合情報の triple グラフ (体系図) を図 1 に示す。ポリエチレンを例にしてあるが、逐一書き下すと①「ID 番号 CU010001 の構成単位をもつポリマーはポリエチレンの名を持つ」、②「この構成単位は、ID 番号 J0000002 の重合パスを持つ」、③「この重合パスは “エテン_&_1,3-ブタジエン_付加重合/高分子反応” の名を持つ」、④「その原料モノマーは ID 番号 M0101001 と M0301021 である」、⑤「これらはそれぞれ、エテン、1,3-ブタジエンの名を持つ」となり、①～⑤の全てが主語・述語・目的語の Triple により記述され、連鎖的につながる。重要なことは図 1 により PoLyInfo のポリマー情報が最終的にモノマーに帰結する事であり、そこを窓口にしてモノマー主体の日化辞と結合できる点である。このデータベース間の

結合も、W3C の規格である skos (Simple Knowledge Organization System) [2]に基づいて、triple で表される。すなわち⑤の後に⑥「PoLyInfo に収蔵されているこれらのモノマーは、日化辞の ID 番号 J1.939I と J4.043F のモノマーと等価と考える」と概念的に定義することで DB 間の統合が出来上がる。

【結果】図 1 に基づいて作成した Triple は、①が 34,778 triples、②が 31,768 triples、③が 89,847 triples、④と⑤が合わせて 54,280 triples、⑥が 16,809 triples となり、合計で 227,482 の関係をセマンティックに書き下すことに成功した。ここで、⑥の数が現在日化辞 RDF と共有しているモノマー数となる。

これを triple ストアにアップロードし、RDF のクエリ言語である SPARQL[3]で、日化辞にあるモノマーからできる PoLyInfo のポリマー (英名) をリスト化するタスクを課したところ、直ちに 39,907 のポリマーが得られた。

【まとめ】PoLyInfo の重合情報を RDF triple で記述することで、日化辞との統合に成功した。今後 PoLyInfo 全体の RDF 化を目指すと共に、社内情報などの非構造データ及びデータベースとの統合へ拡張し、その有効性を検証する予定である。

【謝辞】本研究は、内閣府「戦略的イノベーション創造プログラム (SIP)」の「スマートバイオ産業・農業基盤技術 革新的バイオ素材・高機能品等の開発を加速するインフォマティクス基盤技術の開発」の支援を戴いた。

【文献】

[1] <https://www.w3.org/>

[2] <https://www.w3.org/TR/skos-primer/>

[3] <https://www.w3.org/TR/rdf-sparql-query/>

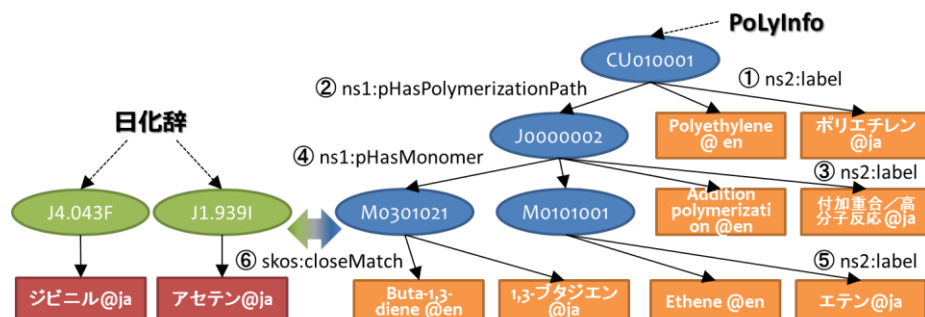


図 1 設計した PoLyInfo の重合情報のトリプル図と skos による日化辞との統合