Robust ontology development in NeOn

by Mari Carmen Suárez & Asunción Gómez Pérez

Expert ontology engineers are a scarce resource and, as ontology engineering enters the mainstream, there is now a strong need for practical methodologies, which can assist a variety of user types with ontology design tasks. To address this need NeOn proposes a scenario-based methodology (“The NeOn Methodology”), which provides guidance with respect to all key aspects of the ontology engineering process, e.g., including collaborative ontology development, the reuse of ontological and non-ontological resources, and the evolution and maintenance of networked ontologies. Crucially, this methodology is tightly coupled with the NeOn Toolkit - http://neon-toolkit.org, which provides several plugins to support concretely the various activities characterising the ontology engineering process.

In contrast with other approaches to providing methodological guidance for ontology engineering, the NeOn Methodology does not prescribe a rigid workflow, but instead suggests pathways and activities for a variety of scenarios. These scenarios cover commonly occurring situations, e.g., when existing ontologies need to be re-engineered, aligned, modularised, or integrated with non-ontological resources, such as databases, folksonomies, or thesauri. As can be expected from an international project involving six European nations, the NeOn Methodology also supports ontology localisation, to facilitate the adaptation of an ontology to different languages and cultures.

Another important aspect of the NeOn Methodology is its provision of a library of ontology design patterns. The Library of Ontology Design Patterns (ODP) is an open and modular architecture on top of the Eclipse platform. The ODP portal, found at http://ontologydesignpatterns.org, contains 67 patterns divided into six main categories: Structural, Correspondence, Content, Reasoning, Presentation, and Lexicon-Syntactic. The portal is meant to define a focal point for the ontology engineering community, and provides support for submitting design patterns, retrieving them from the repository, posting modelling issues, discussing and reviewing existing solutions, and finding educational material on ontology design.

The integration between the ODP portal and the NeOn Toolkit is provided by the XD plugin, which implements the eXtreme Design method, supporting pattern-based design in ontology engineering. XD knows good practices, and provides them to an ontology developer, in response to a specific modelling issue. XD runs like a wizard, using diagrammatic and graphical interfaces, and also assists a developer in adding annotations and publishing the resulting ontology in a repository.

The NeOn project started in March 2006 with the aim of delivering a major advance in the state of the art of semantic technologies, in particular by developing robust solutions, which would improve both the technological and the economical viability of large-scale, network-aware, ontology-based applications.

Three and a half years later, the project is entering its final six months and it now is a good time to reflect on our original tenets and goals and assess what we got right and what we did not. In particular, the key assumption at the basis of the project was essentially that the Semantic Web was not a transient fad, but a key technological development, which is going to stay with us and will eventually dramatically influence the way web applications are created in the future.

Well, I would say that there is little doubt we got this one right. Although the Semantic Web still has a long way to go, the picture has dramatically changed over the past few years: key players, such as Yahoo! and even Google (until recently seen as ‘anti-Semantic Web’), have embraced it.

Continued on page 2
Cupboard: sharing and reusing networked ontologies

Ontology engineers need a place to host their ontologies, as well as tools that can allow them to manage and, ultimately, to make them available to others. As ontology users, we need tools that can help us to locate ontologies that are relevant to our task. In order to be able to make an informed choice, users also need information about the quality, provenance and key characteristics of the ontologies available for reuse. Finally, application developers need infrastructural components, which can support the exploration and querying of both standalone and networked ontologies.

Cupboard is an online ontology hosting system that intends to address the needs of these three categories of users. Each user of Cupboard is provided with her/his own Ontology Space, where ontologies and alignments can be uploaded and stored. An Ontology Space provides a summary of the networked ontologies it includes, as well as means to review and attach rich metadata to them. Moreover, advanced search mechanisms are provided so that users can easily find, inspect and explore ontologies available online. In addition to this user ‘facade’, an Ontology Space also provides a virtual infrastructure for Cupboard users to build applications exploiting ontologies and alignments, as well as ontology ratings and metadata. A number of services and APIs are deployed to handle tasks such as ontology search, exploration and querying.

Web 2.0 gets semantics

Sofia Angelotau

The rapid establishment of the content sharing principles of Web 2.0 has led to a plethora of user-friendly web applications, such as Flickr, YouTube and Delicious. Indeed, these are only a few of the many web applications that allow users to upload their content and annotate it with freely selected keywords, i.e. tags, with minimal effort. This paradigm of content sharing has led to a content intensive web with billions of user-generated resources and tags. However, the loose and unstructured nature of the annotations that currently describe the content, combined with the lack of efficient query mechanisms, hamper the retrieval and navigation of the resulting tagged resource spaces. For example a query for “aquatic plants” in Flickr will only retrieve resources explicitly tagged with these two keywords and not those tagged with lotus, water-lily etc., which are indeed aquatic plants.

The first prize at the Billion Triples Challenge, it earned its authors a best paper award.

This approach to ontology summarisation has provided the basis for developing an innovative approach to visualising and navigating ontologies. In particular, it enables ‘middle-out ontology browsing’, where it becomes possible to move through complex information spaces from the most valuable nodes (i.e., key concepts) and then to unfold smaller chunks of the ontological graph to inspect specific sub-parts of an ontology. This approach is similar to map-based visualisation and navigation in Geographical Information Systems, where, e.g., maps are displayed more prominently depending on the current level of granularity.

Visualising and navigating large ontologies

Martin Dzbor

The user studies carried out at the beginning of the NeOn project clearly indicated that the user interaction metaphors used in ontology engineering toolkits are largely inadequate, especially for those users with limited experience. Hence, a key challenge in our work on Human-Ontology Interaction is overcoming these problems and developing novel interactive frameworks for visualising and navigating large and complex ontologies. To this purpose we designed and implemented an innovative solution, based on the idea of identifying ‘key concepts’ in ontologies and using them as landmarks for exploring and making sense of large ontologies.

Key concepts can be seen as a limited number of descriptive ontology elements that best characterise what a particular ontology is about. In our work we elaborated and grounded this informal notion in a so-called metadata categories, and developed an automatic ontology summarisation technique based on this idea. As a result, it becomes possible for a user to get a quick understanding of an ontology of, say, one thousand concepts, simply by being presented with 15-20 key concepts. This work has been received very well by the Semantic Web community, to the extent that, when first reported at the 2008 Asian Semantic Web Conference, it earned its authors a best paper award.

Another innovation, called ‘conceptual zooming’, offers the user the opportunity to calculate ontology summaries, taking into account both the content and the topology of the underlying model. Drawing upon a visual metaphor familiar to Web 2.0 users, where tags with greater popularity are depicted more prominently, our KC-Viz framework allows the user to distinguish between several layers of key concept importance, thus realizing the notion of key concepts at different levels of granularity.

While Cupboard is still a young system (currently in beta testing), a number of uses and applications have already been identified and realised. For example, a Cupboard Ontology Space has been synchronised with the catalogue of design patterns on OntologyDesignPatterns.org. We have also developed a plugin for the NeOn Toolkit, providing a tool for ontology development by reuse, which exploits in particular the ratings users have given to ontologies. Experiments have shown that this plugin effectively facilitates the tasks of finding, selecting and integrating existing formal knowledge structures in an ontology development project.
A GATE-way into NeOn
by Wim Peters & Diana Maynard

In principle, ontologies are in constant flux. At any time they can be extended on the basis of new information, which can be obtained, e.g., from domain experts or textual data. In particular, this relation between ontology and a corpus of documents supports the knowledge management scenarios, where ontologies are used to facilitate intelligent search in a corporate intranet.

GATE, which is one of the most widely used and well known natural language processing (NLP) architectures, has now opened yet another of its many doors towards interoperability with the NeOn Toolkit. As a result, it is now possible for ontology changes to be propagated in two directions, which we call the top-down and bottom-up approaches to ontology change.

In the top-down approach, changes to an ontology lead to changes in conceptual indexes associated with texts. This is important for GATE, which takes care of the production and management of textual metadata in the form of annotations. Texts are annotated with ontological concepts, and the textual elements function as instances of these classes. For example, if a concept is deleted, all textual instances of this concept will either need to be deleted or moved to the superclass of the deleted class. If a user adds new concepts to the ontology, then it may be necessary to return to the text to check whether additional instances can be found that should be used to populate these new concepts in the ontology.

In the bottom-up approach, change may be initiated from the text side. NLP techniques can assist in the development of Semantic Web technology by imposing a linguistically motivated structured order on unrestricted natural language use within documents. By proposing changes in an ontology on the basis of textual evidence, NLP techniques provide the link between unstructured text and the lexico-syntactic patterns used for this purpose. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.

TEXT-DERIVED SUGGESTIONS FOR ONTOLOGY ENRICHMENT

SYNCHRONISATION OF ONTOLOGY CHANGES WITH TEXT ANNOTATION

For this purpose we have developed a generic GATE application called SPRAT (Semantic Pattern Recognition and Annotation Tool). SPRAT is a text information extraction tool, which is based on syntactic and semantic patterns used by this tool. If we obtain additional relevant textual material and/ or find new instances in that text, it may be necessary to modify the ontology to take into consideration this new information. This could include adding new concepts or new relations between existing concepts in the ontology.

Repairing ontologies with RaDON
by Gullin Qi, Qiu Ji & Peter Haase

As the complexity of semantic applications increases, more and more effort is invested in ontologies which are typically drawn from a wide variety of sources. This new generation of applications also tends to rely more and more on networks of ontologies, rather than isolated, monolithic ones. One of the major challenges in managing these networked and dynamic ontologies concerns the handling of potential inconsistencies, which can occur both in individual ontologies and, more importantly, when integrating multiple distributed ontologies that have been created independently from each other.

For inconsistency handling in single, centralised ontologies, several approaches have been proposed in recent years and a number of tools have been developed. However, there are few tools around which can diagnose and repair inconsistencies in networked ontologies.

The goal of the RaDON system is to provide support for repairing and diagnosing not only within single ontologies, but also in networked ones. It supports novel strategies and consistency models for distributed text. At the moment, the lexico-syntactic patterns used for this purpose were used in a number of projects, and we started to use these patterns for this purpose. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.

TEXT-DERIVED SUGGESTIONS FOR ONTOLOGY ENRICHMENT

SYNCHRONISATION OF ONTOLOGY CHANGES WITH TEXT ANNOTATION

For this purpose we have developed a generic GATE application called SPRAT (Semantic Pattern Recognition and Annotation Tool). SPRAT is a text information extraction tool, which is based on syntactic and semantic patterns used by this tool. If we obtain additional relevant textual material and/or find new instances in that text, it may be necessary to modify the ontology to take into consideration this new information. This could include adding new concepts or new relations between existing concepts in the ontology.

Repairing ontologies with RaDON
by Gullin Qi, Qiu Ji & Peter Haase

As the complexity of semantic applications increases, more and more effort is invested in ontologies which are typically drawn from a wide variety of sources. This new generation of applications also tends to rely more and more on networks of ontologies, rather than isolated, monolithic ones. One of the major challenges in managing these networked and dynamic ontologies concerns the handling of potential inconsistencies, which can occur both in individual ontologies and, more importantly, when integrating multiple distributed ontologies that have been created independently from each other.

For inconsistency handling in single, centralised ontologies, several approaches have been proposed in recent years and a number of tools have been developed. However, there are few tools around which can diagnose and repair inconsistencies in networked ontologies.

The goal of the RaDON system is to provide support for repairing and diagnosing not only within single ontologies, but also in networked ones. It supports novel strategies and consistency models for distributed text. At the moment, the lexico-syntactic patterns used for this purpose were used in a number of projects, and we started to use these patterns for this purpose. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.

TEXT-DERIVED SUGGESTIONS FOR ONTOLOGY ENRICHMENT

SYNCHRONISATION OF ONTOLOGY CHANGES WITH TEXT ANNOTATION

For this purpose we have developed a generic GATE application called SPRAT (Semantic Pattern Recognition and Annotation Tool). SPRAT is a text information extraction tool, which is based on syntactic and semantic patterns used by this tool. If we obtain additional relevant textual material and/or find new instances in that text, it may be necessary to modify the ontology to take into consideration this new information. This could include adding new concepts or new relations between existing concepts in the ontology.

Repairing ontologies with RaDON
by Gullin Qi, Qiu Ji & Peter Haase

As the complexity of semantic applications increases, more and more effort is invested in ontologies which are typically drawn from a wide variety of sources. This new generation of applications also tends to rely more and more on networks of ontologies, rather than isolated, monolithic ones. One of the major challenges in managing these networked and dynamic ontologies concerns the handling of potential inconsistencies, which can occur both in individual ontologies and, more importantly, when integrating multiple distributed ontologies that have been created independently from each other.

For inconsistency handling in single, centralised ontologies, several approaches have been proposed in recent years and a number of tools have been developed. However, there are few tools around which can diagnose and repair inconsistencies in networked ontologies.

The goal of the RaDON system is to provide support for repairing and diagnosing not only within single ontologies, but also in networked ones. It supports novel strategies and consistency models for distributed text. At the moment, the lexico-syntactic patterns used for this purpose were used in a number of projects, and we started to use these patterns for this purpose. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.

TEXT-DERIVED SUGGESTIONS FOR ONTOLOGY ENRICHMENT

SYNCHRONISATION OF ONTOLOGY CHANGES WITH TEXT ANNOTATION

For this purpose we have developed a generic GATE application called SPRAT (Semantic Pattern Recognition and Annotation Tool). SPRAT is a text information extraction tool, which is based on syntactic and semantic patterns used by this tool. If we obtain additional relevant textual material and/or find new instances in that text, it may be necessary to modify the ontology to take into consideration this new information. This could include adding new concepts or new relations between existing concepts in the ontology.

Repairing ontologies with RaDON
by Gullin Qi, Qiu Ji & Peter Haase

As the complexity of semantic applications increases, more and more effort is invested in ontologies which are typically drawn from a wide variety of sources. This new generation of applications also tends to rely more and more on networks of ontologies, rather than isolated, monolithic ones. One of the major challenges in managing these networked and dynamic ontologies concerns the handling of potential inconsistencies, which can occur both in individual ontologies and, more importantly, when integrating multiple distributed ontologies that have been created independently from each other.

For inconsistency handling in single, centralised ontologies, several approaches have been proposed in recent years and a number of tools have been developed. However, there are few tools around which can diagnose and repair inconsistencies in networked ontologies.

The goal of the RaDON system is to provide support for repairing and diagnosing not only within single ontologies, but also in networked ones. It supports novel strategies and consistency models for distributed text. At the moment, the lexico-syntactic patterns used for this purpose were used in a number of projects, and we started to use these patterns for this purpose. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.

Reconstructing an ontology is a complex task that requires a deep understanding of the existing data and the relationships between the concepts. In addition, the NeOn Toolkit also includes algorithms for decomposing modules from ontologies, on the basis of users’ specifications. Finally, whether they are manually specified or automatically extracted, modules can be combined using simple set-based operations. For instance, new modules can be created simply by merging two modules or by removing a set of definitions from a module.
FISHERY ONTOLOGY NETWORK

NeOn People

My name is Eva Blomqvist. I currently live in Rome, and I have a “fresh” PhD this year. I am originally from Sweden but moved to Rome in January 2009, where I am currently doing a post-doc at SALT (CNR). My PhD was in cooperation with Linköping University and Jönköping University, in Sweden. I started working with Aldo Gangemi at SALT after meeting him at the Semantic Web summer school in 2006. We then extended our collaboration to an institutional cooperation between CNR and Jönköping University. This is how I got involved in the NeOn project two years ago, even before being employed at CNR.

How did you become interested in ontology engineering?
Actually that was kind of a coincidence. When I finished my Master’s degree at Linköping University, Sweden was in the worst phase of the crisis of the IT business sector, and there were no jobs for people without experience. I ended up working half a year in a shop, selling candy and tobacco, in Stockholm before I finally found a PhD position in Jönköping. My supervisor, Kurt Sandkuhl, suggested that I do my research in the ontology engineering area. At that time I had no idea what an ontology was, since my background was in traditional Software Engineering. However, I quickly learned the basics, and it was really interesting, as I found it very different from anything I had worked with previously. It was challenging, and I liked challenges.

Congratulations on completing your PhD, What was it about?
The title of my PhD was “An automated Ontology Construction based on Patterns”, which essentially concerns constructing ontologies using automatic methods but enriching the results using ontology design patterns. The problem is that using automatic methods (often called ontology learning methods) produce quite sparse and diverse results. Hence, we need some way of adding domain knowledge to the automatic generated and a better structure to such ontologies, and this is where patterns come in. Content ontology design patterns are small, general, and reusable ontologies, and in my thesis I showed how these can be applied on top of the results returned by existing ontology learning methods, in order to produce better ontologies. The framework I proposed is called OntoCase.

What has been the impact of your work?
So far the method is not widely used, so I cannot say that it had any impact on ontology engineering practices yet. However, it was the first attempt to use ontology design patterns automatically, as building blocks for ontologies, so in the research community this was a first step towards a more automated and connected approach to ontology engineering. Currently I am working on integrating the method into a suite of tools for pattern-based ontology design called XD that will be part of the NeOn Toolkit plug-in. As soon as the method is publicly available in a tool, I think it will have a wider impact.

How does your PhD research relate to the major themes of the NeOn project (i.e., reuse, collaboration, ontology networks, etc.)?
My PhD research was all about reuse, and especially the reuse of best practices in the form of ontology design patterns, so in that sense it was closely related. However, the work at Jönköping University was more focused on establishing various ontologies, rather than general networked ontologies, hence the application focus was slightly different.

What are you working on at the moment?
Apart from implementing parts of the OntoCase framework to be included in the XD NeOn Toolkit plugin, as I already mentioned, I am working on more general methods for pattern matching and selection. Last year we conducted a series of experiments that showed that a major obstacle to the widespread use of ontology design patterns is the lack of tool support for finding, selecting, and reusing patterns. So my current focus has shifted from the purely automatic methods presented in my thesis, to more interactive methods, for example by considering in the selection process ideas on how to use any kind of input the users can produce to specify their requirements – this could be anything, from a set of keywords or competency questions, to an initial ontology and try to match to a catalogue of patterns. Based on this matching, a set of suitable patterns can then be presented to the user. Furthermore, the system can even propose how these could be applied to solve the modelling problem at hand.

With which partners do you primarily collaborate?
On the methodological side, and for experimenting with methodologies, our main partner is the UPM group in Madrid. Earlier this year we held a NeOn training session at FAO, which yielded some valuable initial ideas on how to improve selection and use of patterns. Another close cooperation is with AIFB in Karlsruhe, mainly with Johanna Villalba, working on novel ontology learning methods, which are closely related to pattern matching and ontology enrichment.

NeOn has a higher percentage of women than most other projects in the ICT area. Is this something you are aware of? Are there concrete benefits that ensue from this fact?
It is something I have reflected on in the past, but I believe that this is one of the reasons why this project has been a success, not only on the scientific level, but also among the people working in the project. A mix of different genders, nationalities, and research backgrounds, is always a key factor for creating an innovative and interesting environment to work in. For a young researcher like myself it is also reassuring to see that the female senior researchers are successfully combining their research career with family life. It is important to have role models like that. On a personal level, my family life is important to me, and provides the base for a sound family life.

What do you do when you are not working in NeOn?
I still work a bit for my old institution in Sweden as well helping out with supervising Master students, and I am also involved in a new project at CNR. When not working I like to do sports, such as running and biking, although the traffic situation in Rome sets some limits on the biking at the moment. Then of course there are frequent trips back to Jönköping, since my boyfriend still lives there.

Leaving aside your own work, what is in your view the coolest output of the NeOn project?
Actually there is no single tool or method that is my favourite, but I like the overall picture produced. I think this is the best contribution of the project. From the point of view of the project there have been a lot of tools and methods around also before NeOn, but no attempt to unify them all and show how they fit together. After NeOn I hope that we can really see that ontology engineering has become a mature engineering field, with well established methodologies and well-functioning tools that fit together.

NeOn partners are developing a network of interconnected pharmaceutical ontologies, aimed at helping to bridge the gap between different drug terminologies.

Within NeOn, the Semantic Nomenclature case study is developing a number of new ontologies to cover pharmaceutical resources available on the Internet. We are also connecting these new ontologies to other existing ones, such as Drugbank and subsets of SNOMED-CT among others. In this scenario the focus of the Semantic Nomenclature case study is less on developing a final solution for describing pharmaceutical products, than on providing the appropriate infrastructure and guidelines for connecting existing ontologies in the pharmaceutical domain, in the context of an evolving and changing environment. To this extent, the ontologies will be hosted in a public space where domain experts will be able to check, validate, rank and enhance the existing network, and create their own queries export these profiles to other systems as well as share them across a collaborative ontology development environment. NeOn’s also a semantic tool in its own right. The taxonomy of software components on a local NeOn Toolkit installation is generated on the basis of semantic descriptions of those components. In particular, each plugin is described in terms of the concepts defined in the C-ODD Light network of ontologies, a model characterising collaborative ontology design, which has been developed within NeOn.