The Semantics of Negation Detection in Archaeological Grey Literature

Andreas Vlachidis[1] , Douglas Tudhope[1],

[1] University of South Wales, Hypermedia Research Unit, Pontypridd Wales,
CF37 1DL, UK
{andreas.vlachidis, douglas.tuhope}@southwales.ac.uk

Abstract. Archaeological reports contain a great deal of information that conveys facts and findings in different ways. This kind of information is highly relevant to the research and analysis of archaeological evidence but at the same time can be a hindrance for the accurate indexing of documents with respect to positive assertions. The paper presents a method for adapting the biomedicine oriented negation algorithm NegEx to the context of archaeology and discusses the evaluation results of the new modified negation detection module. The performance of the module is compared against a "Gold Standard" and evaluation results are encouraging, delivering overall 89% Precision, 80% Recall and 83% F-Measure scores. The paper addresses limitations and future improvements of the current work and highlights the need for ontological modelling to accommodate negative assertions. It concludes that adaptation of the NegEx algorithm to the archaeology domain is feasible and that rule-based information extraction techniques are capable of identifying a large portion of negated phrases from archaeological grey literature.

Keywords: Negation Detection, Semantic Technologies, Digital Humanities, CIDOC-CRM, Semantic Annotation, Natural Language Processing.

1 Introduction

The latest advances of semantic technologies have opened new innovative ways in which scholars can act and elaborate information via search and browsing software applications that process the "meaning" of words beyond the level of a simple and dry string matching process [1]. Semantic metadata practices enrich information with conceptualisations that enable sophisticated methods for data publishing and pave new ways for information analysis and data integration [2]. The field of negation detection presents a challenging ground for the application of such semantic metadata technologies that ought to be explored and investigated for the delivery of scholarly research solutions that could significantly enhance and advance the ways in which the search of facts and findings is conducted.

Negation is an integral part of any natural language system. It is a linguistic, cognitive and intellectual phenomenon, which enables the users of a language system to communicate erroneous messages, the truth value of a proposition, contradictions, irony and sarcasm [3]. Philosophers, from Plato to Spencer Brown have independently approached negation as a case of heteron (not-being) described as a "positive assertion of the existence of a relevant difference" [4]. Whereas, there is a plethora of studies and theories addressed to the complexity of negation and its characteristics from a philosophical and linguistic point of view, research on automatic detection of negation and representation of its semantics has not been extensive [5].

In recent years, Natural Language Processing (NLP) applications have mainly drawn attention to the automatic detection of negation in biomedical text [6] and in opinion mining and sentiment analysis [7]. The techniques and approaches that are employed to address the issue of negation within NLP vary and cover a wide spectrum of machine learning and ruled-based (regular expressions and syntactic processing) applications. Both machine learning and rule-based approaches have been reported as capable of addressing the task of automatic detection, with the rule-based approaches having an edge and being competitive in the biomedicine domain [8]. However, there is little evidence of research aimed at the study of negation detection in the domain of archaeology, albeit some strong parallels can be drawn between archaeological and biomedicine research questions particularly when dealing with facts and findings [9].

In the scientific research of the biomedicine and the archaeology domain, appreciation and understanding of negated facts is as equally important as the interpretation of positive findings. For example the medical phrase "*The chest X-ray showed no infiltrates*" reveals a significant negated finding which can lead to a correct diagnosis of a cardiac condition. In archaeology "negative results are essential when providing an assessment of the archaeological potential of a specific site" [10], for example the phrase "*No traces of a Roman settlement have been discovered in the area*" can lead to specific conclusions with regards to settlement activity during the Roman period in a particular area. Being able to distinguish such negative assertions in context is highly desirable for the research and analysis of facts and findings especially when those activities are supported by information retrieval systems.

The following sections of this paper discuss the method and evaluation results of the negation detection module of the OPTIMA pipeline [11] for the semantic annotation of archaeological grey literature with respect to the CIDOC CRM ontology. Semantic Annotation refers to specific metadata which are usually generated with respect to a given ontology and are aimed to automate identification of concepts and their relationships in documents [12]. CIDOC CRM is an ISO Standard (ISO 21127:2006) comprehensive semantic framework that makes available semantic definitions and clarifications that promote shared understanding of cultural heritage information [13].

OPTIMA contributed to the Semantic Technologies for Archaeological Research (STAR) project [14], which explored the potential of semantic technologies in query and integration of archaeological digital resources. The output of the pipeline is

delivered in the form of semantic indices, expressed as RDF triples that enable semantically defined information retrieval and cross-searching over disparate archaeological digital resources i.e. grey literature and datasets.

The paper highlights the essential role of the negation detection module for the aims of the pipeline and reveals the contribution of the NegEx algorithm [6] in the applied method. The necessary modifications of NegEx are also discussed which led to the adaptation of a biomedicine oriented algorithm to the negation requirements of the archaeology domain. The evaluation results of the negation module of OPTIMA are encouraging, delivering high Precision (89%) and Recall (80%) scores. The performance of the negation module is evaluated with the Gold Standard method of humanly defined annotations. The paper also discusses the issue of the limited support of ontologies, in particular CIDOC-CRM, in modelling and representation of negated findings and concludes with known limitations and future improvements of the work.

2   Method

The negation detection module of the OPTIMA pipeline is primarily developed to support the task of Named Entity Recognition (NER) with respect to the CIDOC-CRM entities *E19.Physical Object*, *E49.Time Appellation*, *E53.Place* and *E57.Material*. NER is a particular subtask of Information Extraction aimed at the recognition and classification of units of information to predefined categories [15]. Since the aim of the semantic annotation pipeline (OPTIMA) is to deliver semantic indices of archaeological grey literature, it is important to be able to exclude from indexing those occurrences of CRM entities that are negated. Thus, the aim of the negation module is to strengthen the precision performance of the pipeline by discarding negated matches that could harm the validity of results at information retrieval level.

2.1   Relevant Work

NegEx [6] is a specific algorithm targeted at the identification of negated findings in medical documents. The algorithm determines whether Unified Medical Language System (UMLS) terms of findings and diseases are negated in the context of medical reports. NegEx is particularly relevant to the scope of the OPTIMA negation module, due to its rule-based design, the use of pattern matching mechanism and the employment of vocabulary listings.

The design of the algorithm is based on the use of offset patterns that utilise a negation related vocabulary. The vocabulary contains terms and phrases that denote negation, which are invoked by a set of rules. The algorithm makes use of two specific patterns; The first pattern [Pre-Neg] identifies negated UMLS terms in phrases which commence with a negation phrase followed by a window of up to 5

tokens before matching an UMLS term, i.e. *<negation phrase> * <UMLS Term>*. The second pattern [Post-Neg] is a reversed version of the above, which matches negated UMLS terms that are up to five tokens prior to a negation phrase, i.e. <UMLS Term> * <negation phrase>.

There are two main parallels for archaeological reports which support the adoption of the NegEx approach by the OPTIMA negation mechanism. Firstly, the use of pattern matching rules and vocabulary terms allows a smooth integration of the algorithm within the requirements and scope of the OPTIMA pipeline for semantic annotation via based rule-based techniques that are supported by knowledge organisation resources (i.e. thesauri and glossaries). Secondly, the good performance of the algorithm in detecting negations about findings in biomedicine context. In archaeological reports, as in medical reports, authors frequently negate facts about findings [10].

2.2 Adapting the NegEx Algorithm in the Archaeological Domain

The process of adaptation of the NegEx in the archaeological domain addressed a range of modifications relating to the coverage and use of negation glossaries as well as adaptations to the scope and application of the negation rules themselves. The main aim of the adaptation exercise was to apply the NegEx approach to the identification of negation phrases involving the four CRM entities (Physical Object, Time Appelation, Place and Material) which are targeted by the NER phase of the OPTIMA pipeline. Modification of the original pattern matching rules to aim at CRM entities instead of UMLS terms is a straightforward task. However, a range of additional adaptation issues required further examination before porting the original algorithm and glossaries in the archaeology domain.

The adaptation strategy considered the following issues which potentially affect application of NegEx to a new domain: i) the size of the negation window which originally had been set to a span of a maximum five word-tokens to fit particularly to the writing style of medical text, ii) coverage and re-usability capacity of existing negation glossaries to support the negation detection task in a new domain, iii) usefulness of the *pseudo-negation* glossary list for limiting the scope of a negation phrase and iv) review on the relevancy of the assumption that medical narrative is "lexically less ambiguous than unrestricted documents" [16] in the context of archaeological reports. In addition, the reported limitation of NegEx at targeting cases of conjunct negated terms was addressed during the adaptation task as discussed in the next section.

2.2.1 Corpus Analysis to Inform the Task of Adaptation

The main aim of the bottom up corpus analysis was to reveal additional vocabulary evidence which could be used by the negation detection mechanism in order to improve adaptation of the algorithm to the context and the writing style of

archaeological reports. Therefore, it was decided that a negation window expanding beyond the window limit of five tokens could be exercised for surfacing larger spans.

The first stage of corpus analysis extracted from a volume of 2460 archaeological reports, phrases of a maximum of 10 tokens which contained negation moderators and CRM entity matches. Using the existing NegEx [Pre-Neg] and [Post Neg] glossary listings, the following two separate matching grammars were constructed:

```
({Token.string!="."})[0,5]{PreNeg}({Token.string!="."})[0,5]
({Token.string!="."})[0,5]{PostNeg}({Token.string!="."})[0,5]
```

The grammars are almost identical; they only differ on the listing type which they invoke (i.e. PreNeg or PostNeg). The rules translate as: match a span which expands 5 tokens before a glossary match and 5 tokens after a glossary match excluding full stops (to prevent the rule expanding beyond the limits of a potential sentence). A succeeding matching grammar was invoked for filtering out those phrases that did not contain any of the four CRM entities failing within the scope of OPTIMA.

The second stage implemented a separate pipeline which post-processed the negation phrases delivered by the first stage. The aim of this particular pipeline was to reveal the most commonly occurring Noun and Verb phrases of the negated phrases output. Such commonly appearing noun and verb phrases were then analysed to inform the process of enhancement and adaptation of existing glossaries and negation grammars to the context of archaeology.

In total, 29040 noun phrases and 14794 verb phrases were identified. From them 14686 were unique noun phrases and 2564 were unique verb phrases. Examining the list of the most frequent noun phrases and comparing it with the list of the NegEx lists it became apparent that some of the existing entries were not applicable to the archaeology domain and returned no matches. Such entries are rather particular to the medical domain, for example "suspicious", "decline", "deny" and "unremarkable for". Moreover, frequently occurring negation classifiers of archaeological narrative, such as "unknown", "unclear" and "undated" were not part of the initial NegEx lists. The adaptation exercise created new versions of the [PreNeg] and [PostNeg] lists adapted to the archaeology domain by removing the entries that are particular to the medical domain and by including new entries that are relevant to the archaeology domain.

The analysis of the verb phrases result revealed some very interesting vocabulary patterns. Examining the most commonly occurring verb phrases, a pattern emerged relating to use of passive voice utterances. For example the phrase "should not be considered" occurred 134 times, the phrase "was not excavated" 67 times, the phrase "were not encountered" 39 times, etc. Although, NegEx covered some cases of backward matching via the [Post-Neg] list for phrases where a negation classifier is found at the end of a phrase, the algorithm did not consider extensively the use of passive voice expressions apart from the case "being ruled out".

The intellectual examination of the list of the frequently occurring verb phrases isolated a set of passive voice verbs that could be used to enhance the operation of the negation algorithm. The list of verbs constitutes a specialised vocabulary of 31 entries such as "appear", "associate", "compose", "discover", "encounter", etc., which were

composed under a new glossary listing named *Negation-verbs*. The glossary is used by the pattern matching rules discussed in the section below, for identifying negation in phrases, such as "*deposits were not encountered at the machined level*".

An integral part of the NegEx algorithm is the [Pseudo-Negation] list which is responsible for limiting the scope of a match by identifying false negation triggers. Due to the elaborate and unrestricted report style of archaeological grey literature documents, it was decided to expand the negation window of the algorithm to larger phrases containing a maximum of 10 word tokens instead of 5 that are originally set by NegEx. Thus, the inclusion of the [Pseudo-Negation] operation seemed highly relevant for avoiding matches of positive entity assertions that adjoin with negation phrases.

The general principle of the [Pseudo-Negation] operation was adopted as a means to narrow the scope of a negation window. A new list [Stop-Neg] was created that contained 38 new entries originating from the empirical use of English when separating different clauses in a sentence. The lexical resource Wordnet [17] was employed in the construction of a list containing a range of entries such as "but", "nonetheless", "than", "though" and relevant synonyms from the available synset hierarchies of the Wordnet.

The operation of the [Stop-Neg] list prevents matching beyond the scope of a negation phrase and does not exclude identification of conjunct entities. The original NegEx algorithm reported limitations on accurate matching of long lists of conjunct UMLS terms that expand beyond the word limit (5 tokens) of the negation window [6]. The OPTIMA pipeline is equipped with an Entity Conjunction module which delivers matches of the same CRM entities conjunct with "and", "or", "commas" and other forms of hyphenation. Hence, the negation module can exploit conjunct matches, in order to deliver negation phrases that include a list of entities, as for example the phrase "*no evidence of archaeological features or deposits dating to the Neolithic or Bronze Ages*". The inclusion of the [Stop-Neg] list in the negation grammars prevents the match of "*post-medieval spread*" as a negated case in the phrase "*absence of evidence after the Roman period, with the exception of the post-medieval spread*" while the first clause of the phrase is identified as a negated match.

2.3 Negation Detection in OPTIMA

The negation detection module of the OPTIMA pipeline incorporates the four glossary listings, [Pre-Neg], [Post-Neg], [Stop-Neg] and [Verb-Neg] with a set of information extraction grammars. A set of three different pattern matching rules is deployed for each of the four different CRM entity types that fall within the scope of the negation module. The arrangement of the negation rules avoids multiple annotation of the same phrase, even if more than one CRM entities are mentioned in a phrase. The description of grammars given below refers to a unified form of a CRM entity which encompasses all four different CRM types for simplicity.

The grammars deliver a single annotation span, which covers all CRM entities involved in a phrase. For example the phrase "*no evidence of Roman pottery*" delivers a single annotation spanning the whole phrase rather than two separate annotations for "*Roman*" and "*pottery*". Similarly when conjunction of entities is present, the negation span covers all conjunct entities under a single annotation span

The following grammar is targeted at matching cases of negation which commence with a match from the [Pre-Neg] list and end in a CRM entity or a CRM conjunct entity match, for example "*absence of any datable small finds or artefacts*".

```
{PreNeg}({Token,!StopNeg})[0,10]({CRM}|{CRM_Conjuction})
```

The following grammar matches cases of negation which commence with a CRM entity or a CRM conjunct entity and end with a match from the [Post-Neg] list, for example "*wares such as tea bowl are particularly unlikely to exist*".

```
({CRM}|{CRM_Conjuction})({Token, !StopNeg})[0,10]{PostNeg}
```

The following grammar is targeted at matching cases of negation which commence with a CRM entity or a CRM conjunct entity and end with a match from the [Verb-Neg] list, for example "*pottery and tile remains were not observed*".

```
({CRM}|{CRM_Conjuction})({Token,!StopNeg})[0,10]
{Token.string=="not"}({Token})?{VerbNeg}
```


3 Evaluation

The evaluation phase aimed at benchmarking the performance of the negation detection module using standard evaluation methods. Typically the performance of Information Extraction systems is measured in Recall, Precision and F-Measure scores as established by the second Machine Understanding Conference, MUC 2 [18]. The F-Measure score is the harmonious mean of Precision and Recall used to provide a comprehensive view of system's performance. Attempts to improve Recall will usually cause Precision to drop and vice versa. High scoring of F-Measure is desirable since it can be used to benchmark the overall system's accuracy [19].

The evaluation phase has adopted the above measurements including both fully correct and partial matches, as expressed by the following formulae, with N_{key} being the correct answer, and the system delivering $N_{correct}$ responses correctly and $N_{incorrect}$ incorrectly :

$$Recall = \frac{N_{correct} + (Par}{N_{key}} \quad Precision = \frac{N_{correct} + (Part}{N_{correct} + N} \quad F_1 = 2\frac{Precision * Rec}{Precision + Rec}$$

Partial matches are those having different annotation boundaries than the N_{key} definition, either matching only a part or expanding beyond the limits of a key. Partial matches can be weighted with decimal values ranging from 0 to 1 depending on the importance of such matches in the system's accuracy. The evaluation task treated

partial matches as fully correct matches based on the flexible user-centred approach followed during the definition of the $N_{key\ responses}$, which delivered negated phrases that were syntactically complete from a user's point of view. For example, an N_{key} response might be "*No traces of a Roman settlement have been discovered in the area*". However, the negation algorithm is programmed to extract phrases that commence or end with a CRM entity, in this case only the first part is extracted ("*No traces of a Roman settlement*") delivering a partial match. Hence, the match can be treated as fully correct since the N_{key} response is not defined with the algorithm in mind but with what is useful from an end-user point of view.

The set of the N_{key} responses participating in system's evaluation were delivered by the method of manual annotation also known as "Gold Standard" definition, which is typically employed for comparison against system produced annotations. Such manual definitions are usually built by domain experts but their availability is often scarce. In the case of archaeological reports, there was no available gold standard of semantically annotated documents with respect to negated CIDOC-CRM entities. Therefore, the evaluation stage pursued the definition of a gold standard corpus tailored to serve the purposes of the evaluation task.

In total 10 grey literature documents of archaeological excavation and evaluation reports contributed to the gold standard definition. In archaeology, grey literature reports reflect the different stages of a fieldwork project worth recording and disseminating information about. They contain comprehensive explanations, diagrams, summaries and statistics that deliver in depth analysis and discussion usually not possible to be accommodated by traditional publication. The evaluation corpus contained a set of archaeological excavation and evaluation reports, which typically contain rich discussion about the findings and excavation phases over other types of archaeological reports, such as watching briefs and observation reports. In addition, the selection process included reports from a range of different UK archaeological units aiming to cover different reporting styles and practices. The gold standard overall consisted from 300 pages which contained 144 cases of negation.

3.1 Results

Among the 10 documents that participated in the evaluation task, the negation detection module delivered an overall Recall score 80%, Precision 89% and F-Measure 83% (table 1). The Recall score of individual documents presents a fluctuation ranging from 50% to 100% while fluctuation of Precision scores is smaller ranging from 64% to 100%. The good precision performance of the module is also reflected by the standard deviation score which is 0.11(or 11%) with only one document scoring under 80%. On the other hand, the standard deviation of Recall scores is slightly higher 0.15 (or 15%) with half of the documents scoring under 80%.

Table 1. Performance of Negation Detection Module.

Document	Recall	Precision	F-Measure

Aocarcha1-11167	0.74	0.94	0.83
Birmingh2-28160	0.77	1.00	0.87
Essexcou1-10460	0.83	1.00	0.91
Essexcou1-5166	0.76	0.85	0.80
Foundati1-5205	0.87	1.00	0.93
Heritage1-10767	0.50	1.00	0.67
Heritage1-11948	1.00	0.83	0.91
Suffolkc-6115	0.85	0.89	0.87
Wessexar1-25626	0.70	0.64	0.67
Wessexar1-5680	1.00	0.83	0.91
Average	0.80	0.89	0.83

The negation detection module has overall delivered 114 correct and partially correct matches, 14 false positive (falsely identified) matches, while it missed 30 negation answers of the gold standard definition. The number of total false positive matches is half of those matches being missed. This significant difference between the two is also reflected in the Precision and Recall scores where missed matches directly affect recall and false positives precision. Overall, the negation module delivers better precision than recall, indicative of the module's capacity to accurately identify cases of negation while being challenged by the variety in which negation can be expressed in natural language.

4 Discussion

The evaluation results revealed an encouraging performance of the negation detection module which delivered Recall and Precision scores over 80%. Although, the evaluation task had a limited scope and was based on the use of a small scale gold standard definition, it suggests that negation in archaeological text can be addressed with information extraction techniques that use a small set of domain oriented pattern-matching rules. Our results agree with research findings from biomedical text negation [6] reporting that negation phrases typically comply with the Zipf's law regarding the frequency distribution of words in human languages, where a few very common negation patterns can capture a large portion of pertinent negation cases.

According to the evaluation results, use of frequently occurring negation patterns in extraction rules supports the system's precision. The vast majority of automatically identified negation phrases (approximately 9 out of 10) delivered by the negation module were correct. Incorrect cases (false positives) are primarily the result of limitation in the vocabulary used to support the operation of extraction patterns and not due to the incapacity of extraction patterns themselves. For example the phrase "*It is not unusual to find solitary prehistoric cremations*" has a positive meaning which is falsely identified as a negation case.

The OPTIMA algorithm, similarly to the NegEx algorithm [6], employs a specialised vocabulary [Stop-Neg], which limits the scope of negation. The original NegEx glossary of pseudo-negation phrases is enhanced with additional terms (drawn

from ordinary use of English), through a WordNet Synset expansion technique (section 2.2.1) to include a range of entries such as, "but", nonetheless", "though" and their synonyms. The expanded glossary failed to address fully all the cases of double negatives as for example "not unusual", which has a fairly positive assertion. Possibly use of double negatives is avoided in the restricted context of narrative reports of medical records but in the context of archaeological reports such double negated narratives may occur. Fewer false positive cases relate to the operation of matching patterns as for example the phrase "*non-intrusive survey had accurately predicted the ridge and furrow*" where "*non*" applies only to the immediate noun that follows.

The recall performance of the negation module is reasonable (approximately 8 out of 10) but not as high as precision. The capacity of the algorithm to identify all correct cases of negation in text is challenged by the sometimes creative and indirect writing style of archaeological reports. For example the phrase "T*he low quantity and quality of the remains encountered on the site suggests that there is only a minor archaeological implication*" clearly suggests that findings do not have an archaeological interest. However, it is formulated in an indirect style, which does not invoke any negation triggers that could be matched by the module.

Other cases of missed examples concern use of passive voice utterances that do not employ clear negation classifiers but verbs which are loaded with negative sense. For example the phrase "*both these deposits were largely absent*" is missed due to the definition of matching patterns that expect a negation classifier at the beginning of a phrase for example "*Absence of deposits*", or a negation formation at the end of a phrase for example "*deposits were not largely present*". Fewer examples of non-identified cases concern limitations of the NER vocabulary itself. For example the phrase "*there was virtually no artefactual evidence recovered*" is missed because "*artefactual evidence*" is not recognised as a CRM entity.

4.1 Adapting NegEx in Archaeological Grey Literature

The adaptation and redesign process of the NegEx algorithm in the context of negation detection of CRM entities in archaeological grey literature documents has achieved various useful results. The original NegEx algorithm operated within a less ambiguous and more restricted narrative context [6] than the archaeological report narrative. The main challenge of the adaptation process was to address the flexible and sometimes creative writing style of archaeological grey literature documents. Additional work involved review, modification and enhancement of original regular expression rules and glossaries. Based on corpus analysis (section 2.2.1), the adaptation technique gathered valuable information with regards to vocabulary use and writing style, which informed and directed the adaptation process.

A major modification of the original algorithm directly informed from the corpus analysis task, concerned the expansion of the negation window from 5 to 10 tokens, supported by an enhanced [Pseudo-Negation] list with terms and phrases relevant to archaeological narrative. The negation window of the regular expression rules was expanded from 5 to 10 word-tokens to include longer phrases of negation which

might appear in archaeological reports. A new matching pattern, not previously included in the original NegEx algorithm due to the direct reporting style of medical reports, was introduced by the adaption task for detecting negation phrases of passive voice expressions. The redesign process has also improved the performance of the original algorithm in the detection of conjunct terms via usage of CRM instances previously identified by the Entity Conjunction module of the NER phase. Moreover, the corpus analysis task has informed modification of the original [Pre-Neg] and [Post-Neg] glossaries with the inclusion of new archaeology related terms and removal of the less applicable original entries.

The evaluation phase revealed valuable feedback with regards to the adaptation effort. Clearly a larger scale corpus analysis supported by archaeology domain-expert input, could reveal new clues in the definition of pattern matching rules capable of addressing the flexible and creative reporting style of archaeological documents. Similarly, a larger evaluation corpus (Gold Standard) including a greater number of manual annotations and documents, would potentially reveal the system's capacity to address the variety and complexity of negation in archaeological grey literature reports. Our experience demonstrates a pathway and implementation technique for porting NegEx to a new domain, which we hope is applicable to other domains but this remains to be tested.

4.2 Method Limitations

The development approach of constructing a negation detection mechanism based on shallow parsing delivered results that suite the aims of semantic indexing. Shallow parsing analyses a sentence or a phrase to its constituent parts but without analysing their role or their internal structure. The negation detection module has managed to identify with reasonable success a vast range of phrases containing one or more CRM entities which were negated via a lexical classifier. Upon successful identification of a negation phrase, all CRM entities of the phrase were discarded from further indexing. This approach did not harm the quality of the indexing due to the vast amount of CRM entities being delivered by the NER phase. On the other hand, this particular approach might be considered as a blanket practice that does not support the aims of a detailed and meticulous text mining effort.

Looking closer at the following example "*No artefacts were retrieved from this deposit*" it is clear that there is absence of artefacts. However, the same absence does not apply to the deposit itself which does exist but under the current configuration is excluded from indexing. Adoption of deep parsing techniques, which analyse the role and structure of the constitute parts of a sentence, could be sufficient to address such cases of detailed negation assignment on the level of subject clause. However, the semantic annotation of such negated cases with respect to ontology classes may prove a challenging task as discussed below.

Consider the above example "*No artefacts were retrieved from this deposit*". Assignment of the *E19.Physical_Object* class to "artefacts" instance assumes a positive assertion. Similarly an ontological model may define a relationship property

between place and physical object. Again this kind of property assumes a positive assertion which does not cover the cases where an object in not in place.

A specific project which addressed the issue of factual argumentation using the CIDOC-CRM ontology is the Integrated Argumentation Model (IAM) [20]. Although, factual argumentation is a broader epistemological issue that concerns falsification or verification of arguments, the aim of IAM to connect such epistemological aspects with instances of a formal ontology could be potentially useful and applicable to the semantics of negation assertion. The project presented benefits to archaeological reasoning for a particular case (the natural mummy Oetzi) but its applicability in the context of semantic annotation of archaeological text remains untested.

Providing a semantic annotation i.e. assigning classes or properties, to textual instances that are negated is not always viable within the scope of an ontology that assumes only positive assignment. A possible answer to this limitation might be addressed with the introduction of a property for declaring the sense of an instance for example *has_sense*, positive or negative. However, introduction of such property will significantly increase the chain of triples defined by a SPARQL query on an application level even for the simplest queries.

An alternative approach could be the introduction of negative print of all ontological classes and properties in order to accommodate negative assertions. Thus, every class or property of an ontological model would have its equivalent contrasting class in the sense of "matter, anti-matter". This particular approach though, would double the size of an ontological model which could lead to issues relating to the maintenance and version control of an ontological model.

5 Conclusions

The paper presented the results of a negation detection module targeted at identifying negated cases of four CIDOC-CRM entities in the context of semantic indexing of archaeological grey literature for information retrieval. The evaluation results demonstrate the capacity of rule-based information extraction techniques to accurately detect a large portion of negation phrases. The employment of three small scale glossaries that support the operation of a few simple pattern matching expressions has proved sufficient to deliver high Recall (80%) and Precision (89%) scores. Current limitations of the method relate to the employment of shallow parsing techniques that do not support deeper analysis of negation phrases, the capacity of glossaries to cover all possible vocabulary variations, and the adequacy of pattern matching rules to address every single case of negation which can be expressed by a creative and sometimes indirect writing style of archaeological reports. However, such limitations do not restrict application of the work in the context of semantic indexing. Future steps include a large scale corpus analysis and evaluation study aimed at expanding glossary coverage and improving the system's performance with regards to archaeological negation narrative. Longer term aims may involve the

system's generalisation to the broader field of digital humanities and application of semantic modelling solutions capable of addressing negation at an information retrieval level.

The paper has revealed a method for adapting the NegEx algorithm to the domain of archaeological grey literature. Our experience has shown that porting of NegEx to a new domain is feasible. The method of modification of the original algorithm was driven by a corpus analysis task, which enabled enhancement and adaptation of the original resources to the new domain. This particular method has given promising results for the domain of archaeological reports though its applicability to other domain remains to be tested. The issue of accommodating negative assertions by the current ontological modelling approaches was also highlighted. Negated findings and facts are important for the research and information retrieval in particular domains, such as the medical and the archaeology domain. Semantic technologies can provide a valuable support in modelling and retrieval of such negated findings for enabling new forms of research and information exchange.

Acknowledgments

References

1. O'Hara, K., Berners-Lee, T., Hall, W., Shadbolt, N.: Use of the Semantic Web in e-Research. The MIT Press, Cambridge, MA (2010)
2. Bizer, C., Heath, T., Berners-Lee, T.: Linked Data - The Story So Far. International Journal on Semantic Web and Information Systems 5(3), 1--22 (2009)
3. Horn L.R.: A Natural History of Negation. University of Chicago Press, Chicago (1989)
4. Westbury, C.: Just say no: The evolutionary and developmental significance of negation in behavior and natural language. In: 3rd Conference The Evolution of Language, Paris (2000)
5. Blanco, E., Moldovan, D.: Semantic representation of negation using focus detection. In: 49th Annual Meeting of the Association for Computational Linguistics, Portland (2011)
6. Chapman, W.W., Bridewell, W., Hanbury, P., Cooper, G.F., Buchanan B.G.: A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. Journal of Biomedical Informatics. 34(5), 301--310 (2001)
7. Maynard, D., Funk, A.: Automatic detection of political opinions in tweets. In: Making Sense of Microposts Workshop at ESWC 2011. Heraklion, (2011)
8. Goryachev, S., Sordo, M., Zeng, Q.T., Ngo, L.: Implementation and evaluation of four different methods of negation detection. Technical report. Harvard Medical School (2006)
9. Vlachidis, A., Tudhope, D. 2012. A pilot investigation of information extraction in the semantic annotation of archaeological reports. International Journal of Metadata, Semantics and Ontologies. 7(3), 222—235 (2012)

10. Falkingham, G.: 'A whiter shade of grey: a new approach to archaeological grey literature using the XML version of the TEI guidelines. Internet Archaeology. 17, (2005), http://intarch.ac.uk/journal/issue17/falkingham_index.html

11. Vlachidis, A.: Semantic Indexing Via Knowledge Organisation Resources: Applying the CIDOC-CRM to Archaeological. PhD Thesis, University of Glamorgan, UK. (2012)

12. Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., Ciravegna, F.: Semantic annotation for knowledge management: Requirements and a survey of the state of the art. Web Semantics: Science, Services and Agents on the WWW, 4(1), 14--28 (2006)

13. Crofts, N., Doerr, M., Gill, T., Stead, S., Stiff, M.: Definition of the CIDOC Conceptual Reference Model, http://cidoc.ics.forth.gr/docs/cidoc_crm_version_5.0.1_Mar09.pdf

14. Tudhope, D., May, K., Binding, C., Vlachidis, A.: Connecting archaeological data and grey literature via semantic cross search. Internet Archaeology, 30 (2011), http://intarch.ac.uk/journal/issue30/tudhope_index.html.

15. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes, 30(1), 3--26 (2007)

16. Ruch, P., Baud, R., Geissbuhler, A., Rassinoux, A.M.: Comparing general and medical texts for information retrieval based on natural language processing: an inquiry into lexical disambiguation. Med-info 10, 261--265 (2001)

17. Miller, G.A.: WordNet: A Lexical Database for English. Communications of the ACM, 38(11), 39--41 (1995)

18. Grishman, R., Sundheim, B.: Message Understanding Conference-6: a brief history. In 16th International Conference on Computational Linguistics, Copenhagen (1996)

19. Maynard, D., Peters, W., Li, Y.: Metrics for Evaluation of Ontology-based Information Extraction. In Proceedings WWW Conference 2006, Workshop on "Evaluation of Ontologies for the Web", Edinburgh, (2006).

20. Doerr, M., Kritsotaki,A., Boutsika, K.: Factual argumentation-a core model for assertions making. Journal on Computing and Cultural Heritage. 3(3), 1—34 (2011)