

# A Methodology for Biomedical Ontology Reuse

Nur Zareen Zulkarnain<sup>1</sup>(✉), Farid Meziane<sup>1</sup>, and Gillian Crofts<sup>2</sup>

<sup>1</sup> School of Computing, Science and Engineering, Informatics Research Centre,  
University of Salford, Salford M5 4WT, UK

N.Z.Zulkarnain@edu.salford.ac.uk, F.Meziane@salford.ac.uk

<sup>2</sup> School of Health Science, University of Salford, Salford M6 6PU, UK  
G.Crofts@salford.ac.uk

**Abstract.** The abundance of biomedical ontologies is beneficial to the development of biomedical related systems. However, existing biomedical ontologies such as the National Cancer Institute Thesaurus (NCIT), Foundational Model of Anatomy (FMA) and Systematized Nomenclature of Medicine–Clinical Terms (SNOMED CT) are often too large to be implemented in a particular system and cause unnecessary high usage of memory and slow down the system’s processing time. Developing a new ontology from scratch just for the use of a particular system is deemed as inefficient since it requires additional time and causes redundancy. Thus, a potentially better method is by reusing existing ontologies. However, currently there are no specific methods or tools for reusing ontologies. This paper aims to provide readers with a step by step method in reusing ontologies together with the tools that can be used to ease the process.

**Keywords:** Ontology · Ontology reuse · Biomedical ontology · BioPortal

## 1 Introduction

Biomedical systems are an integral part of today’s medical world. Systems such as electronic patient records and clinical decision support systems (CDSS) have played an important role in assisting the works of medical personnel. One area that could benefit from the development of biomedical systems is ultrasound reporting. In ultrasound, reports generated have more value compared to the image captured during the examination [2]. Variations in ultrasound reporting impacts the way a report is interpreted as well as in decision making. Thus, the standardization of these reports is important. In order to achieve this goal, ontologies are used to understand the reports and structure them according to a certain format [16] as well as recognizing the relationships between the parts of the text composing the report.

General and established domains such as medicine have existing ontologies that cover the general concepts in the domain. Examples of these ontologies include the National Cancer Institute Thesaurus (NCIT), Foundational Model

of Anatomy (FMA) and Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT). These ontologies however are often too large to be manipulated or processed in a specific application. Thus, a domain specific ontology is needed to solve this problem. Building a new domain specific ontology from scratch would not be efficient since this will cause redundancy and takes a lot of time. Thus, ontology reuse has been potentially seen as a better alternative. This paper discusses how ontology reuse has been done before and proposes a methodology to reuse ontologies together with the existing tools that can be used to ease the reuse process. The development of the Abdominal Ultrasound Ontology (AUO) as the knowledge base for an ultrasound reporting system developed by Zulkarnain et al. [16] is used in this paper to explain the proposed ontology reuse methodology.

## 2 Related Work

Ontology reuse can be defined as a process where a small portion of existing ontologies is taken as an input to build a new one [3]. The process of reusing large existing ontologies allows their use without slowing down the process of an application. Ontology reuse also increases interoperability [14]. Indeed, when an ontology is reused by several other new ontologies, interoperability between these ontologies can be achieved much easier since they share several features such as classes naming method and concept modelling.

Even though ontology reuse brings a lot of benefits, there are currently no tools that provide adequate support for the ontology reuse process [8, 14] which hinders the effort of ontology reuse. There is also no one specific method agreed in reusing ontologies. Even so, most ontology reuse methodologies that have been used in previous works [1, 4, 5, 11, 12, 14, 15] falls along the line of these four steps: (i) Ontology selection for reuse, (ii) Concept selection, (iii) Concept customization and (iv) Ontology integration.

The first step for ontology reuse is to select the ontology to be reused. Ontology selection is done according to several criteria according to the needs of the new ontology, for example the language of the ontology, its comprehensiveness and its reasoning capabilities. Once the ontology for reuse is chosen, the next step would be to select the concepts that would be reused. One or several ontologies can be selected for reuse depending on the needs of the new ontology. Russ et al. [11] in their work merged two aircraft ontologies where most of its concepts were selected to develop a broader aircraft ontology. Shah et al. [12] on the other hand reused just one ontology; SNOMED CT where he selected the concepts needed then adds other relevant concepts not included in SNOMED CT.

Concepts selected are then translated into the same semantic language and then merged. In Caldarola et al.'s work [4] this includes manually translating metadata to better understand concepts. Alani [1] in developing his ontology has merged several ontologies that contain different properties for one same concept which resulted in additional knowledge representation. Several different concepts have also been selected from different ontologies which are then compared and merged. Finally, the ontology will be integrated into the system or application.

In this research, these four steps serve as a guideline in developing an ontology reuse methodology for the biomedical domain. The methodology proposed in this research will allow for the ontology to be reused from multiple existing ontologies and suggest tools that would help in each step of the methodology. The ontology developed, Abdominal Ultrasound Ontology (AUO), will serve two purposes in this research: (i) it will be used to standardize the development of ultrasound reports and enforce the use of standard terminology and (ii) to analyse the reports written in Natural Language (English free-text) with the aim of automatically transforming them into a structured format.

### 3 The Proposed Methodology

In developing a new ontology by reusing existing biomedical ones, proper planning and execution are important in order to ensure the modularity of the concepts reused. Thus, the ontology reuse methodology developed in this paper, adopted the general four steps mentioned in Sect. 2 and summarised in Fig. 1.

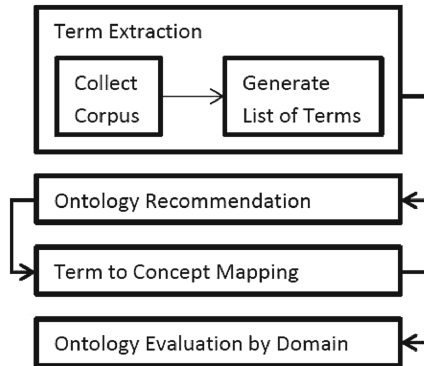


Fig. 1. Ontology reuse methodology

#### 3.1 Term Extraction

The first step in ontology reuse or even in developing one from scratch is to decide on its scope and domain. In this case of developing the Abdominal Ultrasound Ontology (AUO), the scope and domain of the ontology is abdominal ultrasound. 49 sample ultrasound reports have been collected and used as the basis of our ontology corpus. These sample reports were obtained from the Radiology Departments in a large NHS Trust incorporating 4 regionally based hospitals in Manchester and Salford. Once we have our corpus, the next step is to extract relevant terms from the corpus to generate a list of terms for reuse. Two biomedical term extraction applications; (i) TerMine<sup>1</sup> and (ii) BioTex<sup>2</sup> have been used

<sup>1</sup> <http://www.nactem.ac.uk/software/termine>.

<sup>2</sup> <http://tubo.lirmm.fr/biotex/>.

**Table 1.** Comparison of biomedical term extraction using TerMine and BioTex

	TerMine	BioTex
Language	English	English
License	Open	Open
POS Tagger	GENIA Tagger/Tree Tagger	Tree Tagger
Terms found	241 (GENIA Tagger) 232 (Tree Tagger)	761
Extraction type	Multi-word extraction	Multi-word and single-word extraction

for the extraction. All 49 sample reports were submitted to both applications and the results are shown in Table 1.

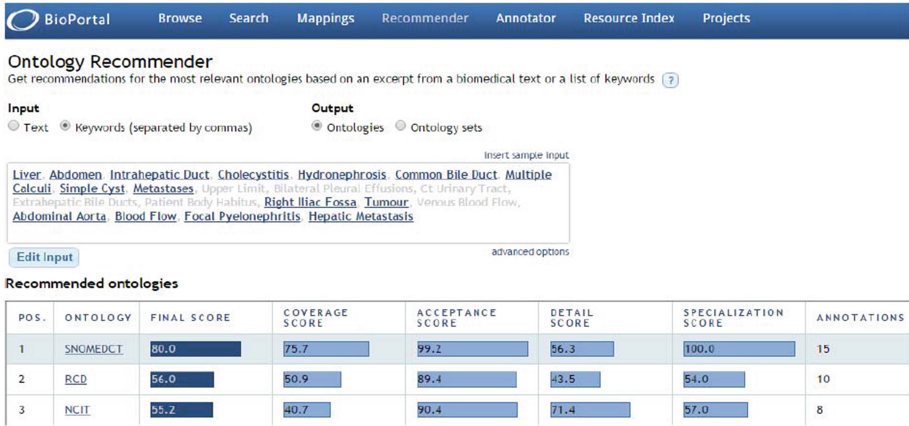
From this comparison, BioTex was chosen as the better biomedical term extractor in this research because of its ability to extract more terms compared to TerMine. BioTex is an automatic term recognition and extraction application that allows for both multi-word and single-word extraction [7]. It is important that the term extractor is able to extract not only multi-word but also single-word terms.

For example, if the sentence “Unremarkable appearances of the liver with no intrahepatic lesions” was submitted to both applications, TerMine will only extract two multi-word terms “Unremarkable appearance” and “intrahepatic lesion” while BioTex will extract not only the two multi-word terms but also “liver” which is a single word term. If single-word terms such as “liver”, “kidney” and “spleen” were not extracted, the ontology developed would be incomplete. Terms which are extracted from BioTex were also validated using the Unified Medical Language System (UMLS) [7] which is a set of documents containing health and biomedical vocabularies and standards.

### 3.2 Ontology Recommendation

The next step after obtaining a list of terms for ontology reuse would be to select the suitable ontology to be reused. Three important criteria were used for selecting the ontology in this research: (i) Ontology coverage - To which extent does the ontology covers the terms extracted from the corpus? (ii) Ontology acceptance - Is the ontology being accepted in the medical field and how often is it used? and (iii) Ontology language - Is the ontology written in OWL, OBO or other semantic languages? Initial review resulted in choosing FMA, SNOMED-CT and RadLex as suitable candidates because of their domain coverage, acceptance in the biomedical community and language which is OWL. In order to verify this, an ontology recommender was developed using BioPortal’s ontology recommender API<sup>3</sup> which is an open ontology library that contains ontologies with domains that range from anatomy, phenotype and chemistry to experimental conditions [10].

<sup>3</sup> <http://bioportal.bioontology.org>.



**Fig. 2.** BioPortal's ontology recommender

BioPortal has an ontology recommender available on its portal that can be used to obtain suggestions on suitable ontology to be reused for certain corpus. The ontology recommender makes a decision according to the following three criteria: (i) Coverage - Which ontology provides most coverage to the input text?, (ii) Connectivity - How often the ontology is mapped by other ontologies? and (iii) Size - Number of concepts in the ontology [6]. When a list of terms is submitted to the recommender, it will give a recommendation of 25 ontologies which are ranked from the highest to lowest scores (see Fig. 2). The final score is calculated based on the following formula:

$$\begin{aligned}
 \text{FinalScore} = & (\text{CoverageScore} * 0.55) + (\text{AcceptanceScore} * 0.15) \\
 & + (\text{KnowledgeDetailScore} * 0.15) + (\text{SpecializationScore} * 0.15) \quad (1)
 \end{aligned}$$

The coverage score is given based on the number of terms in the input that are covered by the ontology. The acceptance score indicates how well-known and trusted the ontology is in the biomedical field. Knowledge detail score on the other hand indicates the level of details in the ontology; i.e. does the ontology have definitions, synonym or other details. Specialization score is given based on how well the ontology covers the domain of the input. An example is given in Fig. 2 where 21 terms were submitted. There is however a limitation in using it on the portal whereby it only allows for 500 words to be submitted. This limitation has prompted us to develop our own recommender by manipulating the data from BioPortal's ontology recommender API. We first develop the recommender that would give 25 ontology recommendations just like how it would be in the BioPortal's recommender. However, it seems that 761 terms were too big for the recommender's server to handle. Because of this, a recommender that would suggest ontology for each term was developed.

A list of terms, in this case the 761 terms that have been extracted, are submitted to the algorithm that would submit each term to BioPortal’s recommender and get ontology recommendations for each term. Then, the frequency of each ontology recommended will be counted and sorted from highest to lowest. The recommender has ranked NCI Thesaurus as the ontology with the highest frequency (341) followed by SNOMED CT (140) and RadLex (37). Figure 3 shows an excerpt of the result from processing 761 terms using the recommender we have developed.

(a)	1	Keywords	Ontology Name	Score
	2	liver	NCIT	0.91
	3	abdomen	NCIT	0.882
	4	gallbladder	NCIT	0.938
	5	us		0 0
	6	pancreas	NCIT	0.847
	7	spleen	NCIT	0.911
	8	kidneys	OMIM	0.779
	9	portal vein	NCIT	0.868
	10	cbd	CCO	0.89
	11	duct	NCIT	0.883
	12	dilatation	SNOMEDCT	0.85
	13	lesion	NCIT	0.879
	14	yr old male	NCIT	0.906
	15	duct dilatation	MP	0.78
	16	portal	NIFSTD	0.855
	17	size	NCIT	0.851

(b)	1	Ontology Name	Frequency
	2	NCIT	341
	3	SNOMEDCT	140
	4	RADLEX	37
	5	MESH	30
	6	LOINC	21
	7	RCD	19
	8	MEDDRA	12
	9	OMIM	8
	10	CRISP	8
	11	SWEET	8
	12	CCO	7
	13	DCM	7
	14	SOPHARM	6
	15	NIFSTD	4
	16	NMR	4
	17	HUPSON	4

**Fig. 3.** (a) Ontology recommendation for each term, (b) Ranking of ontology recommended

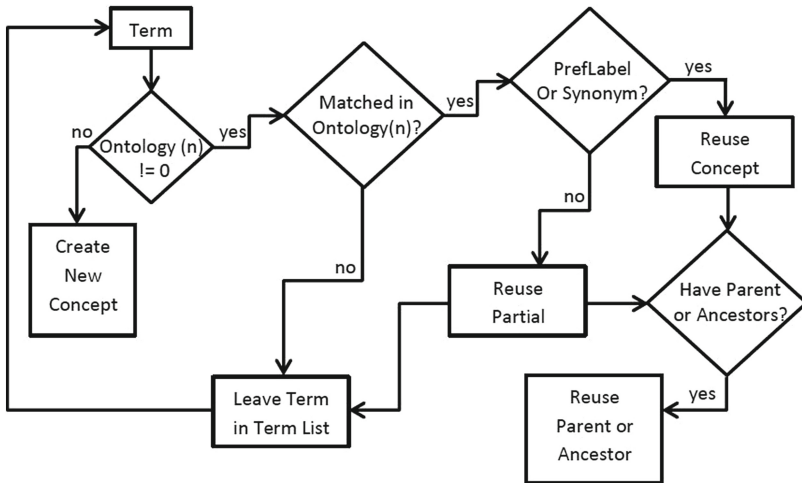
### 3.3 Term to Concept Mapping

Once the ontology for reuse has been selected, the next step in building the abdominal ultrasound ontology is to map the terms extracted to concepts in the ontology which is done by referring to the result from BioPortal’s Search API. The API allows us to insert several parameters to perform concept search which in this case, the parameters used are “q” to specify the term that we would like to search for, and “ontologies” which specifies the ontology where we would like to look for the term. Once these parameters have been submitted, the API will return a concept if there is a match with the term submitted. The concept will be returned with several other properties such as the preferred label, definition, synonym, match type and the terms relationship with its children, descendant, parents and ancestors.

In previous works by Mejino et al. [9] and Shah et al. [13], term to concept mapping was done by referring to the existing ontology and mapping it into the new one by deleting and adding concepts in the ontology to make it complete. Using BioPortal’s API consumes less time and work as the terms are queried according to the provided parameters. This will also ensure the accuracy

of the relationship between concepts and its children, descendant, parents and ancestors since there are links that can be clearly seen in the API result.

There was an intention to auto populate these data into Protégè (the OWL editor that was used in this research) by taking advantage of the option of saving the results in XML compared to JSON. However, there are two reasons why this is not possible at the moment. The first reason was that data from the API does not give the complete properties of a concept. For example, parents and ancestors were provided as links which makes it hard for the data to be manipulated since the properties of the parents and ancestors can only be obtained after the link is visited. The second reason is there are terms which matched several concepts in the ontology. For example, the term “calculus” could mean “branch of mathematics concerned with calculation” or “an abnormal concretion occurring mostly in the urinary and biliary tracts, usually composed of mineral salts”. Thus, it is important to know in which context it is being used in order to adopt the correct meaning.



**Fig. 4.** Term to concept mapping guide

In deciding whether a term should be reused or not, the term to concept mapping guide (see Fig. 4) was used. Firstly, a term from the term list will be queried using the Search API in the ontology with the highest frequency which in this case is NCIT. If there is a match, we will see whether the match is a preferred label, synonym or partial match. A preferred label (PrefLabel) match means that the API found a concept that has an exact match to the term while synonym match means that the term is found as a synonym to the concept. Partial match on the other hand means that there is no exact match for the term but there are at least two concepts that match the term. For example, for the term “intrahepatic biliary”, there are no concepts that match the term exactly.

However, there is the concept “intrahepatic” which is an anatomy qualifier in NCIT and the concept “duct” in NCIT which is an organ that matches.

If the match is a PrefLabel or synonym match, the concept will be reused. If the match is partial, the concepts that make up the term will also be reused. However, the term would still remain in the term list so that it could be compared to concepts in other ontologies. After the concept has been reused, we will find out if the term has a parent or ancestors. If there is, the parent or ancestors will also be reused. Once all terms have been searched, this process will then be repeated for the remaining recommended ontologies.

In this research, all terms are first searched in NCIT followed by SNOMED CT and RadLex. The Abdominal Ultrasound Ontology modelling follows the modelling of NCIT since it is the main ontology being reused. When merging ontologies from SNOMED CT and RadLex into the ontologies reused from NCIT, we would first find a parent that would be suitable for the concept. If no such parent exists, the parent and ancestors of the concept will then be reused. This is done to ensure the modularity of the ontology developed. If no match is found in any of these ontologies, a new concept will then be created with the help of domain experts. The main objective of using this ontology reuse methodology is to achieve as much coverage as possible and reduce the need for domain experts in developing the ontology.

### 3.4 Ontology Evaluation by Domain Expert

Once a complete Abdominal Ultrasound Ontology has been developed using the ontology reuse methodology, it is important that the ontology be evaluated by a domain expert in order to verify that the relationship between the terms as well as their definitions are correct. In evaluating this ontology, we have sat down together with a domain expert and went through the ontology. There are

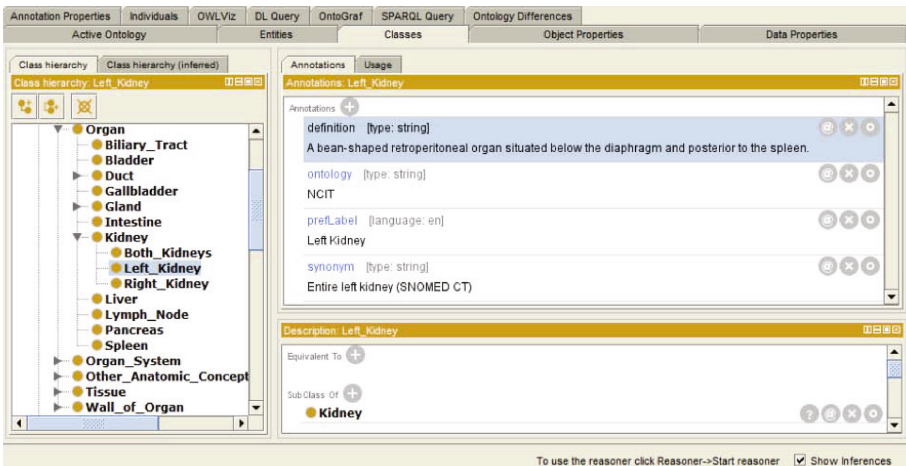


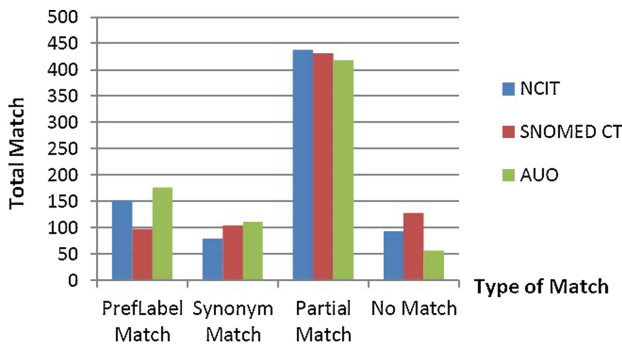
Fig. 5. Snapshot of the Abdominal Ultrasound Ontology



some corrections that need to be done but overall, the domain expert believes that the 92.6% ontology coverage is enough to cover all the important concepts that an abdominal ultrasound report would need. For the other 7.4% terms that have no match in the ontology, some of it were caused by human error whereby spelling mistakes were made by the reporter. As for the rest of it, the domain expert will help in giving definitions and suggestions on where it would fit in the ontology. Out of the 7.4% terms that have no match in the ontology, there are also several terms that the domain expert believes we can omit since these words should not be in an ultrasound report for good practice. Examples of such words are “comet tail”, “NAD”, and “hepatopetal”. Figure 5 shows a snapshot of the complete Abdominal Ultrasound Ontology.

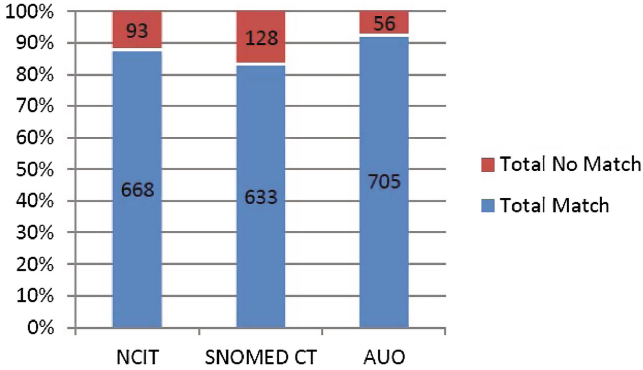
## 4 Result and Discussion

The ontology reuse methodology used to develop the Abdominal Ultrasound Ontology (AUO) has given the highest number of concept match compared to using only one ontology. This can be proved by performing a term to concept matching using the 761 terms extracted from the sample ultrasound report corpus. Figure 6 shows the comparison of total matches according to type (PrefLabel match, synonym match, partial match and no match) between NCIT, SNOMED CT and AUO. Between NCIT and SNOMED CT, NCIT has the higher concept match total with 151 PrefLabel matches, 79 synonyms matches and 438 partial matches. SNOMED CT on the other hand has only 98 PrefLabel matches, 104 synonyms matches, 431 partial matches, and 128 no matches. AUO on the other hand has 176 PrefLabel matches, 111 synonym matches, 418 partial matches, and 56 no matches.



	NCIT	SNOMED CT	AUO
PrefLabel Match	151	98	176
Synonym Match	79	104	111
Partial Match	438	431	418
No Match	93	128	56

**Fig. 6.** Breakdown of total match according to type against NCIT, SNOMED CT and Abdominal Ultrasound Ontology (AUO) (Color figure online)



**Fig. 7.** Percentage of total match and no match in NCIT, SNOMED CT and AUO (Color figure online)

104 synonyms matches and 431 partial matches. The reason SNOMED CT has lower PrefLabel matches compared to synonyms is because of its naming convention. For example, the preferred label for “kidney” is “kidney structure” and “entire gallbladder” for “gallbladder”. When writing report, radiologist often used simpler words like “kidney” and “gallbladder” instead of “kidney structure” and “entire gallbladder” thus, when term to concept matching was performed, SNOMED CT returned more synonym matches compared to PrefLabel.

Compared to NCIT and SNOMED CT, AUO returns the highest total match where it has 176 PrefLabel matches, 111 synonym matches and 418 partial matches. The reason AUO returns the most number of matches is because the ontology reuse methodology selects the best match from different ontologies and merge it into the AUO. Its exhaustive mapping in several ontologies based on the ontology rank has ensured that almost all terms in the corpus are covered by AUO. Whenever possible, a PrefLabel match will be inserted in the ontology. If not, a synonym match will be added then only partial matches are included to ensure the ontology has a wide coverage of the corpus.

From the analysis, it can be concluded that it is better to reuse from several ontologies compared to just one. This is because reusing several ontologies offers better term coverage compared to reusing just one. Figure 7 shows the percentage of total match and no match in all three ontologies. If ontology reuse was done by mapping the 761 terms against NCIT, there will only be an 87.8 % of coverage. If the mapping were done against SNOMED CT, the percentage of coverage would be only 83.2 % which is lower than NCIT. However, the percentage of coverage increases to 92.6 % when several ontologies were reused; which in this case are NCIT, SNOMED CT, and RadLex.

The percentage of no match is also very small (7.4 %) which means that the AUO covers almost all the terms in the corpus. After ontology evaluation with domain expert, the percentage of no match has been reduced to only 5 % after the domain expert included new concepts which before this have no match in any of the other ontologies being reused. The reason there is still 5 % of no match is

because there are several term in the corpus that the domain experts believe are poor usage of terms to describe findings in an ultrasound report. The domain expert believes that this is bad practice and the medical ultrasound experts are now slowly cutting down the usage of such words thus making it irrelevant to be in the AUO. Another reason for the 5 % of no match is spelling errors made by ultrasound reporters. This is not a concern for now but for future work, we could consider using the ontology to also correct and understand these errors.

NCIT has a total of 113,794 classes while SNOMED CT has 316,031 classes. However, there are only 668 and 633 matches respectively for each NCIT and SNOMED CT regarding abdominal ultrasound terminology. On the other hand, AUO has only 509 classes which is less than 0.5 % of either NCIT or SNOMED CT but still managed to have 705 matches which is more than the matches NCIT and SNOMED CT each gets. This is because of the specialization of the ontology. Since the ontology has an intended purpose in an application, it is much better and more efficient to build a domain specific ontology through reuse. It definitely would not be efficient to store a large ontology such as NCIT and SNOMED CT and use only less than 0.3 % of it. This is because it would take a lot of storage space and it will also slow down the application since the application will need to go through the whole ontology to find a match. Thus the better way to develop an ontology based application is to build a new domain specific ontology through ontology reuse methodology.

## 5 Conclusion

Ontology reuse can be beneficial in developing domain specific ontologies for application system whereby it reduces development time and redundancy. The lack of proper methodology and tools in reusing ontology has hindered this effort. Thus, this paper proposed a methodology to reuse ontology together with supporting tools that would make the ontology reuse process much easier. The development of AUO using this methodology has proven that ontology reuse is beneficial in developing a small domain specific ontology which has wide coverage of the terminology used in the application system compared to using a large general domain ontology. It is hoped that the proposed ontology reuse methodology would encourage more usage of ontology in medical system without the development of similar domain ontologies that would cause redundancy.

## References

1. Alani, H.: Position paper: ontology construction from online ontologies. In: Proceedings of the 15th International Conference on World Wide Web, WWW 2006, pp. 491–495. ACM, New York (2006)
2. Boland, G.: Enhancing the radiology product: the value of voice-recognition technology. *Clin. Radiol.* **62**(11), 1127 (2007)
3. Bontas, E.P., Mochol, M., Tolksdorf, R.: Case studies on ontology reuse. In: Proceedings of the 5th International Conference on Knowledge Management IKNOW 2005 (2005)

4. Caldarola, E.G., Picariello, A., Rinaldi, A.M.: An approach to ontology integration for ontology reuse in knowledge based digital ecosystems. In: Proceedings of the 7th International Conference on Management of Computational and Collective Intelligence in Digital EcoSystems, pp. 1–8. ACM (2015)
5. Capellades, M.A.: Assessment of reusability of ontologies: a practical example. In: Proceedings of AAAI 1999 Workshop on Ontology Management, pp. 74–79. AAAI Press (1999)
6. Jonquet, C., Musen, M.A., Shah, N.H.: Building a biomedical ontology recommender web service. *J. Biomed. Semant.* **1**(S1), 1 (2010)
7. Lossio-Ventura, J.A., Jonquet, C., Roche, M., Teisseire, M.: Biotex: a system for biomedical terminology extraction, ranking, and validation. In: 13th International Semantic Web Conference (2014)
8. Maedche, A., Motik, B., Stojanovic, L., Studer, R., Volz, R.: An infrastructure for searching, reusing and evolving distributed ontologies. In: Proceedings of the 12th International Conference on World Wide Web, WWW 2003, pp. 439–448. ACM, New York (2003)
9. Mejino Jr., J.L., Rubin, D.L., Brinkley, J.F.: FMA-RadLex: an application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. In: AMIA Annual Symposium Proceedings, vol. 2008, p. 465. American Medical Informatics Association (2008)
10. Noy, N.F., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.-A., Chute, C.G., et al.: Biportal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, pp. W170–W173 (2009)
11. Russ, T., Valente, A., MacGregor, R., Swartout, W.: Practical experiences in trading off ontology usability and reusability. In: Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management (KAW 1999), pp. 16–21 (1999)
12. Shah, T., Rabhi, F., Ray, P.: OSHCO: a cross-domain ontology for semantic interoperability across medical and oral health domains. In: 2013 IEEE 15th International Conference on e-Health Networking, Applications Services (Healthcom), pp. 460–464, October 2013
13. Shah, T., Rabhi, F., Ray, P., Taylor, K.: A guiding framework for ontology reuse in the biomedical domain. In: 2014 47th Hawaii International Conference on System Sciences (HICSS), pp. 2878–2887. IEEE (2014)
14. Simperl, E.: Reusing ontologies on the semantic web: a feasibility study. *Data knowl. Eng.* **68**(10), 905–925 (2009)
15. Uschold, M., Clark, P., Healy, M., Williamson, K., Woods, S.: An experiment in ontology reuse. In: Proceedings of the 11th Knowledge Acquisition Workshop (1998)
16. Zulkarnain, N.Z., Crofts, G., Meziane, F.: An architecture to support ultrasound report generation and standardisation. In: Proceedings of the International Conference on Health Informatics, pp. 508–513 (2015)

All links were last followed on January 25, 2016.