AGROVOC Web Services: Improved, Real-Time Access to an Agricultural Thesaurus

Boris Lauser, Margherita Sini, Gauri Salokhe, Johannes Keizer, and Stephen Katz

Abstract: Controlled vocabularies, such as the multilingual agricultural thesaurus AGROVOC, are the facilitators of semantic services on the Internet, since they can provide the baseline for connecting distributed resources. The more resources in a particular domain that are described using a certain controlled vocabulary, the more accessible and interoperable they will be. Providers of controlled vocabularies traditionally distribute their complete databases upon request to system developers in order to incorporate them into information systems. This process is cumbersome, and creates duplication of work, since common vocabulary and terminology services (like a query for a related term) have to be redone each time at the local application level. With web services it is now possible to expose such common vocabulary services openly on the Internet, thus avoiding laborious local duplication and ultimately encouraging a wider audience to use these services. The article presents our current work on providing AGROVOC thesaurus web services and our vision for the future, which is to transform this initial service offering into more powerful ontology web services.

Resumé: Les vocabulaires contrôlés, tels que le thésaurus agricole multilingue d'AGROVOC, sont les facilitateurs de services sémantiques sur Internet, puisqu'ils peuvent fournir la base pour connecter des ressources distribuées. Plus il y a de ressources d'un domaine particulier qui sont décrits utilisant un certain vocabulaire contrôlé, plus ils seront accessibles et inter-opérables. Les fournisseurs de vocabulaires contrôlés distribuent traditionnellement leurs bases de données complètes sur demande aux développeurs de système afin de les incorporer dans des systèmes informatiques. Ce procédé est encombrant et redondant, puisque le vocabulaire commun et les services de terminologie (comme une recherche sur un terme apparenté) doivent être refaits chaque fois au niveau de l'application locale. Avec les services de la toile, il est maintenant possible d'exposer tels services de vocabulaire commun ouvertement sur Internet, évitant ainsi la duplication locale laborieuse et encourageant finalement un auditoire plus large à utiliser ces services. L'article présente notre travail actuel pour fournir les services du thésaurus AGROVOC sur la toile, ainsi que notre vision pour l'avenir, qui est de transformer cette offre initiale de service en services en ontologie plus puissants sur la toile.

Resumen: Los vocabularios controlados, como el tesauro agrícola multilingüe AGROVOC, son los facilitadores de servicios semánticos en la Internet, puesto que pueden servir de punto de comparación para conectar los recursos que están disponibles. Entre más recursos haya en un dominio particular que se describen utilizando cierto vocabulario controlado, más accesibles e inter-operables serán. Los proveedores de vocabularios controlados tradicionalmente distribuyen sus bases de datos completas a solicitud a los desarrolladores de sistemas para que éstas sean incorporadas en los sistemas de información. Este proceso es engorroso y genera una duplicación de esfuerzos, ya que los servicios de vocabulario y terminología común (por ejemplo una consulta respecto a un término relacionado) tienen que rehacerse cada vez al nivel de la aplicación local. Con los servicios en la Web, ahora es posible difundir abiertamente dichos servicios de vocabulario común en la Internet, evitando, por tanto, la laboriosa duplicación de esfuerzos al nivel local y, en última instancia, alentando a que una mayor audiencia utilice estos servicios. El artículo presenta el trabajo que se hace actualmente para prestar servicios relacionados con el tesauro de AGROVOC en la Web y la visión que se tiene hacia el futuro, la cual es transformar esta oferta inicial de servicios en eficaces servicios de ontología en la Web.

Introduction

CONTROLLED VOCABULARIES are the facilitators of semantic services on the Internet, since they can provide the baseline for connecting distributed resources. The more resources in a particular domain are described using a certain controlled vocabulary, the more accessible and interoperable they will be. Providers of controlled vocabularies traditionally distribute their complete databases upon request to system developers in order to incorporate them into information systems. This process is cumbersome, and creates duplication of work, since common vocabulary and terminology services (like a query for a related term) have to be redone each time at the local application level. Experience has shown that interoperability between information systems is hardly ever achieved because the necessary mapping between different vocabularies used in the respective systems is too laborious to be done at the local implementation level.

Web services[1] are now available to put complex services, which are offered by providers of controlled vocabularies, online. Such services can easily be incorporated into any application via simple function calls to a Uniform Resource Locator (URL) endpoint and as such, the reimplementation of common functionality in local applications becomes obsolete. Web services provide real time access; hence, outdated databases and system incompatibilities will no longer be an issue if referring to data available via web services rather than to local copies of databases. They provide opportunities

for reducing integration time and expense compared with existing approaches of integration.

AGROVOC and Web Services

The AGROVOC Thesaurus was developed by the Food and Agriculture Organization of the United Nations (FAO) and the Commission of the European Communities, in the early 1980s. AGROVOC is a multilingual, structured and controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment). AGROVOC is currently available online in English, French, Spanish, Arabic, Chinese, Portuguese, Czech, Japanese, Thai and Slovak. Other language versions have been created and maintained by regional collaborating partners and might be considered for future integration and maintenance (Hungarian, Lao, Italian, Hindi and German). For additional information on the AGRO-VOC Thesaurus, see the AGROVOC information sheet at:

> ftp://ext-ftp.fao.org/GI/Agris/
> aims/references/flyers/
> agrovoc_en.pdf

AGROVOC forms an integral part of the Agricultural Information Management Standards (AIMS) initiative

> www.fao.org/aims/

The main objectives of the AIMS initiative are to promote standards in agricultural information management and to provide a one-stop access for system designers, developers and information management specialists. AGROVOC web services have recently gone online as part of the initiative to encourage developers of agricultural information management systems to incorporate AGROVOC into their applications via web services, instead of using local copies of the database. With web services, updates to the thesaurus are immediately available, reducing the time and effort

necessary to regularly download and incorporate the latest version of the thesaurus into applications. Currently, six basic methods to access the vocabulary are available online

> www.fao.org/aims/
> ag_webservices.jsp

including methods to obtain: the numeric code of a concept; labels of any term in all the available languages; broader, narrower and related concepts; definitions and searching for a term. Our work relies on existing standard open source technologies like SOAP and WSDL.

Benefits of Implementing AGROVOC as a Web Service

AGROVOC is updated by FAO on a weekly basis, and the users are informed of the changes via mailing list and the Web site; however, the latest database version for download is available only on a quarterly basis. Traditionally, to ensure consistency and coherence with latest versions of the AGROVOC Thesaurus, all the system developers/ owners that use AGROVOC download the latest version on a regular basis and perform updates of their information systems. The new option of connecting to the AGROVOC Thesaurus via web services makes this work-intensive updating procedure redundant, thus significantly reducing the update and integration costs.

Another big advantage of using web services over the traditional database concerns the timeliness of accessible data. For instance, for an information system that uses AGROVOC for indexing and searching, new concepts might be needed in order to index bibliographical resources in fast changing areas like diseases (avian flu) and so forth. The use of web services will allow immediate access to such new concepts in the thesaurus by sending a simple query requesting back the latest updates. In addition, if new language versions of the AGRO-

VOC Thesaurus are made available, web services functions will automatically and instantly retrieve these new translations.

In summary, designing an information system to use web services will provide it with real-time access to the thesaurus and thus ensure compatibility with the latest version without any additional cost of performing local upgrades.

Future Directions

As part of the AIMS initiative, the Agricultural Ontology Service (AOS)

> www.fao.org/aims/aos.jsp

tries to elevate such terminology services in agriculture to a level allowing for much higher semantic expressivity. One of the main objectives is to provide a framework for distributed maintenance and sharing of common descriptions and definitions of concepts and their relations within the agricultural community, thus increasing the efficiency and consistency with which multilingual agricultural resources are described and associated. Within this framework, AGROVOC will evolve into the AOS Concept Server (AOS/ CS). The Concept Server will be set up in a distributed environment to function as a tool to help structure and standardize agricultural terminology and concepts in multiple languages for use by any number of different systems around the world.

This effort will open a new dimension for services that can be offered from the provider side of ontologies and terminologies. Possible future services range from offering an expanded set of semantic relationships to mappings between ontologies.

In order to be able to offer an extended range of web services, other agricultural ontologies and terminologies need to be accessible via standard web services as well. By being one of the pioneers[2] offering web services for agricultural ontology systems, and by promoting our

THE AUTHORS are all affiliated with the Food and Agriculture Organization of the United Nations, Rome, Italy.

vision of the AOS Concept Server, we hope to convince other providers of agricultural terminology and ontologies to offer standard web services that can be included in the framework of the AOS. By further promoting and getting system developers to implement our standards and services in close collaboration with us[3], we strive to achieve more interoperability between and coherence in agricultural information systems worldwide.

NOTES

1. For more information on web services, go to: http://www.w3.org/2002/ws/

2. The National Agricultural Library of the USA also has recently published standard SOAP web services: http://agclass.nal.usda.gov/agt/agt.shtml

3. To join the AGROVOC Discussion Group, go to: http://www.dgroups.org/groups/ fao/agrovoc/ index.cfm?op=dsp_join

For more on this topic, go to FAO's Agricultural Information and Knowledge Management Papers in the FAO Corporate Document Repository:

http://www.fao.org/documents /advanced_s_result.asp? FORM_C=AND&SERIES=339

CONTACT ADDRESS

Boris Lauser
Information Management Consultant
Library and Documentation Systems
    Division
Food and Agriculture Organization of the
    United Nations
Viale Delle Terme di Caracalla
00100 Rome
ITALY
PH: +39 06 570 54638
Fax: +39 06 570 54049
E-mail: Boris.Lauser@fao.org
Web: http://www.fao.org