

Title:

SKOS and semantic web best practice to access terminological resources: NatureSDIPlus and CHRONIOUS hands-on experience.

AIM

Knowledge Organization Systems (KOS) are pivotal supporting the keyword search in different application contexts. As a consequence we have recently witnessed a growing interest on how to encode and to make available KOS. This growing interest has resulted in specific activities such as the W3C's SKOS recommendation made to encode Simple Knowledge Organization Systems in standardize way, and a particular attention to the semantic web technological progress as means to provide KOS networked in flexible and highly reusable manners.

Based on these considerations, we have deployed semantic web enabled Networked Knowledge Organization System in two European Projects, NatureSDIPlus (ECP-2007-GEO-317007) and CHRONIOUS (FP7-ICT-2007-1-21646).

Common motivation to use KOS in both the projects was facilitating the search of resources (i.e., scientific biomedical literature and geographical information) relying on multilingual terms from controlled, domain specific vocabularies. In NatureSDIPlus, KOS are used as means for facilitating the access to distinct pre-existing multilingual resources related to Nature Conservation. In CHRONIOUS, the concerned issue is to expand two domain ontologies pertaining to Chronic Obstructive Pulmonary Disease (COPD) and Chronic Kidney Disease (CKD) with the aim of indexing certified biomedical documentation.

In this presentation we present our achievements and comment the lessons learnt deploying and making available KOS. We aim at establishing hands-on recipes to publish and to access KOS discussing under a critical perspective the approaches followed in the projects.

In particular the presentation describes the aims and the constraints posed by the two projects motivating the rationale behind the adoption on semantic web technology. It presents strategies and technological solutions selected for each project giving a detailed insight about benefits, lacks and open issues emerged from the project solutions.

METHODS

The experience made in the aforementioned projects is illustrated. It relies on an abstract pipeline that describes the phases to be addressed providing and making accessible KOS. In particular, the following macro-phases are considered:

Resource selection. It is the phase that identifies and selects a list of the relevant resources, i.e., classifications, terminologies and thesauri. Criteria adopted in the selection such as relevance of resource content, issues concerning the copyright constraints and their availability as structured digital form will be introduced.

Resource translation in SKOS (SKOSification) is the phase where the data model of the selected resources is mapped in the SKOS model. The deployed strategies will be described considering that further RDF vocabularies might be also considered to complement the SKOS model whenever information relevant for the final application does not completely fit the SKOS Model;

KOS Publication\Access is fundamental to make the SKOSified KOS available to third parties. The strategies exploited in the two mentioned projects will be presented. In particular we will detail the Linked Data Best Practices\technologies adopted in NatureSDIPlus and the JAVA API developed to grant the access in CHONIOUS.

Knowledge organization systems interlinking: Some specific mappings among KOS, and between KOS and Ontologies, have been defined in the projects relying on a-priori knowledge (e.g., Knowledge about how KOS have been originated) and/or developing appropriate semi-automatic procedure. The interlinking procedure and their results will be shown in the presentation.

KOS advertisement: Once KOS have been interlinked and stable versions have been reached, specific strategies to advertise them in related communities should be foreseen. In this direction there are different actions that can be undertaken to advertise the KOS availability. The actions as registering KOS in existing websites, producing VOID descriptions, submitting the KOS to semantic web search engines as SINDICE will be briefly introduced as strategies that we are going to be deployed.

MAIN FINDING

The above-described pipeline has been instantiated according to the specific constraints and objectives in each project.

In particular, in NatureSDIPlus, distinct existing multilingual resources pertaining to the Nature conservation are provided in a Common Framework according the linked data best practice. The result of SKOSification is deployed as a D2R server at <http://linkeddata.ge.imati.cnr.it:2020/>.

In CHRONIOUS, terminological resources pertaining the Biomedical domain as MESH (<http://www.ncbi.nlm.nih.gov/mesh>), its correspondent translations in Italian, Portuguese, and Spanish are SKOSified and linked to COPD and CKD ontologies. CHRONIOUS represents a typical example where the SKOSification and RDF allow to manage terminologies derived by resources produces by distinct entities. Terminologies are made available as distinct SKOS\RDF that can be plugged in the system. The access to KOS from the CHRONIOS system is ensured embedding these SKOS\RDF through specialized API developed in JENA. Eventually, the API is exploited in the indexing and the search for biomedical scientific literature.

From the technological point of view, an effort has been made to focus on existing open source technologies such as JENA, SPARQL, D2R, MYSQL, attempting to reduce the development to SQL store procedures and simple JAVA programming. That should make quite reproducible the recipes we have deployed: a large number of programmers and practitioners is able to master JAVA and SQL. Moreover, the adoption of open source technology ensures good possibility concerning the customization and it is usually exploitable with very limited costs.