

RSS1.0 と RDF を用いた STP セマンティック Web の構築とその応用

久保 卓也 [1]; 木村 映善 [2]; 村田 健史 [3]; 篠原 育 [4]; 笠羽 康正 [5]; 巨 慎一 [6]; 松岡 大祐 [7]; 島津 浩哲 [8]; # 山本 和憲 [1]; 長谷川 洋 [9]; 小原 隆博 [10]

[1] 愛媛大・理工; [2] 愛媛大 CITE; [3] 愛大・メディアセンター; [4] 宇宙機構 / 宇宙研; [5] 東北大・理; [6] 情通機構; [7] 愛媛大; [8] 情通研; [9] 宇宙研; [10] 情報通信研究機構

An STP Semantic Web Technique and Service using the RSS1.0 and RDF

Takuya Kubo[1]; Eizen Kimura[2]; Takeshi Murata[3]; Iku Shinohara[4]; Yasumasa Kasaba[5]; Shinichi Watari[6]; Daisuke Matsuoka[7]; Hironori Shimazu[8]; # Kazunori Yamamoto[1]; Hiroshi Hasegawa[9]; Takahiro Obara[10]

[1] Ehime Univ.; [2] CITE, Ehime Univ.; [3] CITE, Ehime University; [4] JAXA/ISAS; [5] Tohoku Univ.; [6] NICT; [7] Ehime Univ.; [8] NICT; [9] ISAS/JAXA; [10] NICT

In the Solar-Terrestrial Physics, it is pointed out that both circulation and utilization of observation data among researchers are insufficient. Under such a background, we have been developing a world-wide database that manages metadata of satellite/ground-based observation data files. Note that retrieving metadata from the observation data and registering them to database have been carried out by hand so far.

Our goal is to establish the STP Semantic Web. The Semantic Web provides the secondary information related with observations. The most fundamental issue on the establishment is who generates, manages and provides metadata in the Semantic Web. We developed an automatic metadata collection system for the observation data using the RSS (RDF Site Summary) 1.0.

Using the RSS1.0 as a metadata collection method, the workflow from retrieving metadata to registering them into the database is automated. This technique is applied for several database systems, such as DARTS database system, CDAWeb (Coordinated Data Analysis Web) database system and NICT Space Weather Report Service. The DARTS is a science database managed by ISAS/JAXA and the CDAWeb is managed by NASA/GSFC. We succeeded in generating and collecting the meta-data automatically for the CDF (Common data Format) data provided by the DARTS and CDAWeb. We also succeeded in creating an RDF for both space weather report and real-time global MHD simulation 3D data provided by the NICT.

Once the RDF documents are registered, then a metadata collection agent extracts meta-data to store them in the RDF database. The RDF database provides advanced retrieval processing that are inference processing and near the natural language query processing. Finally, the STP Semantic Web provides automatic processing or high level search for the data which are not only for observation data but for space weather news, physical events, technical terms and researches information related to the STP. As a result, new analytical results and/or events can be mechanically executed, which cannot be discovered up to now since analytical processing from various angles is hard to be conducted simply by person's hands. In the present poster presentation, the outline chart of current and future system will be discussed. We are going to introduce examples of the processing using the RDF database.

本研究で対象とする STP (Solar-Terrestrial Physics) 分野においては、観測データの取得と蓄積は進んでいるものの、それらのデータを第3者に提供するための体制においては不十分な点が多いと指摘されている。この問題を解決するため、我々は STP 観測データファイルの所在や管理権限といったメタデータを取り扱うメタデータベースを開発してきた。しかし、これまではデータファイルからメタデータを抽出し、それらをメタデータベースに登録する作業が全て人の手によって行なわれていたため、人的コストが非常にかかり人為的ミスも起こりうるという問題があった。

我々の目標は STP 分野におけるセマンティック Web を構築することである。この STP セマンティック Web では、観測データだけでなく、それに関する様々な2次情報を提供する。しかしながらセマンティック Web は「誰がメタデータを作成し、誰がメタデータを提供するのか?」という根本的な問題を抱えている。この問題に対し、我々は RSS (RDF Site Summary) 1.0 を用いて観測データのメタデータを自動収集するシステムを開発した。この RSS1.0 は RDF (Resource Description Framework) に準拠しており、RDF はセマンティック Web におけるメタデータの標準記述フレームワークである。

RSS を用いたメタデータ自動収集システムは、ISAS/JAXA が運営している DARTS (Data Archives and Transmission System) のデータベースシステムと NASA/GSFC が運営している CDAWeb のデータベースシステム、情報通信研究機構 (NICT) が提供している宇宙天気情報サービスなどで現在適用されている。我々のシステムを用いることにより、DARTS と CDAWeb が提供している CDF (Common Data Format) ファイルから自動的にメタデータを抽出し、それらをメタデータベースに登録することに成功した。また、NICT が提供している宇宙天気情報サービスと、real-time global MHD simulation 3D data の GFA (Geometry Flipbook Animation) ファイル、HDF5 (Hierarchical Data Format 5) ファイルからも同様にメタデータの抽出と登録を自動化できた。

メタデータ自動収集システムによって生成される RDF 文書は、RDF データベースにも自動的に格納される。RDF データベースは、RDF 内で用いられる語彙の意味を詳細に記述した RDF Schema/OWL (Web Ontology Language) を活かすことで、自然言語に近い形のクエリの発行や推論処理を実現することができる。将来的には、観測データや宇宙天気情報だけでなく、イベント情報や専門用語情報、研究者情報などに対し、自然言語処理に近いクエリの発行や推論処理といった高度な情報処理を提供することを目標としている。これにより、様々な角度からの解析処理を機械的に実行できるため、これまで人の手によっては発見できなかった新たな解析結果・イベントが得られると考えている。本ポスターでは、これまでのシステムと将来的なシステムの概要図を示し、RDF データベースを利用した処理の例について紹介する。