ABSTRACT

EXPLOITING BIOPORTAL AS BACKGROUND KNOWLEDGE IN ONTOLOGY ALIGNMENT

by Xi Chen

Ontology alignment (OA) is the process of taking as input two ontologies and producing mappings between the source concepts and the target concepts. Over the last few years, OA systems have made only minor improvements. To improve performance, some OA systems have included a semi-automatic matching approach which incorporates user interaction to assess low confidence mappings. This research investigates replacing the human expert with an automated expert or “oracle” that relies on specialized knowledge sources in the biomedical domain, BioPortal. BioPortal provides access to different resources including a wide variety of ontologies, classes within ontologies and mappings between the classes of different ontologies. A leading OA system LogMap has been used to evaluate the automated expert on the anatomy and Large Biomed Track of the Ontology Alignment Evaluation Initiative (OAEI). The experimental results are reported and show that the automated expert has a positive impact in the Large Biomed Track with four out of six of the track’s matching tasks having better OA standard performance measure for F-measure. In the Anatomy Track, using the automated expert improves the OA standard performance measure for precision. However, to the detriment of the recall measure, the result is a slight improvement in the F-measure.
EXPLOITING BIOPORTAL AS BACKGROUND KNOWLEDGE IN ONTOLOGY ALIGNMENT

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Computer Science

Department of Computer Science and Software Engineering

by

Xi Chen

Miami University

Oxford, Ohio

2014

Advisor________________________________

Valerie Cross, Ph.D.

Reader________________________________

Dhananjai Rao, Ph.D.

Reader________________________________

Ernesto Jimenez-Ruiz, Ph.D.
# TABLE OF CONTENTS

1 INTRODUCTION ........................................................................................................... 1

2 ONTOLOGY ALIGNMENT TECHNIQUES ........................................................................ 5
   2.1 General architecture of OA systems ........................................................................... 6
   2.2 Matching without external resources ......................................................................... 7
   2.3 Matching using background knowledge source ......................................................... 10
   2.4 State of the art OA systems ....................................................................................... 13
   2.5 Evaluation .............................................................................................................. 19

3 RELATED RESEARCH ................................................................................................. 20
   3.1 Background Knowledge ......................................................................................... 20
   3.2 Background Knowledge Resources Introduction ..................................................... 21
   3.3 Use Cases ............................................................................................................. 24

4 BIOPORTAL AS BACKGROUND KNOWLEDGE ......................................................... 26
   4.1 Overview of the algorithm incorporating BioPortal ................................................ 26
   4.2 Modifications to LogMap Architecture ..................................................................... 29
   4.3 Determining an upper bound .................................................................................... 30
   4.4 Using the BioPortal ................................................................................................ 32
   4.5 Extracting Mappings from BioPortal ........................................................................ 34
   4.6 Extract synonyms from BioPortal ............................................................................ 35
   4.7 Evaluation and dissemination .................................................................................. 38

5 EXPERIMENTS ON “PERFECT ORACLE” ............................................................... 39
   5.1 Anatomy Track (Mouse - Human) ............................................................................ 39
   5.2 Large BioMed track ............................................................................................... 40
   5.3 Summary ............................................................................................................... 46

6 EVALUATION WITH BIOPORTAL ............................................................................. 48
   6.1 Anatomy Track ....................................................................................................... 48
   6.2 Large Biomed Track .............................................................................................. 50
   6.3 Summary ............................................................................................................... 55

7 CONCLUSIONS AND FUTURE WORK ..................................................................... 58

REFERENCES ............................................................................................................... 60
LIST OF TABLES

Table 1: Summary on the error rate experiments ................................................................. 47
Table 2: Performance measure results of OAEI 2013 Anatomy Track using 3 experts .......... 48
Table 3: Three experts’ behaviors when examining only the 251 low confidence mappings ..... 49
Table 4: Outputs of combinations of the three experts’ when examining only the 251 low confidence mappings ................................................................. 49
Table 5: Final results using OAEI 2013 Anatomy Track data set ........................................ 50
Table 6: F-measure for OAEI 2013 Large BioMed tracks ....................................................... 50
Table 7: Precision for OAEI 2013 Large BioMed tracks ....................................................... 51
Table 8: Recall for OAEI 2013 Large BioMed tracks ............................................................ 51
Table 9: Performance measure results of 485 low confidence mappings using 3 experts ....... 52
Table 10: Performance measure results of 1,124 low confidence mappings using 3 experts ...... 52
Table 11: Performance measure results of 3,625 low confidence mappings using 3 experts ..... 53
Table 12: Performance measure results of 834 low confidence mappings using 3 experts ...... 53
Table 13: Performance measure results of 1,264 low confidence mappings using 3 experts ..... 54
Table 14: Performance measure results of 2,489 low confidence mappings using 3 experts ..... 55
Table 15: Evaluation of low confidence mappings for Anatomy ............................................. 55
Table 16: Evaluation of low confidence mappings for small fragments of biomedical track ..... 56
Table 17: Evaluation of low confidence mappings for large ontologies of the biomedical track 56
LIST OF FIGURES

Figure 1: Ontology Matching Process [2] ........................................................................................................... 1
Figure 2: Two simple ontologies and its alignment [2] .......................................................................................... 5
Figure 3: General Ontology Alignment Architecture [17] .................................................................................... 6
Figure 4: using a mediating ontology [24] .......................................................................................................... 12
Figure 5: Main components of YAM++ [14] ....................................................................................................... 14
Figure 6: System architecture of AgreementMaker [7] .......................................................................................... 15
Figure 7: the matching process in GOMMA [27] ................................................................................................. 16
Figure 8: Stage 1: Maximizing recall: computing candidates ................................................................................. 17
Figure 9: Stage 2: Maximizing precision: assessing candidates .............................................................................. 18
Figure 10: Fragment of WordNet Concept Hierarchy [33] .................................................................................... 22
Figure 11: xml result of a category query ........................................................................................................... 23
Figure 12: Number of mappings in different ontologies in BioPortal [16] ............................................................. 27
Figure 13: High level description of LogMap [42] ................................................................................................. 29
Figure 14: Interactive Process in LogMap [42] .................................................................................................... 30
Figure 15: the partial returned content for concept “melanoma” ....................................................................... 33
Figure 16: Performance measure results of LogMap using oracle in OAEI Anatomy Track ..... 40
Figure 17: Performance measure results of LogMap using oracle in OAEI Small FMA-NCI .... 41
Figure 18: Performance measure results of LogMap using oracle in OAEI FMA-NCI ................................. 42
Figure 19: Performance measure results of LogMap using oracle in OAEI Small FMA-SNO .. 43
Figure 20: Performance measure results of LogMap using oracle in OAEI FMA-SNO.............. 44
Figure 21: Performance measure results of LogMap using oracle in OAEI Small SNO-NCI..... 45
Figure 22: Performance measure results of LogMap using oracle in OAEI SNO-NCI ............... 46
ACKNOWLEDGEMENTS

I would like to express my very great appreciation to Dr. Valerie Cross for her constructive suggestions, valuable support and enthusiastic encouragement on both this thesis and my study at Miami University. Without her supervision and constant help this thesis would not have been possible. I would also like to thank Dr. Ernesto Jimenez-Ruiz for serving as my committee member and for the constant support, brilliant comments and suggestions. In addition, I would also like to thank my committee member Dr. Dhananjai Rao for the precious advice on both research and my career. Finally, a special thanks to my family.
1 INTRODUCTION

In the artificial intelligence (AI) field, an ontology is an “explicit specification of a conceptualization” [1]. AI researchers use this term to formally represent knowledge in a specific domain. Ontology alignment is the process of finding mappings between entities in different ontologies. These mappings can specify equivalence, subset or superset relationship between the entities. As illustrated in Figure 1, the inputs are ontology $O$ and $O'$, with an input alignment $A$ which could be an initial alignment provided by a user or could be empty. Additionally, parameters (e.g. thresholds and weights) and resources (e.g. background knowledge) may be provided. The task is to output an alignment $A'$ between $O$ and $O'$.

![Ontology Matching Process](image)

Ontology alignment is a crucial process contributing to the success of the Semantic Web. Fields such as information integration, P2P information sharing, biomedicine and natural language processing also use ontology alignment techniques.

The Ontology Alignment Evaluation Initiative (OAEI) is an annual international campaign for the systematic evaluation of ontology matching systems [12, 13]. Alignment problems in OAEI are organized in several tracks involving different ontology domains. The quality of the alignments computed by an ontology alignment system (OA) is measured in terms of the well-known measures precision and recall with respect to a reference set of alignments (also called a gold standard) [2]. F-measure combines precision and recall and is usually represented as their harmonic mean. OAEI also evaluates mapping computation time and the number of satisfiable
classes obtained when reasoning with the input ontologies together with the computed mappings alignments (a measure commonly called mapping coherence).

Over the past decade, OA systems evolved to produce better alignments by using lexical, structural and logical similarity measures. To improve the OA process, string-based matchers were extended, e.g. ISUB [19], to compute the string similarities of input concepts as well as their synonyms in background or external knowledge sources such as general purpose lexicons, e.g. WordNet [20], and specialized background knowledge, e.g. UMLS Metathesaurus\(^1\) for the Anatomy Track of OAEI. The composition-based approach was also proposed to make use of background knowledge sources such as Uberon\(^2\) and Foundational Model of Anatomy\(^3\) (FMA) as mediating ontologies [21] for the Anatomy Track of OAEI. Here source and target concepts are first mapped to the intermediate background ontology. If an exact common match exists in the mediating ontology, a mapping is made between source and target concepts.

This thesis research also uses background knowledge sources but in a different manner than previous OA research. For OA systems that permit human user interaction in the alignment process, an automated expert or “oracle” replaces the human expert. This automated expert relies on specialized knowledge sources in the biomedical domain, BioPortal [5, 22]. BioPortal provides access to more than 370 biomedical ontologies, synonyms, and mappings between ontology entities so that the OA system does not have to pre-select a specific biomedical ontology for the automated expert. Synonyms and mappings between ontology entities can be queried via the web portal. It also provides a RESTful web service\(^4\) for convenient utilizing its resources. A challenge of using background knowledge sources is determining the best one to use in the alignment process. By using this resource, the full range of the provided ontologies, including Uberon and many of the ontologies integrated in the UMLS Metathesaurus, can be accessed.

\(^{1}\) http://www.nlm.nih.gov/research/umls/
\(^{2}\) http://uberon.github.io/
\(^{3}\) http://sig.biostr.washington.edu/projects/fm/AboutFM.html
\(^{4}\) http://data.bioontology.org/documentation
This thesis research makes the following contributions:

1) BioPortal has not been exploited in the context of the OAEI. This thesis is the first to examine the use of BioPortal as a generalized yet also specialized background knowledge source for the OAEI Anatomy Track and Large Biomedical Ontology Track.

2) The first OAEI interactive track was added in 2013 and simulated a perfect human expert providing answers to interactive OA systems to determine how well they could perform. This thesis research has developed the first automated expert or oracle for the biomedical domain that accesses BioPortal and provides a reliable, i.e., with low error rate, validation criterion using the information (e.g. synonyms and mappings) provided by BioPortal The automated expert makes a judgment on the correctness of mappings between source and target concepts for which there is some uncertainty and decides whether to include them in the final alignment results.

3) This thesis uses a two-step approach to evaluate the performance of the automated expert. First LogMap, a top-performing systems in the last OAEI competition which has an interactive mode for an expert user were used in simulations performed on the Anatomy Track’s and Large Biomed Track’s reference alignments with a variable error rate in response. Simulation were done to understand the upper bound benefits of using an automated expert and for a comparison evaluation benchmark. Then experiments with LogMap and the automated expert were performed using the Anatomy Track and the Large Biomedical Ontology Track.

4) Though there is not a significant improvement in OA standard performance measure for F-measure, the work of this thesis offered BioPortal as an independent framework from the OA. This independent framework can be easily plugged into different OA systems.

The rest of this thesis is organized as follows. Section 2 introduces the general architecture of OA systems, matching techniques used in OA systems and some state of the art OA systems. Section 3 reviews related work using background knowledge. Section 4 describes the thesis work including overview of algorithms, implementation approach and the design of experiments. Section 5 describes the experiments on “perfect oracle”. Section 6 presents the experiment results and analysis of on OAEI 2013 results in both Anatomy Track and Large Biomedical
Ontology Track using BioPortal. Section 7 summarizes the thesis work and discusses possible future research work.
2 ONTOLOGY ALIGNMENT TECHNIQUES

Ontology alignment (OA) is the process of finding relationships (i.e. equivalent, more general superclass and subclass and disjointness) between entities in different ontologies. Classes (or concepts) are the principal entities in an ontology although many OA systems also include properties as entities for alignment purposes. The output of OA is called an alignment. Figure 2 illustrates example mappings in an alignment.

The blue thick arrows show the relationships between ontology $O_1$ and $O_2$, where the symbol = denotes equivalent relationship, $\sqsubseteq$ denotes less general ($\sqsupseteq$ denotes more general) and $\perp$ denotes disjointness which is not shown in the above figure.

OA has its root in knowledge representation and reasoning (KRR), which is a field of artificial intelligence (AI). The following parts of this section discuss the general architecture of OA system, techniques used in OA, the OA evaluation methodology on alignment results and state of the art in OA systems.
2.1 General architecture of OA systems

The general architecture of an OA system illustrated in Figure 3 consists of five components: 1) Selecting features of describing entities; 2) Choosing entity pair to align; 3) Calculating similarity for entity pairs; 4) Aggregating similarity results for a single entity pair; and 5) Alignment Decision.

Iteration is normally involved in the architecture since new mappings can affect previously decided mappings or provide more information for new mappings. Over the past decade, many OA systems have emerged and although each step is implementation specific, most all follow this general OA architecture.

Input:

Two or more ontologies to be aligned and an initial input alignment may be included.

1. Feature Engineering:
This step is to gain as much information as possible, either from the input ontology files or external resources, e.g. labels, properties, derived features, aggregated features, complex axiom, external features and so on.

2. Search Step Selection
Choose the entities of input ontologies to compare, either all entities from input ontologies or entities of the same type (i.e. concepts, relations, and instances).
3. Similarity Computation
This step computes the similarity between entities. For every feature extracted from feature engineering, a similarity is computed. For example, to compute similarity for labels, string similarity methods can be applied. There are typically many different features and methods used to compute similarity and they are referred to as matchers.

4. Similarity Aggregation
The multiple similarity measures produced in step 3 must be aggregated to produce one overall similarity measure, often referred to as a confidence degree for the mapping. One common approach is to use linear weighted combination of the similarity measures.

5. Interpretation
A threshold is used to filter the mappings whose confidence measure is not sufficient to produce it as mapping in the final alignment. More specifically, if the confidence degree of the mapping is greater than the threshold, it is added to the final alignment.

6. Iteration
Iteration occurs when new mappings are discovered because these mappings can affect the similarities of their neighboring entity pairs. It may stop under many circumstances. A common stop condition is when no new mappings are discovered.

Output:
The output is a representation of mappings in the alignment. Below are examples of output mappings:
Book $\sqsupseteq_{1.0}$ Essay Man $\perp_{1.0}$ Woman Football $=_{0.9}$ Rugby

2.2 Matching without external resources
Initial OA systems and a few current ones do the alignment process using only information gathered from the two input ontologies. No other external ontologies or background knowledge
is used in the alignment process. But many of the early OA systems have been updated to use external background knowledge sources. In this section the focus is on the different types of matchers that are used on the information contained in the two input ontologies.

2.2.1 String-based techniques
String-based techniques calculate the similarity score between two strings. The score depends on the strings’ sequences of characters only. They are widely used techniques in ontology alignment systems and can be categorized into three families: edit-based similarity, token-based similarity and hybrid similarity [18].

- **Edit-based similarity.** It calculates a score based on the number of operations need to transform the one string to the other, either by deleting, inserting or replacing a character. For example, **Levenstein distance** [39] can be defined as the following recursive function:

\[
\text{lev}_{a,b}(i,j) = \begin{cases} 
\max(i,j) & \text{if } \min(i,j) = 0, \\
\text{lev}_{a,b}(i-1,j) + 1 & \\
\min \left\{ \begin{array}{l}
\text{lev}_{a,b}(i,j-1) + 1 \\
\text{lev}_{a,b}(i-1,j-1) + 1_{(a_i \neq b_j)} 
\end{array} \right. & \text{otherwise.}
\end{cases}
\]

Where \( i \) and \( j \) are 1-based indices in the string \( a \) and \( b \) respectively, although the values of \( i \) and \( j \) can be 0, which means empty character. The score of string \( a \) and \( b \) is determined by how similar two strings are. The deletion, insertion and substitution of a character will all cause one penalty of the score. **Needleman–Wunsch** [40] algorithm is similar to Levenstein distance except that the deletion, insertion and substitution operations are assigned different penalty scores. **ISUB** is string matcher [19] where both the same and different characters are considered and calculated for the score. The score of ISUB is between 0 and 1. The advantages of ISUB string matcher is its ease of implementation and its proven good performance in ontology alignment.

- **Token-based similarity.** This matcher splits strings into tokens first and then calculates similarity scores between the individual tokens for the sets of tokens for the two strings. In most cases, strings can be tokenized either by special characters, e.g. ‘_’ in
“dorsal_motor_nucleus_vagus”, or by upper case letters, e.g., “MolarTooth”. The algorithms to compute the token similarity score and to compute the similarity score between two token sets impact the final similarity result of a token-based similarity matcher. The similarity score is normally in [0.0, 1.0], where 0.0 means absolutely different and 1.0 means totally the same. The ISUB string matcher can be used to calculate the string similarity score. To compute the overall similarity between two sets, one frequently used method is the set-based Jaccard similarity [43] measure which is simply the ratio of the number of strings in common to the number of total unique strings between the two sets.

2.2.2 Language-based techniques
Unlike string-based techniques which consider a string as a sequence of characters, language-based techniques consider the meaning of the string (i.e. words of the text). The algorithms used in these techniques are mainly inherited from Natural Language Processing (NLP).

- **Stemming.** Stemming removes the suffix of the word, keeping only the stem. For example, matching, matches and matched all become match, and am, is and are turn to be, after stemming.

- **Stopword removal.** Stopwords are those words such as articles, prepositions and conjunctions, which do not have actual meanings. For example, a car becomes car after stopword removal.

However, some word forms cannot be processed using stemming without an external source, e.g. irregular verbs (go and went) and irregular plural (mouse and mice). In this case, a dictionary such as Wordnet [20] must be used to find the base of the word. Using external resources is introduced in section 2.3.

Once the base of the word is identified, the preceding string-based techniques can be applied to calculate the similarity score.
2.2.3 Structure-based techniques
Structure-based techniques exploit the internal structure of the ontologies to perform the alignment. The internal structure consists of the entity name, annotations, properties and relations to other entities. In structure-based techniques, an ontology is considered as a graph or a tree structure. For example, using a graph structure, nodes are entities and edges are relations among them. Each edge is labelled by relation names (i.e. equivalence, more general superclass and subclass and disjointness). The similarity of two nodes is based on their position in the graph. If two nodes are similar, their neighbor nodes must have some similarity. Thus, finding an alignment between ontologies is related to finding a maximum common directed sub graph [2].

2.2.4 Instance-based techniques
Instanced-based techniques exploit the instance classified in the source ontology and target ontology. The similarity of two classes is determined by the number of common instances that they contain, the individual number of instances, and the total number of instances [4]. The most frequently used metrics are Jaccard similarity and K-Statistics [44].

2.3 Matching using background knowledge source
External resources are also referred as background knowledge resources (sources). Unlike the traditional OA process that limits the use of information only provided by the input ontologies, matching using background knowledge resource utilizes external resources to help identify matching between two entities. The use of a background knowledge source can be categorized into three uses, as a reference, as an oracle and as a mediator.

2.3.1 as a reference
A reference background knowledge source could be used in finding and validating a mapping. For example, when trying to map two entities, accessing a reference could provide descriptions and definitions that can be used in a matcher to return a similarity score for the two entities. Many linguistic resources such as lexicons, thesauri and terminologies, can be exploited and used as references for the alignment process.
Lexicons. A lexicon is a set of words used by a human with definitions of these words. It is also called a dictionary.

Thesauri. A thesaurus is a specialized lexicon where similar meanings of words are grouped together. Synonyms are usually contained in a thesaurus.

Terminologies. A terminology is a kind of thesauri where phrases instead of words are included. A terminology is usually domain specific and is hierarchically organized.

These linguistic resources are exploited mainly to find synonyms and thus, increase the certainty of the mapping for two entities. To use these resources, a similarity score needs to be computed between words. A score of 1.0 typically means they are synonyms. The primary reference background knowledge source for ontology alignment in a general knowledge domain has been the English lexical database WordNet. Several OA systems calculate semantic similarity measures between two English lexical concepts in WordNet that are related to the ontology entities and use that in determining a degree of mapping between the ontology entities. A good overview of semantic similarity measures and their use in OA systems is provided in [24].

Semantic similarity algorithms designed for WordNet are generally applicable to other semantic network resources [23, 24], e.g. MeSH (Medical Subject Headings). The known algorithms to compute semantic similarity based on WordNet are Path, WuPalmer and LeacockChodorow, which are categorized as edge or path-based similarity measures, and Resnik, Lin and JiangConath, which are categorized as information-content similarity measures. According to [4], the information-based similarities Resnik and its extension Lin and JiangConath outperform edge-based similarity measures using WordNet as a reference background knowledge source. Similar results have been shown in [24, 25], where Resnik measure outperforms other semantic similarity measures using FMA and Uberon as background knowledge source in the Anatomy Track of the OAEI.
2.3.2 as an oracle
Some OA systems support a semi-automatic process by allowing interaction with a human expert to provide judgment on mappings that have low confidences, i.e. more uncertainty about the mapping. The OA system presents these mappings to the expert and based on the expert’s response decides to keep or discard the low confidence mapping. This new OA capability has just recently been evaluated at the OAEI 2013 competition by introducing an interactive track. The purpose of this track is to determine if such user interaction could improve the final alignment results. This interactive track evaluated the participating OA systems by using an oracle that relied on the reference alignment from the conference track.

Background knowledge sources could serve as input to an automatic expert module that replaces the human expert. This automatic expert queries the background knowledge and analyzes the responses in order to provide a judgment on a low confidence mapping.

2.3.3 as a mediator
Background knowledge sources such as other ontologies that are in the same domain as the input ontologies being aligned can be used as an intermediate or mediating ontology between the two. In this case, typically the concepts in the two input ontologies, the source and target, are first quickly mapped to the mediating ontology. If a source concept and a target concept map to exactly the same concept in the mediating ontology, then a mapping is established between those concepts. Figure 4 demonstrates the use of a mediating ontology in OA process.

Figure 4: using a mediating ontology [24]
$O_S$ is the source ontology and $O_T$ is the target ontology. $O_I$ is the mediating ontology. $M_{SI}$ is the set of mappings between the source ontology and the mediating ontology. These two sets of mappings can be produced by simple matcher such as string matcher. Similarly, $M_{TI}$ is the set of mappings between the target ontology and the mediating ontology. $M_{ST}$ is a set of mappings created based on either the exact match or semantic similarity greater than a threshold on bridge concepts $b_s$ and $b_t$ [24].

### 2.4 State of the art OA systems

One challenge to ontology alignment is the appropriate use of background knowledge [3]. In [4], using the web as background knowledge was shown to help improve the alignment results in terms of recall and precision. Some ontology alignment systems already used background knowledge (BK) to improve the matching results. For example, GOMMA [8] first demonstrated that using Uberon as a mediating ontology could improve the results for the Anatomy Track. AgreementMaker [7] also quickly followed and incorporated the use of Uberon for the Anatomy Track. LogMap [6] uses normalizations and spelling variants from the general (biomedical) purpose UMLS Lexicon. YAM++ [14] uses WordNet as a general purpose resource. WikiMatch [9] and BLOOMS+ [10] use Wikipedia as their background knowledge. Wikipedia is a large knowledge base containing around 23 million articles across various domains and contains a topic hierarchy that can be viewed as an ontology.

The following sections provide overviews of several of the leading OA systems. The LogMap OA system is used in this research for several reasons: speed of execution, open source system, modular architecture and extremely helpful support from it software developer. Since it is being used, its overview provides a more detailed description.

### 2.4.1 YAM++
YAM++ [14] is an acronym for yet another matcher. The main components of YAM++ are illustrated in Figure 5. The workflow of YAM++ is as follows:

1. Ontology Parser component loads and parses input ontologies.
2. Annotation Indexing and Structure Indexing components index class labels and URIs.
3. Candidate Filtering component filters out high confidence candidate mappings.
4. Terminological Matcher components uses information retrieval techniques to produce a set of mappings from the annotations of entities.
5. Instance-based Matcher component discovers new mappings via shared instances.
6. The results of step 4 and 5 are aggregated into element level matching results. Structural Matcher component enhances the element level matching result by investigating structural information from entities.
7. The Combination & Selection component combines and selects the results from the three matchers and produces the final mapping result.
8. Semantic Verification components remove the inconsistent mappings.

2.4.2 AgreementMaker
AgreementMaker [7] is an OA system developed at University of Illinois at Chicago. It supports various matchers. The architecture is shown as Figure 6.
Figure 6: System architecture of AgreementMaker [7]

The most significant feature of AgreementMaker is that its architecture is extensible. Thus new matchers can be easily plugged into it. AgreementMaker uses several concept-based matchers that utilize multiple string similarity measures and a structural-based matcher that searches for shared patterns in the hierarchical structure of the ontologies. Its concept-based matchers are the Base Similarity Matcher (BSM), the Advanced Similarity Matcher (ASM), the Parametric String-based Matcher (PSM), and the Vector-based Multi-Word Matcher (VMM). For the BSM the similarity between two concepts is determined by comparing all the strings associated with those two concepts. The set of strings for a concept include the concept name, label, and comments. PSM is also a string-based matcher but more complex with a combination of a substring measure and an edit distance measure. For VMM, each concept has a virtual document created for it by concatenating the strings of related concepts and annotations. It then transforms the resulting strings into TF-IDF vectors and computes the cosine similarity measure between the vectors for two concepts. The Descendent’s Similarity Inheritance matcher (DSI) is the structural matcher and examines the ancestors of the two concepts in a mapping. Its heuristics is that if two concepts are mapped with high similarity, then the mapping similarity of their descendants should increase. AgreementMaker uses its Linear Weighted Combination (LWC) matcher to produce a single combined alignment by using mapping quality measures to select the best mappings from each matcher.
2.4.3 GOMMA
The Generic Ontology Matching and Mapping Management (GOMMA), is a component-based ontology managing and analyzing system [27]. It supports the management of a variety of ontology-based mappings that describes how these ontologies are related to each other. GOMMA consists of three levels, the tool level, the functional component level and the repository level. The tool level contains tools that are utilized by the GOMMA infrastructure and its functionality. The functional component level can access ontology versions, entity sources and mapping data. The repository level calculates the semantic similarity scores and determines mappings between two input ontologies.

GOMMA has components for ontology alignment, Figure 7 illustrates the matching process in GOMMA.

To find alignments between input ontologies, GOMMA first integrates them into the GOMMA repository. Then, numerous matchers (e.g. Linguistic Matcher, Path Matcher, Annotation-based Concept Matcher and etc.) are selected from Matcher Library. GOMMA can execute these matches in parallel. The results from these matchers are aggregated into a single similarity matrix in Similarity Cube. The final step is to filter most likely alignment according to the selected filter strategy. The alignment results are stored in the mapping pool. In addition, GOMMA supports user interaction to help identify mappings.
2.4.4 LogMap

LogMap [6], which stands for logic-based methods for ontology mapping is an ontology matching system developed at the University of Oxford. It has built-in reasoning (i.e. HermiT reasoner) and it considers the logic-based semantics of input ontologies. In LogMap, an inverted index is used. An inverted index is a data structure that maps words to the location in a sentence. The inverted index makes LogMap extremely efficient when matching large scale ontologies.

LogMap supports both non-interactive mode where heuristics are used to identify the mappings and interactive mode where domain experts can be queried to identify the low confidence mappings. OAEI results reported from recent years show that LogMap achieves high quality results in several of the different tracks.

The main algorithms of LogMap can be divided into two stages: Maximizing recall, which computes candidates and tries to find as many of the mappings as possible and maximizing precision, which assesses candidates and tries to eliminate as many incorrect mappings as possible.

The process of stage 1 is illustrated in Figure 8. In lexical indexation, the inverted index data structure is used to store class labels and URIs extracted in each input ontology. Then the candidate class mapping is computed from the inverted indices. As long as a source entry and a

![Figure 8: Stage 1: Maximizing recall: computing candidates](image-url)
target entry contain the same words, a candidate mapping is created in this step. Though most of the candidate mappings turn out to be incorrect, the recall is maximized while the number of candidates remains manageable. Next, candidate property mapping is computed using the string matcher ISUB. The last step of stage 1 is to extract logic-based module. It ensures all superclasses (super-properties) of a class (property) in a candidate mapping are included. This last step is based on locality of reference since if a candidate mapping exists then there might be possible mappings between the superclasses of the mapped entities.

The process of stage 2 is illustrated in Figure 9. The purpose of this stage is to maximize the precision. Reliable mappings are identified first. Reliable mappings must have high confidence values between the source and target concepts as calculated by ISUB and the source and target concepts’ superclasses and subclasses must be in reliable mappings with each other. Then propositional encoding is performed to detect un-satisfiable classes. Most of these un-satisfiable classes, although not all, are fixed in the ontology repair step. Semantic indexation is performed afterwards, where interval labelling schema [41] is used to index the Horn propositional representation of the ontology modules and the mappings. Next, “non-reliable” candidate mappings are discarded by LogMap’s heuristics. For some mappings that are difficult for LogMap to identify as “reliable” or not, a domain expert can be queried to help in determining reliable mappings. After that, the final class mappings are examined and the repair process is used on unsatisfiable mappings. The last step is to discard candidate property mappings.

Figure 9: Stage 2: Maximizing precision: assessing candidates
**2.5 Evaluation**

OAEI uses Recall, Precision, F-measure, Runtime and the Incoherence Degree to evaluate the quality of the alignments and the performance of the OA systems.

**Recall** is the proportion of the number of correct alignments generated by the OA system to the total number of alignments in the reference ontology. Alignments in the reference ontology are called reference alignment. They are created by domain experts and is known as the gold standard alignment. Recall can be formulated as follows:

\[
R(A, R) = \frac{|R \cap A|}{|R|}
\]

Where \( R \) is set of reference alignment and \( A \) is the mappings found by OA system.

**Precision** is the proportion of the number of correct alignments generated by the OA system to the total number of alignments the OA system generated. It can be formulated as follows:

\[
P(A, R) = \frac{|R \cap A|}{|A|}
\]

**F-measure** is calculated by combining recall and precision using the following formula:

\[
f - \text{measure} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}
\]

**Runtime** is the total running time of an OA system to complete a task.

The **Incoherence Degree** is the ratio between the number of unsatisfiable classes and the size of ontologies plus mappings. A class \( C \) is unsatisfiable in \( O \) if \( O \) entails the axiom \( O \sqsubseteq \bot \). An unsatisfiable class appears because the alignments are given formal semantics, and reasoning with the input ontologies and mappings may lead to undesired logical consequences such as unsatisfiability [45].
3 RELATED RESEARCH

String-based and structure-based similarity measures fail to discover some mappings due to inability to explore the semantics of the ontologies. The OA community then began to exploit background knowledge to improve the OA process [28, 29, 30]. For example, YAM++ uses general purpose background knowledge source WordNet to look up synonyms; AgreementMaker uses Uberon as a mediating ontology for the Anatomy Track; GOMMA and LogMap use UMLS Metathesaurus, specialized biomedical background knowledge for the Anatomy Track of OAEI. WikiMatch [31] and BLOOM+ [32] utilize Wikipedia to exploit its hierarchical categories for the ontology alignment.

The rest of this section is organized as follows. First, the general use of background knowledge (BK) is discussed. Then, frequently used BK resources are introduced. Last, three use cases of BK in OA systems are described.

3.1 Background Knowledge

Background knowledge is the additional resource(s) used in OA process besides source ontology and target ontology. It is a complement to string-based and structure-based similarity measures. String-based similarity measures compute the string similarity score of two entities’ label in the two input ontologies to discover mappings. And structure-based similarity measures exploit the relations to determine similarities based on the structure of the ontologies. The limitation of these two measures is that mappings without any lexical or structural similarity cannot be discovered. For example, there is no lexical similarity between the concepts paper and article. The structural similarity of these two concepts might be missing in input ontologies. In this situation, using background knowledge can help the discovery of the mapping.

However, two issues need to be taken into account. 1) How to select background knowledge? 2) How to make use of semi-structured resources? In [29], Marata et al proposed exploiting online semantic resources to bypass aforementioned two issues. Thus, their method neither needed
manually select background knowledge nor needed transform the background knowledge in an structured resource.

3.2 Background Knowledge Resources Introduction

3.2.1 WordNet
WordNet is a large lexical database of English language [20]. It is one of the most frequent used background knowledge. English words in WordNet are grouped into sets of synonyms (called synsets in WordNet). Each synset represents a lexical concept. As a meaningful sentence consist of a bunch of meaningful words, computers need the information of words and their meanings in order to do natural language processing as human do. Traditionally, this kind of information is given by dictionaries. Though machine-readable dictionaries are available, they are not very convenience of machines. WordNet fills this gap.

WordNet represents a word form by using a string of ASCII characters and a sense by the set of synonyms. There are more than 118,000 different word forms and over 90,000 different word senses in WordNet. For example, a word form “bat” can have several different senses, i.e., the sense of an animal or the sense of a piece of sports equipment. It also contains inflectional morphology for every syntactic category via its interface to its database. WordNet has over 116,000 semantic relations between its lexical concepts.

The WordNet hierarchy is from general to specific and it is illustrated by the fragment shown in Figure 10.
The most general concept “artefact” is in the root node of the hierarchy tree, while the more specific concept such as truck and hatch-back are in the leaf nodes. The hierarchical relations between synsets permits the various semantic similarity measures to be used to determine how similar one concept is to another. For example hatch-back and compact have the high semantic similarities since there is not much semantic distance between them.

The WordNet system is composed of four parts: the source files of WordNet lexicographers; the software that turns the source files into the lexical database of WordNet; The WordNet lexical database; and tools for accessing WordNet database [20].

WordNet provides a cost-free and well-documented open code to the community. It is widely used in a variety of fields of researches. It is an ideal tool for disambiguating the word senses, retrieving information and tagging semantic [34]. A limitation of WordNet is that it originally only was for the English language. Extensions of WordNet for other languages have been produced so that multilingual WordNets are emerging.
3.2.2 Wikipedia

Wikipedia is a free encyclopedia that contains around 23 million articles across various domains. It is a large knowledge base. Articles in Wikipedia are written by volunteers all over the world in 285 languages. Articles in Wikipedia are organized in categories. For example, the entry of the article “music” is under the categories entertainment, music and performing arts. Wikipedia also provides API\(^1\) to conveniently access to wiki features, data and meta-data. Figure 11 illustrates a result in xml format to query the categories of the article “music”.

```xml
<query-continue>
    <categories ga=continue="18539|Pages_containing_links_to_subscription-only_content"/>
</query-continue>

<page pageid="10897656" ns="14" title="Category:All articles with dead external links" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T02:52:10Z" lastrevid="614976617" counter="0" length="563"/>
<page pageid="9329647" ns="14" title="Category:All articles with unsourced statements" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T03:18:35Z" lastrevid="616937812" counter="0" length="571"/>
<page pageid="18159588" ns="14" title="Category:All pages needing factual verification" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T02:24:19Z" lastrevid="592443808" counter="0" length="328"/>
<page pageid="30245994" ns="14" title="Category:Articles with dead external links from January 2011" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T04:55:45Z" lastrevid="547246442" counter="0" length="41"/>
<page pageid="26636568" ns="14" title="Category:Articles with hAudio microformats" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T03:04:12Z" lastrevid="547222213" counter="0" length="253"/>
<page pageid="34209012" ns="14" title="Category:Articles with unsourced statements from January 2012" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T01:52:19Z" lastrevid="547288976" counter="0" length="41"/>
<page pageid="22908191" ns="14" title="Category:Articles with unsourced statements from May 2007" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T02:54:26Z" lastrevid="604207999" counter="0" length="611"/>
<page pageid="693018" ns="14" title="Category:Entertainment" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T01:09:10Z" lastrevid="570603463" counter="1" length="40"/>
<page pageid="691484" ns="14" title="Category:Music" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T01:25:08Z" lastrevid="552634213" counter="1" length="143"/>
<page pageid="25284144" ns="14" title="Category:Pages containing site templates with deprecated parameters" contentmodel="wikitext" pagelanguage="en" touched="2014-07-28T00:00:00Z" lastrevid="606797265" counter="1" length="307"/>
</page>
</query>
</api>

Figure 11: xml result of a category query

\(^1\) http://en.wikipedia.org/w/api.php
3.2.3 UMLS
The Unified Medical Language System (UMLS) is a collection of biomedical vocabularies developed and maintained by US National Library of Medicine [38]. It consists of a set of files and software that can facilitate the interoperability between different computer systems. There are three Knowledge Sources in UMLS:

**Metathesaurus:** It contains more than 1,000,000 biomedical concepts and 5,000,000 concept names from over 100 vocabularies such as MeSH and SNOMED CT.

**Semantic Network:** It is a catalog of broad categories and their relationship. Each concept in Metathesaurus belongs to one or more categories. The links among these categories are their relationship (semantic relations).

**SPECIALIST Lexicon and Lexical Tools:** The SPECIALIST Lexicon contains lexical information (i.e. syntax, morphology and orthography) for more than 330,000 common English words and biomedical terms. Lexical tools are a set of Java programs that can be used for Natural Language Processing (NLP) applications.

3.3 Use Cases

In this part, three use cases of different BK resources in the OA process are briefly presented.

3.3.1 Use of WordNet
Several potential ways exist to use WordNet in the OA process [36]. One possible way is to enrich the synonyms of concepts’ labels by finding in WordNet synset of each term included in the label [37]. Another way is to compute semantic similarity of two concepts either using edge-based semantic similarity measure such as WuPalmer or information-based semantic similarity measure such as Resnik.

3.3.2 Use of Wikipedia
Wikipedia is used by BLOOMS+ [26] to build a set of category hierarchy trees for each concept in the input ontologies. For each class from the input ontologies, after tokenization and stop
words removal, the modified class label is used as a search string. BLOOMS+ calls Wikipedia’s API using the search string. Pages returned by Wikipedia are used to as the root of each category tree. The immediate children of the root are all Wikipedia categories that these pages belonged to. The grandchildren are super categories of each immediate children. BLOOMS+ limits the depth of the category tree to 4 based on their empirical observation. Based on the set of category hierarchy trees, BLOOMS+ is able to compute the similarity score between each concept.

### 3.3.3 Use of UMLS

LogMap stores class labels and URIs from source ontology and target ontology using inverted index, which uses a set of words to represent the label or URI of a class. To enrich the index, each label or URI of a class is preprocessed, e.g. tokenize words, stopwords removal, stemming, normalization, etc. LogMap uses the UMLS SPECIALIST Lexicon for stemming, normalization and spelling variants.
4 BIOPORTAL AS BACKGROUND KNOWLEDGE

The objective of this thesis research is to answer the following questions with respect to the selected Ontology Alignment System (OA) and Background Knowledge (BK) Resource: (1) Does the OA using BK improve the results of the original OA? (2) Does the OA using BK outperform other OA system in the OAEI? (3) Determine the performance of the OA-BK with respect to simulated error rates using a perfect BK, i.e., a reference alignment for OAEI biomedical tracks?

4.1 Overview of the algorithm incorporating BioPortal

The method to extend OA with BK includes the following strategies:

- **Enriching the ontologies with synonyms.** The BK resource is used to enrich the vocabulary of the input ontologies. The main challenge of this strategy relies on its scalability since, when the input ontologies become large, the computation time might grow considerably.

- **Validation of borderline alignments.** Some of the alignments computed by the OA system may not be clear-cut cases; that is, they represent uncertain alignments which should require an extra (semi-automatic) assessment. In this strategy, BK is to function as a component of the “domain expert”. The main challenge relies on the design and implementation of a reliable (i.e. with a low error rate) validation criterion to access the information provided by BK.

4.1.1 Candidate OA

LogMap has been chosen as the candidate OA to be used with the automated domain expert. LogMap has built-in reasoning and considers the logic-based semantics of input ontologies. It is extremely efficient when matching large scale ontologies. It supports both non-interactive mode where heuristics are used to identify the mappings and interactive mode where domain experts can be involved to identify the low confidence mappings. The workflow of LogMap is summarized in Figure 8 and Figure 9 in section 2.4.4.
OAEI results reported from recent years show that LogMap achieves high quality results in several different tracks. For example, in the final results of OAEI 2013, regarding F-measure, LogMap ranked 5th out of 20 OAs in Anatomy Track. However, LogMap currently has very limited use of background knowledge.

4.1.2 Candidate BK
BioPortal\(^1\), a general purpose resource for the biomedical domain, has been chosen the BK resources. BioPortal is a web portal providing access using Web browsers and Web services to its hundreds of biomedical ontologies. It supports different formats of ontologies, including RDF, OWL, OBO format and Protégé frames [5, 15].

The main motivation of using BioPortal is that biomedical ontologies are becoming a mainstream technology in biomedical information systems. For example, BioPortal provides more than 360 ontologies, including the Uberon ontology has already been pre-selected for use in other systems as a mediating ontology. BioPortal has over 6 million classes and more than 30,000 mappings among 20 ontologies. Figure 12 shows the number of mappings between several of these ontologies.

Figure 12: Number of mappings in different ontologies in BioPortal [16]

\(^{1}\) http://bioportal.bioontology.org/
4.1.3 Candidate Algorithms

To better describe the high-level algorithm that incorporates BioPortal to LogMap, this thesis defines three “experts”, the Mapping Expert, the Synonym Expert, and the Heuristic Expert.

**The Mapping Expert:** This expert extracts mappings from BioPortal. Its detail algorithm is described in section 4.4.

**The Synonym Expert:** This expert extracts synonyms from BioPortal to enrich the synonyms of input ontologies. Its detail algorithm is described in section 4.6.

**The Heuristic Expert:** This expert uses LogMap’s heuristic techniques, i.e. the behavior of LogMap in non-iterative mode to identify low confidence mappings. To accept a low confidence mapping, the scope score (i.e. the lexical similarity among the subclasses/superclasses of two entities) needs greater than 0.0 and the lexical similarity of two entities need greater than 0.5.

To identify a low confidence mapping, unless all three experts give the same answers (i.e. all accept the mapping or all discard the mapping), there exists various combinations as candidate algorithms. This thesis work analyzes all possible combinations. The following 4 combinations are selected to illustrate using the three experts.

**The Mapping Expert OR the Synonym Expert OR and the Heuristic Expert:** The logic OR is used to identify a mapping. For each low confidence mapping, it is asked to the Mapping Expert, synonyms expert and Heuristic Expert in sequence. If one of the expert accepts the low confidence mapping, this mapping is accepted immediately. If both Mapping Expert and the Synonym Expert discard the mapping, the Heuristic Expert finally determines whether to accept it or discard it. This is reasonable because BioPortal may not have the information of the low confidence mapping regarding their mapping and synonyms. In this situation, LogMap’s heuristics are used to make the final decision.

**Vote (the mapping, the synonym, and the Heuristic Experts):** If two experts accept (or discard) the low confidence mapping, it is accepted (or discarded). This method assumes that all three experts are equally authoritative. Thus, the majority wins.

**The Mapping Expert AND the Synonym Expert AND the Heuristic Expert:** The logic AND is used to identify a mapping. This combination is most strict since to accept a low confidence mapping, all three experts need to accept it.
The Mapping Expert OR the Synonym Expert: Only the Mapping Expert and Synonym Expert are used. If either accepts a low confidence mapping, the mapping is included in the final alignment. This strategy only trusts using BioPortal since the Heuristic Expert is not involved.

4.2 Modifications to LogMap Architecture

LogMap can work in interactive mode and automatic mode [42]. In interactive mode, LogMap asks a small number of questions to a user (e.g. a domain expert) to help it decide low confidence mappings. In automatic mode, the whole mapping process works automatically by using the Heuristic Expert. Figure 13 illustrates a high level algorithm of LogMap.

Input: $O_1, O_2$: input ontologies; Interact: Boolean value  
Output: $M$: mappings.

1: $(LI_1, LI_2) := \text{LexicalIndexes}(O_1, O_2)$  
2: $M_? := \text{CandidateMappings}(LI_1, LI_2)$  
3: $(O'_1, O'_2) := \text{Module}(O_1, O_2, M_?)$  
4: $M := \text{ReliableMappings}(M_?)$  
5: $M_? := M_? \setminus M$  
6: $(P'_1, P'_2) := \text{PropEncoding}(O'_1, O'_2)$  
7: $M := \text{Diagnosis}(P'_1, P'_2, M, \emptyset)$  
8: $SI := \text{SemanticIndex}(P'_1, P'_2, M)$  
9: $M_? := M_? \setminus \text{Discarded}(LI_1, LI_2, SI, M_?)$
10: if (Interact = true) then  
11: $M_{\text{user}} := \text{InteractiveProcess}(SI, M_?)$  
12: $M := \text{Diagnosis}(P'_1, P'_2, M \cup M_{\text{user}}, M_{\text{user}})$
13: else  
14: $M := \text{Diagnosis}(P'_1, P'_2, M \cup M_?, M)$
15: end if
16: return $M$

Figure 13: High level description of LogMap [42]

Line 1 to line 3, LogMap computes the candidate mappings. The goal is to maximize the recall. The remaining lines to line 9 assess the candidate mappings. The goal is to maximize the precision. Line 11 is the optional call to interactive process. Figure 14 describes this interactive process.
**Procedure** InteractiveProcess

**Input:** $SI$: semantic index; $M_\gamma$: mappings to revise;

**Output:** $M_{user}$: user-selected mappings;

1. $\checkmark$ \text{FilterAmb} := Filter by Ambiguity?
2. \textbf{while} Interact = true and $M_\gamma \neq \emptyset$ \textbf{do}
3. \hspace{1em} $M_\gamma := \text{ComputePartialOrder}(M_\gamma)$
4. \hspace{1em} $\checkmark \langle m, action \rangle := \text{AssessMappingFromList}(M_\gamma)$
5. \hspace{1em} $M_\gamma := M_\gamma \setminus \{m\}$
6. \hspace{1em} \textbf{if} (FilterAmb = true) \textbf{then}
7. \hspace{2em} $\text{Aux} := \text{Ambiguous}(m, M_\gamma)$ \textbf{and} $M_\gamma := M_\gamma \setminus \text{Aux}$
8. \hspace{1em} \textbf{end if}
9. \hspace{1em} \textbf{if} (action = addition) \textbf{then}
10. \hspace{2em} $M_{user} := M_{user} \cup \{m\}$
11. \hspace{2em} $M_\gamma := M_\gamma \setminus \text{Conflict}(SI \cup \{m\}, M_\gamma)$
12. \hspace{1em} \textbf{else if} (FilterAmb = true) \textbf{then}
13. \hspace{2em} $M_{user} := M_{user} \cup \text{Aux}$
14. \hspace{1em} \textbf{end if}
15. \hspace{1em} $\checkmark$ Interact := Stop interaction?
16. \hspace{1em} \textbf{end while}
17. \hspace{1em} \textbf{if} $M_\gamma \neq \emptyset$ \textbf{then} $M_{user} := M_{user} \cup \text{applyHeuristics}(M_\gamma)$
18. \hspace{1em} \textbf{return} $M_{user}$

Figure 14: Interactive Process in LogMap [42]

The symbol $\checkmark$ in Figure 14 indicates the user interaction can be applied. Line 3 organizes mappings in the set $M_\gamma$ in a partial order according to their confidence value (i.e. normally the lowest similarity will be asked to users first). Users can accept or decline the given mapping (line 4). Line 15 indicates that users can stop the interaction, thus the remaining mappings in the set $M_\gamma$ are decided heuristically by LogMap.

To exploit BioPortal as the expert user (described in section 4.1.3), a separate module developed and consisting of the Mapping Expert and synonyms expert replaces “action” in line 4. Thus, the decision to accept or decline a mapping is decided by the automated experts.

### 4.3 Determining an upper bound
The importance of determining an upper bound on evaluation using BioPortal is to provide a maximum value of the best possible improvement in OA performance measures when using the automated expert and to evaluate the approximate error rate of LogMap. Some OA systems such as LogMap have a built-in semi-automatic matching approach to incorporate user interaction to identify mappings to increase their performance. For example, LogMap in interactive mode queries an expert user (or oracle) to assess low confidence mappings determined by LogMap’s matchers.

To evaluate the upper bound of LogMap, a reference alignment is used as the “perfect oracle”; that is, for each low confident mapping, LogMap asks to the “perfect oracle” to provide an accept or decline decision on the mapping. To simulate an oracle with different error rates, the correct answer provided by the “perfect oracle” is randomly reversed based on the probability used for the different error rates (see Algorithm 1)

Algorithm 1: Checks if mapping is in local oracle based on the error_rate

**Procedure** isMappingInLocalOracle

**Input:** e1(entity1), e2(entity2) and error_rate

**Output:** true/false

1: if mapping e1, e1 exists in the reference ontology then
2:     if (a random number in [0, 1) < 1 - error_rate/100) then
3:         return true;
4:     else
5:         return false;
6: endif
7: endif

For each error rate, several experiments need to be run to get the average precision, recall, F-measure. The number of questions is also recorded. The test results are described and analyzed in Section 5.
The results of the error rate experiments provide an estimated error rate of the target OA system. For example, an error rate with the closest F-measure to the F-measure of OA is the proximal error rate of the OA system. The error rate experiment can be also adopted to evaluate a rough error rate using a specific background knowledge, for example, BioPortal.

4.4 Using the BioPortal

BioPortal provides a REST API for convenient access to its rich biomedical ontologies, synonyms, mappings, etc. The BioPortal REST API consists of numerous resources and related information bound together via links. The content returned by the REST API can be either in a JSON format or an XML format. BioPortal REST API uses JSON as its default returned content type. For example, to search the concept “melanoma”, the URL is http://data.bioontology.org/search?q=melanoma&exact_match=true. Figure 15 illustrates the partial returned content.
Figure 15: the partial returned content for concept “melanoma”
In Figure 15, “malignant melanoma” and “Naevocarcinoma” are listed under “synonym:” label can be extracted to enrich the synonyms of input ontologies. The link with label “@id” is its mappings. To further exploit the mappings, the link under the label “mappings” can be navigated to extract more mapping information. Note that Figure 15 merely shows one entry of “melanoma” due to space limitation. There are many more similar entries.

4.5 Extracting Mappings from BioPortal

By extracting mappings from BioPortal, LogMap can better identify which low confidence mappings should be accepted. The Mapping Expert performs the extraction of existing mappings between ontology concepts in different ontologies in BioPortal. For a mapping between entity1 in the source ontology and entity2 in the target ontology to be verified, BioPortal is queried to find existing mappings for each entity within BioPortal ontologies. The results for the source entity and target entity are stored into two files, JsonFile1 and JsonFile2 respectively. The algorithm is described as follows:

Algorithm 2: Verifying mappings by extracting mappings from BioPortal

Procedure VerifyUsingMappings
Input: e1(entity1), e2(entity2) and threshold
Output: true/false

(Level1:)
1: JsonFile1 := call BioPortal RESTAPI to search exact match of e1
2: set1 := extract “@id” links from JsonFile1
3: JsonFile2 = call BioPortal RESTAPI to search exact match of e2
4: set2 = extract “@id” links from JsonFile2
5: if set1 ≠ ∅ and set2 ≠ ∅ then
6: if JaccardIndex(set1, set2) ≥ threshold then
7: return true
8: end if
9: end if
(Level2:)
10:  mappingSet1 := open all the URLs of "mappings" in JsonFile1 and extract "@id" links
11:  mappingSet2 := open all the URLs of "mappings" in JsonFile2 and extract "@id" links
12:  if mappingSet1 ≠ ∅ and mappingSet2 ≠ ∅ then
13:      if JaccardIndex(mappingSet1, mappingSet2) ≥ threshold then
14:         return true
15:      end if
16:  end if
17:  return false;

The Jaccard Index is used to determine how similar two sets A and B are and is defined as follows:

\[ J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]

If only one of the sets A and B is empty, the Jaccard Index is 0 (if both sets are empty, it is defined to return 1). The Jaccard Index is 1 if sets A and B are identical. The Jaccard index is calculated between the two sets of mappings in the files Jsonfile1 and Jsonfile2. A threshold is used in the algorithm to control the mapping validation condition. Level 2 in the algorithm is to further explore the mappings via BioPortal. It opens all the links under the "mappings" label and extracts each of their "@id" label as the mappings. Compared to level 1 algorithms, level 2 might get more mappings. However, the computation time increase significantly.

4.6 Extract synonyms from BioPortal

Synonyms extracted from BioPortal is used to enrich the synonyms of input ontologies. Thus, it helps LogMap better identify low confidence mappings when it is worked in interactive mode. The role of BioPortal using this approach is called Synonyms Expert. The algorithm is described in algorithm 3.
Algorithm 3: Verifying mappings by extracting synonyms from BioPortal

Procedure VerifyUsingSynonyms
Input: $e_1$ (entity1), $e_2$ (entity2), threshold1 and threshold2
Output: true/false
1: if newLexicalSimilarity($e_1$, $e_2$) > threshold1 && newScope($e_1$, $e_2$) > threshold2
2:   return true;
3: end if
4: return false;

To identify a mapping to be good, both the new lexical similarity (algorithm 3.1) and the new scope (algorithm 3.2) of the two entities must greater than the thresholds.

The algorithm for calculating the new lexical similarity is described in algorithm 3.1.

Algorithm 3.1: Calculate the new lexical similarity of entities

Procedure newLexicalSimilarity
Input: $e_1$ (entity1), $e_2$ (entity2)
Output: double
1: set1 := check local dictionary for synonyms of $e_1$
2: if set1 = φ
3:   se1 := ask BioPortal to provide the synonyms of $e_1$
4:   add the mapping $e_1$:set1 to local dictionary
5: end if
6: do the same for $e_2$
7: return getIUSBSimilarity($e_1$, $e_2$)

To get the synonyms from BioPortal, BioPortal REST API is called. By filtering the “synonym” labels in the returned json file, the set of synonyms of the concept can be obtained. The local dictionary enables the queried synonyms stored locally, and avoids repeatedly making expensive
query to BioPortal for the same concepts. For example, the Anatomy Track is still completed in a reasonable time when the local cache is not applied although the running time is doubled. However, when experiments run on the Large Biomedical Ontologies Track, it cannot be completed in a reasonable time. When the number of questions grows, though the total number of synonyms acquired from BioPortal grows approximately linearly, calls to BioPortal grow quadratically in worst case due to repeatedly asking for the synonyms from BioPortal for the enriched synonyms.

After extracting synonyms from BioPortal, the synonyms of the input entities are enriched. The ISUB [19] method described in section 2.2.1 is applied to compute the similarity score between two synonym’s sets. The maximum ISUB score calculated for the two sets of synonyms is kept to represent the final string similarity score of the two entities.

Once the new lexical similarity score is greater than a threshold, the scope (i.e. the lexical similarity among the subclasses/superclasses of e1 and e2) also needs to be computed. The algorithm to compute the new scope of e1 and e2 is described in algorithm 3.2.

**Algorithm 3.2: Calculate the new scope of entities**

**Procedure** newScope

**Input:** e1(entity1), e2(entity2)

**Output:** double

1: get the synonyms of e1’s superclasses and subclasses
2: get the synonyms of e2’s superclasses and subclasses
3: \[\text{score1} := \max(\text{ISUB score of all combinations in two superclasses})\]
4: \[\text{score2} := \max(\text{ISUB score of all combinations in two subclasses})\]
5: \[\text{return } \frac{\text{score1} + \text{score2}}{2.0}\]

To avoid substantial overhead, the number of superclass levels and subclass levels is limited to 3. Again, to calculate the scope score, the ISUB string matcher is used. The intuition is that for two concepts to be considered mapped, their superclasses or subclasses must be mapped.
4.7 Evaluation and dissemination
The target evaluation scenarios are those used in the OAEI evaluation campaign. The results
obtained in previous campaigns by the selected OA are used as a baseline, specifically, the
Anatomy and the Large Biomedical Ontology Track. The ontologies for these two tracks are real
world biomedical ontologies.

The dissemination activities involve three main scientific communities: the Artificial
Intelligence, Semantic Web and Bioinformatics communities. The International Joint Conference
on Artificial Intelligence (IJCAI), the European Conference on Artificial Intelligence (ECAI),
the Extended Semantic Web Conference (ESWC), the International Semantic Web Conference
(ISWC) and the Journal of Biomedical Informatics are prestigious conferences or journals in
these areas, and they have been the target venues to disseminate the results of the proposed
research.
5 EXPERIMENTS ON “PERFECT ORACLE”

This thesis uses datasets from the Anatomy Track and the Large Biomedical Ontology (Large Biomed) Track of OAEI 2013. The task of Anatomy Track is to find mappings (1516 correct mappings) between the Adult Mouse Anatomy (2744 classes) and a part of the Human Anatomy (3304 classes). The tasks of the Large BioMed Track are to find mappings between the Foundation Model of Anatomy (FMA) (78,989 classes), SNOMED CT (122,464 classes), and the National Cancer Institute Thesaurus (NCI) (66,724 classes), as well as finding alignments between small fragments (5% to 36%) of the whole dataset.

Experiments on “perfect oracle” provide a maximum value of the best possible improvement in OA performance measures when using the automated expert. LogMap uses the reference alignment as a “perfect oracle” to assess low confidence mappings it found for each task in the Anatomy Track and the Large Biomed Track. As described in the Algorithm 1 in Section 4.3, to simulate an “oracle” with different error rates, the correct answer provided by the “perfect oracle” is randomly reversed based on the probability used for the different error rates.

5.1 Anatomy Track (Mouse - Human)

In the Anatomy Track of OAEI 2013, there are 251 low confidence mappings found by LogMap. Fig 16 reports the Precision, Recall and F-measure scores when LogMap runs in interactive mode using the “oracle” with error rates ranging from 0% to 40%.

Fig 16 shows that when the “oracle” has an error rate 30%, LogMap achieves 0.878 F-measure score, which is the same F-measure score when LogMap runs in non-interactive mode. That is, the Heuristic Expert has a 30% error rate. To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 30%. Otherwise, the automated expert cannot improve the OA system.

Fig 16 also indicates that the F-measure score increases almost uniformly to the decrement of error rate (i.e. the F-measure score increase about 0.5% per 5% decrement of the error rate).
F-measure score is 0.908 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.908. However, the top systems in the OAEI’s Anatomy Track, which use Uberon as the pre-selected mediating ontology, reach F-measure values greater than 0.92. This limitation is partially due to the techniques implemented within LogMap, but also to the limitation in the number of questions that should be asked to the expert user or oracle.

![Performance measure results of LogMap using oracle in OAEI Anatomy Track](image)

**Figure 16:** Performance measure results of LogMap using oracle in OAEI Anatomy Track

### 5.2 Large BioMed track

Similar experiments are done in the Large BioMed Track. Fig17 to Fig 22 show performance measure results under different error rates for Large BioMed Track.
Small FMA_NCI: Small FMA_NCI represents the small fragment of FMA (3,696 classes) and a small fragment of NCI (6,488 classes). Fig 17 shows that when the “oracle” has an error rate 20%, LogMap achieves 0.915 F-measure score, which is the same F-measure score when LogMap runs in non-interactive mode. That is, the Heuristic Expert has a 20% error rate. To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 20%. Otherwise, the automated expert cannot improve the OA system. Fig 17 also indicates that the F-measure score increases almost uniformly to the decrement of error rate (i.e. the F-measure score increase about 0.5% per 5% decrement of the error rate). The F-measure score is 0.934 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.934.

Figure 17: Performance measure results of LogMap using oracle in OAEI Small FMA-NCI

FMA_NCI: This task consists of 78,989 classes of whole FMA and 66,724 classes of whole NCI ontologies. Fig 18 shows that when the “oracle” has an error rate 20%, LogMap achieves 0.825 F-measure score, which is close to the F-measure score (0.831) when LogMap runs in non-
interactive mode. That is, the Heuristic Expert has an approximate 18% error rate (between 15% and 20%). To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 18%. Otherwise, the automated expert cannot improve the OA system. Fig 18 also indicates that when an “oracle” improves the error rate from 10% to 5%, there is a big jump improvement (2.1%) of the F-measure score (from 0.859 to 0.880). The F-measure score is 0.898 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.898.

![P/R/F WITH DIFFERENT ERROR RATE (FMA-NCI)](image)

**Figure 18**: Performance measure results of LogMap using oracle in OAEI FMA-NCI

**Small FMA_SNOMED**: Small FMA_SNOMED represents the small fragment of FMA (10,157 classes) and a small fragment of SNOMED (13,412 classes). Fig 19 shows that when the “oracle” has an error rate 25%, LogMap achieves 0.796 F-measure score, which is close to the F-measure score (0.792) when LogMap runs in non-interactive mode. That is, the Heuristic Expert has an approximate 27% error rate (between 25% and 30%). To outperform the Heuristic Expert,
the automated expert needs to have an error rate less than 27%. Otherwise, the automated expert cannot improve the OA system. Fig 19 also indicates that the F-measure score increases almost uniformly to the decrement of error rate (i.e. the F-measure score increase about 0.7% per 5% decrement of the error rate). The F-measure score is 0.836 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.836.

**Figure 19**: Performance measure results of LogMap using oracle in OAEI Small FMA-SNO

**FMA_SNOMED**: This task consists of 78,989 classes of whole FMA and 122,464 classes of a large SNOMED fragment. Fig 20 shows that when the “oracle” has an error rate 10%, LogMap achieves 0.710 F-measure score, which is close to the F-measure score (0.711) when LogMap runs in non-interactive mode. That is, the Heuristic Expert has an approximate 10% error rate. To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 10%. Otherwise, the automated expert cannot improve the OA system. Fig 20 also indicates that when an “oracle” improves the error rate from 10% to 5%, there is a big jump improvement (2.7%) of
the F-measure score (from 0.710 to 0.737). The F-measure score is 0.764 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.764.

![P/R/F WITH DIFFERENT ERROR RATE (FMA-SNOMED)](image)

Figure 20: Performance measure results of LogMap using oracle in OAEI FMA-SNO

**Small SNOMED-NCI:** Small SNOMED-NCI represents a small fragment of SNOMED (51,128 classes) and a small fragment of NCI (23,958 classes). Fig 21 shows that when the “oracle” has an error rate 25%, LogMap achieves 0.773 F-measure score, which is close to the F-measure score (0.770) when LogMap runs in non-interactive mode. That is, the Heuristic Expert has an approximate 27% error rate (between 25% and 30%). To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 27%. Otherwise, the automated expert cannot improve the OA system. Fig 21 also indicates that the F-measure score increases almost uniformly to the decrement of error rate (i.e. the F-measure score increase about 0.7% per 5% decrement of the error rate). The F-measure score is 0.809 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.809.
**SNOMED_NCI:** This task consists of 122,464 classes of a large SNOMED fragment and 66,724 classes of the whole NCI ontology. Fig 22 shows that when the “oracle” has an error rate 30%, LogMap achieves 0.693 F-measure score, which is the same F-measure score when LogMap runs in non-interactive mode. That is, the Heuristic Expert has a 30% error rate. To outperform the Heuristic Expert, the automated expert needs to have an error rate less than 30%. Otherwise, the automated expert cannot improve the OA system. Fig 22 also indicates that the F-measure score increases almost uniformly to the decrement of error rate (i.e. the F-measure score increase about 1.2% per 5% decrement of the error rate). The F-measure score is 0.768 when the error rate is 0%, meaning that the upper bound on the F-measure with a perfect expert is 0.768.

![P/R/F WITH DIFFERENT ERROR RATE (Small SNOMED-NCI)](image)

**Figure 21:** Performance measure results of LogMap using oracle in OAEI Small SNO-NCI
Summary Table 1 summarizes the error rate experiments on 7 tasks in Anatomy Track and Large BioMed Track of OAEI 2013. The average error rate of the Heuristic Expert is 23%, with the lowest error rate (10%) in the FMA-SNO task and the highest error rate (30%) in the Human-Mouse task and the SNO-NCI task. The numbers in bold font in the table 1 indicates that even with a perfect expert as the automated expert, LogMap is not able to beat the top systems in the Anatomy Track and FMA-SNO and Small FMA-SNO tasks in the Large Biomed Track. These top systems, however, had access to specifically selected background ontologies. The F-measure tends to improve linearly towards the decrement of the error rate in all tasks except FMA-NCI and FMA-SNO.
<table>
<thead>
<tr>
<th></th>
<th>Appro. error rate of LogMap</th>
<th>Best possible F-measure</th>
<th>Best F-measure in OAEI 2013</th>
<th>The trend of F-measure regarding error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-Mouse</td>
<td>30%</td>
<td>0.908</td>
<td>0.942</td>
<td>0.5% improvement per 5% decrement of the error rate</td>
</tr>
<tr>
<td>Small FMA-NCI</td>
<td>20%</td>
<td>0.934</td>
<td>0.915</td>
<td>0.5% improvement per 5% decrement of the error rate</td>
</tr>
<tr>
<td>FMA-NCI</td>
<td>18%</td>
<td>0.898</td>
<td>0.872</td>
<td>2.1% jump improvement when improves the error rate from 10% to 5%</td>
</tr>
<tr>
<td>Small FMA-SNO</td>
<td>27%</td>
<td>0.836</td>
<td>0.836</td>
<td>0.7% improvement per 5% decrement of the error rate</td>
</tr>
<tr>
<td>FMA-SNO</td>
<td>10%</td>
<td>0.764</td>
<td>0.821</td>
<td>2.7% jump improvement when improves the error rate from 10% to 5%</td>
</tr>
<tr>
<td>Small SNO-NCI</td>
<td>27%</td>
<td>0.809</td>
<td>0.770</td>
<td>0.7% improvement per 5% decrement of the error rate</td>
</tr>
<tr>
<td>SNO-NCI</td>
<td>30%</td>
<td>0.768</td>
<td>0.718</td>
<td>1.2% improvement per 5% decrement of the error rate</td>
</tr>
</tbody>
</table>

Table 1: Summary on the error rate experiments
6 EVALUATION WITH BIOPORTAL

To evaluate BioPortal, LogMap had access to the automated expert and was used on the Anatomy Track and the large biomedical ontologies track. The automated expert has 3 different experts, the Mapping Expert (ME), the Synonym Expert (SE), and the Heuristic Expert (HE). The definition of the ME, SE and HE are described in the section 4.4.

6.1 Anatomy Track

Table 2 summarized the initial results for OAEI 2013 Anatomy Track gained from the experiment using the three experts separately. As can be seen, neither the ME nor the SE has a better F-measure than the HE has, although their precisions are not worse. The performance of the ME is very close to that of the HE.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Expert</td>
<td>0.913</td>
<td>0.842</td>
<td>0.876</td>
</tr>
<tr>
<td>Synonym Expert</td>
<td>0.951</td>
<td>0.799</td>
<td>0.869</td>
</tr>
<tr>
<td>Heuristic Expert</td>
<td>0.913</td>
<td>0.846</td>
<td>0.878</td>
</tr>
</tbody>
</table>

Table 2: Performance measure results of OAEI 2013 Anatomy Track using 3 experts

Table 3 is created from the running results to assist the analysis of the three experts’ behaviors when examining only the 251 low confidence mappings (of which 155 of them are in the reference alignment). The ME identifies 243 mappings, of which, 149 mappings are correct (in reference alignment), and the SE found 119 mappings and 84 of them are correct. The HE considered all the 251 mappings correct, so all the 155 correct mappings are found.
Since the HE has the best Recall but not as good Precision as the SE, various combinations of the experts using logic OR and the majority vote are tested. Note that the Logic AND is not used in Anatomy Track because it cannot get any better Recall than the lowest value 0.542. Neither the Logic OR of the experts including HE is necessary because the results cannot do any better than then the HE, which has the best F-measure for the 251 low confidence mappings. Table 4 shows the experiments with different outputs of combinations of the three experts’ when examining only the 251 low confidence mappings. Not surprisingly, Logic OR of the ME and the SE has the exact same results as Vote (ME, SE, HE) because HE accepts all the 251 low confidence mappings and the Vote (ME, SE, HE) turns to Logic OR of the ME and the SE.

Compared to LogMap’s original heuristics (i.e. the HE), majority voting method (or Logic OR of ME and SE) leaves out 3 less mapping, but only 1 of them is a correct mapping, which results in a better Precision, but the same Recall. So the F-measure improves a little. Table 5 illustrates the final results of OAEI 2013 Anatomy Track using original LogMap and LogMap plugged-in majority vote matcher.
Table 5: Final results using OAEI 2013 Anatomy Track data set

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vote (ME, SE, HE)</td>
<td>0.915</td>
<td>0.846</td>
<td>0.879</td>
</tr>
<tr>
<td>Original LogMap</td>
<td>0.913</td>
<td>0.846</td>
<td>0.878</td>
</tr>
</tbody>
</table>

6.2 Large Biomed Track

Table 6 summarizes the F-measure of OAEI 2013 Large BioMed Track using the original LogMap (i.e. the HE) and different combination of 3 experts. By using the BioPortal as BK, four out of six of the track’s matching tasks have better OA standard performance measure for F-measure. The numbers in bold in table 6 indicates the best F-measure values.

F-measure:

<table>
<thead>
<tr>
<th></th>
<th>Small FMA_NCI</th>
<th>Small FMA_SNO</th>
<th>Small SNO_NCI</th>
<th>FMA_NCI</th>
<th>FMA_SNO</th>
<th>SNO_NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>0.915</td>
<td>0.792</td>
<td>0.770</td>
<td>0.831</td>
<td>0.711</td>
<td>0.693</td>
</tr>
<tr>
<td>ME OR SE OR HE</td>
<td>0.915</td>
<td><strong>0.793</strong></td>
<td>0.771</td>
<td>0.834</td>
<td><strong>0.714</strong></td>
<td>0.692</td>
</tr>
<tr>
<td>Vote (ME, SE, HE)</td>
<td>0.914</td>
<td>0.790</td>
<td>0.771</td>
<td>0.834</td>
<td><strong>0.714</strong></td>
<td>0.692</td>
</tr>
<tr>
<td>SE</td>
<td>0.908</td>
<td>0.776</td>
<td><strong>0.776</strong></td>
<td><strong>0.843</strong></td>
<td>0.704</td>
<td>0.691</td>
</tr>
<tr>
<td>ME</td>
<td>0.912</td>
<td>0.790</td>
<td>0.757</td>
<td>0.834</td>
<td>0.710</td>
<td>0.680</td>
</tr>
</tbody>
</table>

Table 6: F-measure for OAEI 2013 Large BioMed tracks
Table 7 and Table 8 summarize the Precision and Recall of OAEI 2013 Large BioMed Track using original LogMap and different combination of 3 experts. By using BioPortal as BK, the improvement in terms of the precision is more significant than recall.

**Precision:**

<table>
<thead>
<tr>
<th></th>
<th>Small FMA_NCI</th>
<th>Small FMA_SNO</th>
<th>Small SNO_NCI</th>
<th>FMA_NCI</th>
<th>FMA_SNO</th>
<th>SNO_NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>0.949</td>
<td>0.963</td>
<td>0.894</td>
<td>0.872</td>
<td>0.874</td>
<td>0.871</td>
</tr>
<tr>
<td>ME OR SE OR HE</td>
<td>0.949</td>
<td>0.959</td>
<td>0.892</td>
<td>0.864</td>
<td>0.872</td>
<td>0.865</td>
</tr>
<tr>
<td>Vote (ME, SE, HE)</td>
<td>0.951</td>
<td>0.951</td>
<td>0.903</td>
<td>0.842</td>
<td>0.874</td>
<td>0.854</td>
</tr>
<tr>
<td>SE</td>
<td>0.965</td>
<td>0.973</td>
<td>0.957</td>
<td>0.877</td>
<td>0.896</td>
<td>0.867</td>
</tr>
<tr>
<td>ME</td>
<td>0.953</td>
<td>0.951</td>
<td>0.905</td>
<td>0.847</td>
<td>0.863</td>
<td>0.862</td>
</tr>
</tbody>
</table>

Table 7: Precision for OAEI 2013 Large BioMed tracks

**Recall:**

<table>
<thead>
<tr>
<th></th>
<th>Small FMA_NCI</th>
<th>Small FMA_SNO</th>
<th>Small SNO_NCI</th>
<th>FMA_NCI</th>
<th>FMA_SNO</th>
<th>SNO_NCI</th>
</tr>
</thead>
<tbody>
<tr>
<td>HE</td>
<td>0.883</td>
<td>0.672</td>
<td>0.677</td>
<td>0.794</td>
<td>0.600</td>
<td>0.576</td>
</tr>
<tr>
<td>ME OR SE OR HE</td>
<td>0.883</td>
<td>0.676</td>
<td>0.678</td>
<td>0.807</td>
<td>0.604</td>
<td>0.576</td>
</tr>
<tr>
<td>Vote (ME, SE, HE)</td>
<td>0.881</td>
<td>0.676</td>
<td>0.673</td>
<td>0.827</td>
<td>0.603</td>
<td>0.580</td>
</tr>
<tr>
<td>SE</td>
<td>0.859</td>
<td>0.646</td>
<td>0.652</td>
<td>0.812</td>
<td>0.580</td>
<td>0.575</td>
</tr>
<tr>
<td>ME</td>
<td>0.875</td>
<td>0.676</td>
<td>0.651</td>
<td>0.823</td>
<td>0.603</td>
<td>0.562</td>
</tr>
</tbody>
</table>

Table 8: Recall for OAEI 2013 Large BioMed tracks
Small FMA_NCI: This task consists of 3,696 classes of FMA fragment and 6,488 classes of NCI fragment. The total alignments (correct mappings to found) are 2931. There are up to 485 low confidence mappings to identify. Table 9 illustrates the performance measure results of the three experts towards these 485 low confidence mappings.

| Mapping Expert | 0.808 | 0.918 | 0.860 |
| Synonym Expert | 0.862 | 0.792 | 0.826 |
| Heuristic Expert | 0.784 | 1.000 | 0.879 |

Table 9: Performance measure results of 485 low confidence mappings using 3 experts

As can be seen, the Heuristic Expert performs quite well for these 485 low confidence mappings. Any combination of the three experts cannot outperform the Heuristic Expert alone.

Small FMA_SNO: This task consists of 10,157 classes of FMA fragment and 13,412 classes of SNOMED fragment. The total alignments are 8,941. There are up to 1,124 low confidence mappings to identify. Table 10 illustrates the performance measure results of the three experts towards these 1,124 low confidence mappings.

| Mapping Expert | 0.815 | 0.981 | 0.891 |
| Synonym Expert | 0.926 | 0.682 | 0.786 |
| Heuristic Expert | 0.876 | 0.944 | 0.909 |

Table 10: Performance measure results of 1,124 low confidence mappings using 3 experts

Though the Heuristic Expert outperforms for these 1,124 low confidence mappings, the logic OR of three experts find 139 more mappings (51 of them are correct). Thus, the recall is improved, though precision is hurt, causing the 0.1% improvement of F-measure.
Small SNO_NCI: This task consists of 51,128 classes of SNOMED fragment and 23,958 classes of NCI fragment. The total alignments are 18,476. There are up to 3,625 low confidence mappings to identify. Table 11 illustrates the performance measure results of the three experts towards these 3,625 low confidence mappings.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Expert</td>
<td>0.612</td>
<td>0.775</td>
<td>0.684</td>
</tr>
<tr>
<td>Synonym Expert</td>
<td>0.848</td>
<td>0.783</td>
<td>0.814</td>
</tr>
<tr>
<td>Heuristic Expert</td>
<td>0.626</td>
<td>1.000</td>
<td>0.770</td>
</tr>
</tbody>
</table>

Table 11: Performance measure results of 3,625 low confidence mappings using 3 experts

As can be seen, the Synonym Expert outperforms for these 3,625 low confidence mappings. It finds 1527 less mappings, but only loses 491 correct mappings, causing 6.3% improvement in precision, though 2.5% worse in recall. The Synonyms Expert has 0.6% improvement in F-measure.

FMA_NCI: This task consists of 78,989 classes of whole FMA and 66,724 classes of whole NCI ontologies. The total alignments are 2,931. There are up to 834 low confidence mappings to identify. Table 12 illustrates the performance measure results of the three experts for these 834 low confidence mappings.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Expert</td>
<td>0.591</td>
<td>0.941</td>
<td>0.726</td>
</tr>
<tr>
<td>Synonym Expert</td>
<td>0.720</td>
<td>0.861</td>
<td>0.784</td>
</tr>
<tr>
<td>Heuristic Expert</td>
<td>0.655</td>
<td>0.698</td>
<td>0.676</td>
</tr>
</tbody>
</table>

Table 12: Performance measure results of 834 low confidence mappings using 3 experts
As can be seen, both the Mapping Expert and the Synonyms Expert outperform the Heuristic Expert for these 834 low confidence mappings. The Synonyms Expert alone finds 50 more mappings, but it finds 63 more correct mappings. The result is that, both precision and recall are improved. The improvement of the F-measure is 1.2%.

**FMA_SNO**: This task consists of 78,989 classes of whole FMA and 122,464 classes of a large SNOMED fragment. The total alignments are 8,941. There are up to 1,264 low confidence mappings to identify. Table 13 illustrates the performance measure results of the three experts towards these 1,264 low confidence mappings.

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Expert</td>
<td>0.643</td>
<td>0.976</td>
<td>0.775</td>
</tr>
<tr>
<td>Synonym Expert</td>
<td>0.767</td>
<td>0.705</td>
<td>0.735</td>
</tr>
<tr>
<td>Heuristic Expert</td>
<td>0.683</td>
<td>0.936</td>
<td>0.790</td>
</tr>
</tbody>
</table>

Table 13: Performance measure results of 1,264 low confidence mappings using 3 experts

As can be seen, the Heuristic Expert outperforms the other two experts for these 1,264 low confidence mappings. However, using the vote scheme for the three expert, a 0.3% better F-measure is achieved because it finds 18 more mappings but 26 more good mappings.

**SNO_NCI**: This task consists of 122,464 classes of a large SNOMED fragment and 66,724 classes of the whole NCI ontology. The total alignments are 18,476. There are up to 2,489 low confidence mappings to identify. Table 14 illustrates the performance measure results of the three experts on these 2,489 low confidence mappings.
<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping Expert</td>
<td>0.568</td>
<td>0.735</td>
<td>0.640</td>
</tr>
<tr>
<td>Synonym Expert</td>
<td>0.644</td>
<td>0.909</td>
<td>0.754</td>
</tr>
<tr>
<td>Heuristic Expert</td>
<td>0.655</td>
<td>0.909</td>
<td>0.761</td>
</tr>
</tbody>
</table>

Table 14: Performance measure results of 2,489 low confidence mappings using 3 experts

As can be seen, the Heuristic Expert outperforms the other two experts for these 2,489 low confidence mappings. Any combination of the three experts cannot outperform the Heuristic Expert alone. For example, using the vote method for the three method finds 87 more mappings, but only 26 of them are correct mappings.

### 6.3 Summary

Table 15, 16 and 17 summarize the evaluation of low confidence mappings for Anatomy Track and Large Biomed Track. In the following discussion a “YES” indicates that the expert has accepted a low confidence mapping for the final alignment and a “CORRECT” indicates that the accepted low confidence mapping is actually correct, that is, agrees with the reference alignment.

<table>
<thead>
<tr>
<th></th>
<th>Yes</th>
<th>Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>ME</td>
<td>243</td>
<td>149</td>
</tr>
<tr>
<td>SE</td>
<td>119</td>
<td>84</td>
</tr>
<tr>
<td>HE</td>
<td>251</td>
<td>155</td>
</tr>
<tr>
<td>ME OR SE</td>
<td>248</td>
<td>154</td>
</tr>
<tr>
<td>ME OR SE OR HE</td>
<td>251</td>
<td>155</td>
</tr>
<tr>
<td>Vote (ME,SE,HE)</td>
<td>248</td>
<td>154</td>
</tr>
</tbody>
</table>

Table 15: Evaluation of low confidence mappings for Anatomy
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref Align</td>
<td>485</td>
<td>1124</td>
<td>3625</td>
</tr>
<tr>
<td></td>
<td>380</td>
<td>913</td>
<td>2269</td>
</tr>
<tr>
<td>Yes</td>
<td>Correct</td>
<td>Yes</td>
<td>Correct</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ME</th>
<th>SE</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>432</td>
<td>1099</td>
<td>2874</td>
</tr>
<tr>
<td></td>
<td>349</td>
<td>896</td>
<td>1759</td>
</tr>
<tr>
<td></td>
<td>389</td>
<td>673</td>
<td>2095</td>
</tr>
<tr>
<td></td>
<td>301</td>
<td>623</td>
<td>1777</td>
</tr>
<tr>
<td></td>
<td>485</td>
<td>984</td>
<td>3622</td>
</tr>
<tr>
<td></td>
<td>380</td>
<td>862</td>
<td>2268</td>
</tr>
<tr>
<td></td>
<td>458</td>
<td>1110</td>
<td>3368</td>
</tr>
<tr>
<td></td>
<td>368</td>
<td>907</td>
<td>2177</td>
</tr>
<tr>
<td></td>
<td>485</td>
<td>1123</td>
<td>3622</td>
</tr>
<tr>
<td></td>
<td>380</td>
<td>913</td>
<td>2268</td>
</tr>
<tr>
<td></td>
<td>458</td>
<td>1013</td>
<td>3368</td>
</tr>
<tr>
<td></td>
<td>368</td>
<td>893</td>
<td>2177</td>
</tr>
</tbody>
</table>

Table 16: Evaluation of low confidence mappings for small fragments of biomedical track

<table>
<thead>
<tr>
<th></th>
<th>FMA-NCI Low conf.</th>
<th>FMA-SNO Low conf.</th>
<th>SNO-NCI Low conf.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>834</td>
<td>1264</td>
<td>2489</td>
</tr>
<tr>
<td></td>
<td>388</td>
<td>794</td>
<td>1390</td>
</tr>
<tr>
<td>Yes</td>
<td>Correct</td>
<td>Yes</td>
<td>Correct</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ME</th>
<th>SE</th>
<th>HE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>617</td>
<td>1206</td>
<td>1798</td>
</tr>
<tr>
<td></td>
<td>365</td>
<td>775</td>
<td>1021</td>
</tr>
<tr>
<td></td>
<td>464</td>
<td>730</td>
<td>1963</td>
</tr>
<tr>
<td></td>
<td>334</td>
<td>560</td>
<td>1264</td>
</tr>
<tr>
<td></td>
<td>414</td>
<td>1088</td>
<td>1931</td>
</tr>
<tr>
<td></td>
<td>271</td>
<td>743</td>
<td>1264</td>
</tr>
<tr>
<td></td>
<td>655</td>
<td>1225</td>
<td>2349</td>
</tr>
<tr>
<td></td>
<td>382</td>
<td>785</td>
<td>1374</td>
</tr>
<tr>
<td></td>
<td>658</td>
<td>1251</td>
<td>2363</td>
</tr>
<tr>
<td></td>
<td>382</td>
<td>793</td>
<td>1376</td>
</tr>
<tr>
<td></td>
<td>492</td>
<td>1106</td>
<td>2018</td>
</tr>
<tr>
<td></td>
<td>331</td>
<td>769</td>
<td>1290</td>
</tr>
</tbody>
</table>

Table 17: Evaluation of low confidence mappings for large ontologies of the biomedical track
Notice that for all test cases, the HE by itself or with other two experts, performs the best with respect to number of actual correct low confidence mappings determined by the expert. However, the HE also accepts as correct the greatest number of low confidence mapping in all these same cases. In fact for all the cases, except the FMA-SNO (including the small fragment) and SNO-NCI, combining the HE with the other two experts does not improve the number of actual correct low confidence mappings that are determined by the expert.

Also note that for all of the test cases, except the FMA-SNO, the Majority combination produces identical results to that of the ME or SE (ME, SE) pair.

For the FMA-SNO pair, the ME and the ME, SE pair both determine more correct low confidence mappings than the HE. When all three experts are combined, the greatest number of actual correct low confidence mappings is determined. In fact 13 more mappings are found by adding the HE (1123 – 1110) and 6 of those are actually correct (913 – 907).

The test cases for the Large Biomed Track differs from those pairs in the small fragment track in that for none of the cases does the HE determine the most actual correct low confidence mappings. In fact for the first pair, FMA-NCI, HE determines the fewest number of actual correct low confidence mappings. In the other two cases HE determines the second or tied for second fewest number of actual correct low confidence mappings. The HE, however accepts the fewest number of low confidence mappings for all experts and combination of experts except for the FMA-SNO test case where SE accepts the fewest number of low confidence mappings.

For all test pairs of ontologies, adding the HE to the ME, SE pair of experts finds few if any additional actual correct low confidence mappings but increase the number of low confidence mappings that are corrected.

The majority vote combination works best for the FMA-SNO pair since it greatly reduces the number of accepted number of low confidence mappings (by 119, 1225 – 1106) but slightly reduces the number of actual correct low confidence mappings (by 16, 785 - 769).
7 CONCLUSIONS AND FUTURE WORK

Using background knowledge (BK) in the OA process has become prevalent as it can discover additional mappings for entities without any lexical or structural similarity. Many top OA systems in the OAEI use BK to improve the matching results. For example, GOMMA and AgreementMaker use Uberon, LogMap uses UMLS Lexicon, YAM++ uses WordNet, BLOOMS+ and WikiMatch use Wikipedia, etc. Moreover, increasingly OA systems introduce user interaction in the OA process to help identify mappings. In OAEI 2013, the Interactive Matching Evaluation Track was first added to evaluate different interactive matching tools that OA systems used for user interaction.

This thesis work exploits BioPortal as a generalized yet also specialized BK source in OA systems. Experiments are done using a top OA system LogMap in the OAEI Anatomy Track and Large Biomedical Ontology Track. A two-step approach is used to evaluate the performance of integrating BioPortal as the automated expert. First, simulations with a variable error rate on “perfect oracle” were done to understand the upper bound benefits of using an automated expert and for a comparison evaluation benchmark. Then experiments with LogMap and BioPortal as the automated expert were performed using the Anatomy Track and the Large Biomedical Ontology Track.

The following contributions are made by this thesis research:
1) BioPortal has not been exploited in the context of the OAEI. This thesis is the first to investigate the use of BioPortal.
2) Improvements of the OA standard performance measure for F-measure are achieved in 5 out of the 7 tasks in the OAEI 2013 Anatomy Track and Large Biomed Track.
3) Though neither the improvements are significant, nor the improved F-measure scores are better than any other top OA systems in OAEI 2013, the work of this thesis research offered BioPortal as an independent framework from the OA. This independent framework can be easily plugged into different OA systems.
4) This thesis research also helps understanding the limitation of the current matching process in LogMap.
The future work of this thesis research includes the following:

1) More experiments need to be made to further understand the positive impact on the LogMap’s performance using the combinations of the ME, SE and HE. For example, each of the three experts can be assigned a coefficient, which can be trained using the reliable mappings found by LogMap. The decision of a low confidence mapping can be made with a weighted combination of the results of the three experts.

2) This thesis research developed an independent framework using BioPortal as a BK source for LogMap. More evaluation of this framework can be made by plugging this framework into other OA systems such as GOMMA and AgreementMaker.

3) Another approach to using BioPortal would be to integrate its use earlier in the OA process, i.e., not just for the evaluation of low confidence mappings in interactive step but in finding initial candidate mappings.
REFERENCES


[40] Needleman, Saul B., and Christian D. Wunsch. "A general method applicable to the
search for similarities in the amino acid sequence of two proteins." *Journal of molecular


coefficient versus Jaccard's similarity coefficient." *Computers & industrial engineering*


the quality of ontology mappings." *Proc. 3rd International Workshop on Ontology