SWAD-Europe Deliverable 3.14: Developer Workshop Report 7 - Metadata in a multilingual world

Project name:
> Semantic Web Advanced Development for Europe (SWAD-Europe)

Project Number:
> IST-2001-34732

Workpackage name:
> 3 Dissemination and Implementation

Workpackage description:
> http://www.w3.org/2001/sw/Europe/plan/workpackages/live/esw-wp-3

Deliverable title:
> 3.14 Developer Workshop Report 7

URI:
> http://www.w3.org/2001/sw/Europe/reports/dev_workshop_report_7

Author:
> Charles McCathieNevile

Abstract:
> This report summarises the seventh SWAD-E developer workshop, held in Copenhagen on 15-16 July 2004. The workshop explored the topic of Metadata in a multilingual world.

STATUS:
> **Completed** - 3 September 2004. The first draft published 16 July 2004. This version $Id: Overview.html,v 1.7 2004/08/30 11:16:59 danbri Exp $. This report may be updated over the life of the SWAD-Europe Project to link to new work emerging on the topics of the workshop.

Contents

Executive Summary

This workshop brought together developers and users working on the multilingual application of metadata and the semantic web. It discussed the problems and issues invovled in making the Web truly world wide, and the impact of this and the development of a more semantically rich web on each other.

Some tools were presented, some areas of success, and many areas requiring significant further work were identified. Work on glossary tools was directly

advanced in preparation for and as a result of this workshop.

The workshop was jointly organised with CEN-ISSS MMI-DC ensuring rapid flow of information to other relevant European organisations.

1 Introduction

This report is part of the ⊣ SWAD-Europe project ⊣ Work package 3: Dissemination and Implementation.

It describes a developer workshop held in Copenhagen in July 2004, on the topic of metadata in a multillingual world. The workshop was attended by developers based in Europe, with additional participation from the USA and Australia. This workshop was held jointly with the CEN-ISSS MMI-DC workshop group, who investigateand make recommendations on the use of metadata in Europe.

2 Background

A short list of ⊣ background reading for workshop participants is available. Two position papers were provided, one from Thomas Baker who was unable to attend the meeting himself.

Related work -

The Dublin Core Metadata Activity maintains a number of international mirrors of its content, with a variable amount available in translation, including schemas, documents describing usage, etc.

The W3C maintains a collection of documents, and a glossary derived from those documents. A large number of volunteer translators provide translations of various fo these documents, and in some cases terms from the glossary, in order to help ensure consistency of translations.

A number of countries or organisations are working in multiple languages - Fundación Sidar is one, working primarily in Castellano, Catalá, Gallego and Português with collections of documents and tools. Some of Sidar's tools, such as Hera, are using multilingual RDF vocabularies as a base for interfaces and document output.

There are an increasing number of applications of multilingual approaches and technologies to providing accessibiltiy for people with disabilities, including the use of simplified language, provision of visual or other multimedia aids to comprehension.

Finally, many governments and similar organisations (such as the European Commission) are required to work in a number of languages at once, and need to ensure that they are providing and managing information appropriately for this need.

3 Workshop

The workshop was attended by developers from

- Australia
- Denmark
- France
- Spain

- Sweden
- United Kingdom
- United States of America

It was broadcast via the ⊣ #rdfig IRC channel, allowing remote participation, and at relevant points various developers took part in the workshop through this method.

4. Outcomes

Technical discussion -

A number of ⊣ tools and vocabularies were presented or discussed. The ⊣ first day's discussion log, the ⊣ first day's "chumped" highights, the ⊣ second day log and ⊣ second day's highights are all available.

It was clear that in many areas there is a lot of development, and tools are moving towards the level of products developed commercially for end users, while other areas still involved research and development. It is also clear that this is a very large area for exploation, and that most systems are only currently working on a fairly basic level.

In particular, tools dealing with time or location in any complexity tend to be in the early phases of development.

It was clear that the complexity of this area is due in part to the fact that language cannot be readily seperated from its cultural context in many important use cases, and that representing this information in a machine readable way is therefore a very complex problem.

It seems that simple dictionary tools are useful for many cases, and these are relatively advanced in development.

Use cases -

A substantial amount of discussion was devoted to looking at use cases - the things that participants actually want the semantic web to do. These were discussed briefly in the logs as well as being collected in the highlights.

Multilingual glossaries
    This topic was covered in several different cases. Simply providing ⊣ a multilingual glossary for W3C, providing systems for ⊣ relating concepts to different symbolic representations, tools for ⊣ finding meanings by comparing contexts, and similar topics were discussed throughout the workshop.
Extending glossaries
    Various more powerful systems that could incorporate cultural contexts to glossaries are important. From the simple case of ⊣ reading menus, to ways of searching for ⊣ similar information across multilingual knowledge bases.
⊣ Names and addresses
    Internationally, ways of naming people vary widely. This is in addition to the various different forms of address that might be used in different contexts. Therefore, systems for ⊣ matching people's names, and providing intuitive searching (where one person will not have a consistent part of the name as the most obvious metanymic trigger for the different people that they know) need to provide powerful, context-dependent methods for encoding names. Similarly, although perhaps less complex, people's addresses contain different information according to where they live.
Matching to users' locations

Localising software is a common use case. Additionally, being able to ⊣ localise the way a web page is served, according to what the user expects, may be desirable in many instances. Providing ⊣ multilingual systems to create metadata is also helpful. For example, many people may be describing an image using different natural languages.

Work still needed -

A clear outcome of the workshop was the need for simple step-by-step explanations of how to use vocabularies, oriented to developers who want to copy working examples rather than understand the entire theoretical base and then deriving their own tools and code.

Systems that can handle any type of postal address should, in theory, be readily available, but they are not yet in widespread use. Addressing areas such as people's names, or locations of things is still largely in the area of research, although some simple use cases are being met by existing tools.

The outcomes of this workshop, and in particular the lessons learned in the discussions, will be used to inform discussions at the FOAF workshop to take place in Galway, on the topic of how to capture the names of people (and places) in a way that makes sense in the context of a multilingual world wide semantic web.

Tool development -

W3C's glossary system was updated to use the SKOS vocabulary, developed for the SWAD-E project, in preparation for this workshop.

As a planned follow-up to the workshop Sidar's glossary is expected to migrate to SKOS, as a preliminary step to developing intereoperability with the W3C glossary.

General logistics -

Attempting to run a workshop or any similar event in Europe during the two peak months of summer is difficult, and reduces somewhat the opportunity for the broadest possible participation. This effect is speard over at least two months of the year since people take vacaations at different times. As this workshop showed, it is possible to make important and valuable progress during this extended slowdown, but the limited availability of people means that it is unlikely to be possible at the same rate as at other times of year.

Appendix A: Tools presented

⊣ Babylon
>	A simple dictionary tool that allows a user to look up a term in multiple dictionaries.

⊣ CCF demos
>	Tools for producing and manipulating messages in a number of symbolic languages (pictorial representations of words, typically used by people who can not read, write nor speak) and translate them. Developed as part of the WWAAC project and further developed by the Concept Coding Framework group.

⊣ Dublin Core Spanish Mirror
>	A mirror of the Dublin Core project, providing reference documents in Spanish

⊣ Hera
>	A tool for performing evaluations of Web Accessibility. Its interface and

output can be provided in multiple languages, and it can generate a report in RDF with human-readable content included.

⊣ Sidar Glossary

Sidar has a group that coordinates translations of W3C and W3C-related documents into IberoAmerican languages (primarily Spanish, but also Basque (Euskera), Catalan, Gallego, Portuguese and potentially others). They maintain a glossary that provides agreed translations of W3C terms.

⊣ W3C Glossary

A glossary of terms defined in W3C specifications, updated to use SKOS as preparation for this workshop.

⊣ W3C Specifications (etc) in translation

W3C now manages the collection of translations of its documents through RDF