EJASA:DSS – Electronic Journal of Applied Statistical Analysis
http://siba-ese.unisalento.it/index.php/ejasa_dss/index
e-ISSN: 2037-3627
DOI: 10.1285/i2037-3627v5n1p42

Ontological analysis for dynamic data model exploration

By Hobbs, Luca, Fatima

Published: 29 December 2014

Electronic Journal of Applied Statistical Analysis
Vol. 05, Issue 01, December 2014, 42-56
DOI: 10.1285/i2037-3627v5n1p42

Ontological analysis for dynamic data model exploration

Mike Hobbs*[a], Cristina Luca[a], Arooj Fatima[a], and Mark Warnes[b]

[a]*Anglia Ruskin University, Faculty of Science ann Technology, Cambridge, UK*
[b]*Anglia Ruskin University, Department of Education, Cambridge UK*

Published: 29 December 2014

Increasing access to data and computational resources allows use to use more expressive approaches to data analysis. We propose using established statistical metrics to assist the automatic analysis of free text transcripts. The meaningful concepts from a domain and their axiomatic relationships can be captured in an ontology. This provides an aggregate model which describes the domain. However, the fine detail from individual elements and their characteristics are subsumed by the whole. Keeping multiple 'micro models' of the data, along with meta information allows a range of different view points. This can be applied to free text documents that within a domain where significant information is carried by one or a few instances such as in the analysis of interview transcripts. This paper presents a framework that utilises ontological tools to create domain models in a way that it allows for a distributed and parallel implementation necessary for big data analysis.

keywords: Ontology, model, classifier, concept mapping.

1 Introduction

The aim of this paper is to assess the potential for ontologies and associated knowledge based tools to explore the structure and semantics of models that are used to describe qualitative text based research. It is common in many areas of social science research to use a mix of qualitative and quantitative methodologies to explore in detail a few instances that exhibit interesting behaviour. In the analysis of a biographic narrative a single interview can provide significant findings once it has been transcribed and coded

*Corresponding author: mike.hobbs@anglia.ac.uk

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

(typically by hand) (Wenograf, 2001; Holley, 2008). During the investigation the researcher will have gained considerable knowledge of the domain but might benefit from tools that allow a wider collection of interviews and free text resources to be processed and explored.

The enabling technology of Big Data allows large and complex sets of data to be analysed at high volumes, in a variety of data types and at high velocity. The increase in capacity also enables us to consider analytical approaches where size and complexity previously imposed limits on the depth and expressiveness of a model. Lazer et al. (2014) describe some of the pitfalls when relying on data volume without taking sufficient account of existing statistical techniques but also show how large data can be used for previously unsupportable analysis, such as fine grain location data, to enhance the predictive power of the Google Flu Trends model. The ultimate goal of this type of data analysis is to be able to utilise the information from a wide range of internet sources in order to assist exploration, description and predictions of interesting social phenomenon.

This paper outlines a framework of ontological development and shows how statistical measures and machine learning techniques can be used to provide dynamically controlled views of the underlying data and model. We review existing ontology evaluation, validation and quality metrics and present a meta model that uses these inputs to give a description of the ontology evolution.

Our particular interest is with the analysis of Higher Education student satisfaction from surveys, interviews and free text comment from blogs and twitter. Part of our motivation comes from previous work undertaken at Anglia Ruskin University, where over 800 documents were analysed for the concepts of quality and excellence in Higher Education. By inputting definitions into the qualitative data analysis tool Nvivo (2014), patterns of text were matched from a set of documents. This gave a descriptive model which categorised papers with relevant concepts and the frequency of key supporting phrases. The aim for this research is build on this experience and produce a tool that will assist a researcher to build richer models with a greater degree of automation.

2 Background

2.1 Ontologies

The philosophical definition of 'Ontology' is the study concerned with the formal description of the structure of reality. In computer science and knowledge engineering there are different definitions but one commonly agreed on is given by Gruber (1993) 'An explicit formal specification (i.e. description) of a conceptualization (i.e. the concepts, terms and relationships within a domain).' These can be represented in different forms but are typically constructed as a taxonomic hierarchical structure of classes linked by relationships. Guarino (1998) distinguishes between different types of ontology where 'upper-level' is used to describe general, fundamental concepts, which are more akin to the original philosophical view of ontology, and 'domain' ontologies which are used to describe the more specialised context that might be found in a particular area of discourse or research.

In this paper we are concerned with domain ontologies and their ability to express a model to represent significant features and their relationships. As stated by Noy and McGuiness (2001) and repeated in Cimiano et al. (2010), an ontology can be used to communicate the information, structure and semantics of a knowledge domain. The key steps in making use of an ontology are as follows:

- determine the scope and purpose of the ontology
- construct, select or adapt a suitable ontology
- populate the ontology with instances (creating the knowledge base)
- generate applications that use the knowledge base frequently rule based classification.

They can be used and re-used from libraries of pre-existing ontologies or they can be built by manual analysis and design, automated generation, or a combination of the two approaches. Ontologies are also subject to change over time to accommodate new concepts from the domain it describes, or changes in the way these concepts are viewed. A common practice is to start with an existing version of an ontology and adapt it to suit a particular view of the domain or its conceptualization.

The meaningful concepts from a domain and their axiomatic relationships can be captured in an ontology which also provides a lens through which other examples can be examined. Generic ontologies can be used for free text analysis an extensive review is given by Kalfoglou and Schorlemmer (2003). The typical goal of the research process is to create a single ontology validated by the examples from the domain. This provides an aggregate model which can describe and summarise the domain. We propose a finer grained approach that creates separate ontologies to summarise single, or small groups of texts, that can then be combined in different ways to highlight different aspects of the domain.

2.2 Ontology quality

There are many ways to evaluate the quality of an Ontology, one overview separates them into two approaches which look at the quality of their representation alone, or include an evaluation of the associated knowledge base of instances- referred to as content evaluation or technology evaluation. Tartir et al. (2010) developed OntoQA which uses schema metrics and the following knowledge base measures:

Class richness —how well instances are distributed across classes measured by the percentage of non empty classes divided by the total number of classes.

Class connectivity —the number of relationships instances of the class have with instances of other classes.

Class importance —the number of instances found associated with a class along with the classes in its inheritance sub-tree.

Cohesion —the number of connections that instances have with each other for a class.

Relationship richness —a measure of the utilisation of the given classes by the knowledge base, measured by comparing the relationships defined in the schema with those used by the instances for a class.

Hlomani and Stacey (2014) propose a categorisation based in part on Burton-Jones et al. (2005) metric for ontology auditing which uses the concept of Ontology Correctness, evaluated by measures of its performance in terms of accuracy, completeness, conciseness and consistency and Ontology Quality evaluated by measures of its structure such as coupling, cohesion, size and clarity. They also point out that there will be an element of subjectivity in the choice and weighting of the many possible criteria that can be used to evaluate an ontology. Within a particular domain a researcher will try to apply the evaluation criteria most appropriate to the application, so as to capture the most important aspects for that investigation.

In this paper we are concerned with how Ontology correctness can be used as a tool to gather information about a particular knowledge domain. If an ontology correctly describes a domain then there will be a strong correspondence between the concepts described by the ontology and individual examples drawn from the underlying knowledge base.

2.3 Text and Sentiment Analysis

The increasing use of feedback mechanisms and other Web 2.0 user generated content has created a large, unstructured but potentially valuable source of information representing the opinions of users, consumers, patients, students, travellers, holiday makers, diners, etc (Pang and Lee, 2013). A site such as TripAdvisor operates an explicit star rating system but there are many other sources of data that could be useful to the manufacturer, retailer or service provider that do not provide their own degree of satisfaction. Kontopoulos et al. (2013) use an ontology to analyse the sentiment expressed in twitter posts.

Sentiment analysis of text can highlight those concepts that are associated with positive or negative sentiment e.g. the statements 'Tutor is good', 'Tutor is bad' both reinforce the importance of 'tutor' but one has a positive connotation and one negative. Although it is rare to have such a clear classification it is possible to attain an estimation of sentiment for words and phrases extracted from a text as well as a overall view as to the sentiment that the text might be expressing.

Turney (2002) uses an unsupervised learning algorithm to classify sources as 'recommended' or 'not-recommended'. Descriptive phrases are identified, extracted and compared to positive and negative reference words using Point-wise Mutual Information and Information Retrieval (PMI-IR). This was used by Church and Hanks (1990) as part of their work on calculating the 'association ratio' to provide an estimate of the association strength between two words based on the probability of occurring together moderated by the probability of their occurrence in the text i.e.

$$I(x, y) = \log_2 \frac{P(x,y)}{P(x)P(y)}$$

Dumais (2004) describes the development and use of 'Latent Semantic Analysis'

(LSA). This is an unsupervised learning approach that extracts the relations between words from large collections of texts. It builds a document / term frequency matrix which is transformed to take account of the overall occurrence of a term in the collection (using a measure related entropy) and dimensionally reduced in a process similar to a principle components analysis. This measure has been used widely and could form the basis of a scoring system to represent the significance of terms for a particular domain.

Sentinet WordNet 3.0 (Baccianella et al., 2010) builds on previous work to provide a system for opinion mining or sentiment classification in natural texts. It uses WordNet (2010) to generate a set of synonyms which it analysis to give a numerical value indicating the Objective, Positive or Negative perception of the word.

SenticNet (2014) is a system that uses a 30,000 multi-word expressions to form a 'bag of concepts' that can be used to find the polarity (positive / negative) of a phrase or sentence(Poria et al 2014) providing an online demonstration and programmable API. '*I found the tutor to be very helpful*' ⇒ Sentence Polarity - **positive**

Extracting meaning or sentiment from natural language and free text is a large and well established research area. Our framework provides an approach that allows us to utilise existing tools that facilitate the processing for our descriptive model creation.

3 Ontological construction framework

This section describes the key features and mechanisms of our proposed framework, summarised in Figure 1.

3.1 Text Pre-processing and extraction of terms

Source text(s) are analysed and processed to create a sets of 'terms' for each text, including meta data about the source. Each term is tagged with associated keywords, and additional information such as sentiment or pre-classified label. Terms can be weighted by the 'Term Frequency Inverse Document Frequency' (TF-IDF) that is commonly used (Ma et al., 2012) in information retrieval to provide a measure of importance to a term, where low frequency terms are regarded as more significant than high frequency terms.

A tool like SentiNet is used to automatically extract the list of terms from the free text and assign sentiment values (positive, negative, neutral) that can be used to classify concepts and build ontologies to enable researchers to identify differences in sentiment in the domain.

3.2 The upper-level ontology

For many domains there exist comprehensive upper-level ontologies that provide the fundamental concepts and their relationships. Libraries of ontologies, such as the Protégé Library (Protege, 2014), provide examples that cover a wide range of topics and can provide information on the underlying structure of concepts within a research domain.

In our target application area of Higher Education student satisfaction there are a number of suitable ontologies, reviewed by Zemmouchi and Ghomari (2013) who present

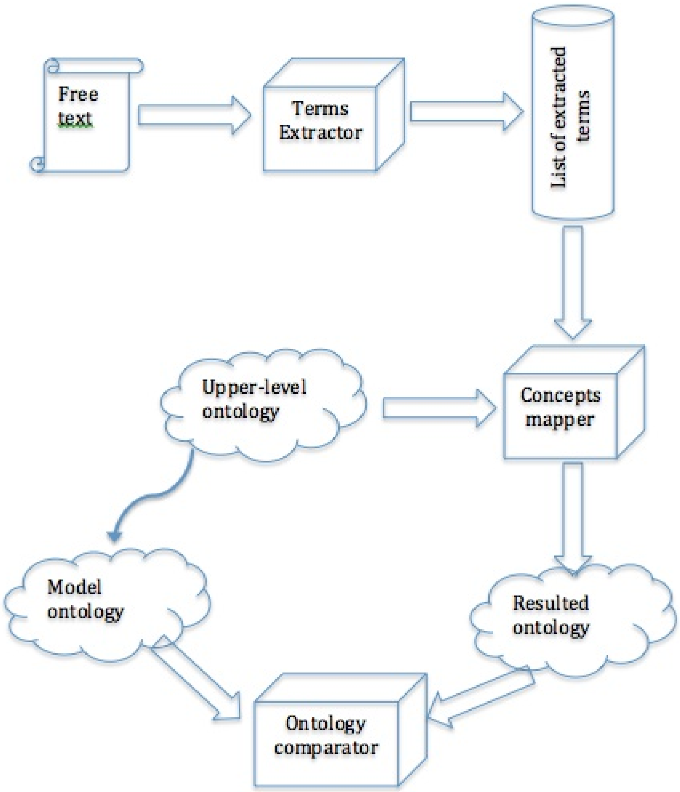

Figure 1: Ontological construction framework

their 'Higher Education Reference Ontology' (HERO). Figure 2 shows the upper-level class hierarchy presented by the Protégé ontology editor which also facilitates the process of creating, editing and merging ontologies.

When modelling a domain a researcher will utilise implicit knowledge about the wider world to describe the salient features relevant to their investigation. An automated process can only summarise the evidence presented to it, so not all the connections between concepts would necessarily be expressed within the text, e.g. 'driving' and 'parking' involve the more general concepts of vehicle, travel etc. A collection of texts can provide evidence for a concept but even when aggregated may not provide a coherent structure. In the proposed framework the purpose of the entry ontology is to provide the class structure for the concepts that might be found in the collection of texts to be processed.

3.3 Matching terms to concepts

There are many suitable evaluation metrics that can provide a measure of comparison between terms in the text and concepts in the ontology. Brewster et al. (2004) evaluate

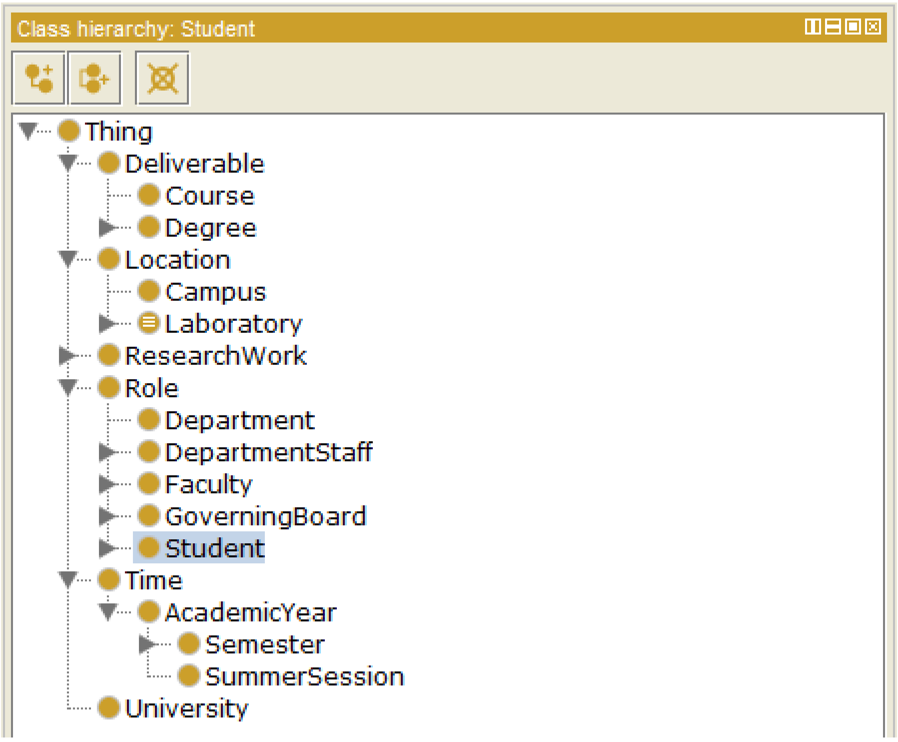

Figure 2: Top level Higher Education Reference Ontology, presented by Protégé ontology editor

concepts described in an ontology against texts in which key terms and concepts had been tagged. They use the concept labels in the ontology to expand the set of terms matched by generating associated terms (hypernyms and synonyms) from WordNet (2010). They then use a distance similarity measure to obtain the degree of fit between all the significant terms from the set of documents being studied and the ontology being evaluated. Our approach is similar but rather than evaluating all the source material at once we use the information to define an ontology for each text. Reviews of semantic similarity measures are given by Harispe et al. (2013) and Budanisky and Hirst (2006).

In Figure 3 the ontology fragment has four levels, so the statement - *'the teacher was enthusiastic'* is more specific at level 4 and hence more predictive than - *'the teacher has good personal qualities'* at level 3, or '*staff are important'* at level 2. Maynard et al. (2008) propose a 'Balanced Distance Metric' (BDM) which takes into account the semantic distance and level within an ontological hierarchy. Additionally for each concept in the ontology we do not have a crisp match/not-match binary value but a degree of membership, so although 'staff' is not an explicit concept label a tutor 'is-a-kind-of' staff, so matches the concept to some degree.

An Ontology can be expressed in a number of languages. OWL is a W3C recommendation as an ontology language. OWL describes classes, properties and relations among

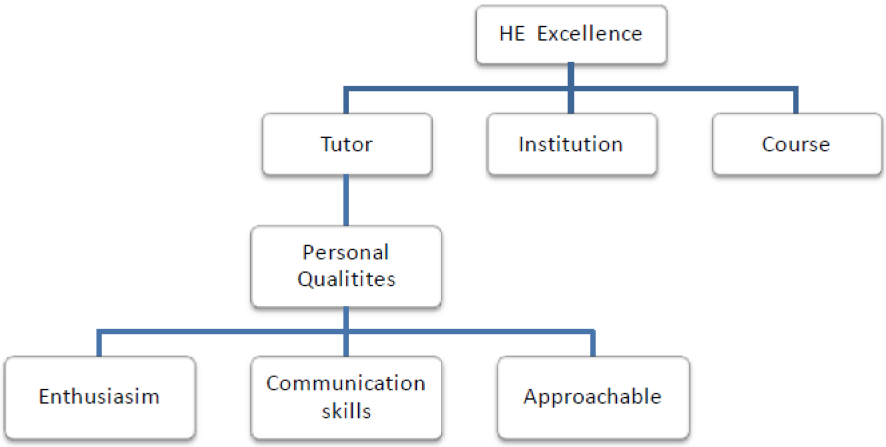

Figure 3: Fragment of a possible domain ontology for higher education excellence

conceptual objects in a way that facilitates machine interoperability of web content. Using the features of OWL, we can take an existing ontology, edit to add / subtract important concepts or merge it with other relevant ontologies. We then extend the class definition to enable each concept to process, store and present data so that the ontology becomes an integral part of the system.

The concept mapper from the proposed framework takes as input the list of terms extracted from the free text and compares them with the concepts and properties defined in upper-level ontology. Using the features of OWL, a term that is not an exact match to a concept label can still be mapped through the properties of the concept, which will include similar and related keywords. Table 1 shows how terms from the text are mapped, by searching the OWL data structures onto properties and classes. For example the term ‘teach’ is mapped to the relation ‘teaches’ (via a lookup from a list of related keywords) which, in this ontology, is a property of the concept Teacher.

Figure 4 describes a fragment of the ontology outlined in Figure 1 showing concepts represented by ovals, concept properties/attributes in rectangles and relations between concepts and properties.

3.4 Missing concepts

It is expected that concepts in the model but not represented by terms in the texts will receive a low level of support and will not form part of the final ontology. However, it is likely that there are many terms extracted from the text that are not represented in the top level ontology.

During ontology construction if a term from the text is not matched in the top level ontology we do not have the information needed to add it to the model. These terms may be random noise or could be important so are allocated to an ’unclassified’ set of terms. This creates a set of ’potential concepts’ that are evaluated in the same way as

Algorithm 1 Algorithm for mapping the terms to the concepts

```
Input: termList, upper-level ontology
for term in termList do        ▷ for every term in termList find the associated concept
                               ▷ ignore the term if it was already found as unclassified
    if term in unclassifiedTermList then
        continue;
    end if          ▷ find the list of concepts and properties associated with the term
    conceptList = filterConcepts(term)
    propertyList = filterProperties(term)
    if conceptList is empty then                           ▷ no concept was found
        conceptList = findConceptFromProperty(properties)
                                 ▷ find the concepts associated with the properties -
                               ▷ if still no concept found then the term is unclassified
        if conceptList is empty then
            add the term to the unclassifiedTermList
        end if
    else                           ▷ get the most relevant concept of the conceptList
        if size(conceptList) > 1 then
            concept = reFilterConcept(conceptList, propertyList)
        else
            concept = get the only element in the list
        end if
    end if
    expression_term = findExpression(term)
    writeOutput(concept, expression)
end for
```

Table 1: Mapping terms to concepts

Terms	Mapping term	Type	Ontology concepts
teach	teaches	owl:Property	(rdfs:domain, Teacher), (rdfs:range, Course)
taught	teaches	owl:Property	(rdfs:domain, Teacher),(rdfs:range, Course)
teaching	teaches	owl:Property	(rdfs:domain, Teacher),(rdfs:range, Course)
teacher	teacher	owl:Class	(rdfs:subClassOf, Person)
lecturer	lecturer	owl:Class	(rdfs:subClassOf, Teacher), OR (owl:equivalentClass, Teacher)

the concepts that are defined by the ontology, so that their level of support from the evidence can be assessed.

If there are a sufficient number of examples of a term then they can be added manually to the domain model and the top-level ontology can be edited to include the concept with appropriate linkage to existing concepts. This process can be done as part of an initial fine tuning of the top level domain ontology, or left until the whole corpus has been analysed. The amount of noise can be reduced by adding a filter that defines a set of 'uninteresting terms'.

3.5 Example

The following paragraph is used to exemplify the proposed framework and algorithms.

> When I visited London I found it very helpful to use the tube train to get around the Capital. It was cheap and efficient. This meant I could get to my lectures on time. I was very interested in my course because the lecturer, Professor Brown, was an expert in the subject. I found that the teaching was good but the lecture was too long. Classes on this course were well supported with the latest equipment and enthusiastic tutors.

The processes included in the proposed framework in Figure 2 are analysed step by step. The list of terms generated by the Terms Extractor from the above text is:

1	$Train$	2	$Capital\_London$
3	$cheap$	4	$efficient$
5	$ontime\_Lecture$	6	$interestedin\_Course$
7	$expert\_Lecturer$	8	$good\_Teaching$
9	$long\_Lecture$	10	$wellsupported\_Classes$
11	$latest\_Equipment$	12	$enthusiastic\_Tutor$

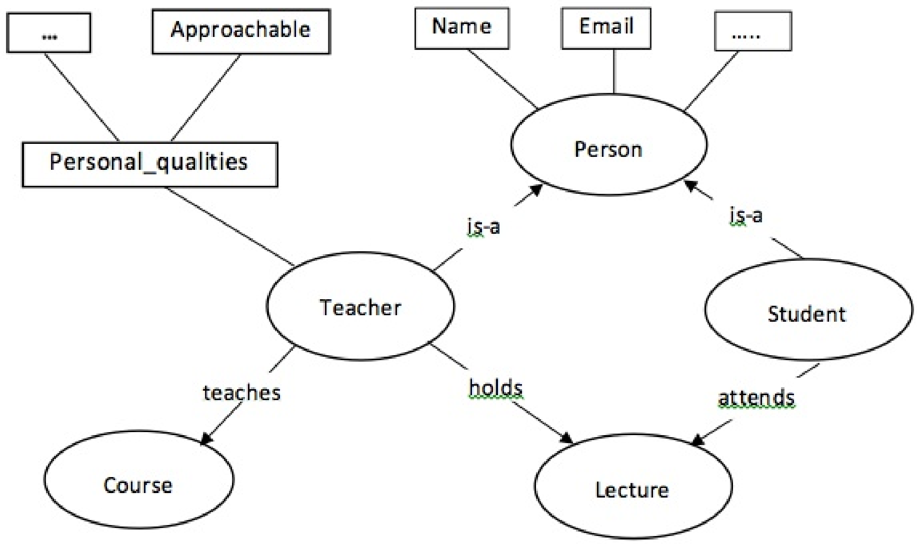

Figure 4: A fragment of conceptual relationships

The list of terms along with the Upper Ontology are inputs to the algorithm that matches terms to concepts. A fragment of the ontology used for this example is presented in Figure 3. Below is a description of how the algorithm is applied to a few terms from the list above.

For term 1. $Train$ and term 2. $Capital_London$ there are no relevant concepts found so, these terms are added to $unclassifiedTermList$.

For term 5. *ontime_Lecture*

- the concept is *Lecture*
- the property is *ontime*
- the expression is $Positive$

For term 8. *good_Teaching*

- the *conceptList* is initilally empty
- the *propertyList* is $\{good, teaching\}$
- the$findConceptFromProperty(\{good, teaching\})$
- returns the concept list $\{Teacher\}$

Therefore:

- the concept is $Teacher$
- the property is *good*
- the expression is $Positive$

3.6 Hypothesis testing

Good quality research normally starts with a hypothesis that is examined and tested by the research process. In this case we would expect a researcher to have some views about what is important in the domain. To test a hypothesis a researcher will construct a model by selecting relevant concepts from the top level ontology. This will also provide a check on the appropriateness of the generic ontology to describe the features of the research. If it cannot represent core concepts of the hypothesis it may need to be amended, merged with another ontology or replaced.

Analysing a large set of texts generates many individual ontologies. These ontologies are combined to highlight a core of common concepts but will also show the differences between individual cases. Once the information has been gathered for each text (or group of texts) the ontologies can be clustered and aggregated. Since all the ontologies take the same structure and concept labels from the top level ontology, concepts can be merged by combining the data held by equivalent concepts.When displaying or analysing the finished domain ontology the degree of support can be used as a dynamic control to change the fidelity of the expressed model.

The model(s) derived from the ontology construction can be compared to the initial hypothesis. If the hypothesis is strongly supported by the evidence it can be found by selecting an appropriate level of fidelity in the final, summary ontology.

3.7 Classification

As well as a descriptive model of a domain a researcher may want to develop a predictive, classification of texts. This could take the form of a particular dependent concept, such as excellence in relation to higher education, or sentiment analysis tools can be used to positive and negative sentiments to the terms extracted from during the pre-processing of the texts. In this case a separate ontology is created for each of the classes required. For example we may want to separate the Excellence concept into positive, neutral, unknown and negative classes. By pre-classifying the terms into these groups we can direct them into separate ontologies which are developed in the same way as before. Although it is likely that there will be more evidence for concepts that are not connected to Excellence this will not swamp the information regarding our target variable as the ontology creation and combination will be separate for each class. The models derived from this process can be used to classify the texts being studied, classify new texts or be used to support a rule based classifier system.

4 Conclusions

In this paper, we outline a process to support the analysis and creation of descriptive models for free texts that is computationally lightweight and amendable to parallel distribution to enable the analysis of large numbers of source documents. The ontological model produced can be represented at different levels of fidelity to enhance common features or to allow the investigation of rarer events. Maintaining data at each step increases the size and complexity of the processing but this is offset by the ability to utilise multiple processing nodes on a fully distributed processing infrastructure.

One of the common characteristics of big data is that attempting to handle all the data in a single process, on standard hardware is likely to exceed the memory limits of the processing architecture and may exceed the total storage capacity and address space. Single 'in-memory' solutions are possible but rely on specialised hardware or expensive commercial services and can themselves become a bottleneck.

A distributed approach to processing allows many separate computers to join together and each may employ a sequential or locally parallel approach to processing. Although many algorithms are successfully reverse engineered to fit a distributed paradigm, the best approach is to design the controlling logic to be inherently parallel from the start. The approach we have proposed here is designed from the outset to focus on small, independent processing steps that can be distributed and amalgamated allowing the process to be scaled up.

For a fully distributed processing paradigm our approach allows the pre-processing of texts to be done individually. A copy of the top level ontology and a number of coded texts can be used to create the individual ontologies in separate processing environments. The nature of the ontology construction and evaluation means that combining ontologies is a commutative process, able to be done in any order. So our framework allows both a locally parallel implementation and a distributed implementation, suitable for a big data application using a cloud computing infrastructure.

References

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In *LREC*, volume 10, pages 2200–2204.

Brewster, C., Harith, A., Dasmahapatra, S., and Wilks, Y. (2004). Data driven ontology evaluation. In *Language Resources and Evaluation (LREC 2004), Proceedings of the International Conference on*, volume 2, pages 24–30. LREC.

Budanisky, A. and Hirst, G. (2006). Evaluating wordnet-based measures of lexical semantic relatedness. *Computational Linguistics*, 32(1):13–47.

Burton-Jones, A., Storey, C., Sugumaran, V., and Ahluwalia, P. (2005). A semiotic metrics suite for assessing the quality of ontologies. *Data and Knowledge Engineering*, 55(1):84–102.

Church, K. W. and Hanks, P. (1990). Word association norms, mutual information, and lexicography. *Comput. Linguist.*, 16(1):22–29.

Cimiano, P., Buitelaar, P., and Vlker, J. (2010). *Ontology construction.* CRC Press: London.

Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology*, 38(1):188–230.

Gruber, T. (1993). A translation approach to portable ontology specification. Technical report, Knowledge Acquisition.

Guarino, N. (1998). Formal ontology and information systems. In *Formal ontology and information systems, 1998., Proceedings of Formal Ontology in Information Systems (FOIS), Conference on 6-8 June*, volume na, pages 3–15. IOS Press.

Harispe, S., Ranwez, S., and Janaqu, S.and Montamin, J. (2013). Semantic measures for the comparison of units of language, concepts or instances from text and knowledge representation analysis, arxi. Technical report, http://arxiv.org/ftp/arxiv/papers/1310/1310.1285.pdf.

Hlomani, H. and Stacey, D. (2014). Approaches, methods, metrics, measures, and subjectivity in ontology evaluation: A survey. *Semantic Web Journal*, na(na):1–5.

Holley, D. (2008). Formal ontology and information systems. In *Using biographic narrative to explore students experiences of online learning in Using Auto/biography in Learning and Teaching Conference on*, volume na, pages 53–58.

Kalfoglou, Y. and Schorlemmer, M. (2003). A method of automated non-parametric content analysis for social science. *American Journal of Political Science*, 18(1):1–31.

Kontopoulos, E., Berberidis, C., Dergiades, T., and Bassiliades, N. (2013). Ontology-based sentiment analysis of twitter posts. *Expert Systems with Applications*, 40(10):4065–4074.

Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176):12031205.

Ma, J., Xu, W., and Sun, Y.-h. (2012). An ontology-based text-mining method to cluster proposals for research project selection, systems, man and cybernetics, part a: Systems and humans. *IEEE Transactions*, 42(3):784–790.

Maynard, D., Peters, W., and Li, Y. (2008). Evaluating evaluation metrics for ontology-based applications: Infinite reflection. In *Evaluating Evaluation Metrics for Ontology-Based Applications: Infinite Reflection Conference on 2008*, volume na, page na.

Noy, N. and McGuiness (2001). Ontology development 101: A guide to creating your first ontology, technicareport smi-2001-00880. Technical report, Stanford Medical Informatics.

Pang, B. and Lee, L. (2013). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135.

Protege (2014). Protege ontology library, protoegesiki. Available at `http://protege.stanford.edu`. Accessed April 2014.

SenticNet (2014). Sentic net website. Available at `http://sentic.net`. Accessed July, 2014.

Tartir, I., Arpinar, B., and Sheth, A. (2010). *Ontological Evaluation and Validation, in Theory and Applications of Ontology*. Springer.

Turney, P. (2002). Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL'02), Philadelphia, Pennsylvania,* *USA*, pages 115–130.

Wenograf, T. (2001). *Qualitative Research Interviewing: Biographic Narratives and Semi-structured Methods*. Sage: London.

WordNet (2010). About word net, website,princeton university. Available at `http://http://wordnet.princeton.edu`. Accessed July, 2014.

Zemmouchi, I. and Ghomari, A. (2013). Process of building reference ontology for higher education. In *Process of Building Reference Ontology for Higher Education Conference on July 2013*, volume 3, page na.