UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE INFORMÁTICA

# Visualization for Biomedical Ontologies Alignment

**Mestrado em Bioinformática e Biologia Computacional**
Especialização em Bioinformática

**Catarina Tavares Martins**

Dissertação orientada por:
Prof. Doutora Cátia Luísa Santana Calisto Pesquita

2016

# Resumo

Desde o início do século, a investigação biomédica e a prática clínica levaram a uma acumulação de grandes quantidades de informação, por exemplo, os dados resultantes da sequenciação genómica ou os registos médicos. As ontologias fornecem um modelo estruturado com o intuito de representar o conhecimento e têm sido bem sucedidas no domínio biomédico na melhoria da interoperabilidade e partilha. O desenvolvimento desconectado das ontologias biomédicas levou à criação de modelos que apresentam domínios idênticos ou sobrepostos. As técnicas de emparelhamento de ontologias foram desenvolvidas a fim de estabelecer ligações significativas entre as classes das ontologias, por outras palavras, para criar alinhamentos. Para alcançar um alinhamento ótimo é, não só importante melhorar as técnicas de emparelhamentos mas também criar as ferramentas necessárias para que possa existir intervenção humana, particularmente na visualização.

Apesar da importância da intervenção de utilizadores e da visualização no emparelhamento de ontologias, poucos sistemas o suportam, sobretudo para grandes e complexas ontologias como as do domínio biomédico, concretamente no contexto da revisão de alinhamentos e interpretação de incoerências lógicas. O objetivo central desta tese consistiu na investigação dos principais paradigmas de visualização de ontologias, no contexto do alinhamento de ontologias biomédicas, e desenvolver abordagens de visualização e interação que vão de encontro a estes desafios.

O trabalho desenvolvido levou, então, à criação de um novo módulo de visualização para um sistema de emparelhamento do *state of the art* que suporta a revisão de alinhamentos, e à construção de uma ferramenta online que visa ajudar o utilizador a compreender os conflitos encontrados nos alinhamentos, ambos baseados numa abordagem de visualização de subgrafos. Ambas as contribuições foram avaliadas em pequena escala, por testes a utilizadores que revelaram a relevância da visualização de subgrafos contra a visualização em árvore, mais comum no domínio biomédico.

**Palavras Chave:** Ontologias Biomédicas, Visualização, Alinhamento, Conflito

# Abstract

Since the begin of the century, biomedical research and clinical practice have resulted in the accumulation of very large amounts of information, e.g. data from genomic sequencing or medical records. Ontologies provide a structured model to represent knowledge and have been quite successful in the biomedical domain at improving interoperability and sharing. The disconnected development of biomedical ontologies has led to the creation of models that have overlapping or even equal domains. Ontology matching techniques were developed to establish meaningful connections between classes of the ontologies, in other words to create alignments. In order to achieve an optimal alignment, it is not only important to improve the matching techniques but also to create the necessary tools for human intervention, namely in visualization.

Despite the importance of user intervention and visualization in ontology matching, few systems support these, especially for large and complex ontologies such as those in the biomedical domain, specifically in the context of the alignment revision and logical incoherence explanation. The central objective of this thesis was to investigate the main ontology visualization paradigms, in the context of biomedical ontology matching, and to develop visualization and interaction approaches addressing those challenges.

The work developed lead to the creation of a new visualization module for a state of the art ontology matching system, that supports the alignment review, and to the construction of an online tool that aims to help the user understand the conflicts found in the alignments both based on a subgraph visualization approach. Both contributions were evaluated, in a small-scale, by user tests that revealed the relevance of subgraph visualization versus the more common tree visualization for the biomedical domain.

**Keywords:** Biomedical Ontologies, Visualization, Alignment, Conflict

# Resumo Alargado

O crescimento da área biomédica foi acompanhado por um grande influxo de informação. Para uma melhor análise e compreensão desta informação surgiu a necessidade de a modelar de maneira a que esta possa ser lida e compreendida por um computador. As ontologias foram criadas com o objetivo de modelar o vocabulário de um determinado domínio e de estabelecer todas as relações possíveis entre diferentes termos, de forma a representar uma realidade que possa ser descodificada por uma máquina.

No entanto, a criação indiscriminada de ontologias levou a uma sobreposição na informação incluída nestas. A duplicação de dados produzida, numa área já por si vasta, como é o caso das ontologias biomédicas, levou a que surgisse a necessidade de agregar a informação partilhada. Para tal, foi imprescindível criar técnicas que permitissem encontrar ligações relevantes entre diferentes ontologias. O conjunto dessas ligações é denominado de alinhamento e é constituído por todas as classes correspondentes em ambas as ontologias. O objetivo destas técnicas consiste na criação de um alinhamento composto por todas as ligações relevantes entre as ontologias. As correspondências podem ser encontradas manualmente ou automaticamente. Contudo, quando o emparelhamento de ontologias é realizado de forma automática algumas das relações entre as ontologias encontram-se erradas ou fazem parte de inconsistências lógicas entre as ontologias. Para alcançarmos o alinhamento ideal é importante desenvolver as atuais técnicas de emparelhamento automático de ontologias, assim como, desenvolver ferramentas para que o utilizador possa contribuir para a correção do alinhamento final. As atuais ferramentas de emparelhamento de ontologias são compostas por algumas das funcionalidades essenciais para a revisão e exploração do alinhamento final mas todas

apresentam lacunas. O maior obstáculo dos sistemas atuais recai sobre a dificuldade destes em lidar com ontologias de grandes dimensões constituídas por múltiplos elementos e múltiplas relações, como é o caso das ontologias biomédicas. Consequentemente, o carregamento e visualização deste tipo de ontologias revela-se bastante complexo e computacionalmente intenso. Para além disso, vários dos sistemas atuais falham no que diz respeito à informação sobre os elementos envolvidos no alinhamento e sobre a confiança das correspondências encontradas. Ambos os fatores são essenciais para que o utilizador do sistema seja capaz de realizar uma revisão acertada e ponderada sobre o alinhamento em mão.

Esta dissertação tem como foco dois paradigmas da visualização de ontologias: a visualização e edição manual de alinhamentos e a visualização de conflitos encontrados, por vezes, nos alinhamentos.

A primeira parte desta tese consistiu na melhoria da interface gráfica do AgreementMakerLight (AML) um dos melhores sistemas atuais de emparelhamento de ontologias. À interface gráfica foram adicionadas as ferramentas necessárias para a revisão de alinhamentos, que dão ao utilizador a oportunidade de avaliar quais as correspondências corretas ou incorretas entre as ontologias e de adicionar possíveis ligações em falta. O AML tem agora as ferramentas necessárias para intervenção humana, incluindo uma lista compreendida de todas as correspondências encontradas entre as ontologias, a possibilidade de adicionar e remover as correspondências e a informação léxica e estrutural das classes das ontologias. As ferramentas incluídas vão assim possibilitar a produção de alinhamentos de melhor qualidade.

Paralelamente, surgiu o desafio de criar uma ferramenta que permitisse ao utilizador entender o conceito complexo de conflito num alinhamento. O alinhamento de ontologias leva, por vezes, à criação de inconsistências lógicas. As inconsistências encontradas têm de ser reparadas automaticamente, manualmente ou uma combinação de ambos. Contudo, a compreensão do conflito pode ser complexa tanto para utilizadores inexperientes como, por vezes, para especialistas. A criação da ferramenta online VizRepair teve como objetivo simplificar a tarefa de compreender os conflitos lógicos que, por vezes, podem ser encontrados ao alinhar ontologias. Para tal, foram usados 10 pares de ontologias do Bioportal às quais foi aplicado o algoritmo de reparação do AML. O VizRepair visa ajudar o utilizador a compreender a incoerência encontrada, de modo a que este possa decidir se esta deve ser mantida ou removida de acordo com a informação fornecida.

A avaliação de ambas as tarefas centrais deste trabalho foi, também, um ponto crucial no desenvolvimento da tese. Para a avaliação do AML e do VizRepair, foram reunidos 7 participantes que realizaram um conjunto de tarefas, previamente construído para explorar todas as funcionalidades das ferramentas, e responderam a um inquérito para avaliar a utilidade e o desempenho de ambas. Os resultados mostraram que as ferramentas apresentam um elevado potencial e utilidade, obtendo valores acima de 60 no inquérito realizado. O AML foi, por sua vez, comparado com outro sistema cujo paradigma de visualização é diferente da representação em grafo. O sistema escolhido pela sua representação em árvore das ontologias e por se encontrar disponível online, foi o AgreementMaker (AM). Ao comparar ambos os sistemas cheguei à conclusão que os utilizadores realizaram mais rápida e eficientemente as tarefas no AML. Como tal, os utilizadores preferem uma visualização em grafo disponibilizada pelo AML em vez da visualização em árvore disponibilizada pelo AM. A visualização disponibilizada pelo AM demonstrou ser confusa e desnecessariamente complexa para os utilizadores. Para além disso, a visualização de parte do alinhamento em vez do alinhamento total, implementada pelo AML, contribuiu para uma exploração do alinhamento mais fácil e rápida e permitiu, consequentemente, que os utilizadores tivessem mais facilidade para concluir as tarefas disponibilizadas. Contudo, os participantes do teste apontaram várias sugestões e limitações a ambas as ferramentas que serão consideradas no futuro para que seja possível contribuir para um desenvolvimento positivo das ferramentas.

A tese desenvolvida teve como objetivo primário compreender qual a melhor forma de representar as ontologias de maneira a que o utilizador compreenda adequadamente qualquer alinhamento que pretenda analisar. A representação da ontologia, em grafo ou árvore, e a representação do alinhamento, global contra parte do alinhamento, são discutidas nesta tese, com o objetivo de compreender quais as que darão a um sistema um perfil mais intuitivo, rápido e completo. A inovação do AML e a criação de um módulo de interação humana de alinhamentos que reuniu as ferramentas necessárias para a revisão fácil e plena de um alinhamento teve como objetivo a construção de alinhamentos de melhor qualidade. Por fim, outro foco da tese, consistiu na exploração de conflitos lógicos presentes em certos alinhamentos e na importância da sua visualização. A construção de uma ferramenta online com o objetivo de desconstruir a complexa tarefa de compreensão dos mesmos foi uma das principais contribuições da tese. No futuro, ambas as ferramentas serão melhoradas de maneira a ir de encontro às recomendações dos utilizadores e para acompanhar o desenvolvimento da área.

# Acknowledgements

My advisor, Dra. Cátia Pesquita was an inspiration and example to follow. The constant help and non-stop motivation not only on my master thesis but also in all the extracurricular work were vital to my growth professionally and as a person. Without her guidance and patience this dissertation would not have been possible.

I would also like to thank professor Daniel Faria to all the help provided and advices and for introducing me to Java and AgreementMakerLight. To my tree consciences Daniela Oliveira, Joana Barros and Maria Fernandes for their friendship and presence, personally or ’in spirit’, that kept me going even in the tougher days. I thank Daniela Oliveira for all the suggestions and help provided to this thesis and, more importantly, for being a true friend for a long time and for all the nerd conversations. To Joana Barros for all the encouragement words, suggestions and perfectionism. Finally, to Maria Fernandes that even far away was always there.

A special thanks to my friends Claudia for reading my mind even far away and for your presence in the good and bad days; Luís, my brother from another mother, for making me laugh and support me every single time; Claudia for all the hours listening to me and for believing in me and in my growth; and Sara for her down-to-earth advice, true friendship, strength and persistence. Without them I couldn’t maintain my mental sanity.

I am thankful to my father for thinking outside of the box and teaching me the importance of details and hard work, and to my mother whose persistence is an example to follow. Both of them and my grandparents taught me the importance of an education and to always aim high and for that I am truly grateful.

My brother also deserves a special thanks for the craziness and unconditional support that only him can provide.
I am deeply grateful to my aunt for never giving up on me and for the support on my career.
To my pets, dead or alive, for sharing their fluffiness with me and for helping me in all the stressful moments the best way they could.
Finally, to my boyfriend Nelson for believing in me when I didn't. Your strength and encouragement kept me going.

# Contents

# List of Figures

# List of Tables

# List of Acronyms

**AM** AgreementMaker.

**AML** AgreementMakerLight.

**BDO** Bone Dysplasia Ontology.

**CCONT** Cell Culture Ontology.

**EFO** Experimental Factor Ontology.

**EP** Cardiac Electrophysiology.

**FMA** Foundational Model of Anatomy.

**GUI** Graphical User Interface.

**IRI** International Resource Identifier.

**MA** Mouse Adult Gross Anatomy Ontology.

**NCBO** National Center for Biomedical Ontology.

**NCIT** National Cancer Institute Thesaurus.

**OAEI** Ontology Alignment Evaluation Initiative.

**OBO** Open Biomedical Ontologies.

**OMIM** Online Mendelian Inheritance in Man.

## List of Acronyms

**OWL** Web Ontology Language.

**RDF** Resource Description Framework.

**SKOS** Simple Knowledge Organization System.

**SNOMEDCT** Systematized Nomenclature of Medicine - Clinical Terms.

**SUS** System Usability Scale.

**UBERON** Uber Anatomy Ontology.

**URI** Uniform Resource Identifier.

# Chapter 1

# Introduction

## 1.1 Motivation

Since the begin of the century, biomedical research and clinical practice have resulted in the accumulation of very large amounts of information, e.g., data from genomic sequencing or medical records. This growth of information was accompanied by a need to model that information in a way that could be easily understood by both humans and computers, to support data sharing, integration and analysis, among others.

During the last decade, biomedical ontologies have been gaining importance in the biomedical field, since they provide a structured model of a domain describing the concepts and relationships between them. Despite their advantages, applying ontologies to biomedical data still has some challenges, specially when considering their size, complexity and vocabulary. The number of ontologies has been steadily rising, and currently there are over 400 ontologies listed in Bioportal (Whetzel *et al.*, 2011), an online repository for biomedical ontologies. These include popular ontologies, such as the Gene Ontology (Ashburner *et al.*, 2000) dedicated to gene products functional aspects, or the Systematized Nomenclature of Medicine - Clinical Terms (SNOMEDCT) that collects clinical terminology and their meaning from around the world.

The disconnected development of biomedical ontologies has led to the creation of models that have overlapping or even equal domains. To ensure interoperability between these ontologies, and to realize their full potential, ontology matching techniques can be used to establish meaningful correspondences between different ontologies. Figure 1.1 illustrates the matching between portions of two biomedical ontologies.

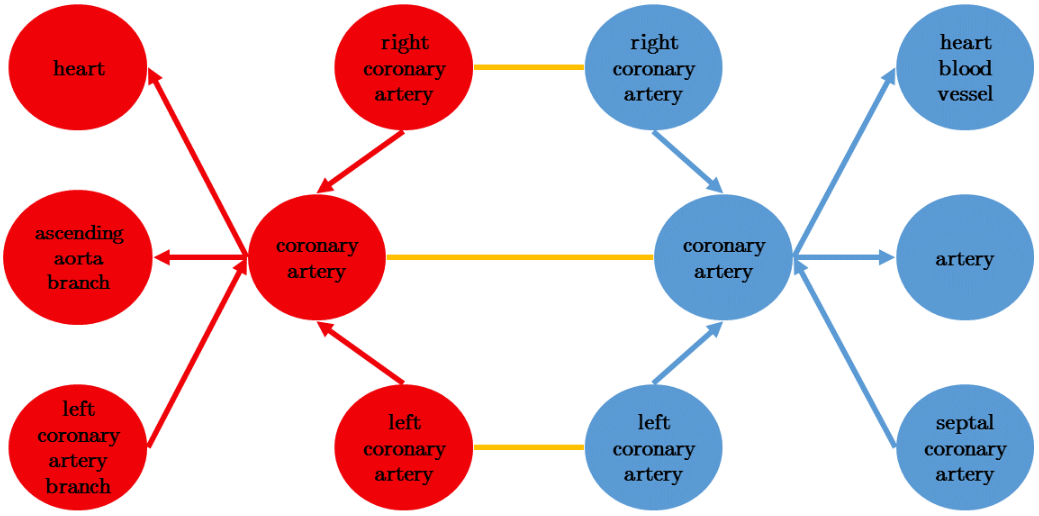


Figure 1.1: Example of a graph representation of an alignment portion. Each ontology is represented by a colour (red or blue) and the mapping between them is displayed in yellow.

Given the complexity of ontologies and the ontology matching process, ontology matching systems can provide visualizations that intend to give the user a better understanding of the alignment itself and the possibility to detect and correct errors. This can be specially relevant in the case of erroneous mappings or incoherent mappings (e.g. mappings that give rise to logical incoherence in the ontologies). This is particularly relevant in the biomedical domain where users are very frequently not ontology experts but domain specialists. The importance of human intervention in order to build better alignments (interactive matching) is well recognized (Shvaiko & Euzenat, 2013). In order to allow this kind of intervention the graphical user interface should give the users the necessary tools for the tasks, allowing them to explore the alignment and ontologies, and make the adjustments that better fit their needs, with the ultimate goal to achieve an

optimal alignment.
However, many of the current ontology matching systems lack these important features or provide visualization strategies that have not been assessed in the context of biomedical ontologies, which pose additional challenges given their large size and complexity.

## 1.2 Objectives

The central objective of this thesis was to investigate the main ontology visualization paradigms, in the context of biomedical ontology matching, and to develop visualization and interaction approaches addressing those challenges. With that purpose, we focused our attention in two main challenges:

- Visualization for manual review of the alignment, where different visualization strategies were compared and tools to support the manual review were developed;
- Visualization of mapping incoherences to support their understanding, where novel visualization strategies were developed and evaluated.

## 1.3 Contributions

The focus of this thesis was to study certain ontology visualization paradigms and contribute to the field expanding on some of the limitations found. The two main targets of this work were the user intervention, to the creation of an optimal alignment, and the visualization of alignment incoherences. The contributions of this project are dissected in this thesis and include:

**Contribution 1:** Extension of the requirements for the visualization of biomedical ontologies alignment: the list of requirements proposed extends the state of the art aiming to perform a better evaluation of the current ontology matching systems with a visualization module.

**Contribution 2:** AgreementMakerLight mapping reviewer: a new addition to AgreementMakerLight graphical user interface that makes it possible for the user to decide which mappings should be part of the final alignment and which ones are absent and must be added. This feature aims to help the user make an informed decision and to reach an optimal alignment, overcoming the problems that a machine is not able to resolve.

**Contribution 3:** Improvement of AgreementMakerLight graph viewer: the graph viewer was improved in order to give first hand information about the mappings. The visualization provided helps the user to better understand the connection between the elements of different ontologies side by side with a representation of that connection. The aim was to present the user with the necessary information to help him decide if the mappings are correct or incorrect.

**Contribution 4:** Development of a webtool to visualize incoherences in Bioportal ontologies: the tool developed aims to contribute to the field by using the power of visualization to explain the incoherences that can sometimes be present in certain alignments. In this case, we used ten pairs of Bioportal ontologies and used the results of the AgreementMakerLight repair module. The tool represents the logical incoherences found in the alignment of those ontologies in a graph format and it is available at `http://xldb.di.fc.ul.pt/biotools/vizrepair/`.

**Contribution 5:** User Evaluation of AgreementMakerLight and VizRepair: the evaluation aimed to access the utility and performance of the tools. The evaluation consisted on the performance of a set of tasks previously chosen with the aim of manipulating all the features present in each tool. Moreover, the users answered a System Usability Scale questionnaire where they would score the overall usability of the tools. The results show that AgreementMakerLight graph format is preferable to a tree format and a whole alignment perspective is unnecessarily complex when compared to the visualization of a section of the alignment. The users also gave a good score to VizRepair and highlighted its applicability to the field.

## 1.4 Overview

All the chapters of this thesis describe the work developed with the intent to further improve the field through the addition of new tools. Chapter 2 starts by describing the essential definitions and theoretical foundations required for a thorough comprehension of this thesis. Chapter 3 overviews several ontology matching systems graphical user interface (GUI) and evaluates their performance taking into account a set of requirements that are thought to be necessary to create a good GUI. Chapter 4, besides describing AgreementMakerLight (AML) system, also includes all the enhancements made to its GUI such as the Alignment reviewer and the improvements to the graph viewer. The next focus of this thesis is discussed in Chapter 5, illustrating the problem of the incompatibilities occasionally found in the alignments and how it could benefit from a tool that visually represented the problem. The result was a webtool called VizRepair that shows the conflict sets computed by AML repair algorithm in Bioportal ontologies. Both, chapter 4 and 5 also include an evaluation procedure implemented to access the utility and performance of both tools and a discussion about the evaluation results. The last chapter overviews the complete work elaborated in this thesis, discusses the limitations found and possible future work.

# Chapter 2

# Basic Concepts

In this section, we will give an overview of the basic concepts required for an in-depth comprehension of this dissertation including basic ontology concepts, ontology matching, alignment repair and visualization of ontologies and alignments.

## 2.1 Ontology

An ontology is a practical representation of a reality in a way that is easier for a computer to understand (Euzenat *et al.*, 2007), it typically provides the necessary vocabulary to describe a certain domain and specifies the meaning of the classes in that vocabulary.

The main components in an ontology are the classes and their relations (e.g., 'part of' and 'is a') (Hoehndorf *et al.*, 2015). A class is considered as an entity that refers to a set of entities in a certain domain, for example the class 'Gene' refers to the set of all the genes. They are defined by their properties, features and relations. The properties characterize the classes and can comprise the ontology structures, data instances, semantics, or labels. Both classes and relations are identified by a special identifier generally an Uniform Resource Identifier (URI). Besides the URI, there are other identifiers: the Internationalized Resource Identifier (IRI) is the preferred identifier in semantic web and database identifier strings are, less frequently, used by certain biomedical databases.

A set of labels, in one or many languages, is also provided by the ontology and it represents the set of terms associated with the classes and the relations involved. There are several types of labels like primary labels, secondary labels and synonyms.

To provide even more information, ontologies also have textual definitions and descriptions that yield examples, background information and links to external sources.

The structural framework of the ontology is defined by the formal definitions and axioms (for example, 'equivalentTo' or 'subClassOf') represented. Ontologies are modelled using formal languages, for instance, Open Biomedical Ontologies (OBO), Flatfile Format (Mungall *et al.*, 2011), Resource Description Framework (RDF) (Klyne & Carroll), Simple Knowledge Organization System (SKOS) (Miles & Bechhofer, 2009) or Web Ontology Language (OWL) (Grau *et al.*, 2008). These are usually possible to translate into each other and provide different levels of complexity.

## 2.2 Ontology Alignment

The number of biomedical ontologies has been increasing and their disconnected development has led to overlapping information which is neither ideal nor needed by the community. This introduces the need to discover meaningful relations between ontologies with the same or similar domains, to provide interoperability and it can be accomplished by the use of ontology matching techniques. The set of those correspondences is called an alignment and the correspondences are usually called mappings.

One of the main goals of ontology matching is the reduction of the heterogeneity between the ontologies. There are four relevant kinds of heterogeneity (Euzenat *et al.*, 2007). The first corresponds to syntactic heterogeneity and it is present when there are two ontologies represented in different languages. However, it is sometimes possible to implement a translation between the set of ontologies. The next type of heterogeneity takes into account the different name variations between equal entities in different ontologies, e.g. 'Paper' and 'Article' can be considered synonyms in certain domains. This type of heterogeneity is called terminological. The conceptual or semantic heterogeneity corresponds to the differences between the modulation of the same domain of interest like the use of different axioms when defining concepts or the use of totally different concepts to express the same reality. Finally, the semiotic or pragmatic heterogeneity stands for the interpretations that different people make when considering the same entity. This kind is the most difficult for a computer to detect and to translate due to its abstract content. Generally, the different categories of heterogeneity occur together. Most ontology matching systems focus on terminological and semantic heterogeneities.

All in all, the matching process results in a set of correspondences between pairs of entities, an alignment, belonging to a pair of ontologies. The process can be extended by the addition of an input alignment; matching parameters, e.g. weights, thresholds; and the use of external resources, e.g. common knowledge and domain specific thesauri. There are different techniques that can be used to solve the matching problem and they can be classified as (Euzenat *et al.*, 2007):

- Element-level techniques;
- Structure-level techniques.

Element-level techniques establish correspondences by analysing either the entities or instances independently and do not take into account the relationships between the elements or their instances. In contrast, structure-level techniques consider the entities or their instances as a structure.

The previous classification can be subdivided in syntactic, external or semantic techniques. Syntactic techniques interpret the input as its structure. External techniques make use of external resources of domain or common knowledge to explain the input. Finally, techniques based on semantics use formal semantics like models to understand the input and the results.

In the Element-level, the main techniques are:

- String-based techniques - equal concepts correspond to similar strings (sequences of letters in an alphabet). Often used to match names and name descriptions, for example, 'Book' is similar to 'Textbook' but not 'Volume'.
- Language-based techniques - relies on natural language processing techniques.

- Constraint-based techniques - they take into account the internal constraints of the entities definitions, such as types, cardinality (or multiplicity) of attributes, and keys and are generally used to find clusters instead of accurate correspondences between entities.
- Linguistic resources - use of lexicons or domain specific thesauri (e.g. WordNet) to match words based on linguistic relations, such as synonyms and hyponyms.
- Alignment reuse - use of an alignment previously computed as an external resource.
- Upper level and domain specific formal ontologies - use of upper level ontologies and domain specific formal ontologies as an external source (e.g. Uberon).

Furthermore, the Structural-level includes the following techniques:

- Graph-based techniques - they consider the ontologies as labelled graphs structures and the comparison of the nodes is based on the element position in the graph.
- Taxonomy-based techniques - if two classes are connected by a is_a relation they are similar therefore their neighbourhood may also be similar to each other.
- Repository of structures - stores the similarity between the ontology and not the alignment and uses that fragment in future alignments to analyse the similarity.
- Model-based techniques - based on the semantic interpretation of the ontology.
- Data analysis and statistics techniques - take advantage of a portion of the population to access what is correct or incorrect.

The evaluation of the techniques used can be accomplished by comparing the results of a matching process with a reference alignment that is usually manually curated by experts. The comparison is accompanied by a mapping confidence that corresponds to the level of confidence by which we can say a certain mapping is correct or incorrect. The mapping confidence generally appears as a percentage.

### 2.2.1 Alignment Repair

A matching system usually comprises several matching techniques aiming to achieve the best alignment between the different ontologies. The alignments produced by large ontologies can lead to desired and undesired entailments. These undesired entailments are commonly associated with erroneous mappings in the ontology, but can be occasionally related to conflicting descriptions of the classes involved. We can divide the undesired entailments in the ones that are causing unsatisfiable classes, or incoherences, that are easily detected by the use of logical reasoning (including automatic), and the ones that require domain knowledge in order to verify if they are correct. The incoherences are usually characterized by a group of mappings involved in the conflict that are called conflict sets. To better understand the definition of a conflict focus your attention on figure 2.1.

This figures shows that, in the National Cancer Institute Thesaurus (NCIT), the class 'property or attribute' is disjoint with 'biological process', which means they cannot be related by subsumption and the same applies to their progeny. But when this ontology is matched with the Bone Dysplasia Ontology (BDO) a connection is created between these two classes. Their descendants match with the BDO ontology classes 'position', 'angle' and 'orientation'. The class 'obtuse angle to' from BDO is related to all the classes of this ontology mapped with NCIT. This creates a connection between the classes 'position', 'angle' and 'orientation'. Since they can not be related, this represents an inconsistency in the alignment.

The alignment can be automatically repaired by the removal of the necessary mappings to maintain a final coherent alignment. The ontology matching systems LogMap and ALCOMO (Meilicke, 2011) were the first to integrate repair

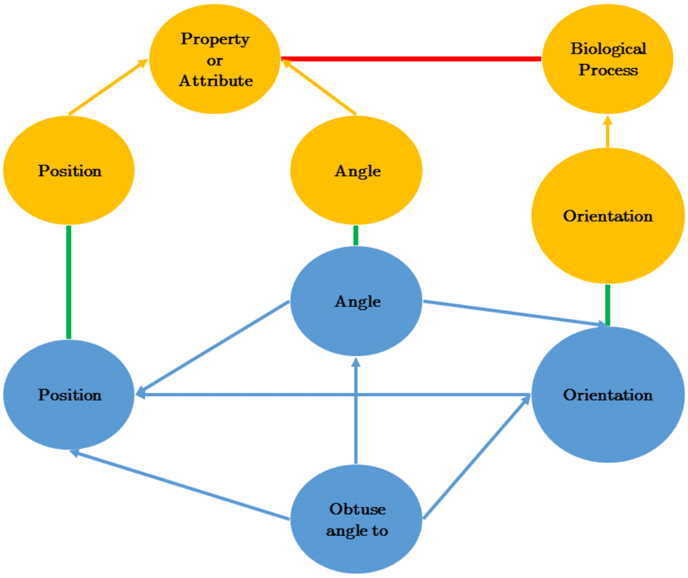


Figure 2.1: Example of a conflict set between National Cancer Institute Thesaurus (yellow) and Bone Dysplasia Ontology (blue). The arrows correspond to 'subclass of' relationships, the green lines refer to mapping correspondences and the red to the disjoint relationship.

techniques regarding the biomedical field. There are two main approaches (Santos *et al.*, 2013):

- Local - determining the minimal impact in small subsets of the matched ontologies, which makes it more efficient, but has a major impact in the input alignment. This approach is applied by the LogMap system.
- Global - determines the minimal impact but, in contrast with the previous approach, it takes into account the whole alignment with all its classes and relations. The results are generally better but when handling large ontologies a more holistic approach is not always the best. ALCOMO uses this kind of approach.

The development of these alignment repair techniques provide a logical solution and aim to help in the construction of better quality alignments thus improving the data. However, these repair algorithms aim to eliminate incoherences by removing or altering mappings. This is not always the best option given that the choice of mappings to eliminate is based on a change minimization strategy and in same cases we can lose crucial information. For example, a protein and its corresponding gene sometimes are considered related to each other, and although they are semantically distinct they share some kind of relationship. However, occasionally it creates a conflict between the real mapping equivalence, for example, the same protein. The repair could, in this case, remove the mapping between the proteins and with that we would lose crucial information (Faria *et al.*, 2014a).

### 2.2.2 User Involvement in Alignment

The user involvement in all the ontology matching tasks is vital for the construction of better quality final alignments. This process is called interactive matching and users can have different roles in it, e.g. by recommending relevant background knowledge or by adjusting the matching parameters (Shvaiko & Euzenat, 2013). To support these tasks the user must be given the necessary frameworks to easily explore and edit the matching algorithms and alignments in use.

There have been different areas of research regarding user involvement in the matching process. The strategies, proposed by Shvaiko & Euzenat (2013), include query logs to enhance mapping candidate generation (Elmeleegy *et al.*, 2008; Nandi & Bernstein, 2009); design of time matcher interaction (Do & Rahm, 2007; Noy & Musen, 2003); graphical visualization of alignments based on cognitive studies (Falconer & Storey, 2007); environment for manually designing complex alignments through the use of connected perspective (Mocan *et al.*, 2006); explicitly specification of structural transformations using a visual language (Raffio *et al.*, 2008); learning from the user by adjusting system parameters or experimenting with alignment selection strategies (Von Hippel, 2005).

The overwhelming size of the data in biomedical ontologies is one of the biggest challenges for the implementation of interactive approaches. The large amounts of classes involved and complex relationships can become confusing and users can feel overwhelmed by the influx of data. The purpose of systems with user involvement is to represent the information visualized in the most natural and complete way as possible. In the future, the use of these techniques aims to help with the vast amounts of data and construct higher quality alignments.

## 2.3 Ontology and Alignment Visualization

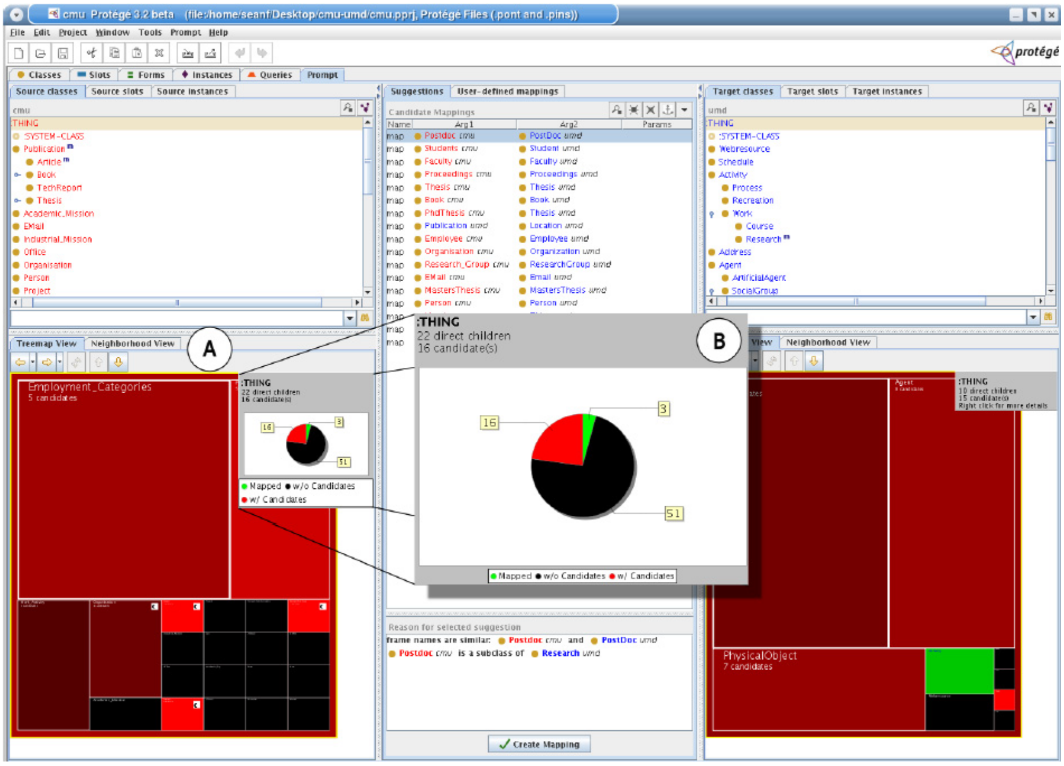
Ontology visualization supports the design, management and browsing of all the information comprised within an ontology. According to Katifori *et al.* (2007) the ontology visualization techniques can be divided in the groups displayed in Table 2.1. This categorization is not strict and some groups can have elements of the other groups. Typically, two main paradigms are considered: trees and graphs. (Fu *et al.*, 2013)

Trees are suitable when representing hierarchical relations but are confusing when multiple inheritance is involved because it would be necessary to the duplicate the classes to solve the problem. (see Figure 3.1)

Graphs, by contrast, are able to manage both hierarchical and non-hierarchical relations that trees are unable to handle. However, when the number of nodes is very high it is difficult to analyse the content of the alignment creating an hairball of nodes and relations. (see Figure 3.2).

In Fu *et al.* (2013) it was concluded that, trees are better for supporting list-checking activities as the evaluation of mappings and graphs surpass trees when the aim is the overview of a mapping (Fu *et al.*, 2013).

Treemaps, are an alternative ontology representation. In this case, the nodes of the hierarchy are represented as nested rectangles and the size and colour of the rectangles corresponds to data property (see Figure 2.2 for an example). Histograms maybe associated with a certain node. The main goal of this kind of representation is to provide an overview of the class hierarchy. Although, treemaps are suited for the visualization of large ontologies, this kind of ontology representation does not allow a detailed analysis of all the elements involved. Thus, it does not benefit the alignment review and a graph or tree representation is favourable. In addition to the visualization of an individual ontology, it is also possible to visualize the intertwine of information between ontologies. The result of the alignment can be visualized in order to better understand the process itself by supporting navigation and inspection of mappings, as well as interactive



cmu Protégé 3.2 beta (file:/home/seanf/Desktop/cmu-umd/cmu.pprj, Protégé Files (.pont and .pins))

Candidate Mappings

| Name | Arg1 | Arg2 | Params |
|---|---|---|---|
| map | Postdoc *cmu* | PostDoc *umd* | |
| map | Students *cmu* | Student *umd* | |
| map | Faculty *cmu* | Faculty *umd* | |
| map | Proceedings *cmu* | Proceedings *umd* | |
| map | Thesis *cmu* | Thesis *umd* | |
| map | Book *cmu* | Book *umd* | |
| map | PhdThesis *cmu* | Thesis *umd* | |
| map | Publication *umd* | Location *umd* | |
| map | Employee *cmu* | Employee *umd* | |
| map | Organisation *cmu* | Organization *umd* | |
| map | Research_Group *cmu* | ResearchGroup *umd* | |
| map | EMail *cmu* | Email *umd* | |
| map | MastersThesis *cmu* | MastersThesis *umd* | |
| map | Person *cmu* | Person *umd* | |

Reason for selected suggestion

frame names are similar: Postdoc *cmu* and PostDoc *umd*

Postdoc *cmu* is a subclass of Research *umd*

Figure 2.2: Example of a Treemap representation of COGZ extracted from (Granitzer *et al.*, 2010).

Table 2.1: Ontology Visualization Techniques.

| Technique | Ontology Representation | Examples |
|---|---|---|
| Indented List | Tree<br>(only 'is a' relationships) | Protégé<br>(Noy *et al.*, 2000)<br>OntoEdit<br>(Sure *et al.*, 2002) |
| Graph | Interconnected nodes (possibility to adjust information detail | OntoViz<br>(Sintek, 2003)<br>Tree Viewer<br>(Kleiberg *et al.*, 2001) |
| Zoomable | Nodes are represented nested inside their parents (zoom-in child nodes to enlarge them) | CropCircles<br>(Parsia *et al.*, 2005)<br>Gopher VR<br>(Andrews *et al.*, 1997) |
| Space-filling | Using the whole of the screen space and subdividing the space for a node among its children | TreeMaps<br>(Shneiderman, 1992)<br>BeamTrees<br>(Van Wijk & Van de Wetering, 1999) |
| Focus and context or distortion | Distort the view in order to focus the attention | 2D hyperbolic tree<br>(de Souza *et al.*, 2003)<br>3D hyperbolic tree<br>(Munzner, 1998) |
| Information landscapes | Color- and size-coded 3D objects | File System Navigator<br>(Strasnick & Tesler, 1996)<br>Harmony Information Landscape<br>(Eyl, 1995) |

matching. For instance, allowing the users to add new mappings or classify them as incorrect or correct. Apart from the choice of the individual ontology representation it is also important for the visualization of the alignment, the mappings representation. According to Thomas *et al.* (2009) when representing ontology mappings the following rules must be taken into consideration:

- ability to express the mapping relation;

- enable sharing and reuse of existing mappings, in order to minimize the effort of creating them and facilitate the information share;
- computational efficiency.

Several aspects must be taken into consideration in the design of the mapping relation. The decision to represent a part or the whole ontology is an important factor, accompanied by the kind and amount of relations, number of neighbours and possibility to add or remove mappings and mapping relations.

Moreover, the alignment should also be possible to save in a compatible format with the management tools available in order to be possible to visualize, share and reuse the mappings created in a matching process.

# Chapter 3

# State of the Art

This section describes several ontology matching systems and their visualization approaches.
According to Ivanova *et al.* (2015) the visualization of alignments can support different tasks:

- Explanation of matching results;
- User involvement in the matching process (for example, marking a mapping as correct or incorrect or adding and removing a mapping);
- Social and collaborative matching;
- Infrastructure and support that allow a big data input without slowing down the system.

The requirements above should be a guideline for current and future ontology alignment visualization systems. To provide a fine grained analysis of the state of the art, I extended the above requirements:

1. Ontology representation: depending on the ontology representation some tasks can be more efficient so this is a crucial feature. Usually the ontology is represented as a graph, a tree or both, but some variations can appear. The advantages and disadvantages of both this representations were enunciated in Chapter 2.

2. Mapping Representation: the absence or presence of this kind of representation is very important for the user comprehension of the alignment and it is a very important feature when dealing with huge amounts of mappings.

3. Level of detail adjustment: possibility of zooming in and out, and the specification of the neighbourhood to be visualized can help the user to understand the alignment as a part of a whole.

4. Alignment Overview: usually the visualization of the total alignment is a very difficult task, specially in biomedical ontologies where the number of elements to represent is very high. Nevertheless, it is generally important to see the whole picture.

5. Alignment Edition: give the user the power to remove or add mappings to accomplish the ideal final alignment.

6. Mapping Confidence: it is important to show the user the confidence of the mappings in order for him to make a better decision if either or not the mapping is correct or incorrect.

7. Mapping Reason: the reason why the mapping is created must be as explicit as possible, giving the user the best background on how the mapping was created. More information equals a better decision making that results in a better final alignment.

8. Element Information: in order to make a decision about the correctness of a mapping the user must not only have access to the overview of the mapping but also to the information about each element involved in the correspondence including information about possible disjoint classes.

9. Save and Load: one other important feature is the possibility to load a previous created alignment or to save an alignment providing the opportunity to save and perhaps reuse the user work.

10. Collaboration: the possibility to exist a connection between different users could greatly benefit the alignment review process, specially in large ontologies, allowing the possibility to divide the task between users or exchange comments.

11. Support for very large ontologies: some ontology matching systems are not capable of handling very large ontologies and become slower or not even support their loading.

The requirements here enumerated were used as evaluation parameters to the tools in review. The results of this evaluation can be seen in the last section of this chapter.

## 3.1 Ontology Alignment Systems with Visualization

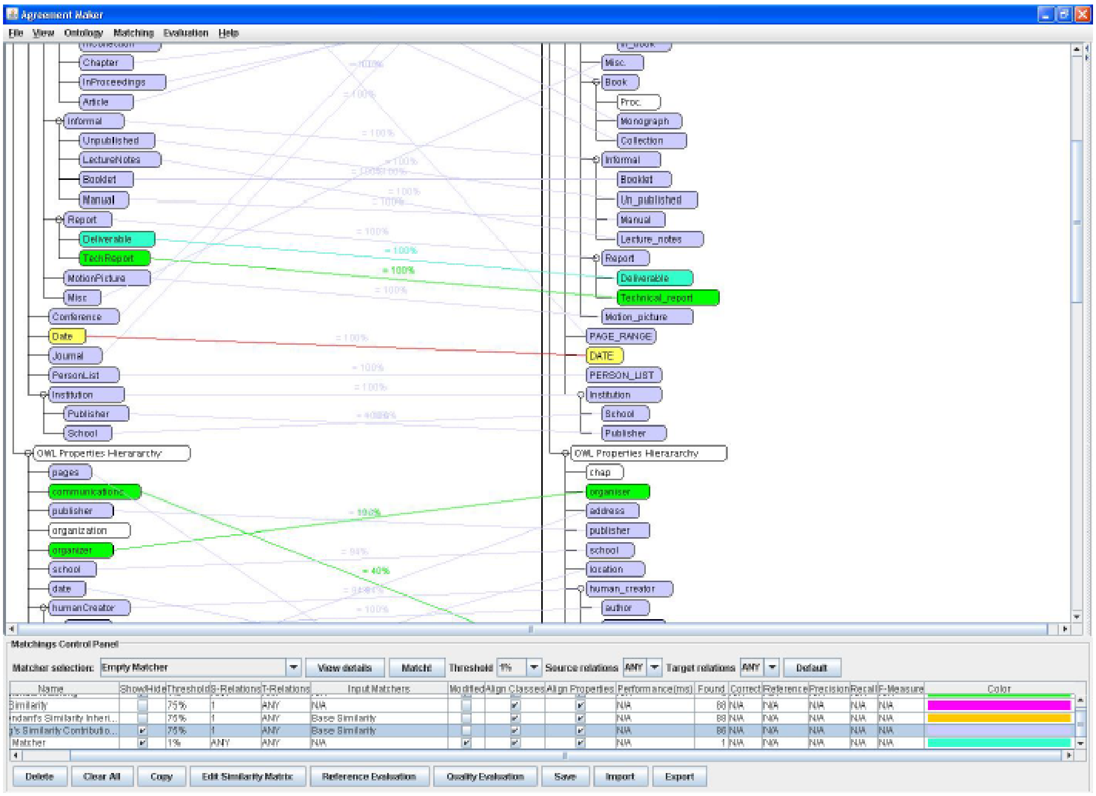
### 3.1.1 AgreementMaker

The AgreementMaker (Cruz *et al.*, 2009) representation of the ontologies consists of two trees side by side in a scroll-enabled pane. The mappings are then represented by a straight line with an indication of the similarity score of the correspondence (Figure 3.1. The system allows the visualization of different alignments simultaneously by using different colours in the mappings. The users can see the properties of a class by clicking on a node.

In the past, non-hierarchical ontologies could not be visualized by this system. However, AgreementMaker overcame that problem by allowing the duplication of specific subtrees. Although, the system already difficulty to handle with ontologies with thousands of classes effectively, was aggravated by the solution found because showing duplication data makes the task even more difficult.

### 3.1.2 AlViz

Alviz (Lanzenberger & Sampson, 2006) was built as a multiple view tab plug-in for Protégé. One of the views consists of a tree allowing the expand and collapse



Matcher selection: Empty Matcher
Threshold 1%
Source relations ANY
Target relations ANY

| Name | Show/Hide | Threshold | S-Relations | T-Relations | Input Matchers | Modified | Align Classes | Align Properties | Performance(ms) | Found | Correct | Reference | Precision | Recall | F-Measure | Color |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Similarity | ☐ | 75% | 1 | ANY | N/A | ☐ | ☑ | ☑ | N/A | 88 | N/A | N/A | N/A | N/A | N/A | 🟪 |
| ndard's Similarity Inheri... | ☐ | 75% | 1 | ANY | Base Similarity | ☐ | ☑ | ☑ | N/A | 88 | N/A | N/A | N/A | N/A | N/A | 🟧 |
| 's Similarity Contributio... | ☑ | 75% | 1 | ANY | Base Similarity | ☐ | ☑ | ☑ | N/A | 86 | N/A | N/A | N/A | N/A | N/A | 🟪 |
| Matcher | ☑ | 1% | ANY | ANY | N/A | ☑ | ☑ | ☑ | N/A | 1 | N/A | N/A | N/A | N/A | N/A | 🟩 |

Figure 3.1: Example of graphical user interface of AgreementMaker from (Cruz *et al.*, 2009).

of the branch nodes representing the hierarchical structure of the ontologies. To overcome the problems that trees have with large and complex ontologies they added a graph view. This view creates node clusters depending on the selected level of detail. The colours in the graph and tree represent the relations between the nodes: red for equal, syntactically equal is coloured orange, when an entity in one ontology is broader than the entity in other ontology is represented by a blue, a narrower-than relation appears violet, green is for similar entities and yellow for different entities (Figure 3.2).

There are other features implemented by this system like its zooming function, selection/highlighting function, several alignment functions, and tracking buttons for questionable associations, all important for the user exploration of the alignment. Besides that, the user can reject the alignment result by adapting the automatically derived associations and assigning them manually. All changes made to the alignment can be redone or undone using the functions for that purpose.

The major drawback of this tool is the absence of details about each individual concept involved in the alignment which is an impediment to tires grained analysis.

The last update was on April 2013 but the code is unavailable in sourceforge and the Protégé version is not currently working so we based our evaluation in the system paper.

### 3.1.3 COMA 3.0

COMA 3.0 Community Edition (Massmann *et al.*, 2011) is the successor of COMA++ shows the ontology as indented graphs in side-by-side scroll-enabled panels (Figure 3.3). Node clicking allows the visualization of the main label and the path to the root node in the form of coma separated labels. The mappings have different colours depending on their similarity scores. Moreover, it allows the user to add and remove mappings and to match just fragments of the ontologies.

The system does not allow for the visualization of non-hierarchical relations, nor of multiple inheritance. The last update to this system occurred in 2013.

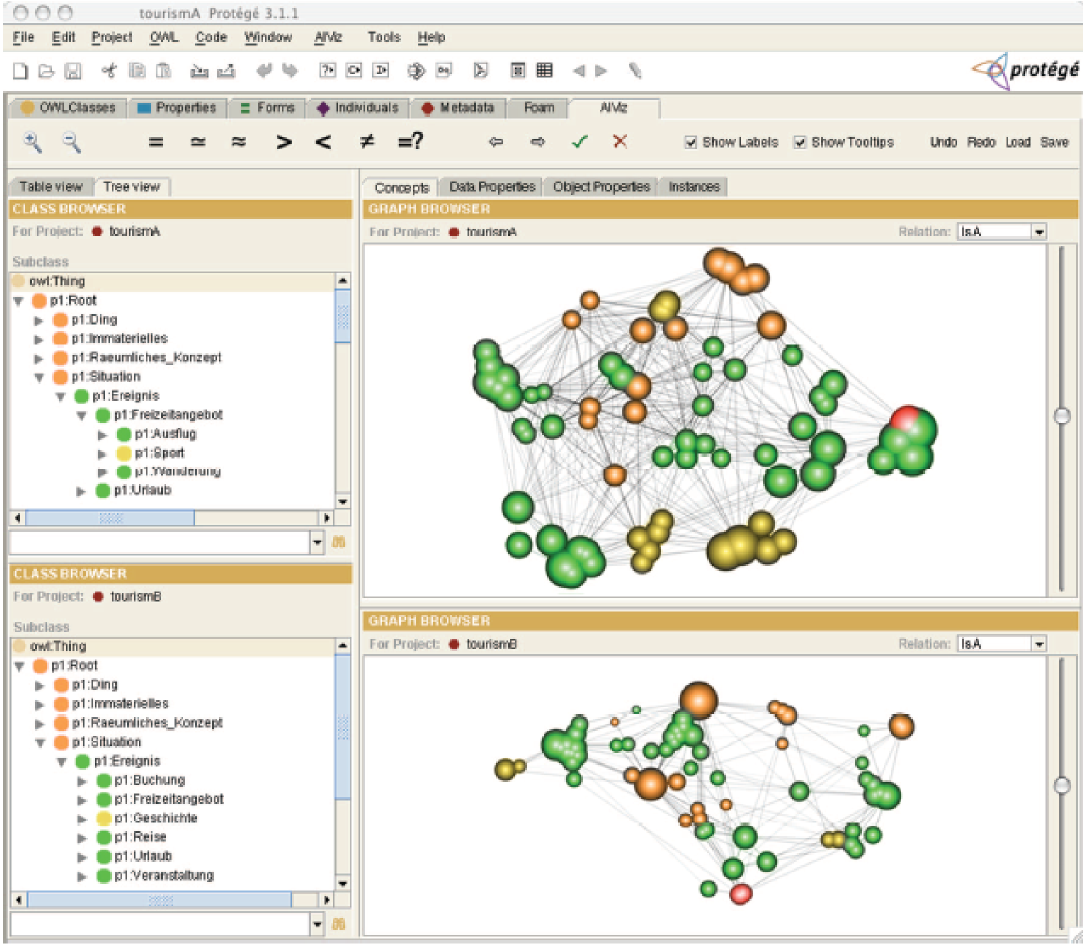
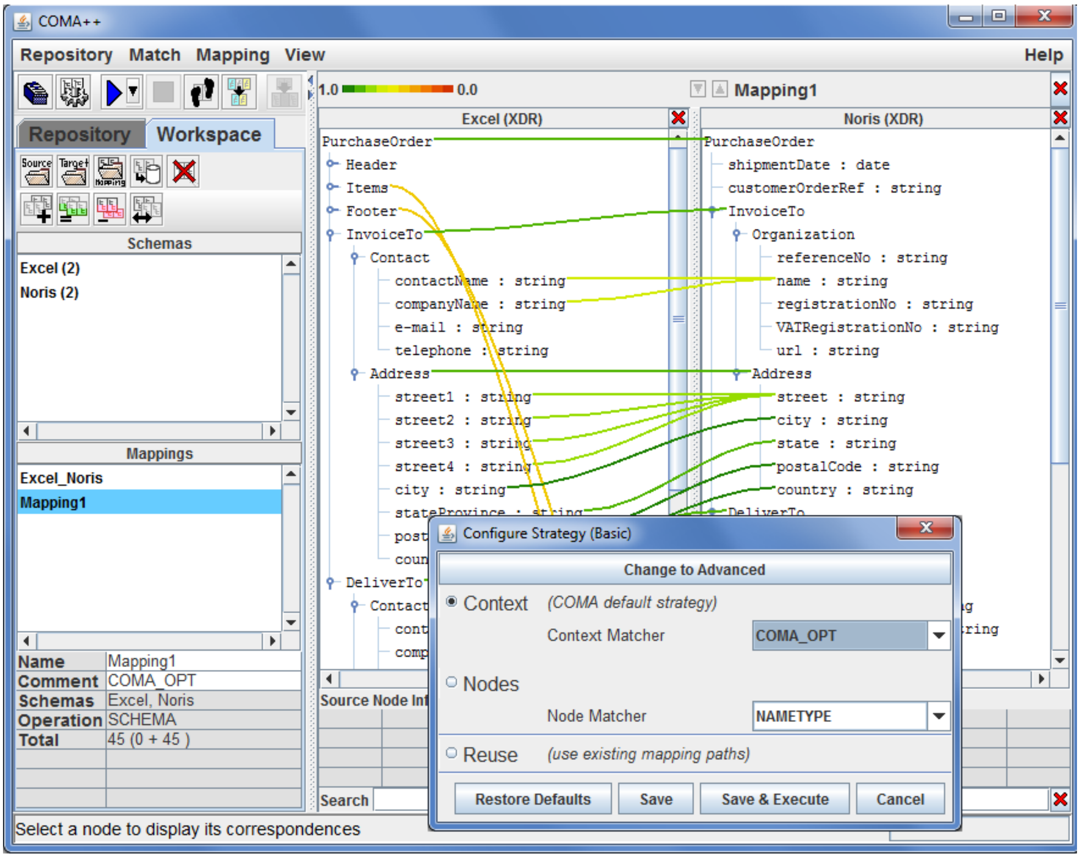


Figure 3.2: Example of graphical user interface of Alviz from (Lanzenberger & Sampson, 2006).



Name Mapping1
Comment COMA_OPT
Schemas Excel, Noris
Operation SCHEMA
Total 45 (0 + 45 )

● Context (COMA default strategy)
Context Matcher COMA_OPT
○ Nodes
Node Matcher NAMETYPE
○ Reuse (use existing mapping paths)

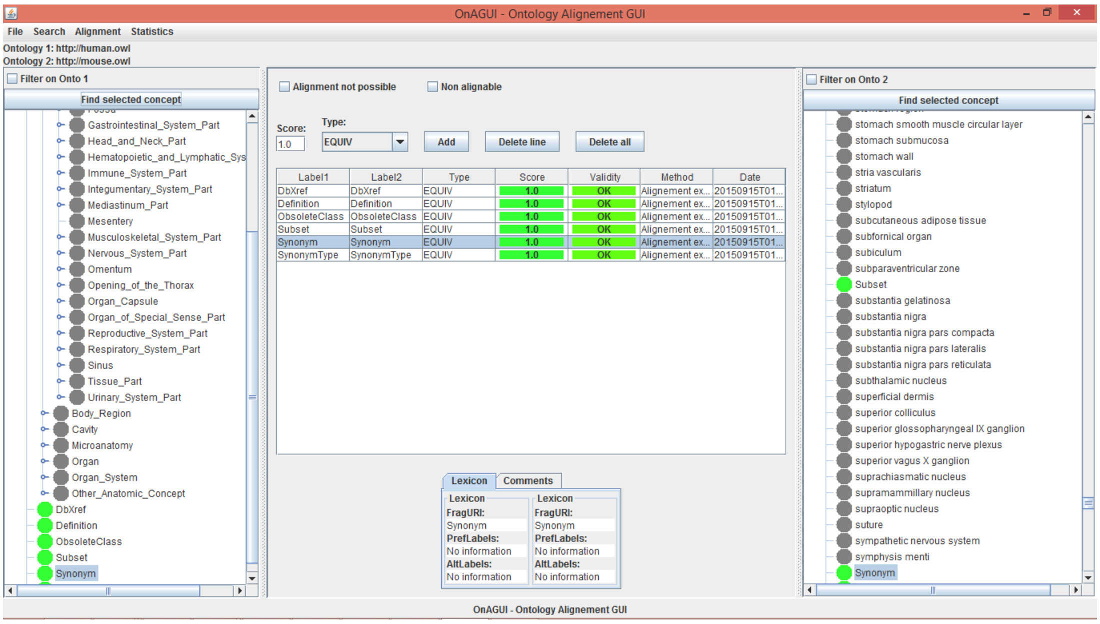
Select a node to display its correspondences

Figure 3.3: Example of graphical user interface of COMA 3.0 from http://dbs.uni-leipzig.de/Research/coma.html.

### 3.1.4 OnAGUI

One of the most recent system in the bunch is OnAGUI (Ontology Alignment Graphical User Interface) (Mazuel & Charlet, 2009). OnAGUI provides the user with the necessary tools to edit an ontology (SKOS, OWL) or an alignment, and to experiment with different ontology matching algorithms.

The system represents each ontology as a tree and gives a list of all the mappings involved in the alignment (Figure 3.4). It also grants the user the opportunity to add and remove a mapping and to leave a comment on a certain mapping.



Ontology 1: http://human.owl
Ontology 2: http://mouse.owl

☐ Alignment not possible ☐ Non alignable

Score: 1.0
Type: EQUIV

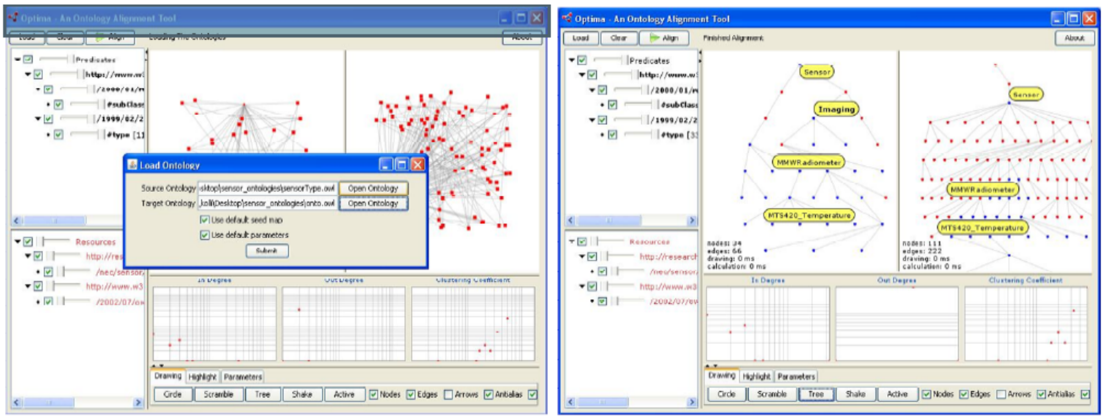
| Label1 | Label2 | Type | Score | Validity | Method | Date |
|---|---|---|---|---|---|---|
| DbXref | DbXref | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |
| Definition | Definition | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |
| ObsoleteClass | ObsoleteClass | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |
| Subset | Subset | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |
| Synonym | Synonym | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |
| SynonymType | SynonymType | EQUIV | 1.0 | OK | Alignement ex... | 20150915T01... |

| Lexicon | Lexicon |
|---|---|
| FragURI: | FragURI: |
| Synonym | Synonym |
| PrefLabels: | PrefLabels: |
| No information | No information |
| AltLabels: | AltLabels: |
| No information | No information |

Figure 3.4: Example of graphical user interface of Ontology Alignment Graphical User Interface.

### 3.1.5 Optima

Optima represents ontologies as graphs (Thayasivam & Doshi, 2011) but, it also enables the visualization in several other formats such as trees and circles (Figure 3.5). The display does not allow zooming, limiting the visualization of large ontologies. The mapped nodes are highlighted and their labels appear when clicked. Moreover, when a node is double-clicked the matched node label appears in the other ontology. The mappings have no graphical representation, nor any listing. Furthermore, the matching technique employed by Optima is also unsuitable to handle large ontologies.



Source Ontology: -sktop\sensor_ontologies\sensorType.owl
Target Ontology: _kolli\Desktop\sensor_ontologies\onto.owl
☑ Use default seed map
☑ Use default parameters

```
nodes: 34
edges: 66
drawing: 0 ms
calculation: 0 ms
```

```
nodes: 111
edges: 222
drawing: 0 ms
calculation: 0 ms
```

Figure 3.5: Example of graphical user interface of Optima.

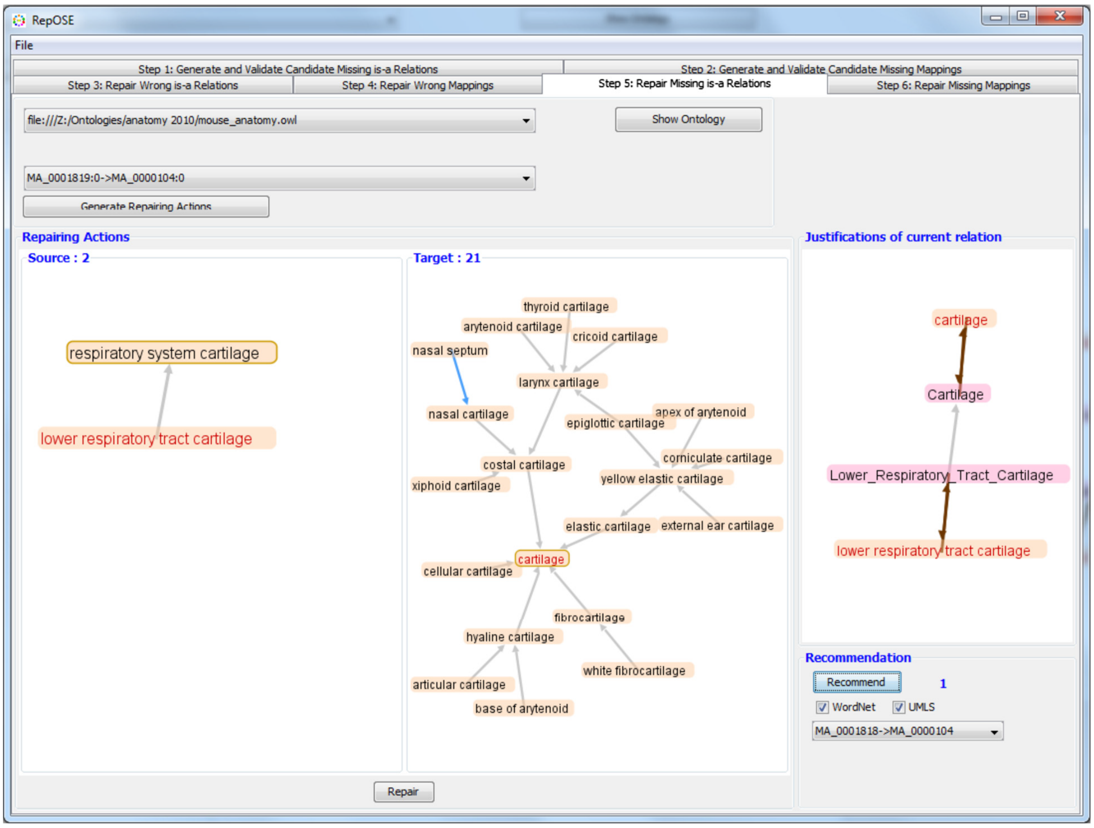
### 3.1.6 RepOSE

RepOSE (Lambrix & Ivanova, 2013) consists on a framework for debugging the 'is_a' structure and mappings between different ontologies in a graph format (Figure 3.6). Besides an overview of each ontology it is possible to visualize potentially wrong mappings. The user can then decide if he wants to apply a repair in the mapping selected or if he wants to repair the whole alignment. Moreover, the system also recommends possible missing mappings giving the user the ability to include or exclude them from the alignment. In the end, it is possible to save the resulting alignment. The many possible editions of the alignment should be accompanied by an undo/redo feature.

The description of this system was based on the most recent article but the version provided by the authors corresponds to the Legacy version. The choice made aimed to give the reader the most updated information.

### 3.1.7 Yet Another Matcher for Ontology Matching

Yet Another Matcher for Ontology Matching (YAM++), represents the ontologies as trees (Duyhoa & Bellahsene, 2013) and the connections between the ontologies



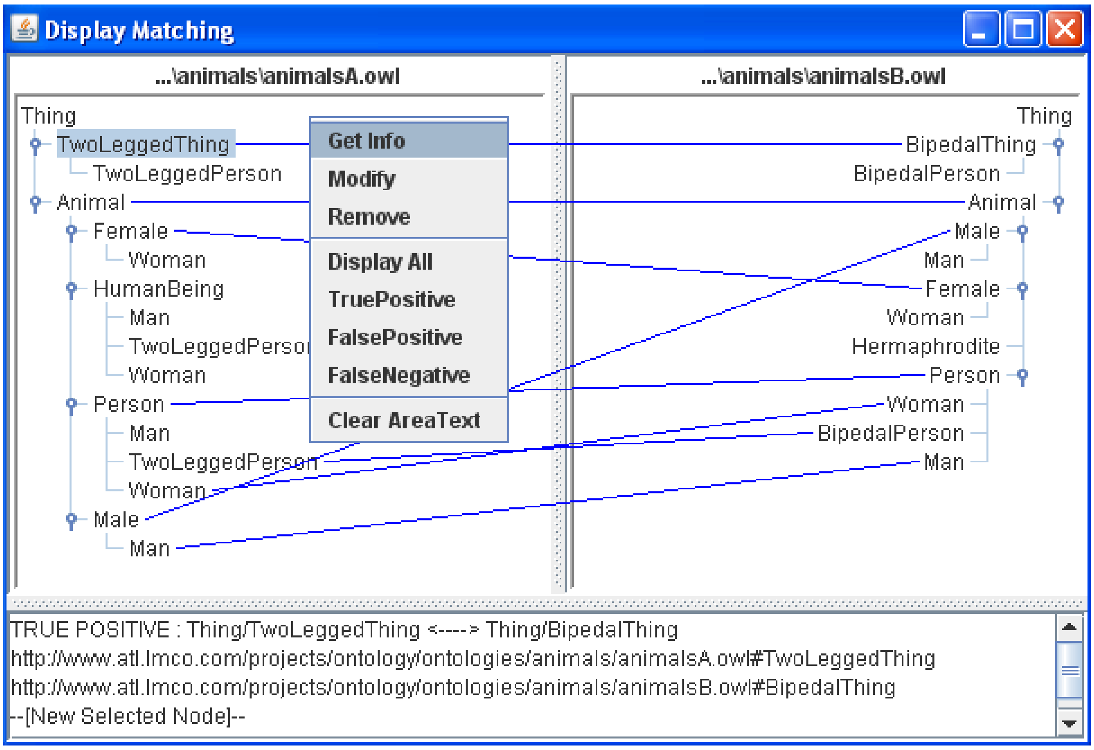
file:///Z:/Ontologies/anatomy 2010/mouse_anatomy.owl

MA_0001819:0->MA_0000104:0

MA_0001818->MA_0000104

Figure 3.6: Example of graphical user interface of RepOSE.

as lines (Figure 3.7). The user can then judge a mapping as correct or incorrect and by that add or remove it from the alignment.
This review was based on the limited information given by the article describing the system due to the unavailability of a functional tool.



```
TRUE POSITIVE : Thing/TwoLeggedThing <----> Thing/BipedalThing
http://www.atl.lmco.com/projects/ontology/ontologies/animals/animalsA.owl#TwoLeggedThing
http://www.atl.lmco.com/projects/ontology/ontologies/animals/animalsB.owl#BipedalThing
--[New Selected Node]--
```

Figure 3.7: Example of graphical user interface of YAM ++.

## 3.2 Overview

The following table compiles the evaluation performed with the previous ontology matching systems and also includes the evaluation of the ontology matching system used as the basis of this thesis, AML (this system is described in detail in Chapter 4). Overall, the systems struggle in loading and showing the very large alignments that result from the matching of very large ontologies, such as biomedical ones. The alignments are composed of several labels and relationships

and the task of showing this intricate structure is often difficult. An overview of those alignments can be very difficult to present and the majority of the tools lacks this feature. One other missing feature is the collaboration module.
All the tools allow the load and save of the alignments. The alignments are represented, by most of the systems, as trees and represent the mappings as lines. Only Optima, Repose and AML have graph representations. The edition of the alignment is possible in the majority of the systems but some do not give information regarding the elements, mapping reasons or confidence, not giving the user all the possible information for him to make the right decision when manually reviewing the alignment.

Table 3.1: Ontology matching visualization systems overview.

| | AgreementMaker | AlViz | COMA 3.0 | OnAGUI | Optima | RepOSE | YAM++ | AgrrementMaker Light |
|---|---|---|---|---|---|---|---|---|
| Ontology Representation | tree | tree and graph | indented graph | tree | multiple | graph | tree | graph |
| Mapping Representation | line | none | line | list | highlights nodes | graph edge | line | graph edge |
| Level of detail adjustment | collapse | collapse | collapse | collapse | No | Yes | collapse | Yes |
| Alignment Overview | No | No | No | No | Yes | Yes | No | No |
| Alignment Edition | Yes | Yes | Yes | Yes | Yes | limited | Yes | No |
| Mapping Confidence | Yes | No | Yes | Yes | No | No | No | Yes |
| Mapping Reason | No | Yes | No | Yes | No | Yes | No | No |
| Element Information | Yes | No | No | Yes | No | No | Yes | No |
| Save and Load | Yes | Yes | Yes | Yes | Yes | Yes | ? | Yes |
| Collaboration | Yes | No | No | Yes | No | No | No | No |
| Support for very large ontologies | Poor | Poor | Poor | Poor | Poor | Limited | Poor | Yes |

Granitzer *et al.* (2010) conducted a similar evaluation to the one here presented. The requirements considered in their evaluation were:

- Detailed mapping information provided;
- Ontology navigation and exploration;
- Overview of alignment results;
- Selectable level of detail and area of interest;
- Filtering.

The evaluation was then implemented in three types of ontology alignment representation: linked trees, graphs and treemaps. For each type, several systems were evaluated. AgreementMaker (Cruz *et al.*, 2009), COMA++ (Massmann *et al.*, 2011), COGZ (Falconer & Storey, 2007) and PROMPT (Noy & Musen, 2003) have a linked tree representation; Optima (Thayasivam & Doshi, 2011) and Alviz (Lanzenberger & Sampson, 2006) were used for graph representation and Treemap (Shneiderman, 1992) and COGZ (Falconer & Storey, 2007) for treemap. The results show that the requirements were closer fulfilled by treemap and graph-based interfaces than tree-based interfaces. The authors suggested a combination of treemaps, to provide an overview of the alignment, and graphs to better navigate and explore the ontologies. An indispensable point to all systems is the need of a table displaying detailed mapping information. One lacking feature in all the systems evaluated in this paper is the filtering capabilities.

The results of the survey conducted in this thesis also support the authors findings related to the lack of information about the elements involved in the alignment, an extremely important feature for the alignment review. Granitzer *et al.* (2010) pointed out in their paper that the filtering capabilities were missing in all the systems evaluated. However, AgreementMaker, AlViz and Optima, also in review in this thesis, have now the filtering capabilities that were unavailable at the article's evaluation time.

The current state of the art lacks the capability to manage very large ontologies and in special biomedical ontologies. Biomedical ontologies are not only large in size but they also comprise multiple labels and multiple relations between the elements increasing the overall complexity. The representation of this kind of ontologies is always accompanied of several computational problems starting from

the loading of large amounts of data, to the visualization of that data. Many decisions should be taken into account when building a visualization framework but we should always consider the great amount of data and its intricate relations.
The user must be given as much information as possible in order for him to understand the problem at hand and contribute to its solution. Nevertheless, majority of the systems do not include the element information nor mapping related information, like the mapping confidence.
This thesis sets its star point at the limitations of the current state of the art and proposes several solutions to overcome them.

# Chapter 4

# Ontology Alignment Visualization

One of the objectives of this thesis was to investigate the challenges in visualization support for manual ontology alignment revision. In this task, the user is given an alignment and asked to remove incorrect mappings and add missing ones. To understand the impact that the visual representation of ontologies and the mappings between them can have on this task, two competing strategies for ontology visualization were assessed: graph versus tree. Treemaps, although successful to support other tasks, do not allow fine-grained visualization and were thus not considered. This was accomplished by improving the graphical user interface of AgreementMakerLight to allow alignment revision, which is graph-based, and comparing it to a popular tree-based system, AgreementMaker.
This chapter includes not only a description of AML framework and current GUI but also the novel additions to the system I designed and implemented. These aimed at adding alignment revision capabilities to AML. A small-scale usability study was conducted to evaluate and compare the graph and tree based strategies of AML and AM.

## 4.1 AgreementMakerLight

AML is an automated ontology matching system based on AM but designed to handle large ontology matching problems (Faria *et al.*, 2014b). The system comprises several ontology matching techniques in particular lexical strategies and

the use of external resources as background knowledge. Furthermore, the alignment can be repaired by an existing module that ensures the alignment coherence. AML is available on github `https://github.com/AgreementMakerLight` in two different formats: a runnable Jar and an Eclipse project.
The Ontology Alignment Evaluation Initiative (OAEI) has contributed to the development of systems like AML by providing a discussion forum for developers and a platform for an annual evaluation of their tools. Yet few provide a user interface although user involvement is an indispensable tool for the development of high-quality alignments. AML performance in this initiative has been improving over the years and it is currently one of the best systems available with top results in all the tasks evaluated (Faria *et al.*, 2015).

### 4.1.1 AgreementMakerLight Framework

The AML framework can be divided in three main modules: ontology loading, ontology matching, and alignment selection and repair (Figure 4.1).
The first module reads the ontologies and parses their information into specific AML data structures, like the Lexicon, which comprises all the class names and synonyms ordered by their matching score. Regarding the ontology matching module, AML contains several ontology matching algorithms and matchers. Some matchers take advantage of background knowledge, a key feature in AML. Finally, the last module is responsible for ensuring that the final alignment has the correct cardinality and that it is coherent.

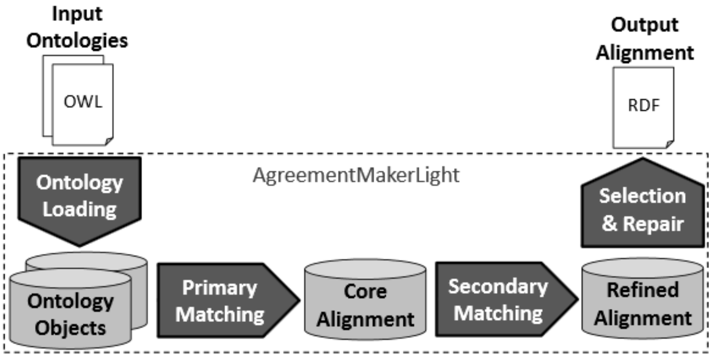
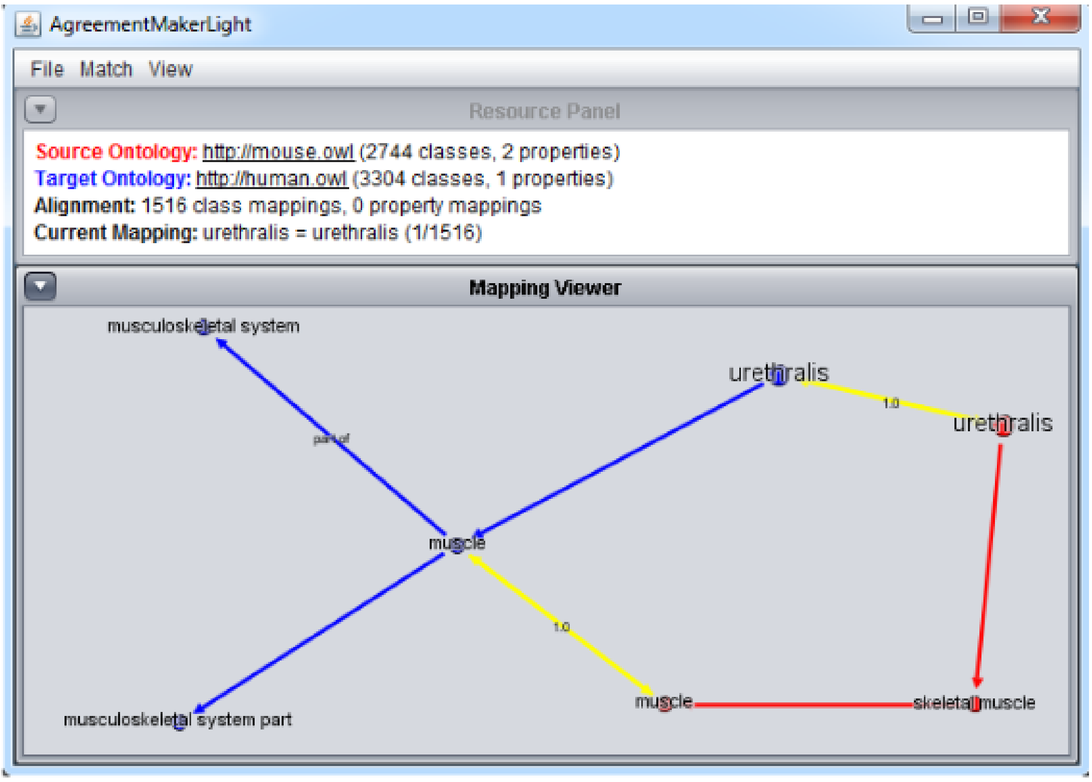


Figure 4.1: AgreementMakerLight ontology matching framework.

### 4.1.2 AgreementMakerLight Graphical User Interface

The graphical user interface of AML (see Figure 4.2) comprises two main areas:

- Resource Panel;
- Mapping Viewer.



**Source Ontology:** http://mouse.owl (2744 classes, 2 properties)
**Target Ontology:** http://human.owl (3304 classes, 1 properties)
**Alignment:** 1516 class mappings, 0 property mappings
**Current Mapping:** urethralis = urethralis (1/1516)

Figure 4.2: AgreementMakerLight Graphical User Interface.

#### 4.1.2.1 Resource Panel

The Resource Panel shows the information about the ontologies by specifying the current source and target ontologies and their number of classes and properties. In addition, it also shows the number of mappings involved in the alignment and the current mapping in display.

#### 4.1.2.2 Mapping Viewer

The mappings and their neighbours are displayed in this section in a graph format implemented by the use of Gephi API (Bastian *et al.*, 2009). The graph represents the different elements using a colour code: blue for source ontologies, red for target ontologies and yellow represents the mappings. The confidence of the mapping, some types of relationships within the ontologies and the nodes main label are also shown.

Moreover, there is also the possibility to expand or restrict the neighbourhood by navigating through the View Menu in the Options section. The neighbourhood of a mapping includes the classes that are at a predefined distance from the mapped classes, and any mappings between them. The neighbourhood radius accepted by AML goes from 0 to a maximum of 5 edges.

The navigation through the alignment can be accomplished in three different ways:

- selection of the next/previous mapping option in the View Menu;
- searching a certain mapping containing a certain term of interest, which is supported by an auto-complete function.

All the alignment changes are recorded and can then be saved by the user in RDF format.

## 4.2 Design and Implementation of Alignment Revision in AML

The role of the user as a reviewer should be supported by all ontology alignment visualization tools. AML interface was improved in order to enhance its review capabilities. The addition of several new features allows the user to make a better decisions about keeping or discarding mapping and, therefore, to achieve a better final alignment. Since AML already possessed graph visualization for selected mappings, this work focused on adding the capabilities necessary to support alignment revision, namely: (1) the addition of the Alignment Reviewer pane, to support mapping browsing and deletion or addition; and (2) the improvement of the Mapping Viewer pane, to provide more information to the user. The new features were implemented in the AML Java project that was already built using the Swing library, that is part of the Oracle's Java Foundation Classes. The Swing library provides an API for the construction of a GUI for Java Programs.

### 4.2.1 Alignment Reviewer

A mapping list was added to the system and can be found in the Alignment Reviewer pane. Each mapping is represented by the main labels of its classes and a symbol representing the type of mapping relation (e.g., = for equivalence). When selecting an individual mapping, the user is presented with a pop-up window which displays two types of information: lexical information and structural information. The lexical information consists on the labels, synonyms or formulas related to the classes involved in the mapping. The structural information section includes the superclasses and the possible disjoints of the mapped classes. The following example (Figure 4.3) shows the lexical information that can be displayed when analysing a certain mapping in detail.

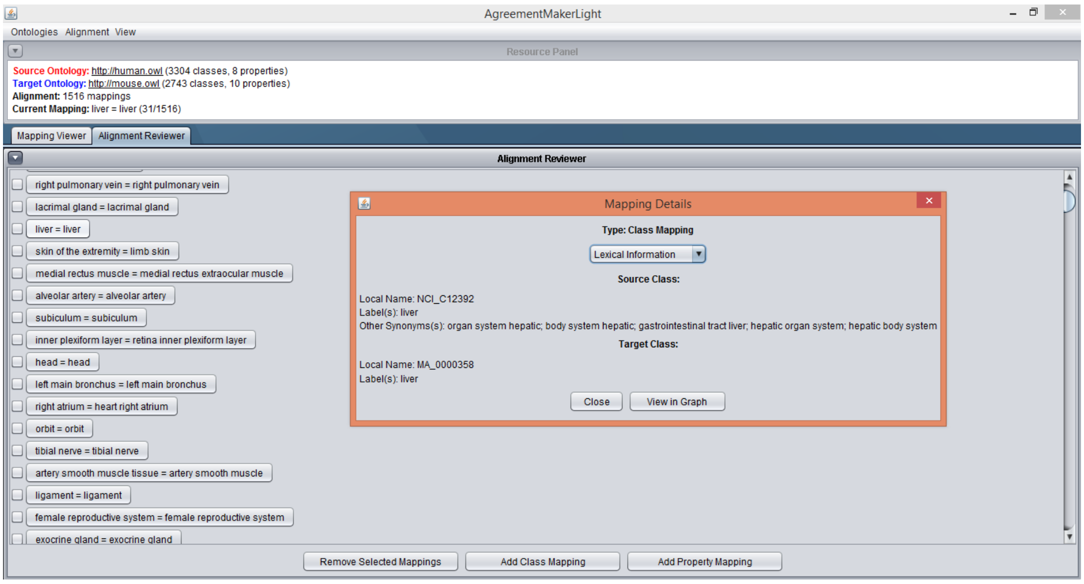
All this information aims at giving relevant background to support revision decisions. This pane does not use visualization strategies, but rather quick summaries of relevant information that can be used to quickly assess mappings.

### 4.2.2 Editing the Alignments

Other improvements were conducted to support the manual revision of the alignment by evaluating if the mappings are correct or incorrect and adding or removing a certain mapping accordingly.

Removing a mapping is a simple process where the user can tick the checkbox next to a mapping to select it for deletion. More than one mapping can be selected at a time, and when the user is satisfied with his/her selection he/she can click the button to actually remove the mappings from the alignment. This does not only remove the mappings from the mappings table and visualizations but also alters the alignment structures of AML.

Mappings can be added either between properties or classes. When adding a class



**Source Ontology:** http://human.owl (3304 classes, 8 properties)
**Target Ontology:** http://mouse.owl (2743 classes, 10 properties)
**Alignment:** 1516 mappings
**Current Mapping:** liver = liver (31/1516)

☐ right pulmonary vein = right pulmonary vein
☐ lacrimal gland = lacrimal gland
☐ liver = liver
☐ skin of the extremity = limb skin
☐ medial rectus muscle = medial rectus extraocular muscle
☐ alveolar artery = alveolar artery
☐ subiculum = subiculum
☐ inner plexiform layer = retina inner plexiform layer
☐ head = head
☐ left main bronchus = left main bronchus
☐ right atrium = heart right atrium
☐ orbit = orbit
☐ tibial nerve = tibial nerve
☐ artery smooth muscle tissue = artery smooth muscle
☐ ligament = ligament
☐ female reproductive system = female reproductive system
☐ exocrine gland = exocrine gland

**Type: Class Mapping**

**Source Class:**

Local Name: NCI_C12392
Label(s): liver
Other Synonyms(s): organ system hepatic; body system hepatic; gastrointestinal tract liver; hepatic organ system; hepatic body system

**Target Class:**

Local Name: MA_0000358
Label(s): liver

Figure 4.3: AgreementMakerLight updated Graphical User Interface.

mapping (Figure 4.4) the user can choose the class from a drop-down menu or search for it in a text-box with auto-complete capabilities. The search looks for main labels. The mapping relation can also be specified and the user can choose between equivalence, superclass, subclass, overlap or unknown.

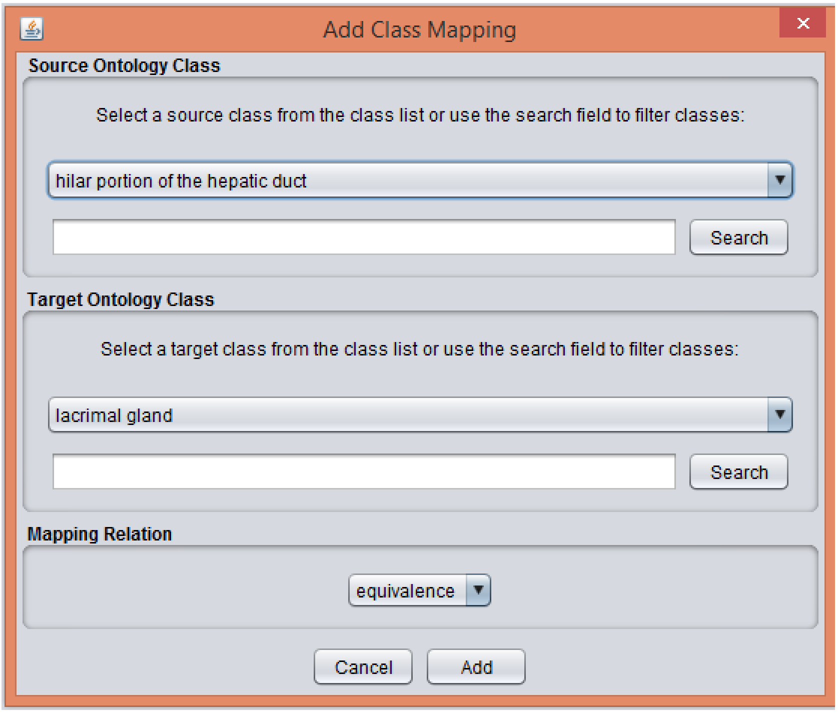
In contrast, the property mapping (Figure 4.5) has other features to choose from. In this case, the user has to specify the property type, then select the corresponding source and target property from a drop-down menu and finally select the relation types.

The user can download the manually repaired alignment in the end of his evaluation, either as RDF or as a tab-separated text file.

### 4.2.3 Mapping Viewer Update

The mapping viewer was subjected to several changes in order to improve the exploration of the alignment and also to give a quicker access to the information of each element involved in the mapping displayed (see Figure 4.6).

The previous version of the Mapping Viewer only displayed the graph representation of the selected mapping and its close neighbourhood. However, and given



Select a source class from the class list or use the search field to filter classes:

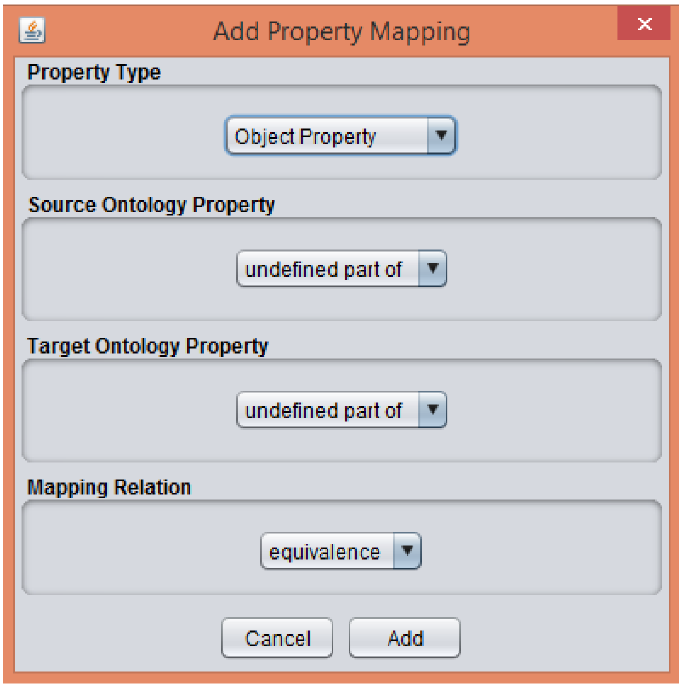
hilar portion of the hepatic duct

Select a target class from the class list or use the search field to filter classes:

lacrimal gland

equivalence

Figure 4.4: AgreementMakerLight add class mapping feature.



Object Property

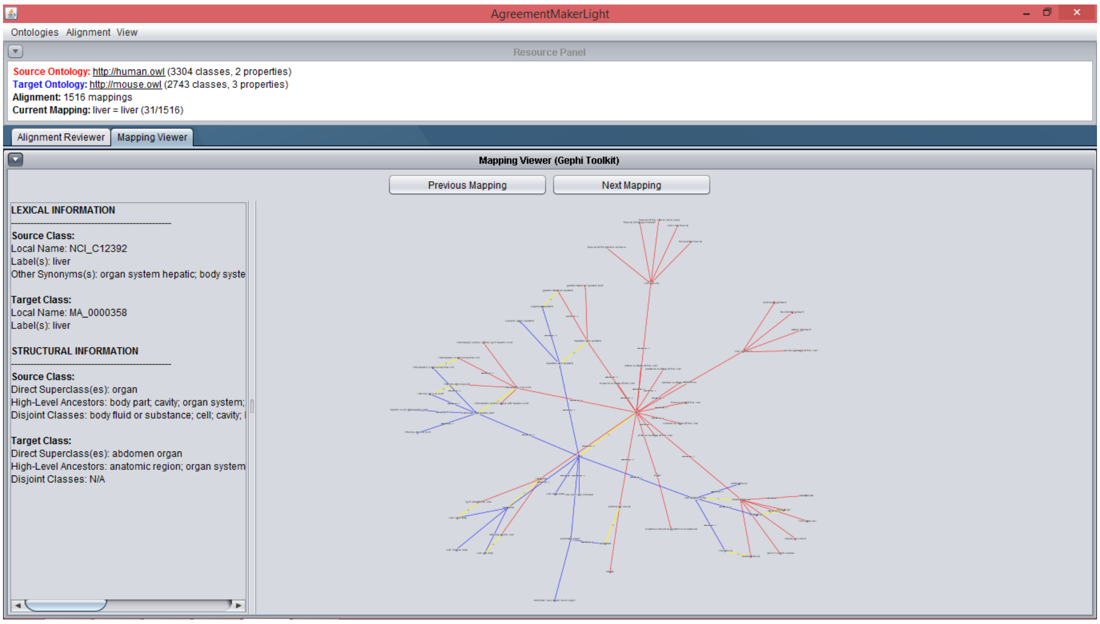
undefined part of

undefined part of

equivalence

Figure 4.5: AgreementMakerLight add property mapping feature.

that sometimes the decision process needs not just the local structural information, but also a more global perception, a mapping information panel was added alongside the graph visualization. The mapping information gives not only lexical information like synonyms or formulas, but also structural information that comprise information about ancestor and disjoint classes. The user is exposed to the information without leaving the mapping viewer and without having to scroll the mapping list in the alignment reviewer tab, which aims at making it easier for him to comprehend more efficiently the elements being visualized. This way we are combining the local structure given by the graph with the lexical and global structural information of the mappings in parallel.
Besides that, a new way to navigate through the alignment was added by the addition of previous and next arrows directly in the Mapping Viewer tab.



**Source Ontology:** http://human.owl (3304 classes, 2 properties)
**Target Ontology:** http://mouse.owl (2743 classes, 3 properties)
**Alignment:** 1516 mappings
**Current Mapping:** liver = liver (31/1516)

**LEXICAL INFORMATION**
--------------------------------------------------
**Source Class:**
Local Name: NCI_C12392
Label(s): liver
Other Synonyms(s): organ system hepatic; body syste

**Target Class:**
Local Name: MA_0000358
Label(s): liver

**STRUCTURAL INFORMATION**
--------------------------------------------------
**Source Class:**
Direct Superclass(es): organ
High-Level Ancestors: body part; cavity; organ system;
Disjoint Classes: body fluid or substance; cell; cavity;

**Target Class:**
Direct Superclass(es): abdomen organ
High-Level Ancestors: anatomic region; organ system
Disjoint Classes: N/A

Figure 4.6: AgreementMakerLight new Mapping Viewer tab.

## 4.3 Evaluation and Discussion

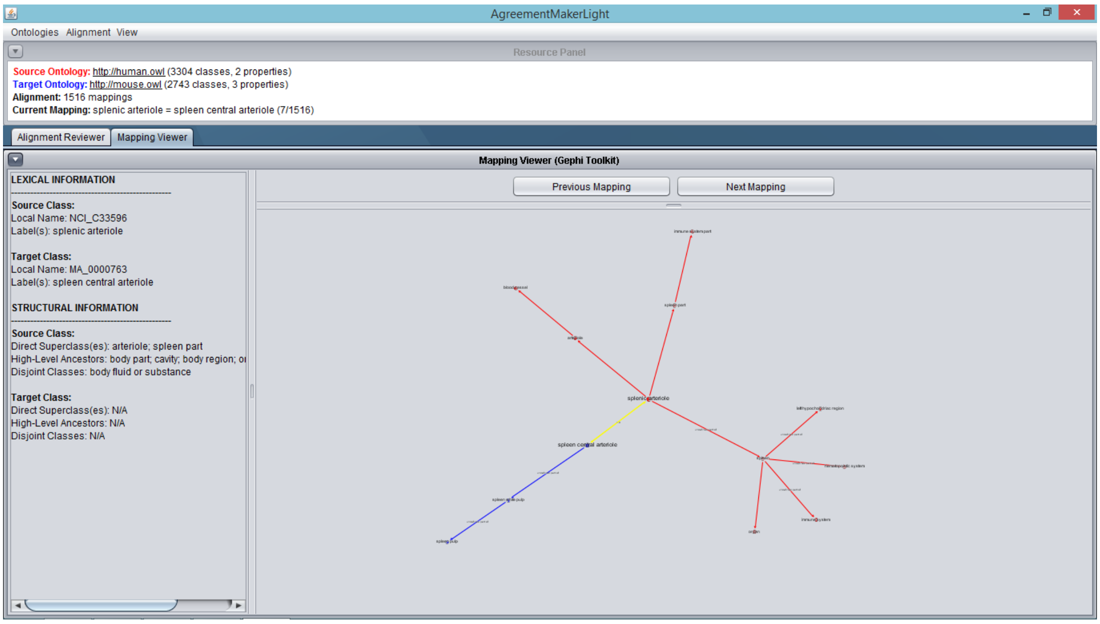
In order to establish the importance of the changes performed we invited users to participate in a small test and to answer a standard user usability questionnaires.

We were able to gather seven individual samples to evaluate our tool. While we acknowledge that this is a relatively small sample, it was not possible to gather a higher number of participants, given the fact that users need some understanding of ontologies and ontology alignment. The users were graduate students in the area of bioinformatics and all knew the basic ontology concepts. In the beginning of the test, the participants were given a small background presentation about ontology alignment necessary to better comprehend the tasks that they were going to execute.

The questionnaire chosen for this purpose was System Usability Scale (SUS) (Brooke, 1996). The questionnaire is composed of a ten-item scale that gives a global assessment of the usability of a user based interface. The Likert scale used allows the user to evaluate its degree of agreement or disagreement towards a certain statement on a 5 point scale. The overall question scores are comprised in Appendix A.

Besides the questionnaire, the participants were given tasks in order to better evaluate the performance of the system's graphical user interface. The tasks were chosen to compel the user to use all the new features of AML and evaluate their performance. The tasks are summed up in Table 4.1. The human and mouse ontologies (Bodenreider *et al.*, 2005) and their reference alignment from OAEI 2015 were the dataset chosen for the evaluation. The choice was based on the size of these two ontologies, considered very large but still able to be loaded by AM, and given that their domain is not too complex for a non-biomedical expert to grasp. Moreover, a reference alignment for these ontologies has been created as part of the OAEI evaluation.

Users were asked to complete the tasks in AML and also AM in order to evaluate which visualization format (graph or tree) was best to help the user in the alignment exploration and edition. The participants were separated randomly into two groups, and four participants started by evaluating AML and the other tree started by AM. AM was the chosen system to comparison because it displays a different visualization paradigm and its available. The 2012 version of AM was used due to issues in setting up the testing environment using the more recent version.



**Source Ontology:** http://human.owl (3304 classes, 2 properties)
**Target Ontology:** http://mouse.owl (2743 classes, 3 properties)
**Alignment:** 1516 mappings
**Current Mapping:** splenic arteriole = spleen central arteriole (7/1516)

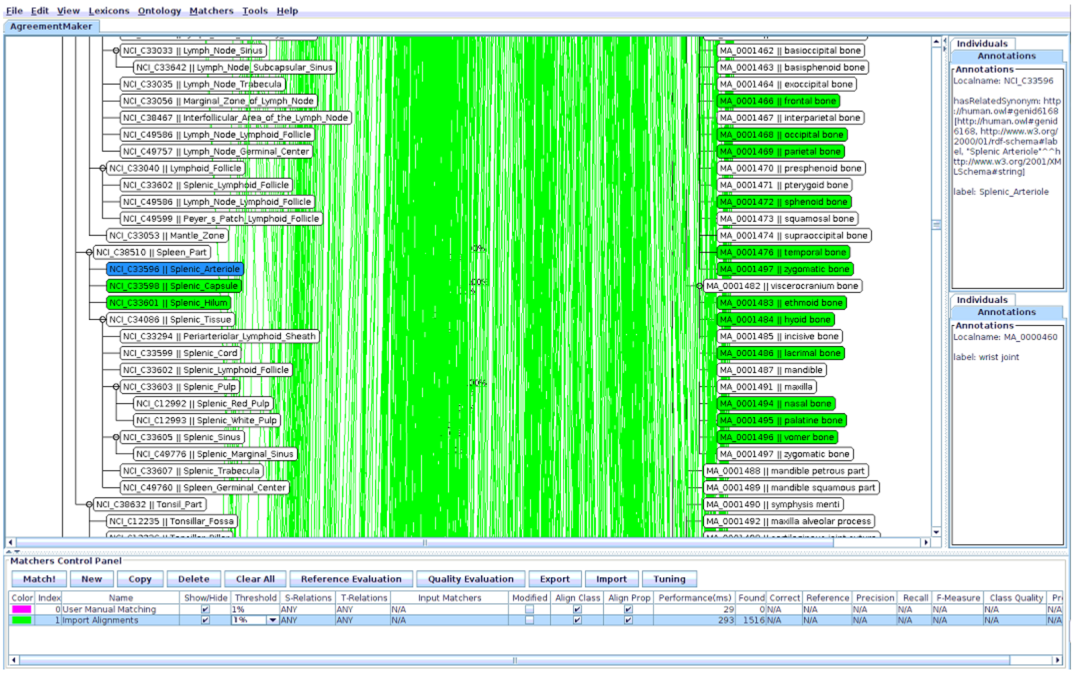
**LEXICAL INFORMATION**
------------------------------------------------
**Source Class:**
Local Name: NCI_C33596
Label(s): splenic arteriole

**Target Class:**
Local Name: MA_0000763
Label(s): spleen central arteriole

**STRUCTURAL INFORMATION**
------------------------------------------------
**Source Class:**
Direct Superclass(es): arteriole; spleen part
High-Level Ancestors: body part; cavity; body region; o
Disjoint Classes: body fluid or substance

**Target Class:**
Direct Superclass(es): N/A
High-Level Ancestors: N/A
Disjoint Classes: N/A

Figure 4.7: Visualization of the mapping 'splenic arteriole' - 'spleen central arteriole' in AgreementMakerLight.



**Annotations**
Localname: NCI_C33596

hasRelatedSynonym: http://human.owl#genid6168 [http://human.owl#genid6168, http://www.w3.org/2000/01/rdf-schema#label, "Splenic Arteriole"^^http://www.w3.org/2001/XMLSchema#string]

label: Splenic_Arteriole

**Annotations**
Localname: MA_0000460

label: wrist joint

| Color | Index | Name | Show/Hide | Threshold | S-Relations | T-Relations | Input Matchers | Modified | Align Class | Align Prop | Performance(ms) | Found | Correct | Reference | Precision | Recall | F-Measure | Class Quality | Pr |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 🟪 | 0 | User Manual Matching | ☑ | 1% | ANY | ANY | N/A | ☐ | ☑ | ☑ | 29 | 0 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |
| 🟩 | 1 | Import Alignments | ☑ | 1% | ANY | ANY | N/A | ☐ | ☑ | ☑ | 293 | 1516 | N/A | N/A | N/A | N/A | N/A | N/A | N/A |

Figure 4.8: Visualization of the mapping 'splenic arteriole' - 'spleen central arteriole' in AgreementMaker.

Table 4.1: Evaluation Task of AgreementMakerLight and AgreementMaker Graphical User Interface.

| Number | Task |
|---|---|
| 1 | Remove the mapping<br>'splenic arteriole' - 'spleen central arteriole' (Figure 4.7 and 4.8). |
| 2 | Add the mapping<br>'nerve plexus' - 'nerve plexus'. |
| 3 | Search for the mapping<br>'vitreous humor' - 'vitreous humor'. |
| 4 | Find the disjoint classes with the source class<br>in the mapping 'dura mater' - 'dura mater'. |
| 5 | Find the mapping 'white matter - 'white matter'. |
| 6 | Maximize the radius of the graph. |
| 7 | Find the definition of the target class<br>in the mapping 'suprarenal artery' - 'suprarenal artery'. |

### 4.3.1 Tasks Evaluation

The users had several difficulties when performing the tasks in AM finding it slow and usually not very intuitive to search or visualize a mapping in particular and to add and remove a mapping. Moreover, information about the disjoint classes was not provided in the element information. The tasks were more easily executed by AML but not without its challenges. The search in the Alignment Review list proved to be very cumbersome, with the users having to scroll through all the mappings in order to find the desired mapping. The button responsible for the graph radius maximization was, also, not very intuitive to the users. They thought that zooming in and out was the correct way of defining the radius. The most crucial complain was the lack of interactivity of the graph. A list of suggestions was compiled in Table 4.2.

### 4.3.2 System Usability Scale Evaluation

The System Usability Scale questionnaire comprises ten questions that are scored by the level of user agreement (5) or disagreement (1). The scores given by the participants will be converted in a final value. The final value is calculated in

Table 4.2: List of user suggestion.

| Suggestion | Frequency |
|---|---|
| Search in the Alignment Reviewer tab | 6 |
| Text and numbers in the graph are small | 1 |
| When adding a mapping the search should appear first and the list second | 1 |
| Rename 'Search Alignment' for 'Search Mapping' | 6 |
| Interactive graph | 6 |

two phases. First the scores of the even question numbers are subtracted by five and the odd numbers are subtracted by one. In the second phase, all the values obtained in the previous phase are summed up and then multiplied by 2.5. The resulting value corresponds to the final score that is not considered a percentage and it should only be considered in terms of their percentile ranking. According to Brooke (1996) a score above 68 is considered above the average.

AML scored 75.36 in the questionnaire, a result above the average, showing the promising application of AML system although some issues should be overcome in the future. A simple statistical overview can be seen in Table 4.3. The result analysis show that the questions with a higher variability of answers were 1, 4, 7 and 8 with a standard deviation above 1. The subject of those questions was related to future use, the need of assistance to use the system and the complexity of the system. The results support the difficulty that some users had to find some features to accomplish some of the tasks given and by the fact that they were not experts in the area. Nevertheless, overall the participants thought the system was easy to use and had a consistent and well integrated framework reinforced by good average scores in all the questions.

The same questionnaire was applied to the performance of AM on the same tasks and the statistical evaluation of the questionnaire is displayed in Table 4.4. The score for AM was lower than the AML score falling in 38.93 percentile. The overall results show a high variability in question 4, 7, 8, 9 and 10 with standard deviations above one. All the questions comprehended issues about the complexity of the system itself, the confidence the system granted, if it was easy too learn and also the necessity to learn more things to use it. The results

Table 4.3: Questionnaire statistical evaluation of AgreementMakerLight.

| Question | Average | Median | St. Dev. |
|---|---|---|---|
| I think that I would like to use this system frequently. | 3.86 | 4 | 1.07 |
| I found the system unnecessarily complex. | 2.00 | 2 | 0.82 |
| I thought the system was easy to use. | 3.43 | 3 | 0.98 |
| I think that I would need the support of a technical person to be able to use this system. | 2.14 | 2 | 1.35 |
| I found the various functions in this system were well integrated. | 4.43 | 4 | 0.53 |
| I thought there was too much inconsistency in this system. | 1.43 | 1 | 0.53 |
| I would imagine that most people would learn to use this system very quickly. | 4.14 | 5 | 1.21 |
| I found the system very cumbersome to use. | 1.71 | 1 | 1.11 |
| I felt very confident using the system. | 3.71 | 4 | 0.76 |
| I needed to learn a lot of things before I could get going with this system. | 2.14 | 2 | 0.90 |
| Final Score | 75.36 | | |

reinforce the different expertise demonstrated by the users and the difficulties that the system encompasses in the task completion. The average mean shows that the users found the system unnecessarily complex and difficult to use with scores of 3.29 and 2.43, respectively.

The users were split into two groups to minimize the influx of the order of the tool tests. These results are summarized in tables 4.5 and table 4.6 and show that the system that was firstly used had slightly higher average and median values in almost all the questions than the one evaluated in second. In AML case the one question that did not follow the pattern where the ones relative to the system complexity, inconsistency and difficult to learn how to learn how to use it. on the other hand, am questions were related to support of a technician or previous study to learn how to use the system or its usage difficulty. These results indicate that the novelty factor may impact user opinion positively.

The difference between the scores of both systems favours the AML system when comparing to AM. The graph format in AML is suitable to handle large

Table 4.4: Questionnaire statistical evaluation of AgreementMaker.

| Question | Average | Median | St. Dev. |
|---|---|---|---|
| I think that I would like to use this system frequently. | 2.00 | 2 | 0.82 |
| I found the system unnecessarily complex. | 3.29 | 3 | 0.76 |
| I thought the system was easy to use. | 2.43 | 2 | 0.79 |
| I think that I would need the support of a technical person to be able to use this system. | 3.43 | 4 | 1.40 |
| I found the various functions in this system were well integrated. | 2.29 | 2 | 0.49 |
| I thought there was too much inconsistency in this system. | 3.29 | 3 | 0.76 |
| I would imagine that most people would learn to use this system very quickly. | 2.71 | 3 | 1.11 |
| I found the system very cumbersome to use. | 3.57 | 4 | 1.27 |
| I felt very confident using the system. | 2.71 | 3 | 1.25 |
| I needed to learn a lot of things before I could get going with this system. | 3.00 | 3 | 1.29 |
| Final Score | 38.93 | | |

Table 4.5: Questionnaire statistical evaluation of AgreementMakerLight taking into account the order.

| AgreementMakerLight | | | | | | |
|---|---|---|---|---|---|---|
| | First | | | Second | | |
| Question | Average | Median | St. Dev. | Average | Median | St. Dev. |
| 1 | 4.00 | 4.5 | 1.41 | 3.67 | 4 | 0.58 |
| 2 | 2.00 | 2 | 0.82 | 2.00 | 2 | 1.00 |
| 3 | 3.50 | 3.5 | 0.58 | 3.33 | 3 | 1.53 |
| 4 | 2.25 | 2 | 1.26 | 2.00 | 1 | 1.73 |
| 5 | 4.50 | 4.5 | 0.58 | 4.33 | 4 | 0.58 |
| 6 | 1.25 | 1 | 0.50 | 1.67 | 2 | 0.58 |
| 7 | 4.00 | 4.5 | 1.41 | 4.33 | 5 | 1.15 |
| 8 | 1.50 | 1.5 | 0.58 | 2.00 | 1 | 1.73 |
| 9 | 3.75 | 3.5 | 0.96 | 3.67 | 4 | 0.58 |
| 10 | 2.00 | 2 | 0.82 | 2.33 | 3 | 1.15 |

Table 4.6: Questionnaire statistical evaluation of AgreementMaker taking into account the order.

| AgreementMaker | | | | | | |
|---|---|---|---|---|---|---|
| | First | | | Second | | |
| Question | Average | Median | St. Dev. | Average | Median | St. Dev. |
| 1 | 2.67 | 3 | 0.58 | 1.50 | 1.5 | 0.58 |
| 2 | 3.33 | 4 | 1.15 | 3.25 | 3 | 0.50 |
| 3 | 3.00 | 3 | 1.00 | 2.00 | 2 | 0.00 |
| 4 | 2.67 | 2 | 1.15 | 4.00 | 4.5 | 1.41 |
| 5 | 2.33 | 2 | 0.58 | 2.25 | 2 | 0.50 |
| 6 | 3.67 | 4 | 0.58 | 3.00 | 3 | 0.82 |
| 7 | 3.67 | 4 | 0.58 | 2.00 | 2 | 0.82 |
| 8 | 3.00 | 3 | 2.00 | 4.00 | 4 | 0.00 |
| 9 | 3.67 | 3 | 1.15 | 2.00 | 2 | 0.82 |
| 10 | 2.67 | 3 | 1.53 | 3.25 | 3 | 1.26 |

amounts of data and to better represent the hierarchical relations between the ontology classes. AM tree format was unnecessarily complex for the users and the relationships between the classes were not very intuitive. The tree format was not the best representation for the large amount of information visualized, which was reflected in the questionnaire results. The difficulties experienced in the tree format would be magnified if the ontologies were comprised of forty thousand classes, such as Gene Ontology. The ontologies chosen for the evaluation were only an example and, although being considered very large are only comprised of around three thousand classes.

AML also performed better with the high number of classes and relationships because, it only shows a mapping and a part of its neighbourhood (that can be specified by the user) instead of showing the whole alignment, like AM. The whole alignment visualization can help with an overview of the information but it tends to be very cumbersome, confusing and computationally difficult.

Ivanova *et al.* (2015) also did a recent evaluation with 8 participants (3 master and 5 PhD students) with background in Computer Sciences, using the SUS in three different systems: COMA (Massmann *et al.*, 2011), SAMBO (Lambrix & Tan, 2006) and CogZ (Falconer & Storey, 2007). The scores were respectively, around 60, around 50 and around 70. Although the set of tasks proposed by Ivanova was larger than the one we used, AML's score of 75.38 is still a good indicator of its above average performance.

The results reinforced the fact that a graph format handles better with large amounts of classes, with complex hierarchies between them, and the visualization of only a section of the alignment instead of the whole alignment, helped with the interpretation of what was being visualized. The tree format showing the whole alignment implement by AM seems to be too cumbersome and more difficult to understand than the AML graph format displaying just a section of the alignment, as the user opinion demonstrated. The partial representation of the set of mappings between the ontologies helps in the understanding and exploration of the alignment giving a more natural and intuitive interface.

The AML GUI also benefits from an alignment review module that allows the user to edit the final alignment by adding or removing mappings. Moreover, the user has access to detailed element information that aims to help in the decision making and, in the end, to improve the quality of the final alignment. These led to improvements in the listed criteria in table 3.1, specifically to the alignment edition and element information.

The evaluation brought up not only system assets but also some limitations to the current AML GUI and future work will focus on overcoming the limitations found and improve on the current state of the art.

# Chapter 5

# Alignment Repair Visualization

The aim of all matching systems is to achieve the best alignment between different ontologies. Nevertheless, incoherences between the mappings can arise due either to erroneous mappings or incompatibilities between both ontologies (Meilicke & Stuckenschmidt, 2008). The example in Figure 5.1 can help better understand this concept.

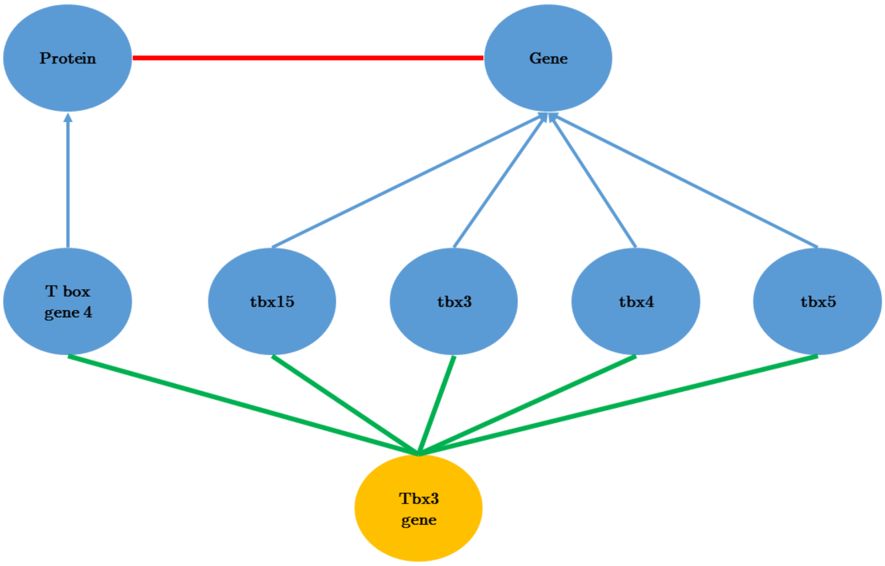


Figure 5.1: Example of a conflict set between National Cancer Institute Thesaurus (yellow) and Bone Dysplasia Ontology (blue). The arrows correspond to subclass of relationships, green lines correspond to mapping correspondences and red to the disjoint relationship.

As you can see in this example, 'protein' and 'gene' from the Bone Dysplasia Ontology are disjoint classes. That said, an equivalence or subsumption cannot be created between the descendants of these classes without creating a logical incoherence. However, when aligned with the ontology National Cancer Institute Thesaurus a connection is created between the descendant of the disjoint classes, 'protein' and 'gene', and 'tbx3gene'. This connection creates a conflict in the alignment, because the mappings 'gene'-'tbx15', 'gene'-'tbx3', 'gene'-'tbx4', 'gene'-'tbx5' and mapping 'protein'-'T box gene 4' cannot co-exist without forming a logical incoherence. To solve the incoherence, manual or automated repair techniques can be employed. Automated repair may not always be the best option because when removing a mapping we can lose useful information. It has been argued that, ultimately, the decision of whether or not the conflicts should be repaired should fall upon the user (Pesquita *et al.*, 2013). In order to support this decision, users must be given the necessary information to make the right call. This thesis proposes that the best way to do that is by giving the user the chance to visualize the conflicting mappings and their relevant context.

This chapter describes the design, implementation and evaluation of a webtool for the visualization of conflict sets between 10 pairs of ontologies from Bioportal, the largest repository of biomedical ontologies and alignments.

## 5.1 Incoherence in Biportal mappings

Bioportal (Whetzel *et al.*, 2011) is a web portal with access to more than 400 biomedical ontologies (in OBO format or OWL format) developed by the National Center for Biomedical Ontology (NCBO). The user can browse through the different biomedical ontologies looking for a specific class or their neighbours, moreover, it is possible to download different versions of the ontology which allows to see how it evolved along the years. In addition to the visualization of the ontology itself, Bioportal also maps the classes between different ontologies. There are two ways by which this is done: the mappings can be automatically generated or manually added by experts.

To address this issue, (Faria *et al.*, 2014a) applied both AML (Faria *et al.*, 2014b) and LogMap (Jiménez-Ruiz & Grau, 2011) to detect and repair the incoherences in 19 pairs of ontologies from Bioportal and their mappings, and discovered that 11 in 19 had logical errors involving in average 22% of the mappings. The purpose of the algorithms is to eliminate incoherences by removing or altering mappings, and although they can provide logically sound solutions, these may not always be correct, since the choice of which mappings to eliminate is based on a change minimization strategy. A good example of this problem can be seen in Figure 5.2 extracted from (Pesquita *et al.*, 2013). The example shows that different strategies result in different mappings removed but generally there is a risk to lose the correct mapping. In this case, the correct mapping 'Gingiva'-'Gingiva' from the alignment of FMA with NCIT was removed by ALCOMO repair algorithm. The lost of the correct mapping is detrimental to the construction of a coherent final alignment. A solution suggested by the authors in this paper is the enrichment of the Bioportal mappings with annotations about potential logical conflicts with other mappings. This would facilitate the discovery of erroneous mappings or potential errors in the ontologies.

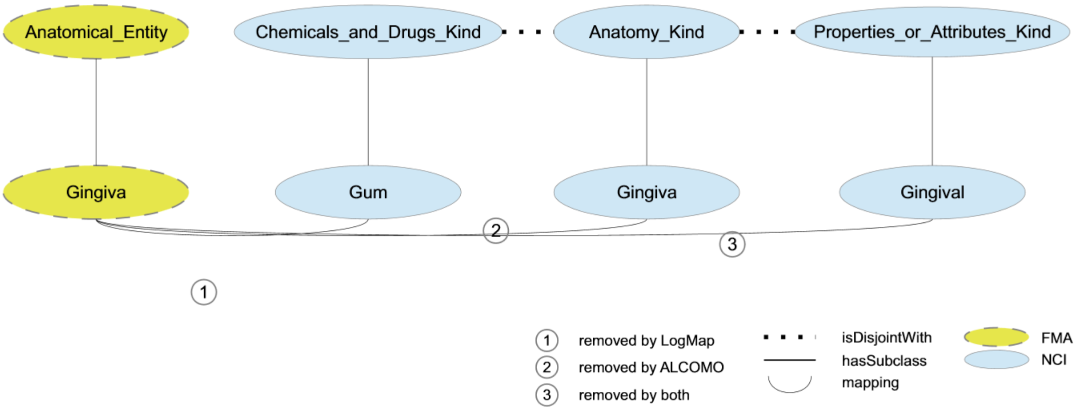


Figure 5.2: Example of repair implemented by ALCOMO and LogMap system in (Pesquita *et al.*, 2013) in the pair of ontologies FMA and NCIT.

The AML repair algorithm (Faria *et al.*, 2014a) computes the core fragments, that only contain the necessary classes and relations to detect the incoherences and the minimal set of classes that need to be checked for incoherences. Then, a full depth-first search will be implemented in each class of the list created previously in order to find all the mapping sets that will lead to an incoherence. The conflicted sets created must be repaired always aiming to remove the least amount of mappings as possible. The heuristic used consists on the iterative removal of the mapping shared by the highest number of conflicting sets, and the lowest confidence values are treated as a tie breaker. The output comprises a set of repaired mappings, an approximate mappings repair, the set of conflicting mappings and the set of all mappings involved in at least one conflicting set.

Regarding the LogMap algorithm it uses a different strategy based on approximate mapping repair techniques. The strategy is based on the assumption that subclasses of an unsatisfiable class are also unsatisfiable so before repairing an unsatisfiable class the repair must, first be on its superclasses. The repair also aims to execute the smallest size repair as possible, in other words, the one with the lowest aggregated confidence taking into account the confidence values assigned to mappings.

In this thesis, it is proposed that the understanding of a complex situation like mapping incoherences is more easily and efficiently understood if the user visualizes the problem at hand. Aiming to address this challenge, we created a web-based visualization tool that identifies the incoherences caused by the mappings between Bioportal ontologies pairs detected by the AML repair algorithm (Santos *et al.*, 2013), in order to support their identification and correction by expert users.

## 5.2 Challenges in visualizing mapping incoherences

The challenges in the visualization of biomedical ontologies alignments were covered in the previous chapters and comprised scalability and complexity issues related to the size and complexity of these ontologies. However, the visualization of mapping incoherences poses additional challenges and constraints. The approach followed here was to identify the minimal information necessary to visualize and understand a conflict set.

Expanding on (Pesquita *et al.*, 2014) and (Ivanova *et al.*, 2015), and taking into account the definitions of logical incoherence in ontology alignment provided in

(Faria *et al.*, 2014a), we arrived at the following minimum set of information to show when displaying a set of conflicting mappings comprises:

- the classes involved in the mappings
- the mappings between classes
- the disjoint axiom involved in causing the incoherence
- the relations between the mapped classes and the classes involved in the disjoint axiom.

All these requirements were taken into consideration during the development of the webtool here presented.

## 5.3 Web application

The backend of the tool is supported by a relational database that stores all the ontology and alignment information. This corresponds to the ontologies and alignments shown in Table 5.1, as well as a list of the conflict sets. Having data stored in a database allows for faster retrieval of the information to draw the graph, rather than parsing the ontology files, since we only require a subset of the information for each conflict set.

The webtool starts to ask the user to select the pair of ontologies that he or she wants to visualize. The pairs of ontologies available can be seen in Figure 5.1. These correspond to the conflicts computed by AML repair algorithm on 10 pairs of ontologies from Bioportal and their mappings in (Faria *et al.*, 2014a). Following the selection of the pair of ontologies the user is presented with a list of mappings involved in each conflict (i.e., the mappings that taken together cause an incoherence) related to the alignment of those ontologies (Figure 5.3). Due to space constraints, the list only comprehends 10 mappings for each conflict set at most, the rest are hidden and can only be displayed if the user wishes to. It is also possible to click on all the elements involved in the mapping to display the lexical information of the element including labels and URIs.

**Select the Alignment:** BDO_NCIT

| Conflict Set | Source Class Name | Target Class Name | |
|---|---|---|---|
| 1 | sarcoma | sarcoma | View in graph |
| | neurofibrosarcoma | malignant mouse peripheral nerve sheath tumor | |
| 2 | orientation<br>position | orientation<br>position | View in graph |
| | angle | angle | |
| 3 | receptor activator of nf kappa b ligand | tumor necrosis factor ligand superfamily member 11 | View in graph |
| | rankl | tumor necrosis factor ligand superfamily member 11 | |
| 4 | hyperplastic<br>size | hyperplasia<br>size | View in graph |
| 5 | recql4 | atp dependent dna helicase q1 | View in graph |
| | recq protein like 4 | atp dependent dna helicase q1 | |
| 6 | pinealoma | mouse pineocytoma | View in graph |
| | neoplasm of the nervous system<br>neoplasm of the central nervous system | nervous system neoplasm<br>central nervous system neoplasm | |
| | malignant neoplasm of the central nervous system | malignant central nervous system neoplasm | |
| | uniparental disomy | uniparental disomy | |

Figure 5.3: VizRepair graph visualization example of the list of mappings.

Table 5.1: Total and conflicting mappings in the ontologies used. Bone Dysplasia Ontology (BDO), Cell Culture Ontology (CCONT), Experimental Factor Ontology (EFO), Cardiac Electrophysiology Ontology (EP), Foundational Model of Anatomy (FMA), Mouse Adult Gross Anatomy Ontology (MA), National Cancer Institute Thesaurus (NCIT), Online Mendelian Inheritance in Man (OMIM), Uber Anatomy Ontology (UBERON).

| Ontology1 | Ontology2 | Total Mappings | Conflicting Mappings |
|---|---|---|---|
| BDO | NCIT | 1636 | 1374 |
| CCONT | NCIT | 2097 | 1136 |
| EFO | NCIT | 2507 | 1541 |
| EP | FMA | 78489 | 109 |
| EP | NCIT | 2465 | 307 |
| MA | FMA | 961 | 22 |
| OMIM | NCIT | 5178 | 1078 |
| SDO | EP | 80 | 3 |
| UBERON | FMA | 1932 | 121 |
| ZFA | EFO | 427 | 11 |

From the list view the user can choose a conflict set to be visualized. The webtool shows a graph-based representation (supported by linkurious.js) for the selected set of conflicting mappings. The graph layout implements a ForceLink layout derived from ForceAtlas2, a force-directed layout algorithm (Jacomy *et al.*, 2014). ForceLink can run on background and stops automatically the optimal node position is found. The algorithm works by repulsing forces between the nodes and attraction forces on the edges, creating a network that allows the node dispersion. This fact contributed to the choice of this layout as the aim was to visualize large numbers of nodes without overlap.

To allow users to understand the conflict we need to show the relations between the classes involved in mappings and the classes involved in the disjoint axiom(s), however in many cases this would result in showing the several classes that compose the path from the mapping to the disjoint axiom. To reduce this visual clutter, we compute the transitive closure between these classes and display them as directly linked. Usually, the conflict is only understood by also visualizing descendants. This happens when classes have more than one parent, and the conflict is caused by only one of the paths. To support this, we need to find common descendants in the ontology opposite to the one that bears the disjoint axiom, or both, if both contain involved disjoints (Figure 5.4).

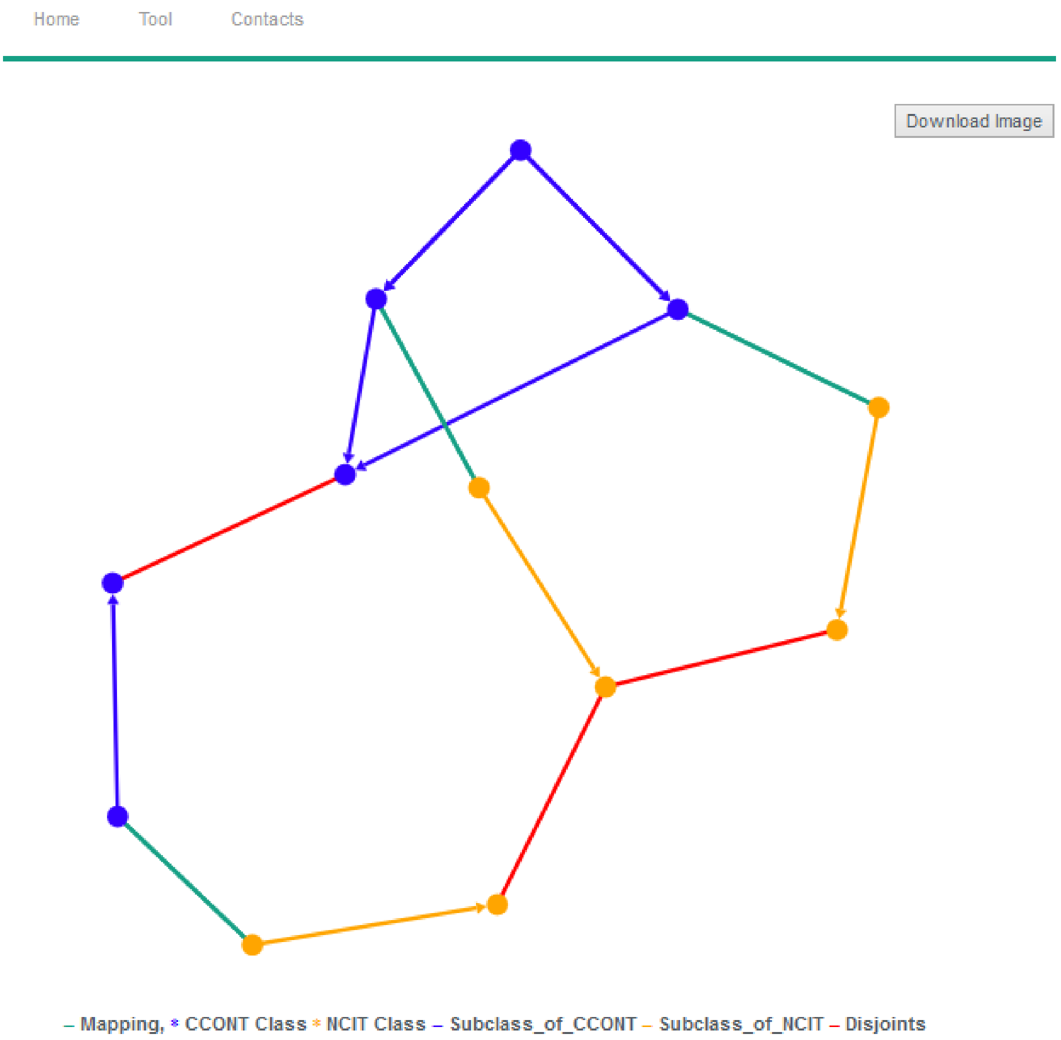


Figure 5.4: Example of VizRepair visualization with common ancestors.

In certain conflict sets, there is a very large number of mappings, which hinders the visualization and understanding. To address this an optimized graph visualization was developed taking into account the following:

- all disjoint axioms involved in the conflict are shown

- each class involved in the disjoint axiom should have at least one mapped class associated.

Then the algorithm iteratively adds more mappings related to each disjoint axiom class until a threshold number of mappings is reached. This threshold can be set as an input parameter.

Figures 5.5 and 5.6 illustrate the improved visualization by filtering the graph, showing that the source of the conflict can still be found (conflict set 14 of the pair of ontologies BDOxNCIT) even if the graph is much simplified.

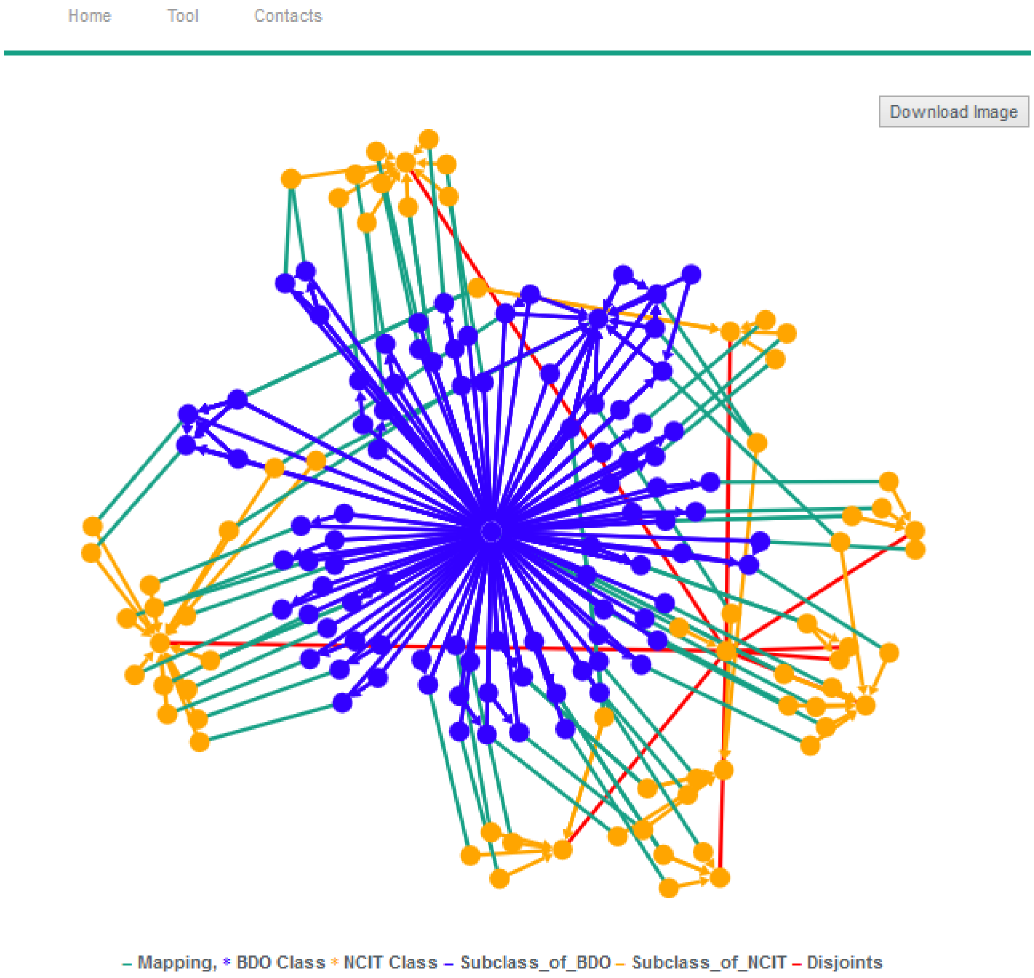


Figure 5.5: VizRepair graph visualization example of an unfiltered graph.

Figure 5.6: VizRepair graph visualization example of a filtered graph.

## 5.4 Evaluation and Discussion

Like it was described in the previous chapter, with our web tool we also built an evaluation task in order to access if the users were able to comprehend the problem in question and to explore future updates to the tool aiming to optimize its capabilities. In addiction, the SUS questionnaire was also applied. The overall question scores can be seen in Appendix A. The users from the evaluation conducted in the previous chapter were the same in this task. A set of 7 individuals comprised graduate students in the area of bioinformatics with knowledge of the basic ontology concepts.

In this case, we asked 7 individuals if they could understand three different conflict sets from the pair of ontologies BDO and NCIT. The sets were chosen taking into account their size, 10 is the smallest and 40 the biggest, in order to see how the users handled with different conflict sizes. Besides that, the user was also asked if the task was easier if less or more mappings were present in the graph showing 4, 6 or 10 mappings instead of the 8 currently implemented. All the tasks implemented are comprised in Table 5.2 and a graph representation of all of them, except the first, can be seen in Appendix B.

Table 5.2: Evaluation Task of VizRepair.

| Number | Task |
|---|---|
| 1 | Understand the conflict set 10 (5 mappings) (see Figure 5.7) |
| 2 | Understand the conflict set 31 (6 mappings) |
| 3 | Understand the conflict set 40 (16 mappings,threshold 8) |
| 4 | Understand the conflict set 40 (16 mappings,threshold 4) |
| 5 | Understand the conflict set 40 (16 mappings,threshold 6) |
| 6 | Understand the conflict set 40 (16 mappings,threshold 10) |

The tasks were performed without problems to most of the users but some required a first explanation of the mappings conflict concept. The list of user suggestions is comprised in Table 5.3. Almost all the participants suggested that they should be allowed to chose the number of mappings and not limit their visualization to a restricted number. The current visualization algorithm also does not always results in the best disposition of the nodes and its relations

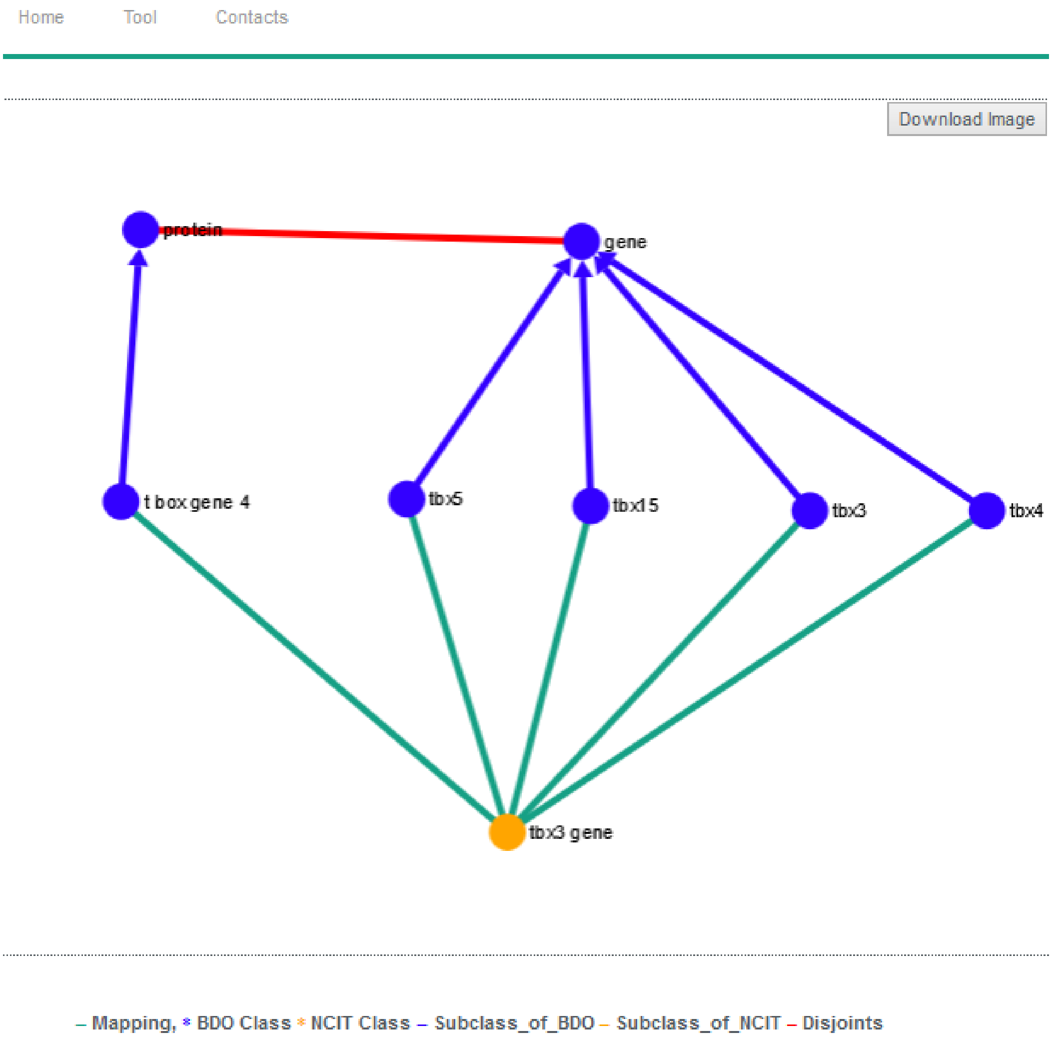


Figure 5.7: Graph visualization of conflict set 10.

making it sometimes difficult for the user to visualize the conflict. A click and drag option in the graph nodes currently implemented helped overcome this problem but is not the best solution. One other user also suggested that the possibility to see the element information in parallel to the graph is a feature that should be added in the future.

Table 5.3: List of user suggestion.

| Suggestion | Frequency |
|---|---|
| Possibility to choose the number of mappings | 6 |
| Give more emphasis to the arrows | 1 |
| Visualization algorithm overlaps information | 5 |
| Bigger caption | 5 |
| The label names should appear independently of the zoom | 1 |
| Element Information in parallel to the graph | 1 |

The questionnaire resulted in a overall score of 75.4 percentile, above the average that usually falls bellow 68. Table 5.4 shows that in certain questions the users occasionally were not in agreement with each other. That is represented by values of standard deviation above one which means that the responses to these questions differed amongst the users. The variability occurred in the questions related to the future use, overall usability and technical support and knowledge of the subject. The result implies that the users may have different backgrounds related to the ontology concepts here in study (e.g. disjoint classes, incoherences, etc.) that may have interfered with the overall performance of the tasks. In contrast, the users seem to find the tool easy to use and quick to learn, considering the system well integrated.

The current tool allows for the development of the field by the visualization of a complex concept that is mapping incoherences in a more interactive and simple way. The visualization of those incoherences allow the user to better understand the problem at hand and to analyze how is the best way to proceed. The results of the evaluation support the usefulness of the tool and its performance. In the future, we aim to overcome several of the limitations found in the evaluation.

Table 5.4: Questionnaire statistical evaluation.

| Question | Average | Median | St. Dev. |
|---|---|---|---|
| I think that I would like to use this system frequently. | 3.71 | 4 | 1.11 |
| I found the system unnecessarily complex. | 1.57 | 1 | 0.98 |
| I thought the system was easy to use. | 3.86 | 4 | 1.21 |
| I think that I would need the support of a technical person to be able to use this system. | 2.00 | 2 | 1.15 |
| I found the various functions in this system were well integrated. | 4.43 | 5 | 0.79 |
| I thought there was too much inconsistency in this system. | 1.57 | 1 | 0.79 |
| I would imagine that most people would learn to use this system very quickly. | 4.14 | 4 | 0.90 |
| I found the system very cumbersome to use. | 1.71 | 1 | 0.95 |
| I felt very confident using the system. | 3.57 | 3 | 0.79 |
| I needed to learn a lot of things before I could get going with this system. | 2.71 | 2 | 1.25 |

# Chapter 6

# Conclusion

The number of biomedical ontologies has been increasing in the last decade, but their disconnected creation demands support for the integration of all the overlapping and related information. Ontology matching techniques can be used to ensure interoperability and to establish meaningful correspondences between different ontologies. To understand the alignments created by these techniques some systems are supported by a visualization module. However, biomedical ontologies composed of a large number of elements and intricate relationships between those elements impose several restrictions to the visualization capabilities and algorithm development. Therefore, the investigation of the main paradigms of the ontology matching visualization becomes crucial to the expanding area of biomedical ontologies.

This thesis investigated whether a visualization strategy based on subgraphs can be used in support of two important matching related user tasks: alignment review and logical incoherence explanation.

Although user tests were small-scale, the subgraph visualization proved to be more successful than the typical tree-based visualization.

To address the alignment review challenge we took advantage of AML, one of the top ontology matching system, and improved on its graphical user interface by expanding its capabilities to now support a manual review of the alignments and to give more information about the elements involved in each mapping. The information given can be lexical (e.g. labels, synonyms, etc.) or structural (e.g.

high level ancestor, disjoint classes , etc.) and combined with the graph visualization can be used to make a better assessment of whether or not the mapping should be removed or maintained. The user feedback demonstrated that the tool is overall intuitive, easy to learn and simple.

In order to facilitate the understanding of mapping incoherences found in certain alignments, this thesis also developed an online tool with the aim of showing those conflicts. The tool, called VizRepair, comprises 10 pairs of ontologies from Bioportal repository and their conflicts computed by AML repair algorithm. The possibility of visually displaying a complex problem like the mapping incoherences seems to be facilitated by its visualization. The evaluation of the tool confirms that by the positive user responses.

Nevertheless, both implementations have their limitations. AML is not always very intuitive and some users had difficulties finding the new features including the remove mappings button or the option to select the neighbourhood radius. The library used to show the graph is also limiting not allowing the user to interact with the graph. In the future, it is necessary to substitute the current library with one that allows the user to drag nodes, display the element information by clicking on the nodes and even allowing him to manually add and remove a mapping by clicking directly on the graph. In addition, it would also be important to show the conflicts directly in the graph or in a parallel one by using the repair algorithm already integrated in the system.

VizRepair had also its drawbacks but it received less complaint than the previous implementation. It could benefit if the element information was also displayed in parallel with the graph and its layout algorithm also limits the visualization not giving the best disposition for the nodes visualized. The tool could improve with the implementation of a different layout algorithm that would be able to draw two independent graphs with connections between them. Future work could look into extending the semi-bipartite graph algorithm layout purposed by Xu *et al.* (2010) for displaying the Gene Ontology and its annotated proteins. This type of graph is a extension of the bipartite graph, where one of the subgraphs has edges between its elements. An extension for alignment visualization would have to consider edges within both subgraphs. Many of the bipartite graph algorithms, including the purposed by Xu *et al.* (2010), are based on the Sugiyama method.

In this approach, composed of a bipartite graph (graph whose nodes can be divided into two disjoint sets connected by edges) with two parallel straight lines, consists of positioning the nodes of a set in one of the lines and the nodes of the other set in the other line. Then, the nodes of both sets are connected by an edge (Sugiyama *et al.*, 1981). In the future, besides the algorithm, we would also like to have a direct link to Bioportal and an automatic repair option should be available for all the alignments visualized.

The current thesis proposes an advancement in the field by adding new features to a top ontology matching system and by evaluating visualization approaches in the biomedical ontology alignment context and proposing that a subgraph visualization supported by additional information is a better solution than the typical tree-based visualization.

# Appendix A

# Evaluation Results

Table A.1: Results of the System Usability Scale (SUS) questionnaire for AgreementMakerLight(AML) and AgreementMaker(AM). The users with a * performed the evaluation first on AML.

| Question | AML | | | | | | | AM | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | User 1* | User 2* | User 3 | User 4 | User 5* | User 6* | User 7 | User 1* | User 2* | User 3 | User 4 | User 5* | User 6* | User 7 |
| I think that I would like to use this system frequently. | 5 | 4 | 4 | 4 | 2 | 5 | 3 | 1 | 2 | 3 | 2 | 1 | 2 | 3 |
| I found the system unnecessarily complex. | 3 | 2 | 1 | 2 | 1 | 2 | 3 | 3 | 4 | 2 | 4 | 3 | 3 | 4 |
| I thought the system was easy to use. | 3 | 3 | 3 | 5 | 4 | 4 | 2 | 2 | 2 | 4 | 3 | 2 | 2 | 2 |
| I think that I would need the support of a technical person to be able to use this system. | 4 | 2 | 1 | 1 | 2 | 1 | 4 | 4 | 5 | 2 | 2 | 5 | 2 | 4 |
| I found the various functions in this system were well integrated. | 5 | 4 | 5 | 4 | 5 | 4 | 4 | 2 | 3 | 2 | 2 | 2 | 2 | 3 |
| I thought there was too much inconsistency in this system. | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 3 | 2 | 4 | 4 | 4 | 3 | 3 |
| I would imagine that most people would learn to use this system very quickly. | 4 | 2 | 5 | 5 | 5 | 5 | 3 | 1 | 3 | 3 | 4 | 2 | 2 | 4 |
| I found the system very cumbersome to use. | 1 | 2 | 1 | 1 | 1 | 2 | 4 | 4 | 4 | 1 | 5 | 4 | 4 | 3 |
| I felt very confident using the system. | 3 | 3 | 4 | 4 | 4 | 5 | 3 | 2 | 2 | 5 | 3 | 1 | 3 | 3 |
| I needed to learn a lot of things before I could get going with this system. | 2 | 3 | 1 | 3 | 2 | 1 | 3 | 3 | 3 | 1 | 3 | 5 | 2 | 4 |

Table A.2: Results of the System Usability Scale (SUS) questionnaire for VizRepair.

| Question | VizRepair | | | | | | |
|---|---|---|---|---|---|---|---|
| | User 1 | User 2 | User 3 | User 4 | User 5 | User 6 | User 7 |
| I think that I would like to use this system frequently. | 5 | 5 | 4 | 3 | 2 | 4 | 3 |
| I found the system unnecessarily complex. | 3 | 1 | 1 | 1 | 1 | 1 | 3 |
| I thought the system was easy to use. | 2 | 3 | 5 | 4 | 5 | 5 | 3 |
| I think that I would need the support of a technical person to be able to use this system. | 3 | 2 | 1 | 2 | 1 | 1 | 4 |
| I found the various functions in this system were well integrated. | 5 | 5 | 5 | 4 | 5 | 3 | 4 |
| I thought there was too much inconsistency in this system. | 1 | 1 | 1 | 3 | 1 | 2 | 2 |
| I would imagine that most people would learn to use this system very quickly. | 4 | 3 | 4 | 5 | 5 | 5 | 3 |
| I found the system very cumbersome to use. | 1 | 1 | 1 | 3 | 3 | 1 | 2 |
| I felt very confident using the system. | 3 | 4 | 5 | 3 | 3 | 4 | 3 |
| I needed to learn a lot of things before I could get going with this system. | 2 | 4 | 2 | 4 | 2 | 1 | 4 |

# Appendix B

# User test tasks of VizRepair

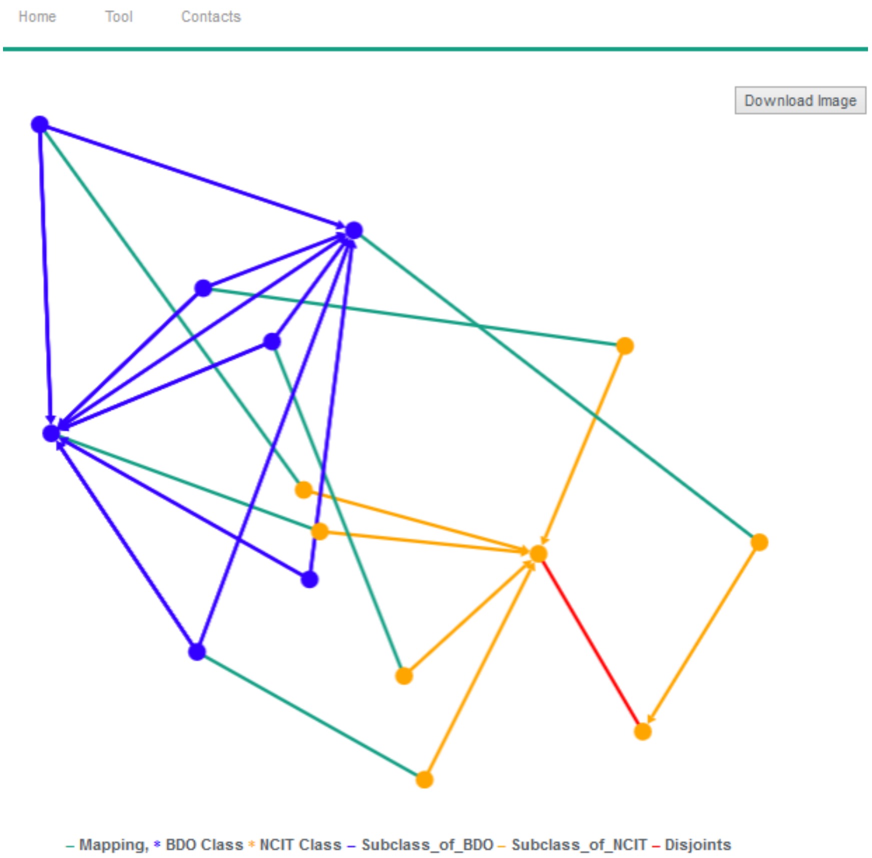


Figure B.1: Graph visualization of conflict set 31.

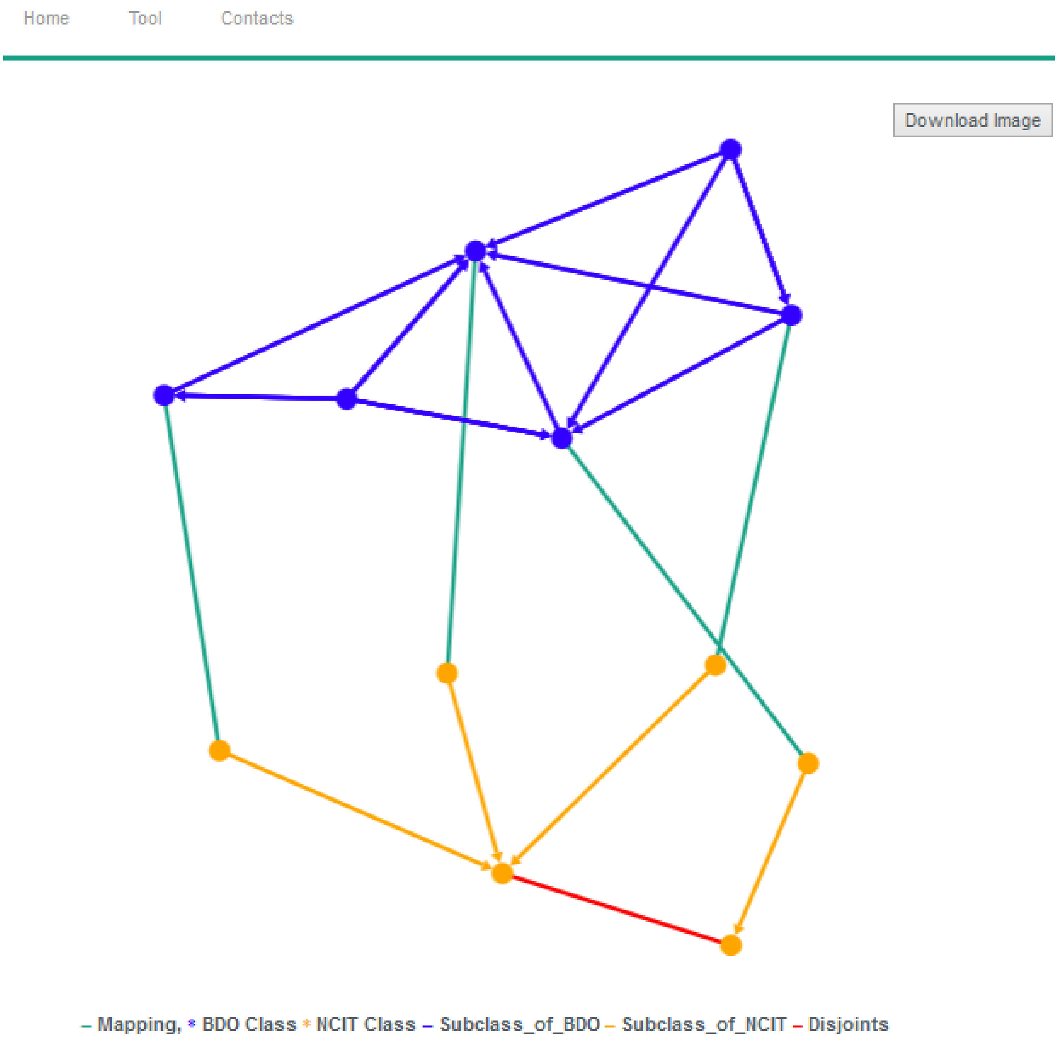


Figure B.2: Graph visualization of conflict set 40 (16 mappings,threshold 4).

Figure B.3: Graph visualization of conflict set 40 (16 mappings,threshold 6).

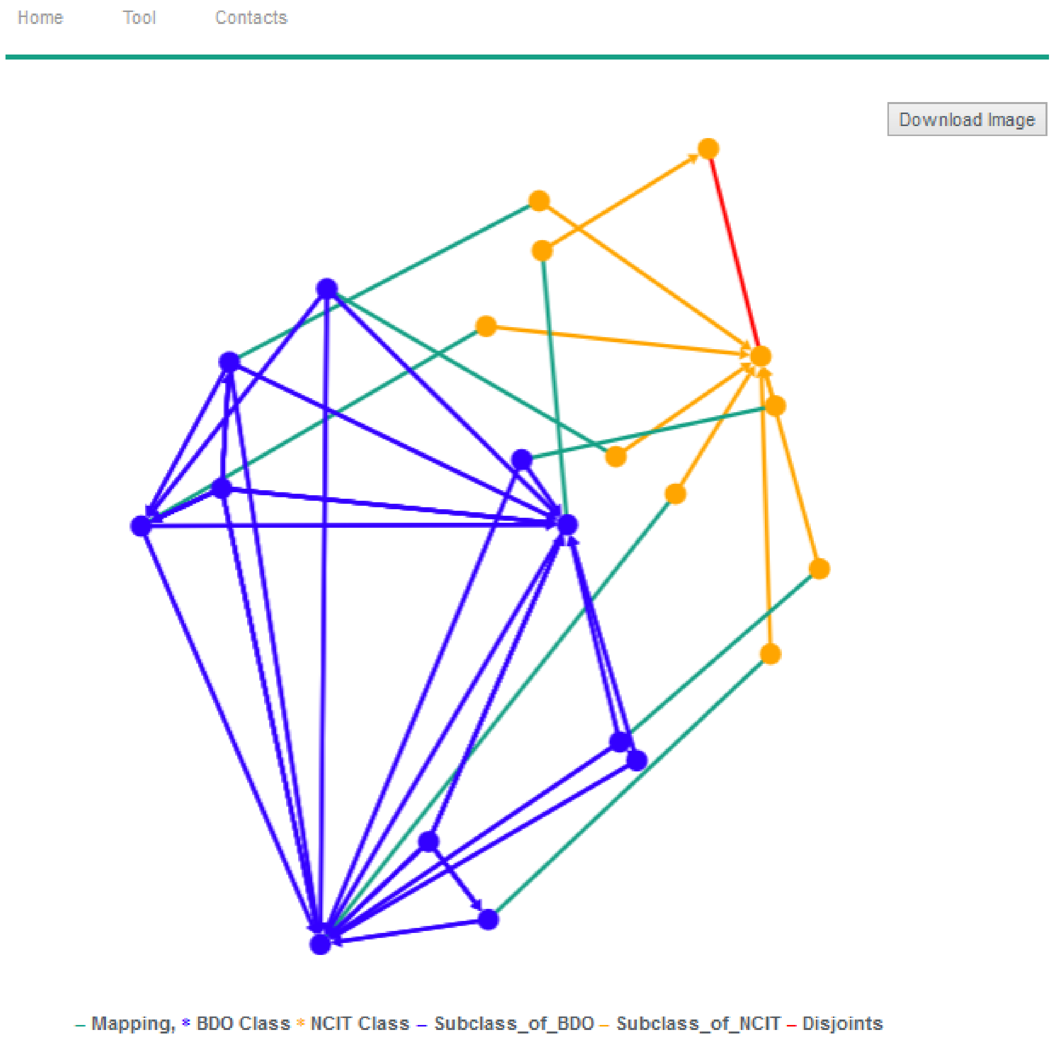


Figure B.4: Graph visualization of conflict set 40 (16 mappings,threshold 8).

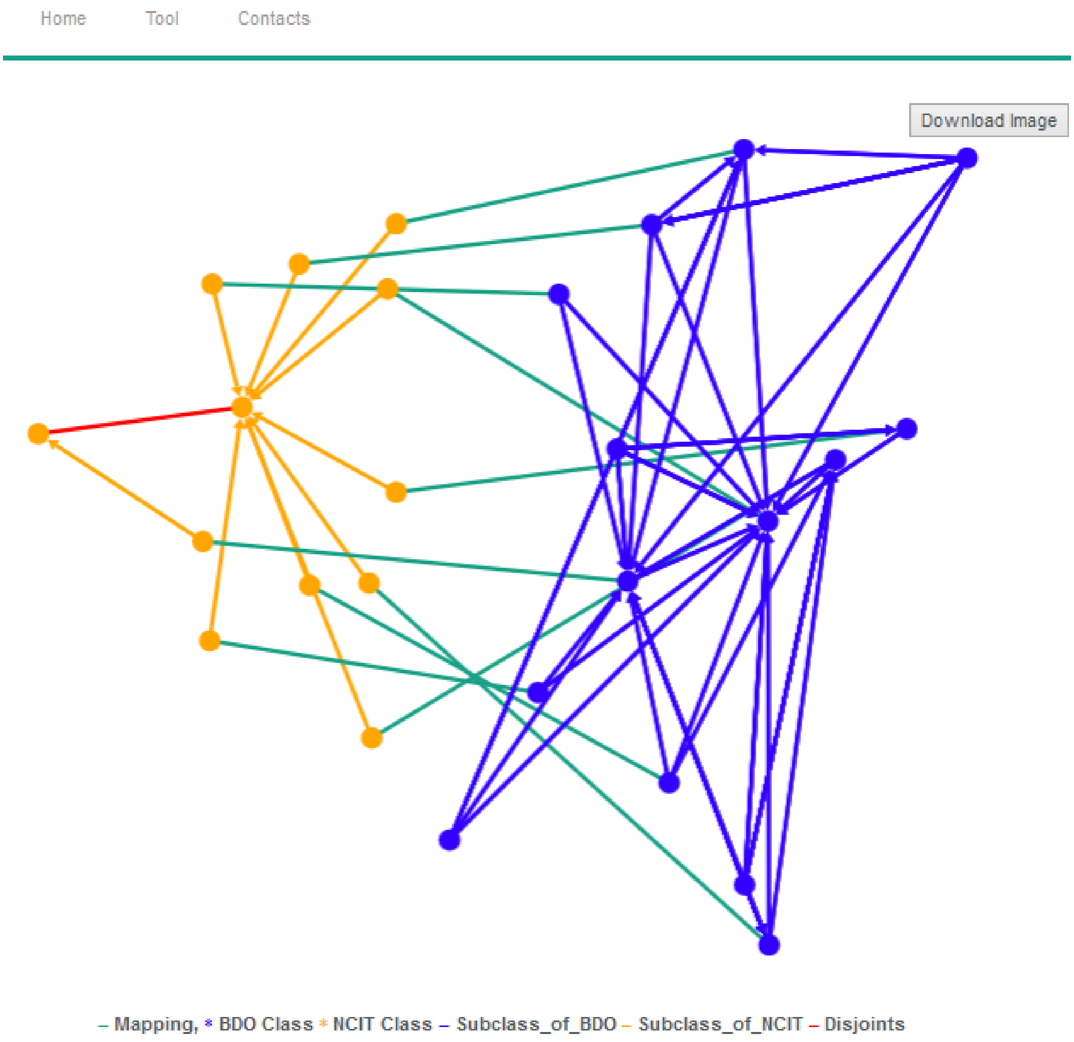


Figure B.5: Graph visualization of conflict set 40 (16 mappings,threshold 10).

# References

ANDREWS, K., WOLTE, J. & PICHLER, M. (1997). Information pyramidstm: A new approach to visualizing large hierarchies. In *Proceedings of the IEEE Visualization'97*, 49–52. 16

ASHBURNER, M., BALL, C.A., BLAKE, J.A., BOTSTEIN, D., BUTLER, H., CHERRY, J.M., DAVIS, A.P., DOLINSKI, K., DWIGHT, S.S., EPPIG, J.T. *et al.* (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, **25**, 25–29. 1

BASTIAN, M., HEYMANN, S., JACOMY, M. *et al.* (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, **8**, 361–362. 36

BODENREIDER, O., HAYAMIZU, T.F., RINGWALD, M., DE CORONADO, S. & ZHANG, S. (2005). Of mice and men: Aligning mouse and human anatomies. In *AMIA Annual Symposium Proceedings*, vol. 2005, 61, American Medical Informatics Association. 41

BROOKE, J. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, **189**, 4–7. 41, 44

CRUZ, I.F., ANTONELLI, F.P. & STROE, C. (2009). Agreementmaker: efficient matching for large real-world schemas and ontologies. *Proceedings of the VLDB Endowment*, **2**, 1586–1589. xv, 21, 22, 31

DE SOUZA, K.X., DOS SANTOS, A.D. & EVANGELISTA, S.R. (2003). Visualization of ontologies through hypertrees. In *Proceedings of the Latin American conference on Human-computer interaction*, 251–255, ACM. 16

DO, H.H. & RAHM, E. (2007). Matching large schemas: Approaches and evaluation. *Information Systems*, **32**, 857–885. 13

DUYHOA, N. & BELLAHSENE, Z. (2013). Yam++ results for oaei 2013. In *OM'2013: 8th International Workshop on Ontology Matching collocated with ISWC'2013: 12th International Semantic Web Conference*, 8–p. 27

ELMELEEGY, H., OUZZANI, M. & ELMAGARMID, A. (2008). Usage-based schema matching. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, 20–29, IEEE. 13

EUZENAT, J., SHVAIKO, P. *et al.* (2007). *Ontology matching*, vol. 18. Springer. 7, 8, 9

EYL, M. (1995). *The harmony information landscape: interactive, three-dimensional navigation through an information space*. na. 16

FALCONER, S.M. & STOREY, M.A. (2007). *A cognitive support framework for ontology mapping*. Springer. 13, 31, 47

FARIA, D., JIMÉNEZ-RUIZ, E., PESQUITA, C., SANTOS, E. & COUTO, F.M. (2014a). Towards annotating potential incoherences in bioportal mappings. In *The Semantic Web–ISWC 2014*, 17–32, Springer. 13, 50, 51, 53

FARIA, D., PESQUITA, C., SANTOS, E., CRUZ, I.F. & COUTO, F.M. (2014b). Agreementmakerlight: a scalable automated ontology matching system. *DILS 2014*, 29. 33, 50

FARIA, D., MARTINS, C., NANAVATY, A., OLIVEIRA, D., SOWKARTHIGA, B., TAHERI, A., PESQUITA, C., COUTO, F.M. & CRUZ, I.F. (2015). Aml results for oaei 2015. In *ISWC International Workshop on Ontology Matching (OM), CEUR Workshop Proceedings*. 34

FU, B., NOY, N.F. & STOREY, M.A. (2013). Indented tree or graph? a usability study of ontology visualization techniques in the context of class mapping evaluation. In *The Semantic Web–ISWC 2013*, 117–134, Springer. 14

Granitzer, M., Sabol, V., Onn, K.W., Lukose, D. & Tochtermann, K. (2010). Ontology alignment—a survey with focus on visually supported semi-automatic techniques. *Future Internet*, **2**, 238–258. xv, 15, 30, 31

Grau, B.C., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P. & Sattler, U. (2008). Owl 2: The next step for owl. *Web Semantics: Science, Services and Agents on the World Wide Web*, **6**, 309–322. 8

Hoehndorf, R., Schofield, P.N. & Gkoutos, G.V. (2015). The role of ontologies in biological and biomedical research: a functional perspective. *Briefings in bioinformatics*, bbv011. 7

Ivanova, V., Lambrix, P. & Åberg, J. (2015). Requirements for and evaluation of user support for large-scale ontology alignment. In *The Semantic Web. Latest Advances and New Domains*, 3–20, Springer. 19, 47, 52

Jacomy, M., Venturini, T., Heymann, S. & Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *Plos One*. 55

Jiménez-Ruiz, E. & Grau, B.C. (2011). Logmap: Logic-based and scalable ontology matching. In *The Semantic Web–ISWC 2011*, 273–288, Springer. 51

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C. & Giannopoulou, E. (2007). Ontology visualization methods—a survey. *ACM Computing Surveys (CSUR)*, **39**, 10. 14

Kleiberg, E., Van De Wetering, H. & Van Wijk, J.J. (2001). Botanical visualization of huge hierarchies. In *infovis*, 87, IEEE. 16

Resource description framework (rdf): Concepts and abstract syntax. `https://www.w3.org/TR/2004/REC-rdf-concepts-20040210/`. 8

Lambrix, P. & Ivanova, V. (2013). A unified approach for debugging is-a structure and mappings in networked taxonomies. *J. Biomedical Semantics*, **4**, 10. 27

LAMBRIX, P. & TAN, H. (2006). Sambo—a system for aligning and merging biomedical ontologies. *Web Semantics: Science, Services and Agents on the World Wide Web*, **4**, 196–206. 47

LANZENBERGER, M. & SAMPSON, J. (2006). Alviz-a tool for visual ontology alignment. In *Information Visualization, 2006. IV 2006. Tenth International Conference on*, 430–440, IEEE. xv, 21, 24, 31

MASSMANN, S., RAUNICH, S., AUMÜLLER, D., ARNOLD, P. & RAHM, E. (2011). Evolution of the coma match system. *Ontology Matching*, **49**. 23, 31, 47

MAZUEL, L. & CHARLET, J. (2009). Spim-alignmentgui-un logiciel d'aide à la réalisation d'alignements entre ontologies. In *IC*. 26

MEILICKE, C. (2011). *Alignment incoherence in ontology matching*. Universität Mannheim. 11

MEILICKE, C. & STUCKENSCHMIDT, H. (2008). Incoherence as a basis for measuring the quality of ontology mappings. In *Proc. 3rd International Workshop on Ontology Matching (OM) collocated with ISWC*, 1–12. 49

MILES, A. & BECHHOFER, S. (2009). Skos simple knowledge organization system reference. *W3C recommendation*, **18**, W3C. 8

MOCAN, A., CIMPIAN, E. & KERRIGAN, M. (2006). Formal model for ontology mapping creation. In *The Semantic Web-ISWC 2006*, 459–472, Springer. 13

MUNGALL, C.J., RUTTENBERG, A., HORROCKS, I. & OSUMI-SUTHERLAND, D. (2011). Obo flat file format 1.4 syntax and semantics [draft]. Tech. rep., Technical report, Lawrence Berkeley National Laboratory. Available: https://oboformat.googlecode.com/svn/trunk/doc/obo-syntax.htmll. Accessed 2015 Sep 21. 8

MUNZNER, T. (1998). Exploring large graphs in 3d hyperbolic space. *Computer Graphics and Applications, IEEE*, **18**, 18–23. 16

NANDI, A. & BERNSTEIN, P.A. (2009). Hamster: using search clicklogs for schema and taxonomy matching. *Proceedings of the VLDB Endowment*, **2**, 181–192. 13

NOY, N.F. & MUSEN, M.A. (2003). The prompt suite: interactive tools for ontology merging and mapping. *International Journal of Human-Computer Studies*, **59**, 983–1024. 13, 31

NOY, N.F., FERGERSON, R.W. & MUSEN, M.A. (2000). The knowledge model of protege-2000: Combining interoperability and flexibility. In *Knowledge Engineering and Knowledge Management Methods, Models, and Tools*, 17–32, Springer. 16

PARSIA, B., WANG, T. & GOLBECK, J. (2005). Visualizing web ontologies with cropcircles. In *Proceedings of the 4th International Semantic Web Conference*, 6–10. 16

PESQUITA, C., FARIA, D., SANTOS, E. & COUTO, F.M. (2013). To repair or not to repair: reconciling correctness and coherence in ontology reference alignments. In *OM*, 13–24. xvi, 50, 51

PESQUITA, C., FARIA, D., SANTOS, E., NEEFS, J.M. & COUTO, F.M. (2014). Towards visualizing the alignment of large biomedical ontologies. In *Data Integration in the Life Sciences*, 104–111, Springer. 52

RAFFIO, A., BRAGA, D., CERI, S., PAPOTTI, P., HERNÁNDEZ, M. *et al.* (2008). Clip: a visual language for explicit schema mappings. In *Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on*, 30–39, IEEE. 13

SANTOS, E., FARIA, D., PESQUITA, C. & COUTO, F. (2013). Ontology alignment repair through modularization and confidence-based heuristics. *arXiv preprint arXiv:1307.5322*. 12, 52

SHNEIDERMAN, B. (1992). Tree visualization with tree-maps: 2-d space-filling approach. *ACM Transactions on graphics (TOG)*, **11**, 92–99. 16, 31

Shvaiko, P. & Euzenat, J. (2013). Ontology matching: state of the art and future challenges. *Knowledge and Data Engineering, IEEE Transactions on*, **25**, 158–176. 2, 13

Ontoviz tab: Visualizing protégé ontologies. `http://protegewiki.stanford.edu/index.php/OntoViz`. 16

Method and apparatus for displaying data within a three-dimensional information landscape. US Patent 5,528,735. 16

Sugiyama, K., Tagawa, S. & Toda, M. (1981). Methods for visual understanding of hierarchical system structures. *Systems, Man and Cybernetics, IEEE Transactions on*, **11**, 109–125. 65

Sure, Y., Angele, J. & Staab, S. (2002). Ontoedit: Guiding ontology development by methodology and inferencing. In *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, 1205–1222, Springer. 16

Thayasivam, U. & Doshi, P. (2011). Optima results for oaei 2011. In *Proc. of 6th OM Workshop*, 204–211. 26, 31

Thomas, H., O'Sullivan, D. & Brennan, R. (2009). Evaluation of ontology mapping representation. In *Proceedings of the Workshop on Matching and Meaning*, 64–68. 16

Van Wijk, J.J. & Van de Wetering, H. (1999). Cushion treemaps: Visualization of hierarchical information. In *Information Visualization, 1999.(Info Vis' 99) Proceedings. 1999 IEEE Symposium on*, 73–78, IEEE. 16

Von Hippel, E.A. (2005). *Democratizing innovation*. MIT Press, Cambridge, MA, April. 13

Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T. & Musen, M.A. (2011). Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic acids research*, **39**, W541–W545. 1, 50

Xu, K., Williams, R., Hong, S.H., Liu, Q. & Zhang, J. (2010). Semi-bipartite graph visualization for gene ontology networks. In *Graph Drawing*, 244–255, Springer. 64