# CS Honours: Ontology Engineering Miniproject
# Hoot Ontology Verbalisation


Mitch Myburgh
MYBMIT001
mybmit001@myuct.ac.za

Roscoe Jampies
JMPROS002
jmpros002@myuct.ac.za

Bayanda Chakuma
CHKBAY001
chkbay001@myuct.ac.za

Tumelo Makgaka
MKGTUM005
mkgtum005@myuct.ac.za


March 27, 2016


**Abstract**

Ontologies are a structured logical representation of a particular subject domain. Expressing this logic in a verbalisation can help users to understand the complex structure simply by reading sentences in their language of choice. Past attempts have been successful in representing ontologies in English and work has been done in verbalising Zulu ontologies. This report aims to take a subset of the OWL 2 DL Language functionality and verbalise it in a variety of languages including English, Tswana and Afrikaans; through a web based interface for easy processing of ontologies.


## 1 Introduction

Ontologies are used to represent a particular subject domain in a structured and logical manner. This is often difficult for the average user to understand. This makes verbalising or expressing an ontology in a language used by the user of vital importance. We have thus designed this web application called "Hoot", which attempts to provide users with a service which allows them to upload OWL ontologies and select the language they wish to verbalise it in. The verbalisation process takes the complex underlying mathematical relationships expressed in the OWL 2 DL Language and converting it into human readable text in a variety of languages.

### 1.1 Problem Statement

There are currently several approaches of verbalising an ontology into various languages. These include using a grammar based approach and a template based approach. This project will focus on creating a web based application that uses a template based approach to verbalise an OWL ontology in simple English, technical English, mathematical representation, Setswana and Afrikaans.

### 1.2 Research Objectives

**Create a web based application to verbalise an ontology**

- The resulting application allows the user to upload an OWL file and choose a language for it to be verbalized in.

# 2 Literature Review

## 2.1 Introduction

The review is structured as follows: Controlled natural language will be discussed and a comparison of three languages will be provided, then verbalisation will be defined with an example, lastly multilingual verbalisation will be discussed.

## 2.2 Controlled Natural Language

Natural languages denote human language such as English that can be used to explain human knowledge. They are constructed out of words and phrases that are specific and can easily be understood by humans. They have a major advantage of being very expressive however its informal structure can result in ambiguity, vagueness and potential inconsistencies when it is used to represent knowledge. Formal languages, such as description logics, are considered to be more effective in representing knowledge because they have a well-defined syntax, their meanings are unambiguous, and they support formal methods, such as reasoning. The disadvantage of formal languages is that domain specialists are usually unfamiliar with the languages which makes it hard for them to use. Controlled natural languages have been suggested as a solution to this problem.[7][13] They can be regarded as mediators between natural languages and formal languages because they are used to bridge the gap between them. Controlled natural languages are engineered languages which use a selection of vocabulary, morphological forms, grammatical constructs and semantic interpretations[12].[9] Conducted a comparison of three controlled natural languages namely: Attempto Controlled English (ACE)[14], Ordnance Survey Rabbit (Rabbit)[9], and Sydney OWL Syntax (SOS)[14].They found that although there are clearly differences between the three controlled natural languages, there is incredible overlap between them and therefore much common ground to build from.After their comperison of the three Languages, they concluded that there are sufficient commonalities between the three Controlled Natural Languages described in their paper to provide a good base from which to proceed.

## 2.3 Ontology Verbalisation

Verbalization is the process of writing the semantics captured in axioms into natural language sentences, which enables domain experts to be able to participate in the modelling and validation processes of their domain knowledge [11]. It involves automated generation of text in a controlled Natural language from ontologies written is some formal logic. The traditional approach to verbalisation is to produce text that can be read by domain experts and to allow for easier editing of the ontology and the result is often a controlled language. An example can be found in [11].

$$\forall x(Account(x) \rightarrow \exists y, z(Person(y) \wedge Company(z) \wedge (OwnedBy(x, y) \vee OwnedBy(x, z)))$$

can be verbalized into:

Each Account must be OwnedBy a Person or a Company, or both.

Many applications have adopted the use of ontologies to represent the knowledge base. However, domain experts and end users, who are meant to populate the ontologies, are often not familiar with the formal notations and require a means to understand the ontology [15]. Even the graphical interfaces used, for example Protégé, require a good understanding of logics which limits the domain experts use of the ontology. Ontology verbalisation provides a way for ontology engineers and domain expects to access or edit the knowledge in the ontology.[13]

## 2.4 Multilingual Verbalisation

The understanding that verbalisation of ontologies into natural language is one of the approaches crucial to making ontologies accessible to end users and domain subjects, Makes providing verbalisation in a multilingual manner important. South Africa has eleven official languages but still major software support is focused only on one language: English.[10] Examined verbalizing ontologies in isiZulu, their focus was based on natural language generation, grammar options and preferences ,which would be used to inform verbalization of knowledge representation languages and could contribute to machine translation. They provided novel verbalization patterns for simple subsumption, disjoin classes, conjunction, and basic options with quantification. They found that the main features complicating verbalization in isiZulu were the 17 noun classes with embedded semantics in the term, the agglutinative nature of isiZulu, and contextual knowledge about the position of the symbol in the axiom. Their conclusion was that using a plain template to verbalise isiZulu was infeasible and using a grammar engine to verbalise formally represented knowledge in

### 2.5 Conclusion

The review presents an introduction to Control Natural Languages although there are over forty of them only a comparison of three was discussed. The concept of ontology verbalisation and multilingual ontology verbalisation was also introduced and it was noted that almost all Controlled Natural Languages are written in English and that not enough work is being done towards verbalising Ontologies into a natural language that is not English also, the use of templates for natural language generation is not applicable to all languages for example isiZulu. A grammar engine would be more suited in such cases[10]

## 3 Implementation

A web based interface was created for the purpose of uploading, parsing and displaying the verbalised OWL file. The file can be expressed in simple English, technical English, mathematical representation, Afrikaans and Setswana, provided the OWL file only uses the supported functionality as expressed on the *About* section of the website.

The software was built using the Express[1] Node[2] framework, which enables quick prototyping of production ready web interfaces.

### 3.1 Major Software Artifacts

The major software artifacts produced are split into 3 major parts:

1. The interface which comprises of an upload form and display logic for the processed files.
2. The OWL file parser that processes the uploaded OWL file and constructs the tree formation for the display, and obtains the verbalisation from the language modules.
3. The language modules for simple English, technical English, mathematical representation, Afrikaans, Tswana.

### 3.2 Interface

The Hoot Ontology Verbalisation interface is a web based interface providing upload functionality for OWL files, premade OWL files are included and can be used for testing when no suitable OWL files are available. The interface structures the data in a tree structure using the same format and symbols as Protégé to allow users to easily interact with Hoot if they have previous experience with Protégé, however the interface is still easy to use if the user has no ontology experience.

Figure 1 shows the interface of Hoot with an OWL file open.

The interface provides two ways to display the data. A category mode, based on the display used in Protégé, which allows easy scanning as seen in Figure 1. The second mode is a paragraph mode which appends all the sentences to create a paragraph as seen in Figure 2. This method was used because reading a number of short sentences is easier for the user than reading long and complex compound sentences.

### 3.3 Language Modules

The OWL file is processed by a central parser that extracts the data and determines what type of sentence to produce, passing this information to the respective language modules to build a grammatically correct sentence to describe the data. The parser also builds a tree structure from the data for ease of display.

[1]Express is available at `http://expressjs.com/`
[2]Node is available at `https://nodejs.org/`

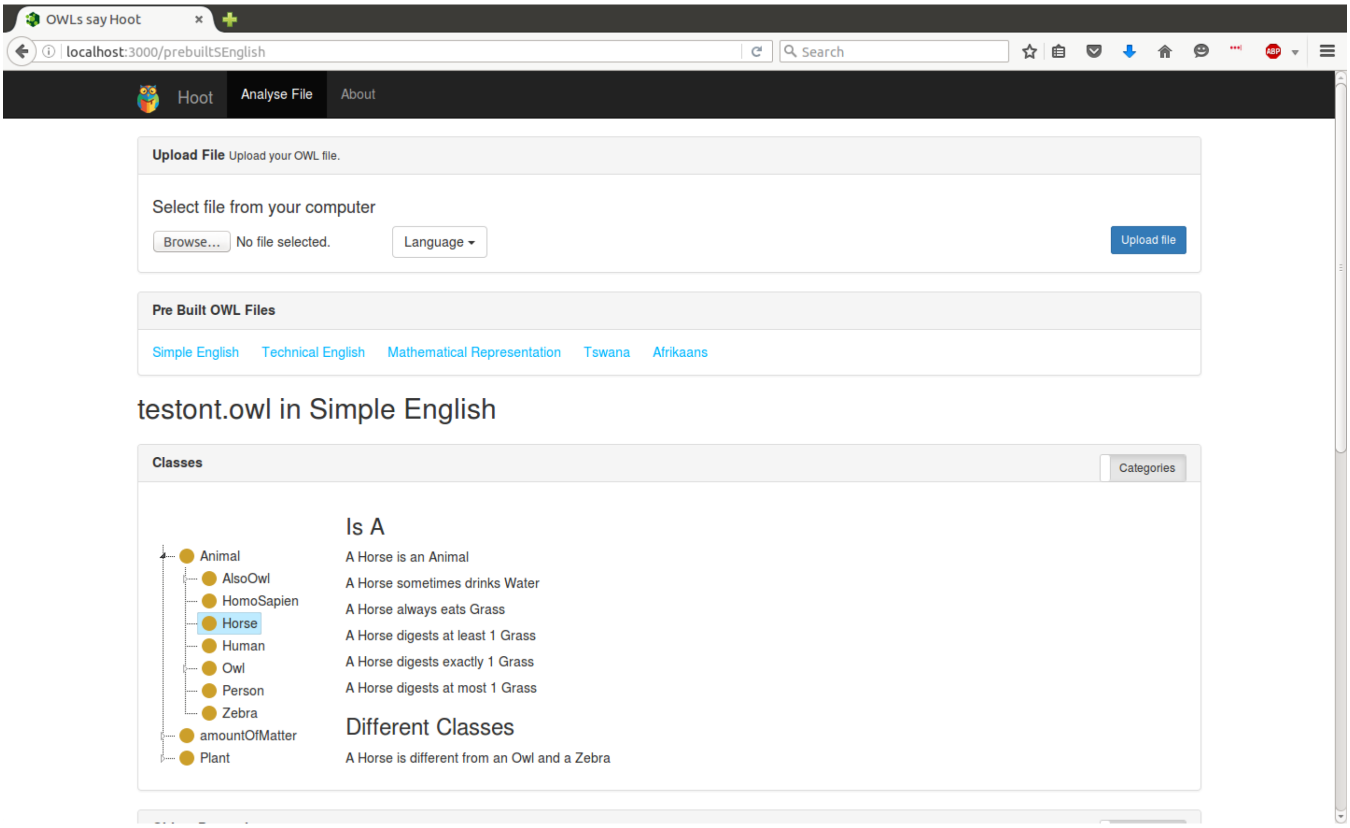


Figure 1: The main Hoot interface displaying the tree structure and some of the verbalisations in simple English

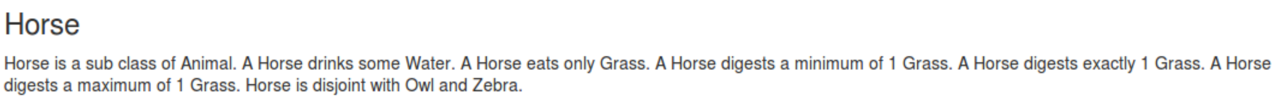


Figure 2: A paragraph verbalisation describing the Horse Class in technical English

### 3.3.1 English

The simple (see Figure 1) and technical English (see Figure 2) language modules build sentences with a largely template based approach. The inputs are processed: removing camel case, capitalising them and prepending the correct indefinite article. A limited amount of dictionary lookups are required to convert the form of verbs (i.e. *drinks* to *drunk*) and as such only the verbs available in the dictionary can be used with certain relations[3]. The input is then used in a standard template and returned to the parser for insertion into the tree.

### 3.3.2 Mathematical Representation

The mathematical representation also makes use of a templating approach to verbalising the ontology, simply dropping unprocessed inputs into a predefined template and returning the string to the parser. An example of the mathematical representation is in Figure 3

### 3.3.3 Afrikaans

The simple English representation was used as a blueprint for creating the Afrikaans representation. The template created for the simple English representation was translated and modified to represent the uploaded ontology in Afrikaans. The text inputs of the ontology are processed in the exact same manner and then translated. Figure 4 shows the verbalisation of a named entity in Afrikaans.

[3]For a full list for supported verbs see the *About* section of the website

### Characteristics

$\forall x\ \exists! y{:}drinks(x,y)$

$\forall y\ \exists! x{:}drinks(x,y)$

$\forall x,y\ drinks(x,y) \Rightarrow drinks(y,x)$

$\forall x,y\ drinks(x,y) \Rightarrow \neg drinks(y,x)$

$\forall x,y,z\ drinks(x,y) \land drinks(y,z) \Rightarrow drinks(x,z)$

$\forall x\ drinks(x,x)$

$\forall x\ \neg drinks(x,x)$

### Disjoint With

$drinks \sqcap (eats \sqcup isPartOf) \sqsubseteq \bot$

Figure 3: A mathematical representation of the drinks relation

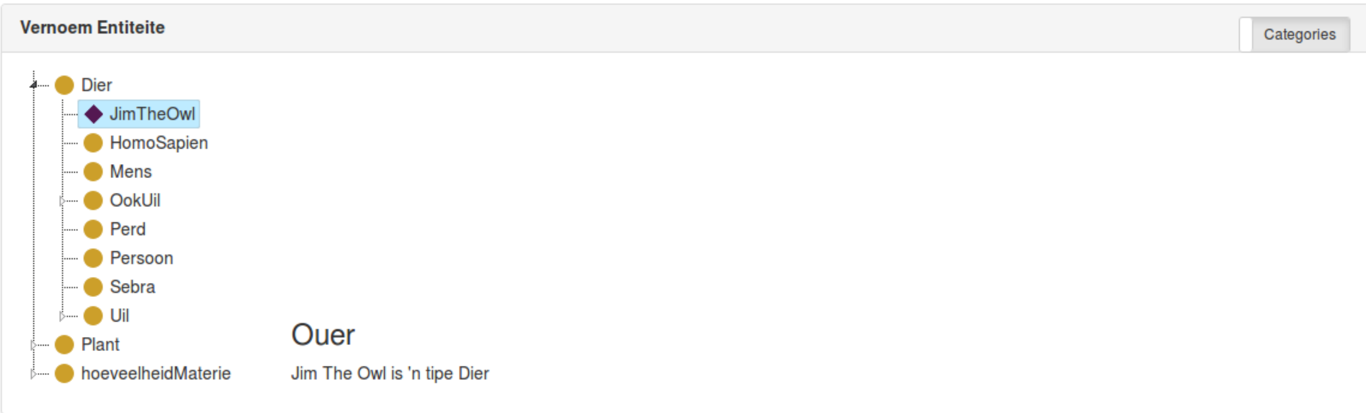


Figure 4: The Afrikaans verbalisation of a Named Entity

#### 3.3.4 Setswana

The Setswana language module was constructed by copying and adapting the simple english module in order to ensure uniformity and consistency. The adaptation involved translating the classes , individuals , cardinalities and relationships , along with modifying the sentence structure in certain sentences order to ensure it's grammatically correct and to contextualize it to the Setswana languagae. An example of such a change is with the sentence "What is the time". A Setswana translation of this with out a grammar correction would be "Ke eng nako" , which is not as grammatically correct as "Nako ke mang" , the preferred translation. Figure 5 shows the verbalisation of a relation in Setswana.

## 4 Conclusion

Ontology verbalisation is a complex task that requires a knowledge of not only the logical structure of the OWL Language but also the grammatical structure of the language of verbalisation. The templating approach used in Hoot, while easy to implement, often can produce grammatically incorrect outputs due to the wide range of possible inputs. The language modules can however be extended to account for more complex languages.

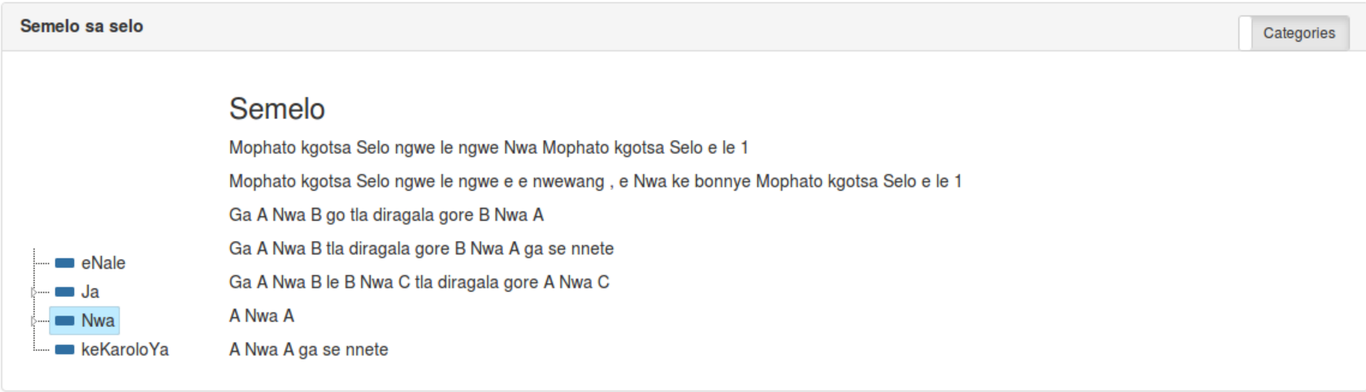


Figure 5: The Setswana verbalisation of a Relation

## 5 Future Research

Due to the modular nature of the language modules adding languages is simple. Languages that can be expressed through a simple templating structure are particularily easy to add, often by simply translating the simple English module. More complex languages can also be added as a language module as complex processing can be done within the module to produce the correct text string. The extension of "Hoot" to work with further languages could help to engage more users with the knowledge stored in Ontologies.

"Hoot" only handles a subset of the features of OWL 2 DL, as listed on the *About* page; but could be extended to handle all the features of the language.

## 6 Who Did What

Mitch did the interface, parser, Abstract, simple English, technical English, mathematical representation language modules and their respective writeups. Roscoe coded the Afrikaans language module, and wrote it's writeup and the introduction. Tumelo and Bayanda developed the Setswana language module and wrote its writeup. Bayanda completed the literature review and Tumelo wrote the Conclusion and Future Research.

## References


[1] N. Leavitt, "*Recommendation technology: will it boost e-commerce?*",Computer(Volume:39,Issue: 5 ),(2006)

[2] Georgi Kobilarov, Tom Scott, Yves Raimond, Silver Oliver, Chris Sizemore, Michael Smethurst, Christian Bizer, Robert Lee, "*Media Meets Semantic Web How the BBC Uses DBpedia and Linked Data to Make Connections*",(2009)

[3] C. Maria Keet, "*Lecture Notes Ontology Engineering*", Department of Computer ScienceUniversity of Cape Town, (2015)

[4] Ian Horrocks, Oliver Kutz, Ulrike Sattler "*The Even More Irresistible SROIQ*",Kr 6,(2006)

[5] Aldo Gangemi, "*Ontology design patterns for semantic web content.*", The Semantic WebISWC 200, (2005)

[6] Mariano Ferndndez, Asuncion Gomez-Perez, Natalia Juristo, "*METHONTOLOGY:From Ontological Art Towards Ontological Engineering*",AAAI Technical Report SS-97-06,(1999)

[7] Stevens, Robert, et al, "*Automating generation of textual class definitions from OWL to English*", Journal of Biomedical Semantics 2.2 (2011): 1.

[8] Bouayad-Agha, Nadjet, Gerard Casamayor, and Leo Wanner, "*Natural language generation in the context of the semantic web.*",Semantic Web 5.6 (2014): 493-513.

[9] R. Schwitter, K. Kaljurand, A. Cregan, C. Dolbear, G. Hart, "*A comparison of three controlled natural languages for OWL 1.1.*",Proc. of OWLED 2008 DC. Washington, DC, USA metropolitan area, on 1-2 April 2008.

[10] Keet, C.M., Khumalo, L, "*Basics for a grammar engine to verbalize logical theories in isiZulu.*",8th International Web Rule Symposium (RuleML14), A. Bikakis et al. (Eds.). Springer LNCS 8620, 216-225. August 18-20, 2014, Prague, Czech Republic

[11] Jarrar, Mustafa, C. Maria, and Keet Paolo Dongilli, "*Multilingual verbalization of ORM conceptual models and axiomatized ontologies.*",2006.

[12] Adam Wyner, Krasimir Angelov, Guntis Barzdins, Danica Damljanovic, Brian Davis, Norbert Fuchs, Stefan Hoefler, Ken Jones, Kaarel Kaljurand Tobias Kuhn, Martin Luts, Jonathan Pool, Mike Rosner, Rolf Schwitter, John SowaShow less , "*On Controlled Natural Languages: Properties and Prospects.*",2010.

[13] Jarrar, Mustafa, C. Maria, and Keet Paolo Dongilli, "*Multilingual verbalization of ORM conceptual models and axiomatized ontologies.*",2006.

[14] Fuchs, Norbert E., and Rolf Schwitter, "*Attempto controlled english (ace).*",arXiv preprint cmp-lg/9603003 (1996).

[15] Schwitter, Rolf, "*Controlled natural languages for knowledge representation.*",23rd International Conference on Computational Linguistics: Posters. Association for Computational Linguistics, 2010..