# Machine Learning based Vocabulary Management Tool Assessment for the Linked Open Data

#### Ahsan Morshed

Intelligent Sensing and Systems Laboratory, CSIRO, Hobart, Australia 7001

Ritaban Dutta
Intelligent Sensing and Systems Laboratory,
CSIRO, Hobart, Australia 7001

#### **ABSTRACT**

Reusing domain vocabularies in the context of developing the knowledge based Linked Open data system is the most important discipline on the web. Many editors are available for developing and managing the vocabularies or Ontologies. However, selecting the most relevant editor is very difficult since each vocabulary construction initiative requires its own budget, time, resources. In this paper a novel unsupervised machine learning based comparative assessment mechanism has been proposed for selecting the most relevant editor. Defined evaluation criterions were functionality, reusability, data storage, complexity, association, maintainability, resilience, reliability, robustness, learnability, availability, flexibility, and visibility. Principal component analysis (PCA) was applied on the feedback data set collected from a survey involving sixty users. Focus was to identify the least correlated features carrying the most independent information variance to optimize the tool selection process. An automatic evaluation method based on Bagging Decision Trees has been used to identify the most suitable editor. Three tools namely Vocbench, TopBraid EVN and Pool Party Thesaurus Manager have been evaluated. Decision tree based analysis recommended the Vocbench and the Pool Party Thesaurus Manager are the better performer than the TopBraid EVN tool with very similar recommendation scores.

# **General Terms**

Linked Open data system, Vocbench, TopBraid EVN, Pool Party Thesaurus Manager, Ontologies, Vocabulary.

#### Kevwords

Principal component analysis, Bagging Decision Trees, Feature, Clustering, Decision Making.

#### 1. INTRODUCTION

The Linked Data is a term that can be defined as the textual or graphical description of a conceptualization. It can be used to share and reuse the knowledge by using the Universal Resource Identifier (URI) and Resource description framework (RDF). In RDF, vocabularies, related semantics and relationships among concepts or terms of a particular domain have been used. In addition, vocabularies are very practical for explaining metadata terms and structuring domain knowledge in a structured and standard way. This type of standardization facilitates reuse and enables the applications to cooperate with one another more efficiently. Using Linked Data, developers can build more intelligent systems that can understand each other more thoroughly and generate more knowledge [1]. However, it is hard to maintain the vast amount of unstructured data such as PDF documents, video, audio etc for an organization. In order to solve this issue data are transformed in a way that machine can understand as well as data can be uploaded on the web for enabling semantic distribution. It is a door opener for connecting several pieces of information from the web cloud.

There is a need of centralized tool for the data engineers to manage all available information in one place. The centralized tool should contain the vocabulary management as well as Linked data publishing functionalities. Many organizations are currently spending significant amount of money for developing their in-house vocabulary management tools to gain all desired functionalities (i.e. FAO, UN is developing SKOS compatible VocBech since 2008). But the big challenge is that even after spending massive budget on new tool development, desired level of customer satisfaction and reliability is not achieved. It might be more efficient to use already existing tools rather than developing own tools. But it's very difficult to judge and evaluate a tool against the organizational requirements so the lack of confidence and confusion contribute to the overall operational costs.

This issue has been addressed by developing a machine learning based vocabulary management tool assessment framework. Idea behind this work was to develop a mechanism to provide recommendation about available tools. It is recommended to have a compatible tool of Linked data management before starting any development as otherwise development process could be very inefficient. The key novelty of this work includes application of unsupervised machine learning based approach for clustering the evaluation properties and to give a recommendation for the end users. Using PCA and Bagging Decision Trees a new evaluation method has been developed for this study for assessment of the three different Vocabulary Management Tools. Furthermore, our analyses are based on qualitative and quantitative measurements.

# 2. VOCABULARY MANAGEMENT TOOLS

Vocabulary such as AGROVOC is constructed by defining concepts, properties, and their relationships. Idea behind AGROVAC was to share knowledge by using similar words, meanings, and relationships among abstractions of a particular domain. The development of vocabulary has been shifted from Artificial Intelligence experts to domain experts. Currently vocabularies on the Internet vary from large classifications of web sites such as Yahoo [16] to grouping of products for sales on Amazon.com [3, 17]. Schema.org has been created by the giant Search Engine for categorize the information.WWW consortium [11] and the US defense have been creating vocabulary presenting languages such as SKOS, RDF and DAML [11] to facilitate the communication among various types of web sites on the Internet.

These standards aim to standardize the vocabularies so that domain experts can create the knowledge on the similar grounds.

The field of agriculture has created large number of common vocabularies, and semantics to create standardized application such as Tagging or Searching. Although the research in the

area of vocabulary is progressing rapidly, there is still a lack of research in selecting an appropriate vocabulary construction tool and managing tool and overall standardization. Some research has been done and reported by Stojanovic et al. [6], who propose a criterion to evaluate and assess ontology editors but mainly to discuss ontology evolution. Also, Denny [7] proposes criteria such as version, release date, source, graph view etc., which gives an overview of the editor but does not help the developer to rank quantitatively. Furthermore, Jürgen et al., [8] has evaluated the ontology editors based on both technological and ontology-related properties and Lambrix et al., [9] evaluate ontology-merging tools for bioinformatics. However, none of these works rank vocabulary editors quantitatively and qualitatively. As each editor has its advantages and shortcomings, having a readily available knowledge regarding the tools can save considerable amount of time and effort to vocabulary development as well as evolution process. In this paper, evaluation mechanism based on evaluation criterions has been proposed to help the vocabulary manager to identify a vocabulary construction tool best suited for their specific requirements. Editors such as Vocbench, TopBraid and Pool Party are evaluated based on quantitative and in qualitative measures to illustrate the use of evaluation criteria. These tools are well organized and acquired names within a short time.

#### 2.1. Vocbench

Vocbench is a web-based multilingual vocabulary management tool developed by FAO, Rome, Italy and hosted by research partner MIMOS Berhad, Malaysia. It is used to manage the AGROVOC thesaurus but it can be used for any kind of vocabularies, classification system and glossaries. It supports OWL and SKOS formats. It is linked open data compatible tool with 21 languages (see Figure 1).



Fig 1: VocBench representation.

# 2.2. TopBraid EVN (TopBraid Enterprise Vocabulary Net)

TopBraid is a web-based tool for simplified development and management of interconnected controlled vocabularies which



Fig 2: TopBraidEVN representation.

is developed by TopQuadrant, USA.

It supports business, stakeholders who need to collaborate on defining and linking enterprise vocabularies, taxonomies, thesauri and Ontologies used for information integration, customization and search. It is based on SKOS format and eclipse framework (see Figure 2).

# 2.3. Pool Party Thesaurus Manager (PPT)

Pool Party is a web base tool which is developed by Semantic Web Company, Austria that supports metadata management and linked data publishing.

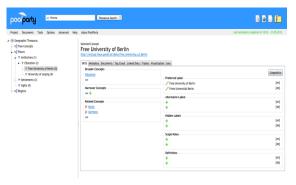


Fig 3: Pool Party representation.

PPT metadata management is based on W3C's Semantic web standard SKOS and can be combined with text mining and linked data technologies. It has functionalities such as data integration, semi-automatic annotation of document (see Figure 3).

#### 3. EVALUATION CRITERIA

Upon conducting research on the vocabulary construction tools, it was found out that the following criteria ( see Table 1) are essential for a vocabulary editor. This section provides simple descriptions of the criteria adapted for this study. The survey for this study was designed around these features to evaluate the tools. Table 1 represents the complete set of evaluation criteria.

## 3.1 RANKING TECHNIQUE

Aspects can be ranked on a Likert scale of 0-5 with ranks ranging from 0 as Very Poor, 5 as Very Good and so on. Based on the total rank of the sub aspects, an average rank for each main aspect can be provided by applying the Mean formula. "The arithmetic mean is what is commonly called the average: When the word "mean" is used without a modifier, it can be assumed that it refers to the arithmetic mean. The mean is the sum of all the scores divided by the number of scores. The formula in summation notation is:  $\mu = \Sigma X/N$  where  $\mu$  is the population mean and N is the number of scores" [10]. The average rank ( $\mu$ ) can be a decimal value and it was decided that decimal values would not be rounded since this might bias the comparison of the three tools.

#### 3.2 SCENARIOS AND TEST CASES

The following scenarios were used because they are easy and common for vocabulary construction, as each scenario's domain knowledge remains same from country to country [18]. For instance, a patient receiving treatment for sickness from a doctor is universal. A student enrolling in the university is also same from university to university. So a

model was developed that facilitate reuse of the domain knowledge and benefit the research community.

- A patient receives treatment for his/her sickness from a doctor.
- A scientist looks for past research papers for his/her research.

Table 1. Description of the criterions used for the evaluation survey.

CRITERIA	DEFINITION
Functionality (C1)	Degree to which the vocabulary tool permits the addition, modification, clone and deletion of concept, term, properties, relationships.
Reusability (C2)	Degree to which the existing vocabulary, concept, properties and data can be reused from a new editor version.
Data Storage (C3)	Degree to which, the vocabulary can be stored in the SKOS or RDF format. Degree to which the editor allows the SPARQL queries over the triple store.
Complexity (C4)	Degree to which the editor allows easy installation, run on various platforms, and provides user-friendly interface.
Association (C5)	Degree to which the editor is compatible with various industrial databases. Degree to which the editor allows one to add and match to the vocabulary. Degree to which the editor is having Linked Data compatible functionalities.
Maintainability (C6)	Whether the editor allows the addition of a new parent concept or child concept to an existing vocabulary. Whether the editor permits the modification of parent concept to child concept, or child concept to child concept. Whether the editor allows the deletion of existing parent concept, or existing child concept.
Resilience (C7)	Whether the editor allows undoing and redoing of the addition of a new parent concept or child concept. Whether the editor permits undoing and redoing of the modification of an existing parent concept or an existing child concept. Whether the editor allows undoing and redoing of the deletion of an existing parent concept, or an existing child concept.
Reliability (C8)	Degree to which the editor verifies and corrects the minor errors made during the vocabulary development.
Robustness (C9)	Degree to which the parent concept property values are transparent in all of child concepts. Degree to which the editor allows representation of various relations between classes to facilitate efficient query searches. Degree to

	which the editor allows declaration of user defined types. Degree to which the editor allows formatting of the form for data entry and data presentation purposes.
Learnability (C10)	Amount of knowledge required constructing a basic ontology. Usefulness of the help manual in the learning process. Amount of learning time required in creating a basic ontology.
Availability (C11)	Degree to which, the editor is accessible online. Amount of financial funds required in the purchase of the editor.
Flexibility (C12)	Degree to which the editor allows easy browsing of the vocabulary model. Degree to which the editor allows efficient queries. Degree to which the editor provides developer with easy to use design tools. Degree to which the editor provides the web services.
Configurability (C13)	Degree to which the editor provides easy configuration.
Multilinguality (C14)	Degree to which the editor allows multiple languages.
Visibility (C15)	Degree to which the editor permits the visual representation of vocabulary while using an external program as plug-in. Degree to which the editor has a built-in capability to draw visual model of the vocabulary. Degree to which the editor permits the conversion of RDF and HTML pages.
Authentication (C16)	Degree to which the editor has higher security.

# 3.3 SURVEY AND DATA

Sixty users participated in a conducted survey for this assessment study. Participating users were asked to provide scores against the defined criterions as well as general feedback regarding overall comparative user experiences. The evaluation process initiated with the construction of two models by applying the two scenarios discussed earlier. VocBench was utilized since it was readily available for the installation and it was provided with a very useful user manual. Its website illustrated many useful examples and this in turn facilitated the learning and construction process. PPT was made available from Semantic web company, Austria. The company sent an evaluation right for a short period. The creation of model using the three editors lasted for approximately 120 hours. Each editor was installed on three computers and later examined for learning purposes. Individual models were created and populated with test data accompanied with the scenarios described before. Finally, inference, import and export using different ontology

languages, as well as browsing capabilities were analyzed to ensure that all criteria aspect be answered with great care.

# 4. PRINCIPAL COMPONENT ANALYSIS

PCA is a linear supervised multivariate statistical method, based on the Karhunen-Lowve expansion, used in classification models to produce classification results for enose pattern recognition techniques. The method consists of expressing the response vectors  $\mathbf{r}_{ij}$  in terms of linear combinations of orthogonal vectors, and is sometimes referred to as vector decomposition. Each orthogonal vector, principal component (PC), accounts for a certain amount of variance in the data with a decreasing degree of importance. The scalar product of the orthogonal vectors with the response vector gives the value of the  $\mathbf{p}^{th}$  principal components:

$$X_p = \alpha_{1p}r_{1j} + \alpha_{2p}r_{2j} + ... + \alpha_{ip}r_{ij} + ... + \alpha_{np}r_{nj}$$

The variance of each PC score,  $X_p$ , is maximized under the constraint that the sum of the coefficients of the orthogonal vectors or eigenvectors  $\alpha_p = (\alpha_{1p}... \ \alpha_{jp}... \ \alpha_{np})$  is set to unity, and the vectors are uncorrelated. PCA is in essence a data reduction technique for correlated data, such that a n-dimensional problem can be described by a two or three dimensional plot. The overlapping clusters have significant effect on over all classification and it also increases the rate of misclassifications.

The PCA method consists of expressing the response vectors in terms of a linear combination of orthogonal vectors. Each orthogonal (principal) vector accounts for a certain amount of variance in the data, with a decreasing degree of importance. PCA based data clustering was used to investigate how the response vectors from the different sensors are clustered into multisensor space. In this scenario

different criteria were considered instead of sensors. The objective of this analysis was to establish a list of least correlated attributes which are contributing towards most data variance. This clustering approach was used on normalized data. Individual feature or criteria data column was normalized based on the maximum and minimum values for the particular column to avoid any unnecessary data value imbalance. Three principal components were kept, which accounted for 100% of the variance in dataset (PC no 1, PC no 2 and PC no 3 accounted for 90.67, 8.85 and 0.41% of the variance, respectively). Three categories appear to be evident representing three vocabulary tools. It was very clear that the first principal component captured most of the information variance from the dataset. The PCA method was used to help us to get a better understanding of the nature of our data. For information all the 'load values' for first three PCs are included here. Figure 4 shows the clustering results acquired from PCA method. It was evident from the PCA cluster plot that VocBench and PPT had great deal of similarity in terms of cluster position, whereas TopBraidEVN had different cluster position [20-23]. Based on PCA method all criterion were sorted according to their cross correlation. The loadings associated with the four least correlated criteria, namely, Data Storage (C3), Complexity (C4), Robustness (C9) and Multilinguality (C14) were (-0.5066 0.0754 0.7806),  $(-0.5135\ 0.0823\ 0.7808)$ ,  $(-0.2828\ 0.5490\ -0.5490)$  and (-0.4754 -0.8164 -0.2987) respectively. Next four criteria from the sorted list were Authentication (C16), Functionality (C1), Resilience (C7) and Learnability (C10). Although these later four criteria did not carry any information variance the loadings associated with the next four criteria were (-0.5613 0.1810 - 0.6947),  $(-0.5478 \ 0.1608 - 0.0663)$ ,  $(0.4828 \ 0.5490)$ -0.5490) and (-0.1754 -0.8164 -0.2987) respectively.

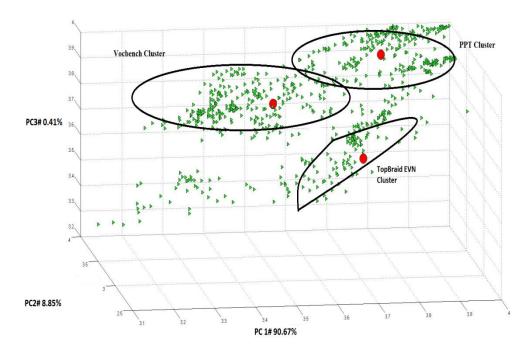


Fig 4: PCA clustering based on survey data. Red dots represent the cluster Centroid where as green points represent projected scores calculated from the criteria based survey data.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16
VocBench	2	3	1	2	2	3	4	5	3	4	3	1	5	2	5	4
TopBraidEVN	4	3	4	3	4	4	4	4	4	2	3	2	4	3	4	3
PPT	5	3	4	4	5	5	4	4	5	4	5	2	3	5	4	4

Table 2. Example of survey data collected from one user and used in the evaluation work.

The correlation coefficients using the standard MATLAB function 'corrcoef' [16] were calculated and these correlation results were also indicative that Data Storage (C3), Complexity (C4), Robustness (C9) and Multilinguality (C14) were less correlated criteria which could be used for overall tool evaluation. Least correlated features or criterions from PCA method were used in the bagged decision trees based evaluation process. Since a reasonable correlation exists between tools, it can be assumed that the blind categories established by PCA are consistent with the characteristics of the three different tools. However, the inter category boundaries are complex in shape and some patterns that belong to different categories apparently appear to have very similar scores. This effect may have been enhanced by our experimental approach, which has focused, in this initial work, on practicality of vocabulary tool usage. In the next stage recommendation from this PCA method took pivotal role to guide the design and implement the Bagging Decision trees to determine the preferable vocabulary management tool.

## 5. BAGGING DECISION TREES

An automatic evaluation method based on Bagging Decision Trees has been used to identify the most suitable editor. The Bagging Decision Trees method have been used widely for credit rating and financial decision making processes where final decision depends on N number of different criteria. Evaluation of the vocabulary tools could be modeled as a similar problem as effectiveness and usability of a tool could only be judged if only all possible associated attributes used were considered to describe a tool. User survey based data is unique in a sense that it reflects quantitative and qualitative measurements about a tool. Bagging Decision Trees algorithm was ideal in this context as it allowed to incorporate all possible criteria into our into our problem space. PCA based analysis provided a valuable estimation regarding whole data set and was able to identified top four least correlated criteria. These selected criteria were used as weighted priority in the modelling of the Bagging Decision Trees although all of the criteria were used. aggregating (bagging) is a machine learning ensemble metaalgorithm to improve machine learning of statistical classification and regression models in terms of stability and classification accuracy. It also reduces variance and helps to avoid over fitting. Although it is usually applied to decision tree models, it can be used with any type of model. Bagging is a special case of the model averaging approach. Given a standard training set D of size n, bagging generates m new training sets, each of size n' < n, by sampling examples from D uniformly and with replacement. By sampling with replacement, it is likely that some examples will be repeated in each. If n'=n, then for large n the set is expected to have the fraction (1 - 1/e) ( $\approx 63.2\%$ ) of the unique examples of D, the rest being duplicates. This kind of sample is known as a bootstrap sample. The m models are fitted using the above m

bootstrap samples and combined by averaging the output (for regression) or voting (for classification) [24-25].

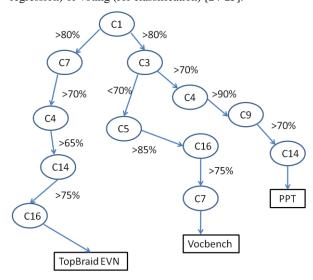


Fig 5: Bagging decision tree representation for evaluation of the vocabulary management tools.

In this paper, the variable importance measure which was used was the PCA based ranking mechanism which estimated the relative impact of a feature by measuring how much the predictive accuracy of the classifier deteriorates when this feature's values are randomly permuted. The idea is that when the feature in question adds little to the predictive power of the classifier, using altered (in this case permuted) values should not impact the classification results. Relevant information, on the other hand, cannot be randomly swapped without degrading the predictions. Now, if two highly correlated features are equally important, they will both rank high in this analysis. In that case, keeping one of these features should suffice for accurate classifications, but one would not know that from the ranking results alone. One would have to check the correlations separately, or use an expert's opinion. "Bagging," in this context, stands for "bootstrap aggregation." The methodology consists in generating a number of sub-samples, or bootstrap replicas, from the data set. These sub-samples are randomly generated, sampling with replacement from the list of customers in the data set. For each replica, a decision tree is grown. Each decision tree is a trained classifier on its own, and could be used in isolation to classify new combination of criteria. The predictions of two trees grown from two different bootstrap replicas may be different, though [26-27].

Process was started with a small number of trees, 35 only, because main focus was to compare the initial trend in the classification error for different leaf sizes. For reproducibility and fair comparisons, process was reinitialized

with the random number generator, which was used to sample with replacement from the data, each time a classifier was built. The data were not segmented into training and test subsets. This is done internally; it is implicit in the sampling procedure that underlies the method. At each iteration the bootstrap replica is the training set, and any customers left out ("out-of-bag") are used as test points to estimate the out-ofclassification errors. MATLAB's implementation of bagged decision trees was used. Regression mode was used so that the end prediction would be a value between 0 and 1 representing the probability of the binary class. The number of features for each tree to sample was left at its default, for regression mode, of 40% the number of features selected from the PC analysis. The two parameters that were modified were MinLeaf and NumTrees. MinLeaf is the minimum number of observations needed per tree leaf. This is recommended to be set at 1 for classification and 5 for regression; however, the optimal values for this parameter were often between 19 and 75 based on testing with a small validation set. The NumTrees parameter is the number of random decision trees trained. The rule of thumb is to use a value of 40 or greater. Values between 40 and 700 were tried. MinLeaf was searched from 1 to 100 in increments of 1 and NumTrees was set at 60 for this parameter search. NumTrees did not appear to affect the optimal MinLeaf value chosen. It is possible that there is a different optimal MinLeaf value depending on NumTrees. Each tree trained by Random Forests resamples from the training data, with replacement. The size of the resampled training data can be set, however, this was left at its default value which was to create a resampled set the same size as the original training set.

The ensemble aggregates the predictions of all the decision trees that are grown for all the bootstrap replicas. If the majority of the trees predict one particular class for a specific tool, it is reasonable to consider that prediction to be more robust than the prediction of any single tree alone. Moreover, if a different class is predicted by a smaller set of trees, that information is useful, too. In fact, the proportion of trees that predict different classes is the basis for the classification scores that are reported by the ensemble when classifying new data. Bagging Decision Trees were designed to produce an average ranking score for each of the vocabulary tool. The final score based ranking mechanism was designed in a way that it considered variable sizes of all beans (bean size was based on number of trees that predicted same class) to make the final decision. It was discovered that VocBench, TopBraidEVN and PPT have a similar functionality with an average rank of 4.3, 3.33, and 4 respectively. Figure 5 shows an example scenario of the bagging decision trees used for this study. Qualitative results from this framework are discussed in the next section.

#### 6. COMPARATIVE RESULTS

In this section describes the quantitative evaluation with respect to each criteria aspect for the three tools, namely VocBench, PPT and TopBraidEVN. Decision tree based analysis recommended the Vocbench and PPT were the better performer than the TopBraidEVN. This conclusion was achieved based on the machine learning analysis on the survey data collected for this study. Average ranking results were very similar with PCA based projected findings.

Functionality: All three editors provided similar functionality. One can add, change and delete concepts, slots, and relations. However, VocBench does not allow multiple vocabularies at a time; it is not possible to SPARQL query.

Reusability: TopBraidEVN and PPT has similar reusability features. One can reuse vocabulary, concepts, and properties. However, there is minor difference between the two, whereas VocBench does not support the import of data to populate.

Data Storage: TopBraidEVN and PPT allow one to store the vocabulary in the RDF format in the triplestore. In contrast, VocBench also supports OWL or SKOS format but it does not have a large triple.

Complexity: The installation process of TopBraidEVN and VocBench is quite simple. One can download the software from the Internet and install it in the matter of few minutes. VocBench requires updating the functionality and asking a developer for installation. There is no clear guideline for installation process.

Association: PPT supports for data integration on the run time. On the other hand, VocBench and TopBraid do not support the direct data integration process.

Maintainability: TopBraidEVN and PPT are very friendly when it comes to improving or maintain the vocabulary. One can add, change and delete concept, properties with few mouse clicks. VocBench on the contrary is less-user friendly when one wants to make changes to the vocabulary. It allows the addition of parent concept and child concept but it does not allow one to change the parent conceptor child concept without first deleting the given super or child concept.

Resilience: PPT fully supports the undo and redo actions taken by the user. TopBraid allows only undo the actions taken by the user and VocBench does not allow one to either undo or redo actions.

Reliability: VocBench is reliable editor since it comes from FAO of UN. However, three editors do not have the verification mechanism, which enables it to identify and correct errors.

Robustness: PPT and TopBraid are very similar when it comes to robustness. They both support the transparency principle, variety of relationship building, and form formatting capability. However, they do not enable the user to define their own attribute types. In Contrast, VocBench is very different from it counter parts when it comes to robustness. Transparency principle is questionable since one cannot upload the vocabulary automatically unless the engagement of VocBench developers.

Learnability: VocBench and PPT two easy to use tools. Vocabulary editors can handle easily by learning in one or two days whereas TopBraidEVN is based on eclipse that makes editors clumsy

Availability: VocBench can be obtained free of cost. PPT and TopBraidEVN can be used for 30 days without any cost, but after the expiry of the trail version, there is a need to pay 1600 Euros to purchase the full fledges PPT. TopBraidEVN on the

other hand is more expensive. It comes with a high price tag of 3450 dollars.

Flexibility: VocBench does not have model browsing capability but it offers one to do simple query searches, and it has user-friendly design tools. TopBraid and PPT in contrast have impressive model browsing capability.

Configurability: VocBench, PPT and TopBraidEVN are not configurable by a expert editor. It is necessary to be asked their developers for configuring the vocabulary.

Multilinguality: VocBench supports 21 languages whereas PPT and TopBraidEVN support fewer languages. But, PPT and TopBraidEVN support automatic translation of concept labels and VocBench does not support it.

Visibility: VocBench does not offer the visual model for Linked Open Data. There is a need to draw the visual model by using an external tool as a plug-in. PPT support the Linked Open Data visualization. Furthermore, it also supports auto tagging for exploring more automatic information from the Linked Open Data cloud. On the other hand, VocBench and TopBraid do not have this kind of functionalities.

Authentication: VocBench has more security functionalities than other two editors. However, users' login and password are common among these three tools.

Qualitative and quantitative analysis based on PCA and bagging decision tree recommended that the Vocbench and the Pool Party Thesaurus Manager are the better performer than the TopBraid EVN tool with very similar recommendation scores.

# 7. CONCLUSION

In this paper, vocabulary construction tools such as VocBench, PPT, and TopBraidEVN were evaluated. Upon evaluating the above-mentioned tools, the following conclusion was reached: VocBench can be used to construct medium size vocabulary and PPT, with its full functionality, can be used to develop medium to large-sized vocabulary and publish them as Linked Open data. TopBraidEVN on the other hand is average for enterprise architecture modeling. The proposed framework is very flexible to cope with the specific needs of vocabulary developers and further research can be conducted to enhance the evaluation criteria and ranking technique. Lastly, for future studies, more tools will be evaluated.

## 8. ACKNOWLEDGMENTS

The authors would also like to thanks the Intelligent Sensing and Systems Laboratory, CSIRO, Hobart and the Tasmanian node of the Australian Centre for Broadband Innovation are assisted by a grant from the Tasmanian Government which is administered by the Tasmanian Department of Economic Development, Tourism and the Arts for providing capability funds to continue this research work. The work described in this paper could have not been possible without the

collaboration of a number of people. The authors wish to thank Yves Jacques, Johannes Keizer, Gudrun Johannsen, Sachit Rajbhadari, and Andreas Blumauer.

## 9. REFERENCES

- [1] T. R. Gruber, A translation approach to portable ontologies. Knowledge Acquisition, 5(1993), 199-220.
- [2] Y.Sure, Onto-To-Knowledge-Ontology based Knowledge Management Tools and their Application, German Journal Kuentliche Intelligenz, Special Issue on Knowledge Management (01/02), 2002.
- [3] Stanford medical informatics Home page, at URL: http://www.smi.stanford.edu/
- [4] Ontoprise®GmbH (1999), Onto Edit tutorial Homepage [Online], at URL: http://www.ontoprise.de/documents/tutorial\_ontoedit.pdf
- [5] Computus (1985) Home page [Online], at URL: www.computus.com
- [6] L.Stojanovic and B.Motik, Ontology evolution within ontology editor, 13th International Conference on Knowledge Engineering and Knowledge Management EKAW02, Sigüenza (Spain)
- [7] M.Denny, Ontology Building: A survey of Editing Tools, Nov 6, 2002.
- [8] Jürgen, York Sure, White paper: Evaluation of ontology-based tool, 13th International Conference on Knowledge Engineering and Knowledge Management EKAW02, Sigüenza (Spain).
- [9] P.Lambrix and A. Edberg, Evaluation of ontology merging tools in bioinformatics, Proceedings of the Pacific Symposium on Biocomputing PSB03, 8:589-600, Kauai, Hawaii, USA, 2003.
- [10] David Lane's Homepage. http://davidmlane.com/hyperstat/A15885.html
- [11] World Wide Web. http://www.w3.org/
- [12] Oracle. http://www.oracle.com/index.html
- [14] M.A.Musen (2000), the Evolution of Protégé: An Environment for Knowledge-Based Systems Development, Retrieved October 10, 2004 at URL: http://www.smi.stanford.edu/pubs/SMI\_Reports/SMI-2002-0943.pdf
- [15] N.Noy, R.W. Fergerson, M. A.Musen (2000), The knowledge model of Protégé-2000: combining interoperability and flexibility, Retrieved October 11, 2004 atURL:http://wwwsmi.stanford.edu/pubs/SMI\_Reports/S MI-2000-0830.pdf website
- [16] Search Engine Yahoo: www.yahoo.com
- [17] Amazon. www.amazon.com
- [18] A.Morshed and R.Singh, Master thesis (No: 05-x-223) Evaluation and Ranking of Ontology Construction Tools, Royal Institute of Technology, 13th Jan, 2005
- [19] Asunción Gómez-Pérez (1999), Evaluation of Taxonomic Knowledge in Ontologies and Knowledge Bases, October 16 Twelfth Workshop on Knowledge

- Acquisition, Modeling and Management, Voyager Inn, Banff, Alberta, Canada.
- [20] Jackson, J. E., A User's Guide to Principal Components, John Wiley and Sons, 1991, p. 592.
- [21] Jolliffe, I. T., Principal Component Analysis, 2nd edition, Springer, 2002.
- [22] Krzanowski, W. J. Principles of Multivariate Analysis: A User's Perspective. New York: Oxford University Press, 1988.
- [23] Seber, G. A. F., Multivariate Observations, Wiley, 1984
- [24] Altman, E., "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy," Journal of Finance, Vol. 23, No. 4, (Sep., 1968), pp. 589-609.
- [25] Basel Committee on Banking Supervision, "Studies on the Validation of Internal Rating Systems," Bank for

- International Settlements (BIS), Working Papers No. 14, revised version, May 2005. Available at: http://www.bis.org/publ/bcbs\_wp14.htm.
- [26] Basel Committee on Banking Supervision, "International Convergence of Capital Measurement and Capital Standards: A Revised Framework," Bank for International Settlements (BIS), comprehensive version, June 2006. Available at: http://www.bis.org/publ/bcbsca.htm.
- [27] Loeffler, G., and P. N. Posch, Credit Risk Modeling Using Excel and VBA, West Sussex, England: Wiley Finance, 2007.
- [28] Merton, R., "On the Pricing of Corporate Debt: The Risk Structure of Interest Rates," Journal of Finance, Vol. 29, No. 2, (May, 1974), pp. 449-70.