Downscaling Linked Data Vocabularies

Daan de Ruijter

Vrije Universiteit

Introduction

In the developed world, linked data offers significant opportunities for creating, sharing and managing information. Having access to powerful hardware and a stable Internet connection with sufficient bandwidth enables the ability to access and interact with the Semantic Web. However, this infrastructure is not available to the same extent in developing countries. Linked data must be "downscaled" to allow its use in these low-resource infrastructures.[1] An example of downscaling linked data is the creation of a communication protocol that allows Semantic data to be sent via SMS.[2] While the data is still stored on the Web (e.g. in Triple Stores), HTTP/SMS converters are put in place that allow phone numbers to request and receive this data.

The ability to access linked data is one prerequisite to reap its benefits, another prerequisite is to be able to understand the various concepts that are present in a linked data resource. In the developed world, something written in English can be understood by the majority of people. In the developing parts of the world however, illiteracy and language barriers are access barriers to the Web.[3] As long as these access barriers are in place, the Semantic Web serves to widen the digital divide rather than to narrow it.[1,4] Given this, downscaling linked data should thus also involve providing linked data in a wider variety of languages.

For demonstrative purposes, consider an RDF file expressed as a SKOS vocabulary. An example of two concepts from a SKOS vocabulary is given in Figure 1. Concepts can have multiple labels with different language tags ("@en" / "@nl"). Because SKOS is an internationally recognized standard, this approach is sufficient to facilitate a vocabulary with multiple languages in the developed world. For a person completely disconnected from the English language however, terms like "broader" and "narrower" are completely foreign. For a

western person, this language barrier might be emulated by imagining Chinese characters to take the place of “broader” (skos:宽). Concept labels for small languages may be added, but if the relations between concepts are not understood the developing world cannot interact with the Semantic Web on a level playing field with the developed world.

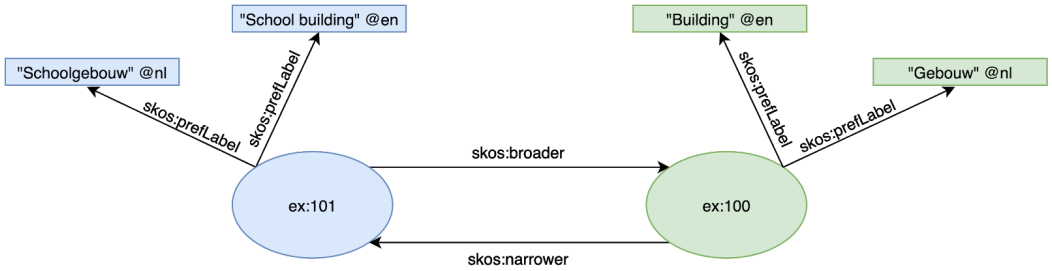

Figure 1. Classic SKOS example

Proposed Method

Similar to how concept labels can be tagged with a language, it would not be far-fetched to assume that vocabularies like SKOS can be defined in multiple languages. However, putting all languages of a vocabulary in a single document would greatly increase its size. Larger file sizes would hinder the accessibility of vocabularies in low-resource infrastructures.[2] In order to facilitate vocabularies in multiple languages without increasing file sizes, converters could be placed in between the requests similar to the HTTP/SMS communication protocol. An example of how vocabularies could be distributed in multiple languages is shown in Figure 2.

Figure 2. SKOS with multiple language support

Agents are able to send a request for concepts to the online resource. The language setting of the agent is sent along with this request. The result that is sent by the online resource is

converted according to this language setting, allowing the agent to interpret the result regardless of English proficiency.

In order not to interfere with the established Semantic Web of the developed world, the different language versions are added as an extension to the vocabulary. This method is reminiscent of the so-called slot-and-filler method, which can be used to produce localized voice-based services.[3] Additional languages can be added to the extension by means of adding their translated vocabularies to the translation table.

Conclusion

Language localization on the Web is commonplace, allowing Semantic Web vocabularies to be localized would increase their availability to non-english speakers in developing countries. Actual implementation of the proposed method would undoubtedly face enormous amounts of technical difficulties. However, no longer having the English based-terms in vocabularies like SKOS, RDF and Dublin Core be considered as access barriers to the Semantic Web would help close the digital divide between the developed and the developing world.

References

1. Guéret, Christophe & Boer, Victor & Schlobach, Stefan. (2014). Let's "Downscale" Linked Data. *Internet Computing, IEEE.* 18. 70-73. 10.1109/MIC.2014.29.
2. Valkering, Onno & Boer, Victor & Lô, Gossa & Blankendaal, Romy & Schlobach, Stefan. (2016). The Semantic Web in an SMS. 697-712. 10.1007/978-3-319-49004-5_45.
3. Gyan, Nana Baah & Boer, Victor & Bon, Anna & van Aart, Chris & Akkermans, Hans & Boyera, Stephane & Froumentin, Max & Grewal, Aman & Allen, Mary. (2013). Voice-based Web access in rural Africa. 122-131. 10.1145/2464464.2464496.
4. Davies, Tim & Edwards, Duncan. (2012). Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *IDS Bulletin*. 43.10.1111/j.1759-5436.2012.00372.x.