VIRTUAL LANGUAGE FRAMEWORK (VLF)
A Semantic Abstraction Layer

Frédéric Hallot[1,2] and Wim Mees[1]
[1]*Department CISS, Royal Militatary Academy (RMA),Av de la Renaissance, 1000 Brussels, Belgium*

[2]*STARLab, Vrije Universiteit Brussel (VUB), Pleinlaan 1,1000 Brussels, Belgium*

Keywords: Semantic, Abstraction, Interoperability, Ontology, Multilingualism, Layer, Framework, SAL, VLF.

Abstract: In a previous paper, we presented the concept of the Semantic Abstraction Layer (SAL) as a theoretical abstraction aiming to solve some recurrent design problems related to semantics and multilingualism. In this paper, after a short recall of what a SAL is, we present the Virtual Language Framework (VLF), which is our implementation of the SAL concept. We present two approaches for implementing the VLF, one centralized and the other decentralized. We discuss their advantages and drawbacks and then present our solution, which combines both strategies. We end with a short description of an ongoing project at the Royal Military Academy of Belgium where the VLF is used in the context of a disaster management information system.

1 INTRODUCTION

Currently the Web is primarily developed by humans (and web applications) for humans. In the paper (Berners-Lee, 2001) that originated the Semantic Web (Daconta, 2003), Tim Berners-Lee announced the emergence of a new, parallel web that would be made by the machines for the machines. A lot of efforts already exists on the Web in order to help humans understand each other across language boundaries (Wikipedia, Wordnet, Babelfish, ..., to cite only a few of them). Unfortunately they are aimed at human people usage, leaving space for interpretation, which humans can do, but not machines, at least not today. The main goal of the PhD research entitled "Multilingual Semantic Web Services" (Hallot, 2005) is to search how we could provide a better comprehension (interoperability) between applications, programs, computer agents used (developed) by people having different cultures, locale, languages. The subject of this thesis initially focused on the Semantic Web and thus on ontologies (computer science). We rapidly discovered though, that semantic interoperability between applications was a cross-cutting concern that spanned most domains of computer sciences and not only ontologies. We pointed out that fact in (Hallot, 2008) where we developed the concept of Semantic Abstraction Layer (SAL). The SAL should provide an indirection layer between every possible combination of computer applications, files, schemas, ... and the concepts and relationships they are working with (see Figure 1).

Figure 1: SAL - An indirection layer (from Hallot, 2008).

The SAL is a theoretical abstraction, pointing the need for applications to be able to share concepts without any ambiguity, without the need for a subjective interpretation. The second important development of the mentioned PhD thesis will be the development of the Virtual Language Framework (VLF). The VLF will provide one implementation of a SAL. The concept of VLF was briefly presented in (Hallot, 2008) and will be developed further in this paper.

Hallot F. and Mees W.
VIRTUAL LANGUAGE FRAMEWORK (VLF) - A Semantic Abstraction Layer.
DOI: 10.5220/0001818203930398
In *Proceedings of the Fifth International Conference on Web Information Systems and Technologies (WEBIST 2009)*, page
ISBN: 978-989-8111-81-4

2 THE VIRTUAL LANGUAGE FRAMEWORK (VLF)

2.1 Introduction

Having defined the Semantic Abstraction Layer, it is now the time to implement it. Our implementation will be called the Virtual Language Framework (VLF).

But before beginning, we would like to point out that it is our intention to formalize the VLF and to develop it as an open source initiative. The formalization of the VLF though will not be part of this paper, but will be the main focus of a later one.

2.2 VLF Goals

The main goal of the VLF should be to promote semantic interoperability: this means sharing concepts across companies, organizations, languages and cultural boundaries. It would also allow user communities to build an ever growing open, language and culture independent reference sets of concepts and relationships.

A second goal, which could be looked upon as a side-effect of the previous one, consists in promoting concept reusability: this would allow companies not having to reinvent the wheel each time they have to develop new Information Systems or Computer Applications.

2.3 VLF Basic Requirements

The VLF must be a framework allowing everyone, every company, organization or even community to create their own Virtual Language(s) (VL). Every VL will be shared and open for access to everyone via web services (SOAP and REST). VL should at least permit to handle concepts and binary relationships.

Taking the basic requirements into account, we will next discuss possible approaches for defining and developing the Virtual Language Framework.

2.4 Possible Approaches for Implementing the VLF

2.4.1 A Centralized Architecture

A first strategy for developing the VLF would be to implement a centralized architecture, leaving the whole responsibility of each VL to those who created it. In this option, we can imagine a VL as two different sets. A set of Unique Concept ID's (UCID) and a set of Unique Relationship ID's (URID) (see Figure 2).

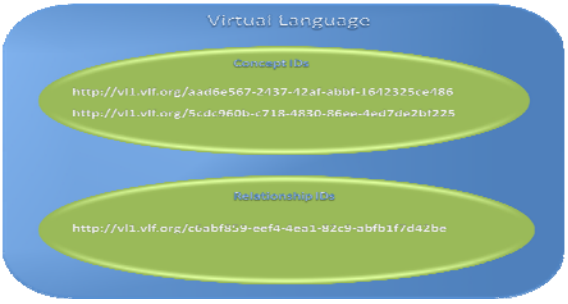

Figure 2: VLF - base sets of a centralized VL.

The VLF should then provide tools (for instance a web based application) in order to manage all the data (see Figure 3) which describes and represents the concepts identified by their UCID.

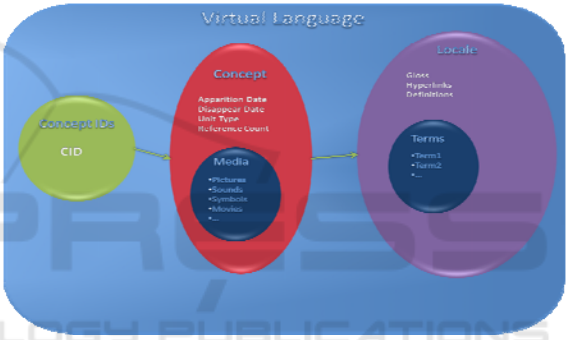

Figure 3: VLF - Concepts data managed by the VLF.

In a similar fashion, relationships should be manageable by the VLF. For the time being, we only intend to implement binary, bidirectional relationships between concepts that can thus be followed in both directions (e.g. Imagine a Relationship that binds the Concept "Professor Meersman" and the Concept "the Database course". Readings must be provided in both directions: Professor Meersman teaches the Database course" and "The Database course is given by Professor Meersman").

The dual reading is the reason why we intend to treat concepts and relationships differently within the VLF. The restriction of relationships to binary ones is sufficient as it has been shown in (Halpin, 1995) that n-ary relationships can be non-loss decomposed in a set of binary ones. Hence RDF triples (Daconta, 2003) and DOGMA Lexons (Jarrar, 2002) exclusively rely on binary relationships in order to build the fact base of their ontologies.

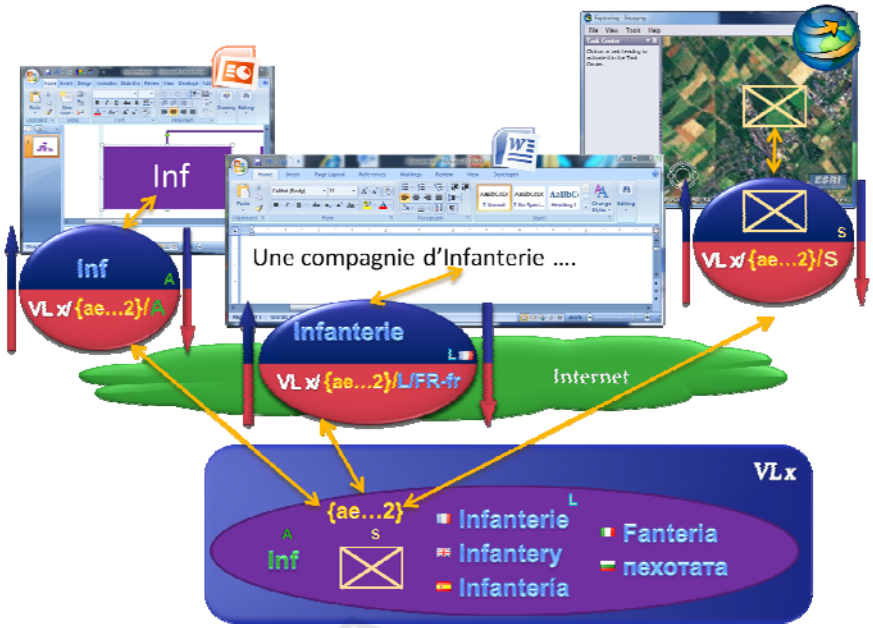

Figure 4: VLF - Application relying on a centralized VLF.

The VLF should provide an external access to the data regarding concepts and relationships, ideally through web services (SOAP and REST). This should allow for the development of applications that share the same UCID's but each choose their preferred representations for the concepts based on the type of application, the locale, restrictions imposed by the type of user interface (e.g. mobile phone versus desktop PC) or even the end user's preference. It is also interesting to note that representations of concepts should not be confined to linguistic labels but to all possible multimedia representations (symbol, picture, movies and even sounds) (see Figure 4).

A centralized control has several advantages:

- It allows companies and organizations to standardize the concepts and relationships used in their applications.
- It helps to control the integrity of the VL data and to provide a certain level of quality assurance.
- It also allows envisaging some access control (parental control, private data ...).
- It finally helps to provide some credibility according to the authority controlling the Virtual Language: this allows users to trust (rely on) the VL data and a certain level of their quality.

It nevertheless implies several drawbacks:

- Defining and maintaining a complete set of concepts is a titanic work, which means that only big companies and organizations could afford to create VL's. Probably neither a company nor an organization exists that could handle the translation of the concepts in every possible locale. Furthermore, what would happen if the "owner" of the VL lost interest in the VL but a large community of users still like to continue using it and having it evolve to fit future needs?
- The proprietary aspect from this centralized strategy could refrain a lot of actors to embrace this technology.
- Hosting a VL asks a consequent investment in terms of hardware and bandwidth, growing with the number of users.
- High availability is essential and backups are critical.

Some big companies and organizations which transcend the linguistic boundaries could be interested in keeping full control over the VL's they develop for their own purpose.

But the main goal of the VLF consists in promoting semantic interoperability and this goal is in this case not met.

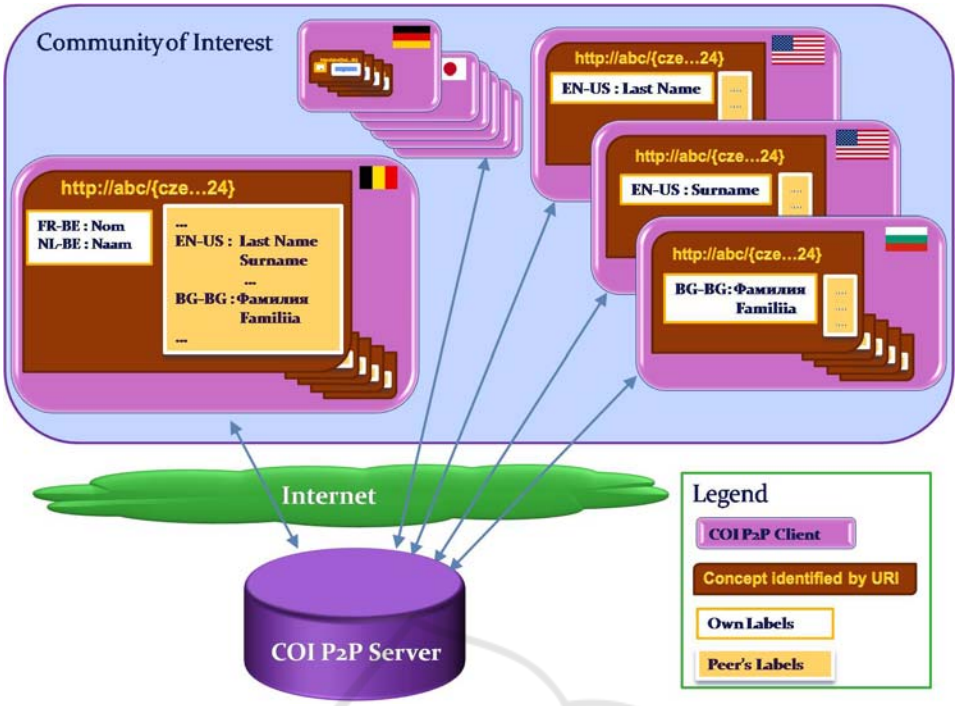

Figure 5: A decentralized strategy.

2.4.2 A Decentralized Architecture

Another possible architecture for the VLF could be a decentralized one. Indeed, at one hand, the important costs associated with the centralized architecture could prevent this to emerge, and at the other hand, communities have proven the ability to conjointly manage big quantities of data by letting each user bring his or her contribution (e.g. Wikepedia, Flickr, social bookmarking sites,...).

We can imagine the VLF as a Framework allowing Communities of Interest (COI) to share the semantic of the concepts in which they have an interest by creating a peer-to-peer network, where each member of the community would eventually define some representation for the concepts and inherit from the representations defined by all other users (see Figure 5).

The decentralized control also has several advantages:

- The community rules: social networking has shown that a lot of good information can grow out of virtual communities where a lot of people each treat a limited amount of information.
- Translations in even very minor languages (with respect to the number of users on earth) would happen a soon as some of its users are a member of the COI. KDE (one of the unix windowing system) has proven to have more translation than Microsoft Windows, because it relies of some small communities that provide the translations).
- High redundancy of the data, which means high availability and few risks of irreversible loss of data.

Nevertheless, it has some drawbacks too:

- There are some difficulties to bootstrap new concepts!
- How to know which COI exists? How do you let know about the concepts for people outside the COI? This means that some search mechanism is needed (flooding: limited horizon filtering the more obscure terms).
- How to deal with several COI at the same time?
- It doesn't prevent that a same concept can exist in different COI, and if so, how is it possible to discover identical concepts in other COI?
- Risk of erroneous interpretations that lead to bad translation, and therefore how to trust translations in languages people don't master.

The goal of promoting semantic interoperability would be achieved, but some big companies or organizations would be very reluctant to accept relying on uncontrolled community provided data for their applications.

Figure 6: Geo-hazards management tool - A VLF Application.

2.4.3 Solution: A Mixed Architecture

As we have seen in the centralized and as in the decentralized approach, there are some pro's and contra's for both architectures. This is why we are proposing a mixed solution, where the VLF would allow creating Virtual Languages which are completely centralized, fully decentralized or offer an in between data management solution allowing at the same time some centrally controlled data and at the same time some delocalized data enrichments. Although more complex to elaborate, this solution should allow us to merge the benefits of both described architectures, and hence to enlarge the potential field of VLF's users. The formalization of the mixed architecture of the Virtual Language Framework will be described in a following paper.

3 VLF APPLICATION

One of the long term goals at the Signal and Image Center (SIC) from the Royal Military Academy (RMA) of Belgium consists in the elaboration of a Crisis Management Information System (CMIS) (Mertens, 2006) (Mertens, 2007). This goal is progressively achieved by starting projects whose objectives are to bring new parts to the CMIS.

One ongoing project in this framework is called "Development of a Geo-hazards Management tool" (project number C4-16). In the context of crisis management, it is important to be able to (geo-) localize the information and knowledge related to the phenomenon of interest (Closson, 2005) (Closson, 2007). In order to collect the information, a field expert is sent to the region of interest. He uses a GPS device in order to determine the location of his measurements and a camera in order to record evidence of the hazards that will help the decision makers to acquire a better knowledge of the situation on the ground. He also takes some notes in order to further describe the situation and keeps track of the position and orientation of each picture (Closson, 2005). Currently, the geographer has to compile manually all the information gathered during one day, in the evening, at his hotel. The elaboration of the CMIS tool (see Fig. 6) has two main goals. First it must help the geographer to integrate all the information he gets from his different sources (devices) in real time when he is in the field. Not only would this make him more productive since it helps him to save time, but furthermore it avoids transcription errors and omissions (due to the delay).

Moreover, the use of the VLF in order to annotate (tag) the geo-hazards should help the expert to collect, store and manage the information at a higher abstraction level than the linguistic one. Indeed, by essence, in many regions of the world, a crisis management information system will most often be used by people from different nationalities, speaking different languages and having a different cultural background. Although it is convenient to expect from all collaborators to speak English, the reality proves frequently that this is far from being the case. For this CMIS tool, two virtual languages will be created, one for the Graphical User Interface and one for the Geo-Hazards terminology.

We initially plan to control the VLF in a centralized fashion, but over time, one can hope that the virtual languages would be adopted and subsequently extended by a user community.

4 VLF EXTENSIONS

We already think about the future extensions of the VLF, once the VLF will have reached a certain level of maturity and use.

A first extension will be the creation of another framework allowing for the development of taxonomies based on the VLF. This framework will logically be called the Virtual Language Taxonomy Framework (VLTF) and would allow us to develop taxonomies at a higher level of abstraction than spoken language, and that would allow to create some classification for which translations in all human languages (even latin) would be straightforward and unambiguous.

Following the same logic, the Virtual Language Ontology Framework (VLOF) will help to create ontologies which will completely rely on Virtual Languages, as well for the concepts as for the relationships. This means that all the facts (lexons (Jarrar, 2002) ,(Spyns, 2004)) of the ontologies will stand at a higher level of abstraction than those from currently designed ontologies which completely rely on linguistic labels.

5 CONCLUSIONS AND FUTURE WORK

Although a lot of work still needs to be done, there is conviction at the Royal Military Academy of Belgium that an efficient use of the VLF will help us to factor most semantic and linguistic matters out of most of the projects relying on information technologies. The definition of the Semantic Abstraction Layer was a very important first step in order to point out the semantic problems that a lot of people within the computer science community are facing. The elaboration of the Virtual Language Framework in order to implement the SAL is a second important step in the direction of semantic interoperability. The formalization of the Virtual Language Framework will be the next important step and will also be published in the form of a paper or article. It is our wish to open the specifications of the VLF in an open source context, in order to let the community collaborate with us in its future development.

REFERENCES

Berners-Lee T., Hendler J., & Lassila O., 2001, *The Semantic We*, Scientific American.com

Closson D., 2005, *Karst system developed in salt layers of the Lisan Peninsula, Dead Sea, Jordan Structural control of sinkholes and subsidence hazards along the Jordanian Dead Sea coast*, Journal Environmental Geology, Springer, ISSN 0943-0105, Issue Volume 47, Number 2, pp. 290-301

Closson D., LaMoreaux P.E., Karaki N.A. and al-Fugha H., 2007, *Karst system developed in salt layers of the Lisan Peninsula, Dead Sea, Jordan*, Journal Environmental Geology, Springer, ISSN 0943-0105, Issue Volume 52, Number 1, pp. 155-172

Daconta M.C., Obrst L.J. & Smith K.T., 2003, *The Semantic Web - A guide to the future of XML, Web Services and Knowledge Management*, Wiley, ISBN 0-471-43257-1.

Halpin T., 1995, *Conceptual Schema & Relational Database Design*, Prentice Hall of Australia, Second Edition, ISBN 0-13-355702-2

Hallot F., 2005, *Multilingual Semantic Web Services*, in On the Move to Meaningful Internet Systems 2005: OTM 2005 Workshops, OTM Confederated International Workshops and Posters, AWeSOMe, CAMS, GADA, MIOS+INTEROP, ORM, PhDS, SeBGIS, SWWS, and WOSE 2005, Agia Napa, Cyprus, October 31 - November 4, 2005, Proceedings, LNCS3762, ISBN 3- 540-29739-1, pp.771-779.

Hallot F., 2008, *Semantic Abstraction Layer (SAL) - An indirection layer to solve semantic ambiguities and multilinguality problems*, WMSCI 2008, The 12th World Multi-Conference on Systemics, Cybernetics and Informatics , June 29th - July 2nd, 2008 Orlando, Florida, USA, Proceedings Volume I, ISBN 1-934272-31-0, pp. 118-123

Jarrar M. & Meersman R., 2002, *Formal Ontology Engineering in the DOGMA Aproach*. in Meersman R., Tari Z. et al., (eds.),On the Move to Meaningful Internet Systems 2002, CoopIS, DOA, and ODBASE; Confederated International Conferences Coopis, DOA, and ODBASE 2002 Proceedings, LNCS 2519, Springer Verlag, pp. 1238-1254.

Mertens K. & Mees W., 2006, *Communication and Information System for Disaster Relief Operations*, Proceedings of the 3rd International Conference on Information Systems for Crisis Response and Management (ISCRAM06), New Jersey, USA

Mertens K. & Mees W., 2007, *Prototype of an Information System for Disaster Relief Operations* , Practitioners proceedings of the 4th International Conference on Information Systems for Crisis Response and Management (ISCRAM07), Delft, The Netherlands

Spyns P, 2004, *Methods to be used in engineering ontologies.* Technical Report 15 2004, VUB - STAR Lab, Brussel.