Ontologies in Neuroscience and Their Application in Processing Questions

Aref EshghiShargh
ORCID: 0000-0002-9443-2212

Doctor of Philosophy
November 2019

School of Computing and Information Systems
University of Melbourne

Submitted in total fulfilment for
the degree of Doctor of Philosophy

Abstract

Neuroscience is a vast, multi-dimensional and complex field of study based on both its medical importance and unresolved issues regarding how brain and the nervous system work. This is because of the huge amount of brain disorders and their burden on people and society. Furthermore, scientist have been excited about the function and structure of brain, ever since it was discovered to be responsible for all our emotions, thoughts and behaviour.

Ontologies are concepts whose origins go back to philosophy and the concern with the nature and relation of being. They have emerged as promising tools for assistance with neuroscience research recently and provide additional data on a field of study. They connect each entity or element to other ones through descriptive relationships. Ontologies seem to suit the complex, multi-dimensional and still incomplete nature of neuroscience very well because of their characteristics.

The first study shines light on applications of ontologies in neuroscience. It incorporated a systematic literature review and methodically reviewed over 1000 research papers from eight databases and three journals. After scanning all documents, 208 of them were selected. Then, a full text analysis was performed on the selected documents. This study found eight major applications for ontologies in neuroscience, most of them consisted of several subcategories. The analysis not only demonstrated the current applications of ontologies in neuroscience, but also their potential future in this field.

The second study was set to represent neuroscience questions and then, classify them using ontologies. For this purpose, a questions set was gathered from two research teams and analysed. This, results in a set of dimensions which represents questions. Then, a question hierarchy was formed based on dimensions and questions were classified according to that hierarchy. Two different approaches were used for the classification including an ontology-based approach and a statistical approach. The ontology-based approach exceeded the statistical approach by 15.73% better classification results.

The last study was designed to tackle and resolve questions with the assistance of ontologies. It first proposed a set of templates that acted as a translation mechanism for changing questions into machine readable code. Templates were based on the question hierarchy presented in the previous study. Second, this study created an integrated collection of resources including two domain ontologies (NIFSTD and NeuroFMA) and a neuroimaging annotation application (Freesurfer). Subsequently, the code created using templates was executed upon the integrated resource (knowledge base) to find the appropriate answer. While processing the questions, ontologies were used for disambiguation purposes too. At the end, all parts created in this study along with the question classification method created in the previous study were merged as different modules of a question processing model.

In conclusion, this thesis reviewed all current ontology applications in neuroscience in detail and demonstrated the extent to which they can assist scientists in classifying and resolving questions. The results of this thesis show that applications of ontologies in neuroscience are diverse and cover a wide range; they are steadily becoming more used in this field; and they can be powerful semantic tools in performing different tasks in neuroscience.

Declaration

This thesis was created based on my original work towards the Doctor of Philosophy degree. Due acknowledgement has been made in the text to all other materials and information used. This thesis follows word limits specified by the University of Melbourne for a PhD thesis.

Preface

- Three different research papers have been created based on the three studies of this thesis and are ready for submission to appropriate journals.
- A research paper that discussed the overall approach of this thesis and included a summary of its findings, was submitted to the 11th Australasian Conference on Health Informatics and Knowledge Management in 2018 (HIKM2018). It won the 'best student paper' award and was published by the Association for Computing Machinery (ACM). Here is the citation for that paper:
  - Eshghishargh, A., Gray, K., Milton, S. K., & Kolbe, S. C. (2018, January). A semantic system for answering questions in neuroinformatics. In *Proceedings of the Australasian Computer Science Week Multiconference* (p. 31). ACM.
- An abstract and poster was presented at the INCF Neuroinformatics 2015 Congress in Cairns, Australia and published by Frontiers in Neuroinformatics:
  - Eshghishargh A, Milton S, Egan GF, Lonie A, Kolbe S, Killeen NE and Lohrey JM (2015). An Ontology-Based Semantic Question Complexity Model and Its Applications in Neuroinformatics. Front. Neurosci. Conference Abstract: Neuroinformatics 2015. doi: 10.3389/conf.fnins.2015.91.00015
- This thesis was funded by an Australian Research Council (ARC) Linkage Grant – LP110100470.
- This thesis was edited by a third-party professional editor. The editor was not familiar with the field or the subject of this thesis.

Acknowledgements

I wanted to take this opportunity to thank all the people who helped me with this thesis including my supervisors, Associate Professor Kathleen Gray, Associate Professor Simon Milton, Associate Professor Andrew Lonie and Professor Gary Egan.

I would especially like to thank Dr Scott Cameron Kolbe from the Florey Institute of Neuroscience and Mental Health for helping me with advising on the neuroscience part of my research; Dr Nolan B. Nichols from Stanford Research Institute for advising me on the implementation and coding sections of my research; Ms Cecily Gilbert from the Health and Biomedical Informatics Centre at the University of Melbourne and advisors from the University of Melbourne library for advising me on the systematic literature review process; Dr Govinda Poudel from Monash University for providing me with sample neuroscience questions.

Moreover, I would like to thank the chair of my research committee, Professor Tim Baldwin for providing invaluable advice throughout my candidature. Moreover, I want to thank Mr Jason Lohrey and Dr Neil Killeen for being part of my research committee.

Table of Contents

1 Introduction

This chapter introduces and explains the outline of the research performed in this thesis. It first discusses the context of the research, then presents research aims and questions. After that, it presents the research design, which sketches how the research is going to be performed. Last of all, there will be a research outline and conclusion.

1.1 Research Context

Neuroscience is an important area of study for two main reasons. First, there are a huge number of diseases and disorders concerning the nervous system. Understanding these diseases and their causes can assist scientists in curing and preventing them. Second, the nervous system is responsible for our thoughts, behaviour, emotions and even views on the world. Therefore, there is a historical curiosity to figure out how the brain and nervous system work.

There are other reasons worth considering. For example, there has been a trend in computer science in recent years to model how the nervous system works. These attempts have shown to be fruitful in designing better computer systems.

Recently, neuroscience has experienced an increase in the amount of data (Landhuis, 2017). This is due to the rapid developments in neuroscience, such as improvements in imaging techniques and modalities, gene expression and extensive experiments. Having huge amounts of data are certainly a benefit for the research. However, it also brings some challenges and difficulties in managing and benefiting from this huge body of data. These challenges include extracting information and working with the data.

A way forward for neuroscience to derive greater insights and gain knowledge from this data are through better ways of representing, classifying, searching and sharing information from resources such as neuroimages, anatomical atlases, pathological data and other documents from various resources.

There is a growing body of literature that recognizes ontologies as effective tools that assist researchers in achieving these goals. Ontologies specify concepts of a domain and their relationships and provide foundations for maintaining, representing, querying and sharing data in a semantic manner which is understandable both for machines and humans. Therefore, they can be useful in neuroscience in multiple ways, including managing neuroscience data and resolving its questions. However, investigating ontology applications and their practicality in neuroscience is still not studied in a detailed way.

This research takes on this goal in neuroscience to explore and investigate all possible and potential applications of ontologies. Furthermore, it seeks enhancements regarding methods for dealing with neuroscience data through using ontologies. Specifically, this research investigates effects of ontologies on processing and resolving neuroscience questions. This means that it will investigate how neuroscience questions can be categorized and mapped to ontologies that assist in resolving them. Investigating ontology applications in neuroscience, along with carrying out experiments towards categorizing and resolving neuroscience questions, will shine a light on the extent of ontology usefulness in neuroscience. Furthermore, it will pave the way for future advancements in neuroscience studies that want to utilize ontologies. In addition, it will demonstrate ontology weaknesses and advantages and direct researchers towards areas for improving ontologies.

1.1.1 Neuroscience and Its Data

Neuroscience aims to decode how the brain processes information and performs tasks. It is an extensive and important area of study and research, driven by a desire for understanding the brain and its implications for health. The nervous system and the way it operates affects human behaviour, thoughts and emotions. Furthermore, the brain has been a mystery since the discovery of its role around 300 years B.C.

In addition, neurological disorders comprise one of the largest groups of human diseases, with more than 2400 individual disorders (Chan, 2010; Köhler *et al.*, 2012). Therefore, the ultimate goal in neuroscience studies is to fully understand the complex behaviour and function of the brain.

In order to understand the brain properly, both biological and anatomical data are necessary. These data which vary in terms of scale, time and space, should be then analysed. The volume of these data are enormous, as the brain contains 80 billion neurons communicating with each other through about 150 trillion synapses (Akil, Martone and Van Essen, 2011). Since the data are so diverse in different aspects, methods for managing data and asking different questions on different levels of structures should be studied right from the level of a cell to the level of neural networks in different spatial and temporal scales, to understand brain function and behaviour.

From massive quantities of data produced in this field continuously, a significant portion is images (neuroimages) as large as several gigabytes (GBs) (Ozyurt *et al.*, 2010). Scientists generally track the attributes and functions of these neuroimages in order to find anatomical and functional changes in the brain and consequently diagnose diseases (Temal *et al.*, 2008).

Therefore, neuroimages are very useful, especially with the advancements in imaging modalities. Different imaging modalities such as Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET) scan, MagnetoEncephaloGraphy (MEG) and ElectroEncephaloGraphy (EEG) can assist scientists by providing high-quality neuroimages (Rubin, Greenspan and Brinkley, 2014; Liu *et al.*, 2015).

1.1.2 The Potential of Ontologies

In this section, a brief introduction to ontologies, alongside reasons for selecting them to use with neuroscience data, is given. Ontologies will be discussed in full detail in chapter 2, which describes the first study.

Neuroscience has been influenced by computer science to a large extent in recent years. Ontologies (Gruber, 1993, 1995) are effective concepts in informatics which have been introduced to neuroscience research and are fast becoming a key instrument in neuroscience studies. Ontology originates in philosophy, and according to the Webster dictionary (Webster Dictionary, 2006), means "a branch of metaphysics concerned with the nature and relations of being". A refined version of an ontology definition in computer science was presented by Studer, Benjamins and Fensel, (1998) - "a formal, explicit specification of a conceptualization". They remain a key element in computer and medical science.

The first concern regarding the application of ontologies in neuroscience is to investigate the reason for using them while there are other tools such as databases. Concepts and tools such as databases can be utilized to organize and retrieve neuroscience data too.

The first answer to this concern lies in the nature of neuroscience and the ability of ontologies to characterize the nervous system and its parts accurately (Hamilton *et al.*, 2012). Organs and parts which are the subject of neuroscience mostly have different types, classes and characterization, and ontologies are a perfect match for representing these characterizations.

Building accurate and detailed models of the brain and nervous system will be very helpful in understanding the brain, its characteristics and how it works. As Ascoli (2013) states, advances in computer science, computational power and access to a huge amount of data is no longer a limitation on the way of building such models. However, the problem of interpreting the data, especially automated interpretation, still remains. Using ontologies can assist in interpreting information regarding the brain knowledge management and the brain itself.

The other answer is ontology's technical benefits, such as being based on description logics (DL) (Baader, Horrocks and Sattler, 2007). Description logics are a group of logic-based knowledge representation formalisms, and allow certain tasks such as using open world assumption, not using unique name assumption and reasoning. Description logics assumes an open world assumption (OWA) allowing incremental completion of the data. This in turn allows things such as an ontology-based system letting users query the data while data are being curated and added by scientists (Heymans *et al.*, 2008). Another benefit lies in the usage of unique name assumptions (UNA) which means that a unique name would be allocated to a property, variable or attribute. Ontologies permit scientists to avoid using unique name assumptions and therefore, allow different names to point to the same element.

Ontologies being created based on description logics, also paves the way for using reasoning. Reasoning is very beneficial to ontologies and can be used to design, maintain and deploy them. A detailed review of ontology characteristics, including the ones mentioned here, will be discussed in chapter 2.

A further useful characteristic of ontology lies in ontology schema and its role. In other technologies like database management systems, the schema rules have already been satisfied before using the database and at the design time. On the contrary, in ontologies, schema is more important, and is active even at the query execution time (Horrocks, 2013).

Not only are ontologies better options than other alternatives such as databases, but also their characteristics makes them better candidates for some other uses concerning data, such as data integration. Even when the data are stored in different databases, using an ontology for data integration is a superior method (Köhler, Philippi and Lange, 2003). Detailed benefits of ontologies over databases and other technologies such as XML will be discussed in chapter 2.

1.1.3 The Background to Ontology Applications in Neuroscience

The application of ontologies in neuroscience will be discussed in detail in chapter 2. However, here a brief overview is discussed in order to create a link between neuroscience and ontologies as the main tool used during this thesis.

Discussions in the above section demonstrated that it is better to benefit from ontologies when the data are sensitive and important, schema have a significant role and the data are yet to be completed (Horrocks, 2013). This is because the characteristics of ontologies discussed in the previous section make them a better choice in these situations in comparison to other tools and technologies.

A diverse set of ontologies have been implemented and used in neuroscience, including Foundation Model of Anatomy (FMA) (Rosse and Mejino, 2003) and Neuroscience Information Framework Standard (NIFSTD) (Fahim T. Imam *et al.*, 2011). The application of ontologies has contributed to many areas in this field.

Ontology can be used via different approaches and techniques to assist neuroscience research. For example, Gupta, Condit and Qian, (2010) provided a fine categorization based on the role of ontologies, which consists of four major groups. They see the overall application of ontologies as an attempt towards resource integration, and divide it into four groups including F-logic based systems, database to ontology mapping systems, integration systems and query expansion systems.

Initiatives such as the Neuroscience Information Framework (NIF) (Gardner, 2008) have emerged and tried to gather neuroscience resources under one umbrella and allow data intensive discovery. NIF have made significant achievements, such as creating an ontology for overcoming the terminology problem by introducing the NIFSTD and federating an enormous number of databases. As pointed out, later in chapter 2 it will be demonstrated that ontologies can be utilized to overcome many other issues in neuroscience.

1.2 Research Questions of this Thesis

The previous section gave a brief introduction to neuroscience, ontologies and reasons for applying ontologies in neuroscience. The current section describes research motivations and problems.

Generally, and as an overarching research aim, this research sets out to further investigate applications of ontologies in neuroscience and use them for presenting, categorizing and resolving neuroscience questions. It seeks to answer the following question: “What is the extent of the power and importance of ontologies in neuroscience, especially in working with neuroscience questions?”. Therefore, this research concerns the use of ontologies in neuroscience and their capability in processing neuroscience questions, including classifying and answering them.

Since the time ontologies have been introduced to neuroscience research, there has not been a broad investigation on their applications. Such an investigation will show benefits and disadvantages of ontologies, implications of using them, shine light on areas where ontology has been of more use and more successful, and possibly point researchers towards new research directions.

Such a study will be extremely useful, since it can enhance current neuroscience data management, integration, augmentation and retrieval. Furthermore, it will assist not only in designing neuroscience models and systems, but also other similar domains such as biomedical science or even distant fields such as geography and cartography. This will be discussed in the Conclusion chapter.

As explained before, answering and resolving neuroscience questions is vital for achieving neuroscience goals such as clinically diagnosing and managing neurological conditions and resolving fundamental questions regarding brain and the nervous system structure and function.

Ontologies can assist in tasks such as representing and modelling questions, as well as classifying and processing them. All these tasks are foundational elements for resolving questions and achieving the goals mentioned above.

Furthermore, investigating the limits of ontologies in creating question processing models or systems that use them to great extent is useful. This is because the expressive power of ontologies will assist in advancements in neuroscience systems.

Based on the above information, the research presented herein has three different aims as follows:

- To investigate the versatility of applications of ontologies in neuroscience in depth.
- To experiment with classifying, representing and categorizing neuroscience questions with the use of ontologies and to examine the result of implementing such a classification.
- To explore the power of ontologies in resolving neuroscience questions and its implications in this field.

In order to achieve these goals, three research questions are proposed:

- What are the applications of ontologies in neuroscience?
- How can questions in neuroscience be represented and classified using ontologies?
- How can questions in neuroscience be resolved using ontologies?

1.3 Research Design of this Thesis

This research consists of three different studies. From now on, the set of tasks, experiments and achievements in this thesis is referred to as 'research'. On the other hand, each of the three different aims stated in the section above, along with their associated question and investigations, are called a study.

The thesis (research) seeks to answer the overall aim discussed in the previous section, while each of the chapters (studies) are shaped around resolving one of the research questions discussed in the Research Questions section.

Each study in this research has a unique method, described in detail in its respective Methods section. However, the overall methodology underpinning this research is a mixed methodology (Johnstone, 2004; Terrell, 2012).

Mixed methods studies are studies that are products of the pragmatist paradigm and combine qualitative and quantitative approaches within different phases of the research process (Tashakkori and Creswell, 2008).

As for the studies, a systematic literature review is performed on ontologies and their applications in neuroscience research in the first study. In this study, research papers were selected from a range of reputable sources including eight databases and three journals. The focus of the study has been on demonstrating current and potential applications of ontologies in neuroscience.

The first study paves the way for the second and third studies. Best available practices and trends that are pointed out through the first study are used in shaping the next two studies regarding application of ontologies in representing, classifying and resolving questions in neuroscience.

Exploring the second aim leads to question classification and representation in neuroscience. The approach used in the second study is based on analysing a set of questions, which were sourced from two research teams and literature and then validated by an expert. The representation can then be used as a foundation for classifying and categorizing questions and later resolving them in the third study.

For the first part of the second study, which is question representation, sourced questions are first pre-processed. Then, they are summarized and represented as parts and named dimensions. After that, a question hierarchy is created based on them and evaluated.

As the second part of this study, questions are classified using two different methods, including an ontology-based approach and a statistical approach. The ontology-based method works by using ontologies to classify questions. The statistical method, which is performed to provide a benchmark for assessing the ontology-based method, consists of applying four different techniques on questions for classification purposes.

The third and final study is an attempt to resolve questions. As its main contribution, it resolves questions by creating a group of templates based on the question hierarchy from the previous study. Templates change questions into code that can be used to interrogate a group of integrated resources. It then creates an integrated resource module. It does so by integrating domain ontologies such as the neuroscience information framework standard (NIFSTD) ontology (Fahim T. Imam *et al.*, 2011) and the neuroscience section of the foundational model of anatomy (NeuroFMA) ontology (Nichols, Mejino Jr and Brinkley, 2011) with a MRI segmentation application called Freesurfer (Fischl, 2012). Freesurfer contains subject-related information.

At the end, the question classification from the second study is combined with the templates to create a question processing module. Then the question processing module is connected to the resource integration module. These two modules along with the query expansion module create a model for neuroscience questions processing.

The topic of the thesis is essentially interdisciplinary, as is typical of biomedical informatics and neuroinformatics research (Lee *et al.*, 2009; Amunts *et al.*, 2016). Since the author of this thesis has a disciplinary background in computer science, it was deemed necessary to add inside-outside legitimation to the thesis, as is accepted practice in mixed methods research (Onwuegbuzie and Johnson, 2006). Accordingly, the research design was informed at key points by advice from neuroscience experts. Those who fulfilled this role are listed in Acknowledgements page at the beginning of this thesis.

Figure 1-1 shows a summary of the research design. The relation of the chapters along with their major contributions are demonstrated in Figure 1-2 while logical relationships between the studies are depicted in Figure 1-3 later in this chapter.

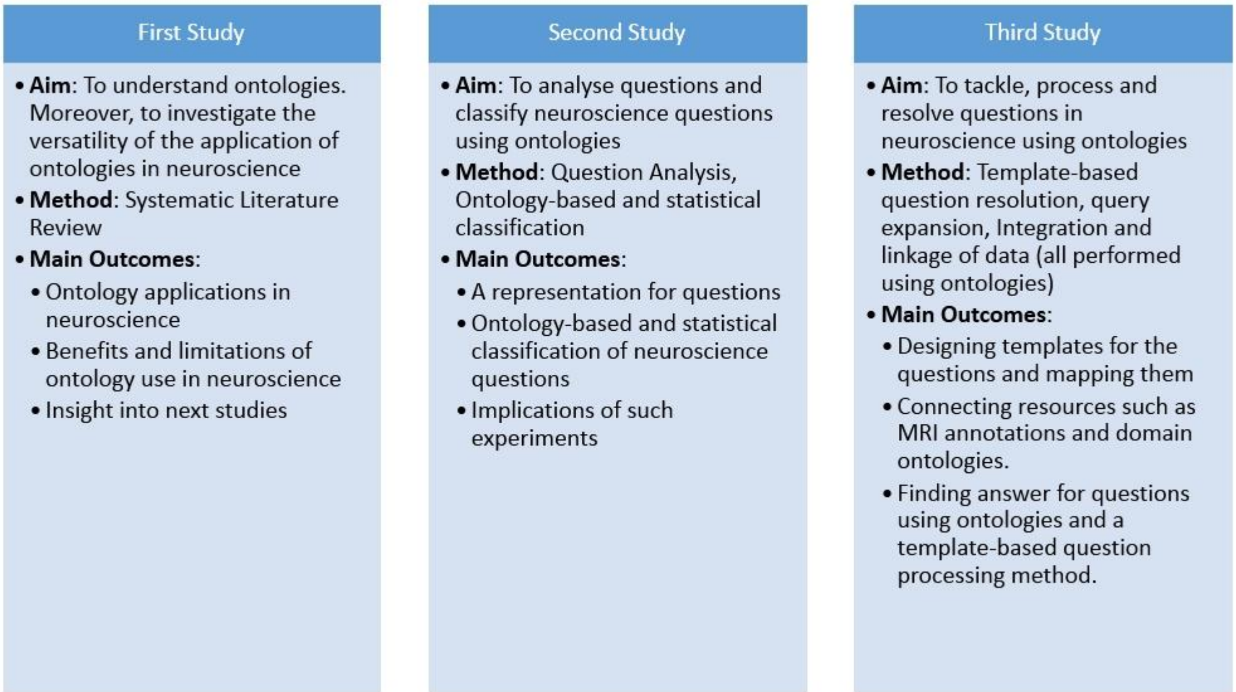

Figure 1-1 A summary of the Research Design

As explained before, studies in this research carry the same theme, but are separate studies. For example, the third study can be viewed as a stand-alone study, despite its contributions to the application of ontologies in resolving neuroscience questions and the fact that it uses outcomes of the second study, such as the question hierarchy and classification. Figure 1-2 demonstrates how chapters are related along with contributions of each chapter.

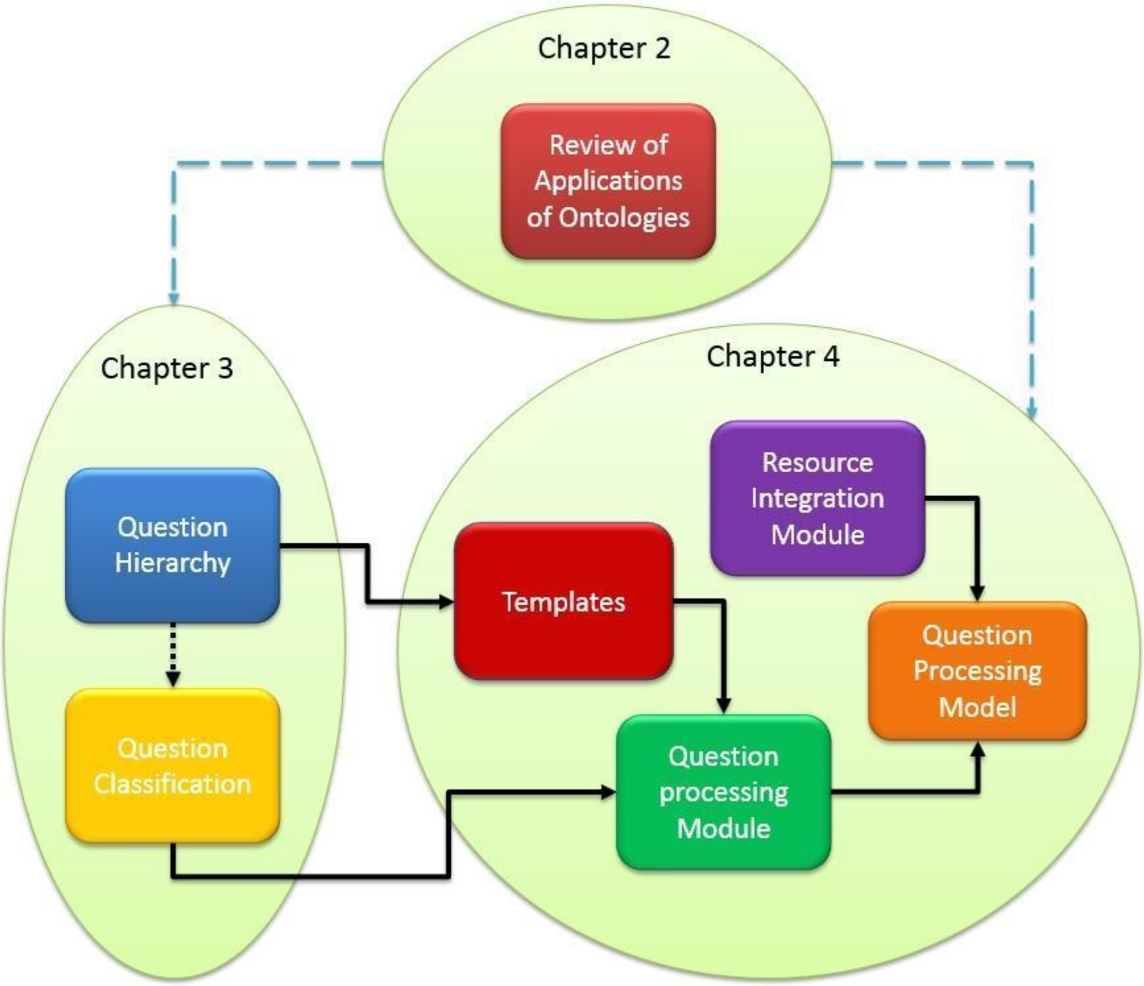

Figure 1-2 How studies are connected in this research, along with contributions of each chapter

1.4 Justification and Significance

As described before, vast and rigorous attempts have shaped understanding of the human brain. There have been ongoing initiatives in exploring neuroscience-related issues through computer-based systems for many years now. These activities have created a relatively new field of research and study called neuroinformatics.

For example, the human brain project (Shepherd, Mirsky and Healy, 1998), which aims to develop tools for searching, integrating and modelling neuroscience data, is one of the major projects that shows the importance of developing computational tools and methods.

Another major project has been the Allen Brain Atlas (Jones, Overly and Sunkin, 2009), which is a web-based gene expression of the adult mouse brain. The United States launched the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) (Insel, Landis and Collins, 2013) initiative to develop tools that scientists need for building better knowledge about the function and structure of the brain.

A project running from 2013 to 2020 is the global collaboration to combat traumatic brain injury (TBI). Traumatic brain injury is a prominent cause of death and disability worldwide with 50-60 million new cases each year. The cost of this injury is something around 400 billion US dollars and there is high chance that an individual will suffer from TBI in their lifetime. Moreover, according to Olesen *et al.* (2012), the cost of brain disorders has been in excess of 798 billion euros just in Europe. This is a huge burden on economic resources. Therefore, developing tools that help scientists in tackling brain disorders seems vital.

At the beginning of the chapter it was stated that computer-based tools are important for overcoming brain disorders and understanding the brain and nervous system. Hopefully by now, it is clear why this claim was made.

It was explained before that performing studies in neuroscience means engaging with huge amounts of data. Whenever scientists are dealing with huge amounts of data such as data produced in neuroscience, the need for computer readable data and interconnected powerful computational methods arises (Martone, Gupta and Ellisman, 2004).

Since neuroscience research includes studies in neurology, psychiatry, neurophysiology and neurosurgery (Liu *et al.*, 2015), the need for efficient and enhanced methods of data processing and management grows in this field. Therefore, research such as that reported in this thesis helps a vast group of scientists who try to resolve neuroscience-related problems. Namely, the first study intends to point scientists towards ontologies and using them for various purposes in their research. This study seeks to find all applications of ontologies in neuroscience and possibly point scientists towards new applications. It also seeks to resolve confusions on ontologies and their role in managing and querying data. The study aims to demonstrate how ontology can be used to study the brain, its disorders and how experiments are performed and managed. By providing this information, the study assists in accelerating knowledge advancements in managing resources and cost, creating and sharing data and applications, extracting and retrieving information and providing easier collaboration among scientists.

The second study intends to demonstrate how questions can be represented and classified with the aid of ontologies. This is important because it can be used to create a question hierarchy that not only helps resolving questions, but also demonstrates which questions can be resolved currently and which ones cannot be resolved at the moment. The question hierarchy will point scientists designing neuroscience information systems towards asking and resolving new and important questions instead of keeping them in a loop of resolving similar questions again and again. It will also assist them in focusing their energy and effort on important issues. The classification part of the study is important since it categorizes questions according to their level of complexity, which helps in resolving them. Furthermore, classification of questions can be used to perform various operations such as tailoring the results according to the background of the user of an information system that takes this approach.

These system users may include neuroscientists, neurophysiologists, neurosurgeons, healthcare technicians working with neurological cases, and many other professions concerned with neuroscience. It can even include their language background since ontologies provide language tags that specifies terms in different languages. This can be used for cross-language searches.

The last study is proposed to contribute to approaches to processing questions in neuroscience using ontologies. It intends to demonstrate how questions can be mapped to predefined templates, how they can be translated to machine language and eventually, how to seek answers for them through an integrated resource platform created by ontologies.

It is envisioned that this study may play a significant role in letting scientists understand ontology role and future in neuroscience and by this, understand the brain and problems concerning it better. Furthermore, it may let them envisage better methods of tackling neuroscience problems.

Through tapping into the part of this research that tackles and processes questions, scientists can ask and find results for questions automatically. These may be questions they could not find an answer for through automated means before and that required time consuming manual work.

1.5 Thesis Outline

The outline of this thesis is as follows:

Chapter 2 answers the research question "What are the applications of ontologies in neuroscience?". It covers a systematic literature review designed to investigate the current state of art in ontology application and other topics related to this research. It first discusses the ontologies as central tools to this research.

This chapter explores what ontologies are, demonstrates their difference to other similar entities such as terminologies and taxonomies, what tools are used for working with them and discusses why they are suitable for neuroscience.

Then, it goes through all current ontology applications via exploiting a systematic method, especially applications used in this research such as disambiguation, classification, and question resolution.

Chapter 3 tackles the second research question which seeks to analyse, represent, and classify questions in neuroscience with the aid of ontologies. This is done by selecting a set of questions, analysing, deconstructing and placing them into different clusters which will be later called dimensions. Dimensions represent questions and have some potential applications that will be pointed out in the chapter.

Furthermore, a question hierarchy will be created using those dimensions and represents questions by categorizing them into specific levels. Then, questions are classified using both an ontology-based method and machine learning (statistical) techniques.

Chapter 4 goes deeper into analysing and resolving questions and aims to answer questions via matching different data (knowledge) resources and ontologies in order to answer the third research question, which is "How can questions in neuroscience be resolved using ontologies?".

It intends to resolve real world questions such as 'What is the white matter volume of the amygdala in the subject?' or 'Which parts of the cortical region have the highest amount of atrophy?' and even more sophisticated questions. It does so by using neuroimage data, domain ontologies, question hierarchy from chapter 3 and techniques such as ontology-based query expansion.

Chapter 5 is the last chapter of this thesis and brings all findings, limitations and future directions of this thesis together. A brief review of each chapter is given, and aims, motivation and research questions are revisited alongside their answers. Figure 1-3 demonstrates a summary of the research outline. The logical relationships between the chapters (studies) are also depicted in this figure.

Appendix A demonstrates information on how databased searches were executed for finding primary studies in chapter 2 in detail and *Appendix B* includes information on platforms, resources and techniques used in this research.

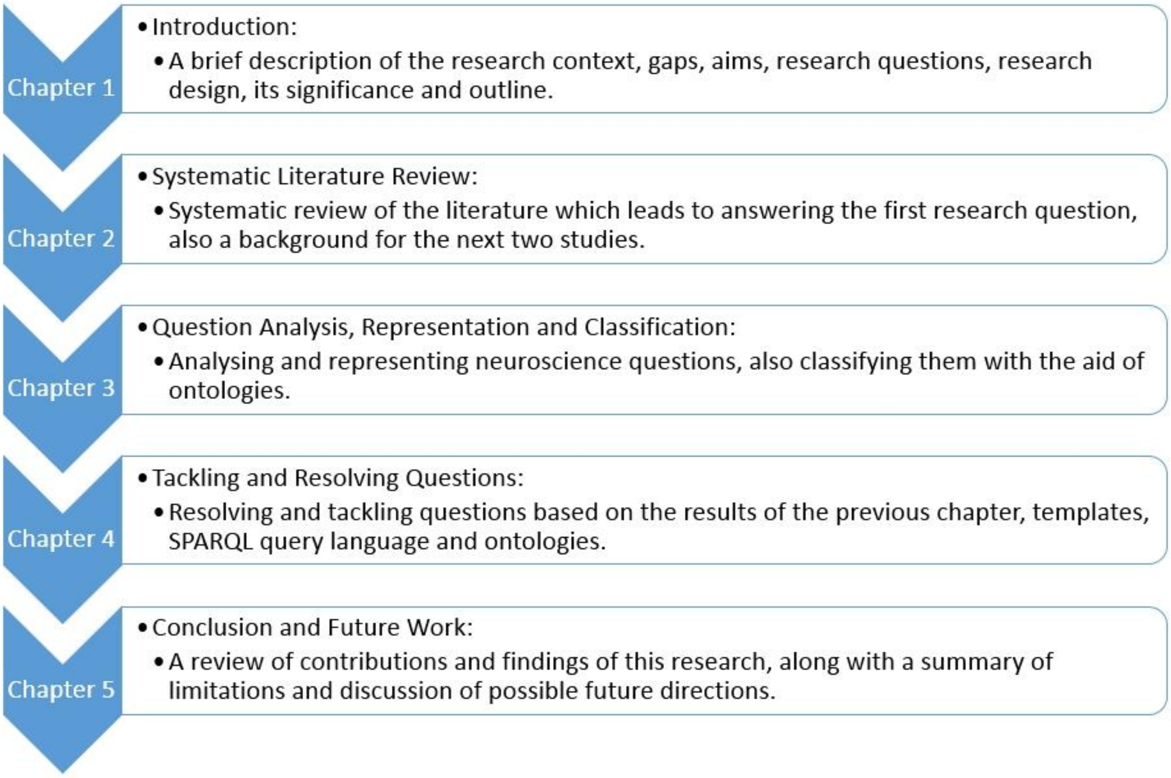

Figure 1-3- A summary of the research outline

1.6 Summary of the Chapter

This chapter gave an overview of the thesis as a whole and placed it in the context of neuroscience. It talked about the field of study and the unknown issues in the field. Furthermore, it introduced ontologies as the main tool of this research and discussed a summary of ontology applications in neuroscience.

In addition, it talked through the aims of this research, stated the research questions and discussed the significance of answering each research question. The structure of the remaining chapters and parts of the thesis was also shown in this chapter.

2 A Systematic Literature Review on the Applications of Ontologies in Neuroscience

As discussed in the Introduction chapter, this research seeks to find out current and potential applications of ontologies in neuroscience and then examine and process neuroscience questions with the aid of ontologies.

The main aim of the study presented in this chapter is to investigate the versatility of the application of ontologies in neuroscience. This is because the application of ontologies in neuroscience can be very useful in interpreting information regarding the brain and brain knowledge management.

Scientists can use potential capabilities offered by ontologies in tackling issues they are faced with, therefore including expanding their knowledge and overcoming obstacles in neuroscience. Ontologies can be used to solve problems and improve the understanding of the brain, nervous system, neuroscience and experiments related to it.

One other reason for the importance of applying ontologies is that they allow automated and computerized studies regarding neuroscience. Being able to express the information in a machine readable format is the goal of many domains, including neuroscience (Ascoli, 2013), and can assist in knowledge management in this field.

As will be discussed later in more detail, there are a few research papers that reviewed applications of ontologies in neuroscience, but most of them considered specific uses of it or did not contain very detailed information.

Therefore, the field can benefit from a comprehensive review based on a rigorous information acquisition and synthesis, alongside maintaining a critical view on the application of ontologies.

Based on the above information, this chapter answers the first question of the research, which is “What are the applications of ontologies in neuroscience?”. This study reviews current uses of ontology in neuroscience and suggests possible ways forward.

This chapter answers the research question defined in the above paragraph in a carefully designed, data-driven manner. Therefore, in order to accomplish the objectives of the study and to investigate the application of ontologies in neuroscience; or in other words, how neuroscience research has used ontologies and benefited from them, a detailed systematic review of the literature was performed.

A systematic literature review of studies will result in a comprehensive answer to the research question of this chapter. In addition, the search for documents performed in this chapter will assist in finding research related to other chapters.

This is because ontologies are the common thread in all three studies of this thesis. Therefore, a vast and rigorous review in this chapter makes the foundation and paves the way for more detailed and specific elaboration of studies performed in the following chapters.

In the remainder of this chapter, first, a background on the concept and notion of ontologies will be provided. Then the methods will discuss how this study was performed using a systematic review protocol. After that, the result section will go through the findings of the research. At the next step, the discussion section further explains the results and finally, the conclusion section finishes this study.

2.1 Background

Before delving into ontology applications, it is better to understand them, as well as rules associated with them. Therefore, first, some information will be provided on ontologies, their origins and characterizations, why and how to use them, and potential tools for working with them. Moreover, systematic literature reviews will be introduced and briefly discussed. Systematic review protocols are also discussed to determine the important elements that must be considered in a systematic review study.

The term ontology comes from philosophy. According to the Webster dictionary, ontology means 'a branch of metaphysics concerned with the nature and relations of being' or 'a particular theory about the nature of being or the kinds of things that have existence'.

In computer science, a widely acclaimed, succinct definition of ontologies that comes from Gruber (1993) was used. His definition of the ontology was 'a specification of a conceptualization'.

This definition was refined later by Studer et al. (1998) as 'a formal, explicit specification of a conceptualization' or as 'ontology is a formal specification of a shared conceptualization'. However, some researchers, such as Chandrasekaran (1999) describe it as 'content theories about the sorts of objects'.

Ontologies provide the structure and specifications of the concepts and relations among them in a specific domain. They are language independent (Larson and Martone, 2009) and the objects represented in ontology are called the 'universe of discourse' (Gruber, 1995).

Chandrasekaran (1999) counts ontologies as vital content theories needed to enable mechanism theories to reach their full potential. The reason for ontologies to be counted as content theories is that they strongly assist in identifying objects, their classes and relations between them.

He states that even though mechanism theories such as neural networks, fuzzy logics and rule-based systems count as the foundational mechanisms of creating intelligent mechanisms, they cannot do well without a good content theory. In fact, in a presence of a good content theory, most mechanism theories can do equally well.

He also mentioned that ontology can represent propositional attitudes such as hypothesise, belief, hope, expectation, desire and fear. There are other things that ontology can represent such as activities and plans. However, these are not in the scope of this research.

There are similar concepts to ontologies such as controlled vocabularies and taxonomies that should not be confused with each other or ontologies. This is because although they are similar, they each have their own definition and usage.

Controlled vocabularies are the backbone of taxonomies and are sets of terms that describe a specific domain and may also represent definitions and identifiers, but lack any kind of explicit relationships (Larson and Martone, 2009). Being the backbone means that a taxonomy structure is shaped on a controlled vocabulary. In other words, a taxonomy is a similar concept to a controlled vocabulary, with class information as additional data. In the same way, ontologies usually have a taxonomy as their backbone. This backbone provides a specification of objects and things in hierarchical formats, such as a tree or a lattice. Generally, if relationship information, which is the relations between concepts is added to a taxonomy, the outcome is an ontology. These definitions will be made clearer via examples given in the next paragraphs.

Figure 2-1 shows the difference between the ontology, taxonomy and controlled vocabulary. Ontology is different from a taxonomy or controlled vocabulary but is constructed on top of them in a hierarchical way.

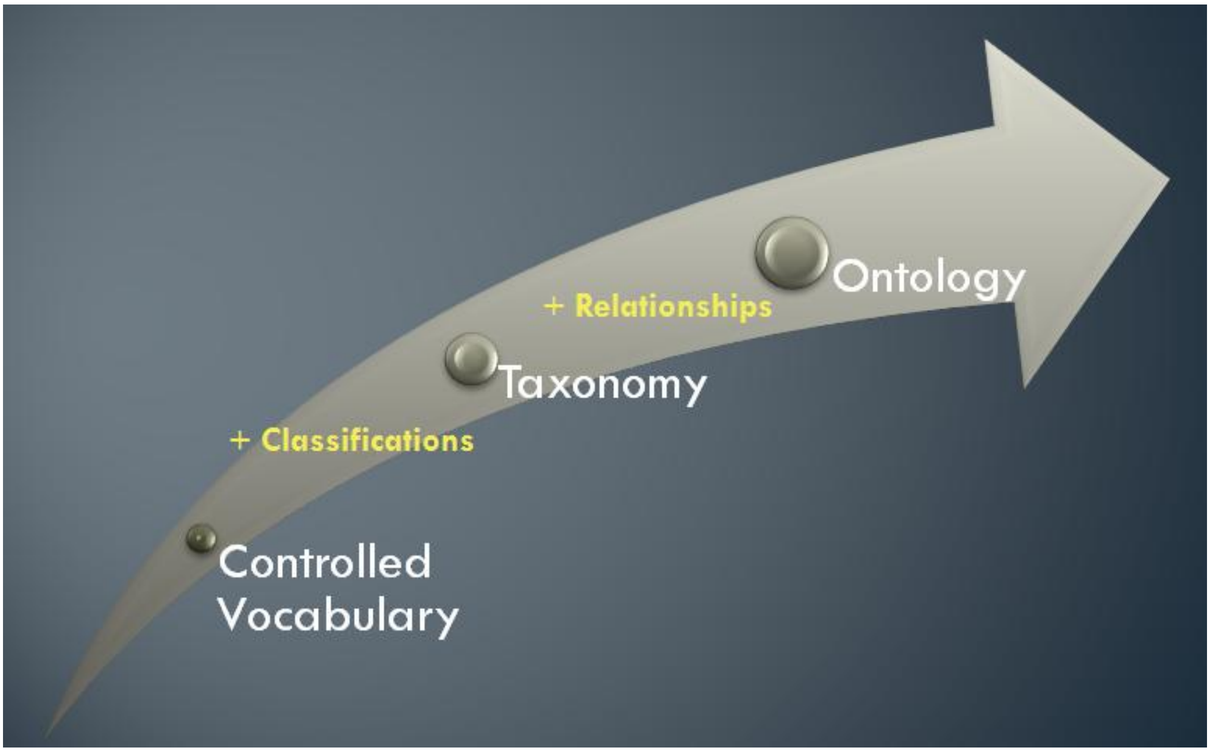

Figure 2-1- Ontology vs Taxonomy and Controlled Vocabulary

The difference between these three concepts lies in how detailed they are and in the level of information they provide. As can be seen from Figure 2-1, moving through the arrow, more details are added to each concept. The main differences between a controlled vocabulary and taxonomy are classifications, and ontology incorporates relationships which a taxonomy lacks.

An example is to think of ontology for the human body. In that case, the human body could be broken down to parts such as the head, neck, hands, legs and many others. Also, each of these parts and limbs has a relationship with other parts. As an explanation, it will be stated that the head is connected to the neck or neck connects the body to the head while each of them have a mass, diameter, bones and skin. All these other properties might also have their own attributes and relationships.

There will be a large list of concepts and relationships in this imaginative ontology. Here, the human body has been conceptualized formally and specifically with details and as a result, one example of ontology of the human body has been created.

There are three types of ontologies based on the level of granularity or abstraction, and usage including foundational (Gangemi *et al.*, 2002) or top-level ontologies (Temal *et al.*, 2008); reference (Larson *et al.*, 2007; Bodenreider, 2008) or domain ontologies (Staab and Studer, 2009); and application (Larson *et al.*, 2007) or task (Staab and Studer, 2009) ontologies.

Foundational ontologies like Basic Formal Ontology or BFO (Smith and Grenon, 2002) and Descriptive Ontology for Linguistic and Cognitive Engineering or DOLCE provide the logical rules and design patterns for other ontologies. They cover abstract concepts and relationships used across domains (Temal *et al.*, 2008).

Reference ontologies are used for multiple applications and define primary and minimal concepts. Minimal means that core ontologies should include only the most reusable and widely applicable categories (Temal *et al.*, 2008).

There are many examples for this type of ontology since they have many applications. Examples of domain ontologies are Subcellular Anatomy Ontology (SAO) (Larson *et al.*, 2007) and the Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003).

Application ontologies such as FMA-Radlex (Mejino, Rubin and Brinkley, 2008) are specifically designed for an application or task (Larson *et al.*, 2007) and conceptualize a specific domain (Temal *et al.*, 2008).

These ontologies can be extracted and built from reference ontologies as described in (Shaw *et al.*, 2008). This task can be achieved in two ways- extracting from a reference ontology or enriching an existing ontology or light-weight ontology (Mejino, Rubin and Brinkley, 2008).

There are many applications for working with and manipulating ontologies. They can be categorized in four groups, including ontology editors such as Protégé (Gennari *et al.*, 2003), SWOOP (Kalyanpur *et al.*, 2006), NeOn ontology engineering Toolkit (Haase *et al.*, 2008) and TopBraid Composer, program API libraries like Jena2 (Reynolds, 2004), ontology reasoners such as Pellet (Sirin *et al.*, 2007), FaCT++ (Tsarkov and Horrocks, 2006), Racer (Haarslev and Müller, 2001) and HermiT (Shearer, Motik and Horrocks, 2008) and finally ontology stores like RDFSuite (Alexaki and Christophides, 2001) and Seasame (Chen *et al.*, 2006).

The mathematical semantics of ontologies is based on the Description Logics (DLs). Description logic is a form of knowledge representation language and can be used in expressing the knowledge of a specific domain in a formal and structured way.

Many ontology languages are based on description logic. Ontology Interface Layer or OIL (Fensel *et al.*, 2001), the DARPA Agent Markup Language or DAML+OIL (McGuinness *et al.*, 2003), and Web Ontology Language or OWL (McGuinness and Van Harmelen, 2004) are some examples of description logic based languages.

Description logics allow reasoning and by this, contribute to ontologies in many ways such as inferring information. As Baader et al. (2007) states, by using description logic as a foundation for ontologies and applying reasoning, some aspects of ontologies will be guaranteed. These aspects include consistency, structure of ontologies, results acquired by reasoning and having no redundancies.

Reasoning can be used both in designing and deploying ontologies, can help inferring hierarchies (Courtot *et al.*, 2011; Horrocks, 2013) and update the hierarchy class of the ontology by checking for implicit subsumption relationships. This helps in finding similarities and differences among data based on ontologies, and discovering or inferring new information from the old data.

Reasoning assists researchers to know if the ontology fulfils their goals or not. A subsumption hierarchy can be calculated by using a subsumption algorithm (Baader, Baader and Sattler, 2001; Baader, Horrocks and Sattler, 2007). It can also assist in checking consistencies of ontologies and to investigate if the ontology is contradictory or not (Baader, Horrocks and Sattler, 2007; Horrocks, 2013). Checking consistencies can be done through consistency algorithms such as tableau-based algorithms.

Many reasoning techniques and algorithms have been designed for working with description logics and as a result, ontologies. Resolution based approaches (Hustadt and Schmidt, 2002), automata based approaches (Calvanese, De Giacomo and Lenzerini, 1999), structural approaches (Baader, Brandt and Lutz, 2005), and tableau-based approaches (Baader, Baader and Sattler, 2001) are some of them.

The most popular approach is the tableau based one. Racer (Haarslev and Müller, 2001), Pellet (Sirin *et al.*, 2007) which uses SPARQL-DL, and FaCT++ (Tsarkov and Horrocks, 2006) are reasoners that use tableau technique for reasoning. HermiT (Shearer, Motik and Horrocks, 2008) is another reasoner that is based on hyper-tableau calculus.

Ontologies can be extended or extracted from and by this behaving as a modularized design (Courtot *et al.*, 2011) that allows further development and curation. This characteristic eases working with large ontologies and parallel work, and accelerates the process. An example of a modular ontology is the Neuroscience Information Framework Standard Ontology (NIFSTD) (Bug *et al.*, 2008).

2.1.1 Why Ontologies?

Before initiating studies about the applications of ontologies in neuroscience, an initial concern or question might be regarding reasons behind using ontologies. While there are plenty of technologies and approaches available in organizing and retrieving data, what is the significance of ontologies for their consideration as an important tool for neuroscience research?

The answer lies in the nature of neuroscience and ontologies. Ontology-based characterizations are flexible, not very restrictive and explicit enough for logical inferences (Hamilton *et al.*, 2012).

For example, studying neurons is essential for understanding the structure and functions of the brain in neuroscience (Bota, 2008). It is known that neurons have a considerable number of properties. On the other hand, ontologies provide flexible characterization methods (Hamilton *et al.*, 2012). Therefore, using ontologies would help substantially in neuron characterization.

Furthermore, ontologies can assist in solving problems in the way of sharing information (Grewe, Wachtler and Benda, 2011). Neuroscience data is usually created in different labs with different procedures and standards.

Thus, while sharing the data, they have to be accompanied with huge chunks of additional information and metadata which describes them and the way for using them. Using ontologies, the data can be easily transferred with minimum size between research teams and scientists.

As discussed earlier, description logics are the foundation of ontologies. They are based on open world assumption (OWA), which allows incomplete information among instances of ontology (called ABox or Assertion Box).

By allowing to be operated with incomplete information, an ontology-based system lets the users query the data while the developer is still completing the data curation (Heymans *et al.*, 2008). This suits neuroscience well since the data on the field is still being completed in many aspects. Therefore, by using open world assumption, ontology helps in reasoning and flexibility; however, it might add to the computation cost.

The other characterization of ontologies is not using unique name assumptions (UNA). Using unique name assumptions means that a unique name would be allocated to a property, variable or attribute. Therefore, while using ontologies a concept can be referred with many different names.

This helps neuroscience and scientists in this field; since many other branches of science research, including medical and biomedical science, deal with neuroscience. Using unique name assumptions assists these scientists from different countries or different backgrounds to use and curate the data, and communicate with each other.

There are some alternatives for ontologies such as databases or XML-based modelling and representation approaches, or digital atlases. However, while databases are very powerful in storing and managing data, they have shortcomings such as not having an implicit, strongly typed relationship for defining the relation of its rows (records) and columns (attributes). Ontologies provide these things. This makes databases a second option for the use in fields such as neuroscience (Larson and Martone, 2013).

Another possible alternative for ontologies would be the eXtensible Markup Language (XML). This has been discussed by Raikov and De Schutter (2012), who show XML being used as model description languages such as NeuroML (Gleeson *et al.*, 2010) and CellML (Lloyd, Halstead and Nielsen, 2004).

However, XML is not as suitable as ontologies for use in the neuroscience domain mainly because of two main reasons. First, it is mostly comprehensible for the human user and not computers. Second, XML is mostly written in natural language, which might lead to ambiguities for computers.

Despite that, XML-based description languages tried to overcome this problem using schema languages, but they still do not offer a strong support for datatypes, processes and complex dependencies between different elements.

Therefore, XML-based description languages are prone to failure, created by different perceptions from different parties. These languages are able to eliminate syntactic issues but lack the ability to overcome semantic issues.

Other possible alternatives could be digital atlases. They are a means to store and manage spatial information of the brain; however, they cannot fully capture functional organisation and relations of various anatomical structures. This is what ontologies can handle. On top of that, they can even assist digital atlases in representing information vividly. This will be discussed later.

Apart from the other technologies, some scientists proposed approaches that do not completely substitute ontologies, but complement or partially substitute them. However, they acknowledge the importance and use of ontologies.

For example, Grewe et al. (2011) discussed open metadata markup language (odML). It includes a terminology used for annotating electrophysiological data and uses a bottom-up approach, which is opposite to the top-down approach introduced by OBO for biomedical ontologies. Of course, it incorporates a terminology which as explained before, is very close to ontologies.

There are some disadvantages to ontologies too. Building and maintaining ontologies is a costly and tough task to accomplish. This is because of hurdles such as the need to know the field the ontology is going to be designed for, the high cost of some of the reasoning-based computations and the management costs.

Therefore, it might be better to use the ontology in domains in which the schema plays an important role, the data is of high importance and whole data is not ready yet. Neuroscience would be a good example of where ontologies can be effective (Horrocks, 2013).

There are not many research papers in neuroscience related to the application of ontologies in neuroscience. However, as demonstrated by Bodenreider (2008), research has shifted from terminologies toward ontology during the past years. In hindsight of this shift, a few research papers discussed the role of ontologies in biomedical science, that can be seen as a similar field to neuroscience.

Rarely detailed research papers have been performed on the application of ontologies in neuroscience. Furthermore, few major research papers used a systematic review protocol towards reviewing ontologies in neuroscience. Nevertheless, the field lacks a vast, rigorous research that demonstrates all applications of ontologies.

Now that the information regarding ontologies, their definition, characteristics and tools for working with them and their importance in neuroscience is explained, it is time to discuss methods this study used and results that arose from incorporating those methods.

To the best of the author's knowledge, this research gives the most comprehensive overview of applications of ontologies in neuroscience so far. This is mainly due to factors such as the accurate and specific method it uses, the wide scope of the documents investigated and the detailed synthesis. Related research papers will be discussed further and in detail later, and in the discussion section.

2.2 Methods

This section describes the methods that assisted in gathering and analysing research papers. The method should be able to accurately direct this study towards finding significant research papers regarding the applications of ontologies in neuroscience.

Therefore, a rigorous systematic review of literature was performed to find research in neuroscience that used ontologies to accomplish their goals. Systematic literature reviews (Kitchenham and Charters, 2007; Higgins and Green, 2008; Grant and Booth, 2009; Kitchenham *et al.*, 2009) identify, evaluate and interpret research related to a specific subject or question.

Reasons for performing a systematic literature review are to summarize current research information and gather evidence on a specific technology or treatment, to find gaps, create a strong background for a research question and to inspect the validity of a hypothesis.

It can be said that a systematic literature review is more reliable and possibly better than other types of literature reviews. Because it defines the search methods prior to starting the actual search, it is not biased. This results in considering research opposite to or contrasting from the author's preferences and presumptions.

In addition, it provides the reader with the search approach, criteria for valid results, the scope and number of results. Therefore, the reader can see more clearly and easily examine the advantages and disadvantages of the research and reach a conclusion.

Different questions can be answered via a systematic literature review. Some of them include assessment of the effect of a software or technology, assessing the frequency of adopting a technology and tracking the effect of a technology on different types of costs.

Systematic literature reviews are multiple stage processes. This means that they have different stages in their review process including planning the review, conducting the review and reporting the review.

The systematic review is called secondary study and studies that it uses for analytical and synthetic reasons, are called primary studies. In other words, from a systematic literature review, scientists draw conclusions based on primary studies in a field.

The review protocol of this study was shaped based on instructions for computer science systematic literature review presented in Kitchenham and Charters (2007) and Kitchenham et al. (2009). This guideline was created according to a review of other systematic review guidelines, experiences of projects performed at universities and consulting domain experts.

It was created based on medical systematic literature reviews, and it was specifically tailored for investigating computer technologies regarding medical research, which matched the goal of this study perfectly.

It is worth mentioning that this guideline is not very different from guidelines in Medical systematic reviews such as Cochrane (Higgins and Green, 2008) which is a review guideline for interventions. It is only different in the sense that it provides translations for medical terminologies and descriptions used in assessing systematic reviews such as population and intervention. The next section will describe the review protocol of the study.

2.2.1 Review Protocol

The review protocol sets the method of handling the review materials which are research papers in this study. It describes the rationale and the way researchers approach the reviewing process.

In this study, some eligibility factors were considered for the primary research. Articles from respected and recognized resources that reported research findings were considered as eligible research. The language was limited to English, and they had to have ontologies included as a part of their approach. Also, the field was limited to neuroscience.

Despite this, if a primary research paper mentioned a study outside search criteria of this study, it was still investigated; for example, if a study is referenced in literature as an important resource, but it is in biomedical science instead of neuroscience.

In order to find appropriate resources, two librarians were consulted from engineering and health informatics backgrounds to make the document search cover possible domain differences.

Based on the information from librarians, some databases, ranging from general to more specific regarding the research topic, were searched. Databases included multi-domain databases Web of Science, Scopus and Science Direct; medical databases PubMed and Medline and computational (informatics) databases Compendex, IEEE and ACM Digital.

Then the search was performed using different elements. These elements included keywords, their synonyms, Boolean operands and filtration mechanisms provided by the databases.

In the first search, general keywords were used and all related documents were selected. The first two keywords were neuroscien*, ontolog*. After that, some filtration strategies provided by the databases were used. For example, in Medline, the results were limited to 37 Mesh headings related to neuroscience and the topic of this study. The full details of the search on each database, alongside the date of coverage, are included in appendix A of this chapter.

As discussed before, looking for documents did not stop at this point. A few journals that seemed related to the domain of research such as Frontiers in Neuroinformatics, Neuroinformatics and Neuroimage were searched too. In these journals, the search process was limited to papers after 2010 to find recent applications of ontologies.

The selection of primary studies was done based on the title of the study, keywords and abstracts. This selection process resulted in the number demonstrated in the screening part of Figure 2-2.

In a few cases, when the initial search mentioned them as relevant papers but information in title, abstract or keywords did not reveal much relevant information, a skim read of the research paper was performed, plus a thorough examination of its conclusion to ensure the relevance of the document before moving to the next phase, which was full-text assessment.

A full-text assessment of the documents revealed the final number of articles involved in the analysis. Figure 2-2 shows the number of selected studies based on different phases of the search for documents.

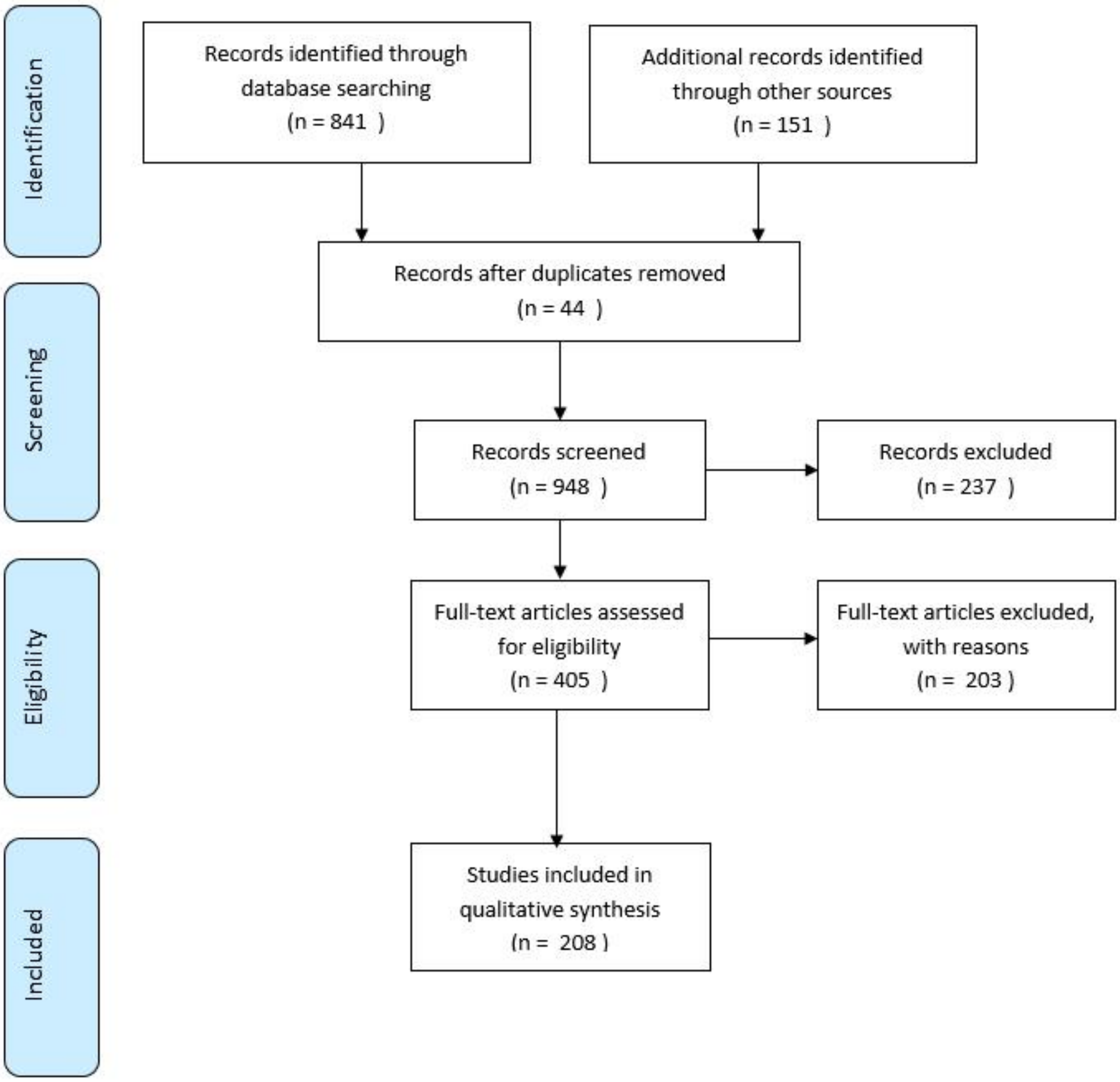

Figure 2-2- Systematic Literature Review Protocol

Extraction of information was performed via several questions, which were created according to the research question of this study. Questions were designed and used for investigating documents according to the method they designed for conducting their research, analysing their data and producing results.

Some of these questions included What was the year and name of the journal or conference?; What was the sub-domain of the paper in neuroscience?; Why did they use ontologies?; What was the ontology?; What reason they used the ontology for? and What technique was used alongside ontologies (if applicable)? An example of techniques which ontologies were used in accordance with include machine learning and semantic similarity.

Data synthesis is to summarize and bring together results of primary studies. As the synthesis part of this review, the data found in search and information extraction phases were combined into a descriptive synthesis and a final mapping of ontology uses was shaped.

In order to do this, findings from primary studies were organised as a list of applications. To create this list, when a primary study was summarized, the applications used in that study were pointed out from the most important to least important.

Then, primary studies were sorted according to the application of ontology in them and placed in related categories. Primary studies that used ontologies in multiple ways have been mentioned in more than one category.

For example, if a study used ontologies for classification and disambiguation purposes, it was mentioned in both related categories. If two research papers used ontologies for resource management, they were put into the same category. The homogeneity of the applications was always kept in mind, so as not to group unrelated research together.

After that, similar applications were color-coded and merged together, and in the cases where an application was a subclass of another one, it was placed as a subcategory. Once categorization finished, a quick reverse-engineered process was performed to make sure all results matched with the categories.

Again, to give an example; in the first stage of grouping the applications, 10 research papers have applied ontologies for *mapping data*, while another 25 papers have mentioned *integration* as their primary application. However, both these groups used ontologies to bring diverse sets of data together and build bigger datasets. Therefore, both groups were referred to as groups that use ontologies for *integration* purposes.

This categorization can provide some extra information, such as the number of neuroscience studies which used ontologies, most popular ontologies in neuroscience, most popular applications of ontologies in neuroscience and last but not least, the status of using ontologies in neuroscience according to each year.

This extra information was used to produce the final result. Some of it has been discussed further in the Discussion section and the Conclusion chapter of this thesis. However, since the goal of this chapter was to investigate the use of the ontologies in neuroscience, the focus in the Result section has been on answering this question.

It is worth emphasizing that this process was data-driven, meaning that there were no predefined application categories. The categories were selected and named based on the information found in the review process of primary studies. In other words, applications were the direct result of analysing the primary studies.

There were not many risks of bias in this research, apart from the fact that the search and synthesis were performed by one individual, and his point of view might have echoed on the work.

However, reviews and comments of professional researchers including people in the neuroscience domain, were sought and reflected. This was done through sending drafts and results to them based on a scheduled routine and asking for their feedback regarding different phases of the research, and then reflecting it on the study.

2.3 Results

The outcome of the full-text assessment of primary studies demonstrates some interesting insights regarding application of ontologies in neuroscience and the way ontologies assist research. This section describes findings of this full-text assessment.

Ontologies have different applications in neuroscience, including search and data discovery, data capture and representation, integration, collaboration, classification and categorization, disambiguation, knowledge management and organization, and finally, data analysis.

2.3.1 Search and Data Retrieval

Ontologies can be beneficial in designing search and data retrieval systems. They assist in searching for specific data, resolving queries or questions and data (text) mining approaches.

Ontology-based methods can be used to allow the performing of sophisticated semantic search and question answering (Tran *et al.*, 2007). They enable search systems to move towards a content based information retrieval which ultimately isolates the current query-string based search task in which users have to create and execute multiple queries in order to achieve the desired answer (Bonino *et al.*, 2004).

Brain disorders comprise a huge number of human disorders (Köhler *et al.*, 2012) and huge budgets are used on them (Olesen *et al.*, 2012). Therefore, being able to study them in every aspect is vital.

Ontologies can be helpful for finding problems caused by nervous system diseases. For example, Merlo et al. (2014) used the synapse ontology to find domains demonstrating synaptic dysfunctions after traumatic brain injuries. Similarly and at a higher level, Gupta et al. (2003) tried to use ontologies in order to constitute a knowledge base of neurological disorders as a formalized disease map.

Ontologies can be applied to answering and resolving queries or questions. In fact, this is one of the most prominent fields that ontologies are used in. The third study of this research discussed this aspect of ontologies in further detail.

Query answering is an important aspect of the semantic web as it can assist in user interaction with data and ontologies (Kollia, Glimm and Horrocks, 2011). Different query languages have been shaped for this purpose including SquishQL (Miller, Seaborne and Reggiori, 2002), SeRQL (Broekstra and Kampman, 2003) and SPARQL (Prud and Seaborne, 2006).

Another use of ontologies can be for query formulation in query answering systems, since the user can create the query by using a step by step query building technique that is made possible by the hierarchal structure of the ontology and terms in the ontology (Jakoniene and Lambrix, 2005; Chen *et al.*, 2006).

Similarly, ontologies can also be applied in computational techniques used for query answering. An example of these techniques would be query expansion that results in better answers for a query (Jakoniene and Lambrix, 2005; Bug *et al.*, 2008).

An example for ontology being used as a means of search and data retrieval is the virtual fly brain browser (Milyaev *et al.*, 2012). It is an online resource dedicated to provide anatomic, genetic and expression of fly brain by use of the BrainName ontology and drosophila anatomy information.

NIF search (Gupta *et al.*, 2008) is an initiative for a search system in neuroscience from the Neuroscience Information Framework (NIF) based on the NIFSTD ontology. NIF is used in the next two studies and will be discussed in more detail later.

Another example would be the (Zao *et al.*, 2014) project, which uses ontologies such as the semantic sensor ontology and the brain-computer interface ontology for semantic searching over linked data.

Other examples that performed ontology-based query answering in part of their project are the semantic synapse project (Samwald and Adlassnig, 2005). This project seeks to build a search engine for neuroscience on the web by using ontologies such as the synaptic ontology. System presented in Birgiolas *et al.* (2015) is a keyword search for NeuroML models and uses NeuroLex (Fahim T Imam *et al.*, 2011) relationships. Veeraraghavan and Miller (2014) represent a faceted search interface for neuroanatomy and uses the FMA for this means and work described by Gertz et al. (2002).

Faceted search (Hearst, 2006) enhances traditional search systems by adding a navigation technique on the data. It is created upon faceted classification, which enables different views on the data on top of the taxonomical view and order.

Ontologies can assist in answering queries and questions by being used as parts of data (text) mining systems. Text mining (Tan, 1999) also known as text data mining, refers to the process of extracting useful and complex data from text documents.

Textpresso for neuroscience (Müller *et al.*, 2008), which is based on the Textpresso (Müller, Kenny and Sternberg, 2004), is a good illustration of using ontologies in text mining. It crawls documents and indexes them based on their ontology and generates answers based on that. Another system that is similar to Textpresso is the system expressed by Li et al. (2007) called an ontology-based mining system for competitive intelligence in neuroscience (OMSCIN).

Other notable examples for application of ontologies in text mining are Groza et al. (2015) that detects and extracts disease information using human phenotype ontology (HPO) and other vocabularies and ontologies; ADO (Malhotra *et al.*, 2014) that uses text mining in part of its work, and the WhiteText (French *et al.*, 2015), that uses the ontology for tasks such as query expansion, user input and disambiguation in a text mining project.

Recent advancements in ontologies allowed the indexing systems to improve, and even resulted in developing new semantic-based approaches towards indexing (Setchi and Tang, 2007). Some search and discovery studies such as (Müller *et al.*, 2008) and (Li *et al.*, 2007) use ontologies for indexing purposes, so they can return results more efficiently.

Ontologies can be used for reasoning, which was discussed in detail in the Background section. One example of using ontologies for reasoning is (Slater *et al.*, 2015), which uses Aber-owl ontology repository (Hoehndorf *et al.*, 2015) and aims to use automated reasoning on BioPortal (Noy *et al.*, 2009) ontologies to get better results.

2.3.2 Data Capture and Representation

Ontologies can be used to represent neuroscience data and allow accurate representation of the reality. Important coordination bodies such as the Open Biomedical Ontologies (OBO) consortium (Smith *et al.*, 2007) suggest ontologies as a means of data representation.

Ontologies play a major role in capturing neuroscience data in an accurate way. Capturing data is one of their most natural applications based on their structure and foundations which was discussed in the Background.

In addition, it is known that information is just processed data. In fact, adding semantic layers and the act of putting data into context turns it into information. Therefore, simply capturing the data based on an appropriate ontology, makes it a step closer to information as ontology adds to the richness of the data by putting it in a clear context and adding metadata.

Generally, the information captured by ontologies can be used for numerous applications, such as resource organization, modelling, integration, annotation and many other applications described in this document.

For example, the Cognitive Atlas Ontology (Poldrack *et al.*, 2011) tries to build an ontology that can be used in systematically answering the question of what mental processes exist and how they are related to the tasks operating them.

Ontologies have the capability of capturing the information of neuroscience experiments. They offer the ability to capture methods for running experiments plus other related data. Tiesinga et al. (2015) emphasises on describing experimental and analysis protocols in order to be able to create a complete model of the brain.

There are a considerable number of research groups that dedicate part of their work to describing experimental protocols. The cognitive paradigm ontology, CogPo (Turner and Laird, 2012), stores information regarding the cognitive experiments.

Another group of researchers sees one of the major problems of electrophysiology that the data in the description of the experiments is unclear (MouÄ�ek, 2014), and suggests ontologies, such as the ontology for experimental neurophysiology (Petr *et al.*, 2013), as a solution. They try to overcome this problem by storing data, metadata, experimental scenarios, user comments and other related information in a structured manner and as an ontology (Jezek and Moucek, 2011).

A study focused on capturing research outcomes is one conducted by Jernigan et al. (2016) that aims to create a data repository with the information related to experimental data. This specific study contains extensive information from 10 different sites with more than 1490 subjects. Another similar project is RegenBase (Callahan *et al.*, 2016), which is a knowledge base for spinal cord injuries biology.

Ontologies also play a role in capturing neuroimaging data. They can capture information regarding neuroimages plus their characteristics. Managing neuroimage information is very challenging since there are many imaging modalities and there is a tendency of data heterogeneity (Prodanov, 2011). An example for this use is the (Poliakov *et al.*, 2007), which captures metadata regarding the fMRI images as instances of an ontology.

Some researchers also created ontologies that not only capture neuroimages, but other forms of multimedia information captured from experiment subjects too. The Stimuli Multimedia Ontology (STIMONT) (Horvat, Bogunović and Ćosić, 2014) is such an effort and proposes the ontology will overcome problems made by storing different types of information in databases.

Another category of capturing neuroscience data is capturing patient data. The Cloudwave (Jayapandian *et al.*, 2015) project is one example that uses an Ontology-driven Patient Information Capture (OPIC) (Sahoo *et al.*, 2012) system.

Ontologies can be used to describe instrumental information too. Describing instruments used for performing the neuroscience assessments is very important. Batrancourt et al. (2014) creates an extension that comprises research instruments in brain studies and cognitive functions, to a pre-existing ontology about neuroimages and region of interests.

Another example is the ASD phenotype ontology (McCray, Trevvett and Frost, 2014) which includes information regarding instruments and their features for scientists working on Autism Spectrum Disorder (ASD).

Ontologies have the capability to capture information regarding the nervous system, brain and its sub-parts. For example, the Foundational Model of Anatomy ontology (Rosse and Mejino, 2003) reveals some information regarding the anatomy of the brain.

The Subcellular Anatomy Ontology (SAO) (Fong *et al.*, 2007) is about entities smaller than a cell and maps macromolecules to subcellular structures, and defines the relationships between a part of a cell to the whole cell or to a higher level entity.

In addition, ontologies such as the Cell Ontology (Meehan *et al.*, 2011), can capture the information concerning neurons and cells. As described in the Background section, ontology is generally well fitted for describing neurons, their types and relationships can be described using ontologies. Neuron Registry (Hamilton *et al.*, 2012) is such an effort. BAMS neuroanatomical ontology (Bota, 2008) and some other ontologies also provide information regarding the cells.

Another notable application of ontologies is in capturing disease information. Capturing disease information would be very effective for diagnosis, research and even curing or eliminating the disease.

For example, the Alzheimer's Disease Ontology (ADO) (Malhotra *et al.*, 2014) is dedicated to storing Alzheimer's disease data. The Mouse Pathology Ontology (MPATH) (Schofield *et al.*, 2013), where most of its information is species-agnostic, captures information related to anatomic pathology. Species-agnostic information is the information that applies to all species.

Other similar projects are the NUSDAST (Wang *et al.*, 2013), that stores information regarding schizophrenia; Maynard et al. (2013), which created the Neurodegenerative Disease Phenotype Ontology (NDPO) for studying neurodegenerative diseases; McCray et al. (2014), which studies autism spectrum disorder and the Epilepsy and Seizure Ontology (EpSo) (Sahoo, Lhatoo and Gupta, 2014).

In addition, gene ontology is used for enrichment analysis, which sets to find anomalies in specific up-regulated gene sets using gene annotation. There are many similar examples, including recent research by Provenzano et al. (2016), which performs comparative analysis on two mouse models of autism and Huang et al. (2015), which uses enrichment analysis for investigating gene regulatory networks related to the profile of autism spectrum disorder (ASD) based on miRNA expression.

The other subgroup is the application of ontologies in annotation. Traditionally, knowledge about a document is recorded as meta-data (Uren *et al.*, 2006). Annotations give the functional description of the data (Larson and Martone, 2007; Bodenreider, 2008). The semantic documents approach is to annotate the documents with the aid of ontologies, and this leads to semantic annotations (Berners-Lee, Hendler and Lassila, 2001).

As Uren et al. (2006) describes, 'semantic annotation formally identifies concepts and relations between concepts in documents'. Annotations have led to what we describe as an intelligent document; a document that 'knows about its own content'.

Using ontologies for annotation will help in having semantic annotations, which leads to standardization, better interoperability and ease of data integration and information retrieval (Uren *et al.*, 2006). Effective annotation depends on the annotator knowledge and attention (Hollink *et al.*, 2003). There are different tools for annotation of both text and images.

Multiple, diverse research papers have used annotation and the application, and importance of annotation has been pointed out a lot in primary studies. The Phenoscape (Edmunds *et al.*, 2016) showed that ontological annotation of phenotypic effects of genetic alterations in model organisms, executed within a knowledgebase, can be applied to generate testable and useful hypotheses about evolutionary changes in morphology.

NeuroCarta (Portales-Casamar *et al.*, 2013) is a knowledge base that tries to strengthen the information on genes and phenotypes in order to shine light on the relationship of genes and diseases. It annotates the information both automatically and manually to add depth to studies.

Examples of research papers using ontology-based annotations include (Shang *et al.*, 2013), which presents an ontology-based entity expansion to improve the annotation of documents; the works presented in (Turner *et al.*, 2013) and (Chakrabarti *et al.*, 2014), which use the CogPo ontology for annotating neuroimages and documents as a part of their approach; NeuroPigPen toolkit (Sahoo *et al.*, 2016), which introduces a signal representation format using an ontology called EpSo (Sahoo, Lhatoo and Gupta, 2014) as a part of its approach with the ability to use metadata and ontology for representing signals; and Domeo (Ciccarese, Ocana and Clark, 2012) which is a software for performing, versioning, sharing and viewing web annotations.

Moreover, research papers such as (Milyaev *et al.*, 2012), (Keator *et al.*, 2013), (Mirbel and Crescenzo, 2009), (Gertz *et al.*, 2002), (Schwartz, Thirion and Varoquaux, 2013), (Li *et al.*, 2007) and (Jayapandian *et al.*, 2015) are among the research using ontology for annotation as part of their approach too.

Another subcategory of data representation is to report and disseminate research. Zaveri et al. (2010) uses the randomized controlled trials (RCT) ontology for reporting clinical research in neurosurgery. Similarly, the RegenBase (Callahan *et al.*, 2016) knowledgebase also tries to address the lack of standards for reporting experiments of spinal cord injuries.

Ontology can be used to demonstrate brain information in atlases. Some research, such as Wu et al. (2016), use ontology for showing the information in a hierarchical order in various atlases.

Calabrese, Badea, et al. (2013) uses ontology for anatomic segmentation and labelling in a magnetic resonance histology (MRH) atlas of the developing rat brain and in other parts of their work. They used a developmental ontology for detecting brain segments while tracking postnatal brain development information. In the same way, Calabrese et al. (2015) uses ontology for segmentation and visualization which is described below.

Visualization is another subcategory of data representation. Most brain atlases use ontologies for displaying and visualizing brain information. Allen Brain Atlas-driven visualizations (ABADV) (Zaldivar and Krichmar, 2014) is an online tool for retrieving and visualizing expression energy data from the Allen Brain Atlas (Hawrylycz *et al.*, 2014).

Ontology is used as a tool for visualization and moving between different structures. Veeraraghavan and Miller (2014) use an anatomy ontology (FMA) in combination with a 3D atlas in order to visualize anatomic structures of the brain.

Majka et al. (2012) uses ontologies such as BAMS and NeuroNames, reconstructing 3D brain structures such as isocortex in atlases. As mentioned above, the rhesus macaque atlas (Calabrese *et al.*, 2015) uses ontology of the mammalian brain for visualization. Nowke et al. (2015) is another research paper that uses ontologies for visualization purposes.

Ontologies also allow information to be modelled the same way the brain does it (Katifori, Vassilakis and Dix, 2010). LOTAR (Riboni, Civitarese and Bettini, 2016) uses ontologies for modelling human activities in order to build a behavioural analysis and anomaly detection system for early detection of cognitive impairment.

Arbib et al. (2014) discusses that modelling the brain needs both structural and functional ontologies and uses the brain operating principals (BOP) ontology in its approach. Another research that uses ontologies for modelling is (Schmitt and Eipert, 2012).

2.3.3 Integration

As pointed out in the introduction chapter, the information in neuroscience is dispersed. There are many resources and experiments created using different methods by different parties involved in neuro-scientific research.

Parties or scientists in charge of organizing resources mostly use local protocols in gathering data, processing them and producing results, and even some of them may use different naming systems. Therefore, integrating these data seems vital.

Integration of data and ontologies results in having a rich set of information. It provides mapping of data and better collaboration between scientists and specialized applications and software.

Ontologies can be used as a ‘glue’ between different data, applications and other ontologies (Jakoniene and Lambrix, 2005; Larson *et al.*, 2007; Bodenreider, 2008). Also, by using ontologies as standard sets of terms, scientists have an easier task in integrating the data. This is because using a common language makes data integration easier.

Since ontology is described explicitly and clearly (formalized) (Larson *et al.*, 2007), coded into machine readable formats (externalized) and put into a recognizable format for other sources and applications (standardized), it can assist with overcoming integration problems.

Furthermore, as discussed in the Background section, ontologies have characteristics such as not using unique name assumption while using open world assumption. These characteristics, alongside reasons discussed above, makes them excellent choices in being used for integration purposes.

However, integration is difficult. This is because the information is coming from different sources, and is captured via different imaging machines using different modalities, under different conditions and through variant processes.

Integration can be used for multiple goals, including creating a larger resource that can be used to answer questions or share information between scientists and applications. Integration can be implemented among two or more ontologies, or one ontology and an external source such as an atlas or a database.

One use of integration is to bring several resources within a single framework with the goal of answering questions. For example, AlzPharm integrates BrainPharm with the SWAN ontology for its query answering purposes.

Another example is Imam et al. (2012), which discusses how the NIF brings many ontologies together and enhances research capabilities. NIF has several modules. The NIF Disco framework presented by Marenco et al. (2010) is a part of the NIF platform, which aims to propagate integrated information on the web.

Another approach towards a web-based platform is work presented in Semantic SenseLab (Samwald *et al.*, 2010). The research builds an integrated framework on the web by mapping SenseLab ontologies to others such as BAMS, SAO, common anatomy reference ontology (CARO), gene ontology (GO) and ontology of biomedical investigations (OBI).

Sometimes scientists integrate ontologies databases or atlases to achieve their goals. The work in (Martone *et al.*, 2008), a part of Cell Centered Database (CCBD) that is a database for integrating cell level imaging data, tries to integrate data captured by different techniques and resolutions.

Similarly, the work described in (Marenco *et al.*, 2007) also uses ontology to map between databases and allow interoperability. Another research that used ontologies for mapping databases and ontologies is (Verbeeck *et al.*, 2010).

Ontology can also be used in works such as Bezgin et al. (2009), which maps and matches the spatial data in atlases with ontological brain regions. This research aims at connecting the macaque atlas to the collection of connectivity data on macaque brain (CoCoMac DB) (Bakker and Wachtler, 2012), and uses the ontology provided by the CoCoMac in order to accomplish this goal.

Sometimes scientists integrate several resources together in order to answer a very important research question. Hastings et al. (2012) brings many ontologies, such as mental functional ontology, chemical entities of biological interest ontology, protein ontology, gene ontology and NIF ontologies together in order to answer a single important research question. However, the approach can be used to answer other questions too.

Sharing information in neuroscience domains such as neuroimaging is the way forward according to Poline et al. (2012). Therefore, some use integration for sharing the data. For example, the Labis (Prodanov, 2011) project presents an integrated information system that enables information sharing with other applications. Mautner et al. (2015) is another one that uses ontology in data sharing.

Temal et al. (2008) create an ontology for sharing medical images and regions of interest (ROIs) called OntoNeuroBase ontology. NeuroLOG (Gibaud *et al.*, 2011) uses the OntoNeuroLOG in order to integrate neuroimaging repositories.

Some research, such as that one done by Subbarao *et al.* (2010) builds ontologies in order to facilitate integration or use pre-existing ontologies for integrating resources. Some research maps different data with the aid of ontology in order to infer new information. For example, Liu *et al.* (2010) creates ontologies called ElectroMagnetic ontologies to achieve a meta-analysis between different studies with event-related potentials data.

With the same goal in mind, Patel et al. (2016) uses the gene ontology for investigating gene relations with imaging phenotypes. Another research is (Kasabov, Jain and Benuskova, 2008), which seeks out to map brain structures to genes by creating a brain-gene ontology (BGO).

The study explained in Nichols *et al.*, (2014) creates mappings between different terminologies and ontologies in neuroscience based on anatomical structures from the foundational model of anatomy (FMA) ontology. In this approach, the foundational model of anatomy ontology works as a hub that connects or glues several terminologies and ontologies together. By doing this, it allows integration and ontology reuse as part of its outcome.

Some other research in which integration is a part of their approach includes (Schofield *et al.*, 2013), (Köhler *et al.*, 2013), (Refolo *et al.*, 2012), RegenBase (Callahan *et al.*, 2016), BrainFrame (Barnes and Shaw, 2009), (McCray, Trevvett and Frost, 2014), (Burns and Turner, 2013), SchizConnect (Wang *et al.*, 2016), (Ježek and Mou, 2011) and NeuroBase (Barillot *et al.*, 2006).

As will be seen later, the third study of this research used the integration capability of ontologies to create a module in the model it presents. It is called the question resolving module and brings ontologies and third-party resources together to work as a question resolution platform.

2.3.4 Collaboration

Ontology can be used to facilitate collaboration among scientists and projects. Ontology is capable of doing so due to its characteristics discussed earlier in this document, namely being standardized, formalized and externalized.

Researchers use ontologies for collaboration purposes such as performing parts of a research in different labs or collaborative research on a specific subject. The Pain Research Forum (Das *et al.*, 2014) created a forum for chronic pain researchers, which provides a community for collaborative work among scientists. Their work is reusable and can be used for creating other similar forums.

Another example would be Channelpedia. This work, which is described in Ranjan *et al.* (2011), is a platform for storing ion channels and collaboration among scientists, also uses ontology for creating its platform.

2.3.5 Classification and Categorization

Classification and categorization are other applications of ontologies. Classification is one of the most common problems in science. Using ontologies gives us the ability to generate new classifications of the concepts through single or multiple sets of properties (Hamilton *et al.*, 2012).

Ontologies can be used for various types of classifications, including classifying parts and organs related to the nervous system such as neurons and their types. For example, BlastNeuron (Wan *et al.*, 2015) uses the neuron morphology ontology for classification of neuron types.

Similarly, the neurodegenerative disease phenotype ontology (NDPO) has been used in Maynard *et al.* (2013) in order to group and compare phenotypes. Kassahun *et al.* (2014) use ontology to automatically classify epilepsy types using machine learning techniques.

Gene ontology is the most used ontology for classification purposes; especially, for the type discussed above. For example, it was used for classification in research such as Patel *et al.* (2016), which uses gene ontology for grouping genes.

Ontologies can classify mechanisms related to the nervous system and neuroscience. The aim of Laird et al. (2015) was to create paradigm classification based on cognitive ontologies for face perception and studying how humans recognize faces.

Ontologies can be used to track changes regarding diseases. One of the uses of common Alzheimer's disease ontology (Refolo *et al.*, 2012) is to create a classification system that allows comparative integration and analysis of Alzheimer's disease research portfolios.

Some other research that used ontology for classification includes Turner et al. (2013), which used the CogPo ontology for classification. Furthermore, classification was one of the goals for creating BAMS ontology (Bota, 2008) too. Other examples are (Lenartowicz *et al.*, 2010) and (Li *et al.*, 2007).

2.3.6 Disambiguation

Disambiguation is very important for researchers, scientists and computers, because having a clear understanding of concepts is vital to do research. It is the same for any computer system: if the information is ambiguous, it will not be machine understandable. Ontologies can be used for disambiguation means by acting like dictionaries; also by providing additional information.

For example, NeuroNames (Bowden and Dubach, 2003) is an ontology that is used for disambiguating neuroanatomical nomenclature. It contains the name of all neuroanatomical entities and structural concepts, and has been used in numerous projects such as the BrainInfo (Bowden *et al.*, 2012). It has a default vocabulary that includes names, synonyms and homonyms in eight different languages.

An example for disambiguating via ontologies is Francken and Slors (2014), which uses ontology for disambiguating common-sense cognitive concepts (CCCs). These are concepts used in everyday life for describing, predicting and interpretation of the everyday behaviour in cognitive neuroscience.

Ontologies can be adopted and used as data dictionaries by showing synonyms, meanings or definitions of the terms. They can even be used for unit conversions (Marenco *et al.*, 2007). This is useful for resolving ambiguous terms in a query, to correctly map two different synonyms, or to make a multi-language search possible.

Examples of using ontology as a data dictionary are Wang *et al.* (2013), which uses NeuroLex for creating a data dictionary regarding a project on schizophrenia; the MNI data sharing ecosystem (Das *et al.*, 2016) also uses ontology as data dictionary in its platforms; (Kennedy *et al.*, 2012) and (Marenco *et al.*, 2007).

Ontologies can assist in user input in situations such as working with interfaces. They do this by controlling or recommending correct terms to users. Examples of ontology being used for user inputs are the WhiteText project (French *et al.*, 2015) and (Marenco *et al.*, 2008), which uses a keyword-based input interface for NIF.

2.3.7 Knowledge Management and Organization

Ontologies can be used in knowledge and resource management; organization of internal or external resources and coordinating resources; procedures and attempts. Here a short description of each of them will be discussed.

Ontologies are useful in knowledge management. So far, one of the best well-known approaches towards knowledge management in neuroscience has been the NIF approach, described by Maryann (2011).

NIF was created to find and integrate neuroscience resources by creating a semantic framework. It is consisted of three different parts, including NIF registry, which is an annotated list of resources related to neuroscience using NIF vocabulary; NIF literature, which is a text index consisting of PubMed-derived index; and NIF database federation, which allows searches for annotated resources.

There are other research that are active in knowledge management, such as BrainFrame (Barnes and Shaw, 2009), SchizConnect (Wang *et al.*, 2016), Neuroviisas (Schmitt and Eipert, 2012), Brain Architectural Knowledge Management System (BAMS) ontology (Bota, 2008) and the project discussed in Rubin *et al.* (2009), which follows a knowledge based approach to assist in organizing, managing and accessing neuroscience information.

Ontologies are useful in resource management too. They can be used in application level and for organization of application modules, or at higher levels, such as managing coordination among scientists and resources to achieve a certain goal.

Ontologies can assist in managing and organizing either internal or external resources. For example, Brede Tools (Nielsen, 2014) use ontology for internal organization of their database and a wiki that assists in organizing topics, brain regions, journals and people. On the other hand, Marenco et al. (2007) use ontology as a map that points towards external resources that are needed.

Ontologies can work as coordinating tools. International Alzheimer's Disease Research Portfolio (IADRP) (Liggins *et al.*, 2014) utilizes the common Alzheimer's Disease Research Ontology (CADRO) (Refolo *et al.*, 2012) for coordinating funds, strategies resources for Alzheimer's disease and avoiding duplications of efforts.

2.3.8 Data Analysis

Ontologies can be useful in working with data and analysing it. For example, Djamanakova *et al.* (2014) uses ontology for creating a systematic reduction system for reducing noise in the data.

Ontologies such as gene ontology (GO) can be used for semantic analysis. An example is (Wiese *et al.*, 2012), which used gene ontology to investigate the relation of a protein and a mutation with a type of brain tumour in children.

2.4 Discussion

Different applications of ontologies were described in the result section. In this section, the aim is to direct the attention of readers to an interpretation of results, summary and some outcomes of this study.

Information provided in this section will demonstrate effects of applications of ontologies in neuroscience. Similar research to this study will be discussed. Furthermore, different applications are mapped together as the final part of the synthesis. Contributions will be discussed in this section too.

This study shed light on the application of ontologies and assisted in having a proper understanding and overview on ontologies and their applications in neuroscience. This understanding and knowledge was made possible through the systematic review presented and its results.

As demonstrated in the results, ontologies can be used for designing search and question answering systems. Furthermore, they can be utilized in data and text mining. They can even be used to find and connect symptoms to brain diseases and disorders and vice versa.

Generally, systems that benefit from ontologies in a part of their approach for search and data retrieval, will be able to produce more detailed and accurate results. However, their speed might be lower than similar approaches.

They can be used for shaping and enhancing queries by tasks such as query formulation, query translation and query expansion. They are also used for creating queries and navigation in the data in faceted search systems. Indexing systems can use ontologies as a part of their semantic based approach.

Processes and tasks allowed by using ontologies help improve neuroscience research and resources. For example, reasoning assists in enhancing neuroscience ontologies by allowing the inference of new information. It can be used to check and correct potential mistakes in ontologies and their entries.

Ontologies can be used in representing and capturing neuroscience data such as experimental data and protocols; instrumental information; subject data that includes neuroimaging and multimedia data and information regarding the patient; and data related to the nervous system and disease information.

They can also be used for reporting clinical research, annotation of neuroscience data, anatomic segmentation and labelling, representing brain information in atlases, visualization and modelling the information.

The other use of ontology is integration. Ontology can map and connect ontologies to each other or to other types of information containers such as databases or brain atlases. The outcome can then be used to answer questions or infer new information. Sharing the information is another outcome of ontology integration.

Ontology can enhance collaboration between both scientists and projects. This way, scientists can share or break an experiment or a neuroscience project among themselves and do it as teams.

Classification studies can utilize ontologies too. Ontologies can assist in classifying parts shaping the nervous system, grouping and comparing nervous system parts, classifying mechanisms regarding the brain and tracking changes in the nervous system.

Ontologies can be used for disambiguation purposes. Using ontologies, disambiguation can be used in circumstances such as resolving a misunderstanding between scientists, resolving uncertainties while answering a research question or resolving disambiguation for users. These are usually done by resolving concepts, unit conversions, synonyms and definitions of neuroscience terms.

Scientists can use ontologies for resource and knowledge management; also, for organization of resources, procedures and scientific attempts. They can be used for coordinating funds, strategies and efforts.

The last application of ontologies discussed in this study was data analysis. Ontologies are useful in manipulating and analysing the data. They can also be used in semantic analysis of disease and tasks related to the nervous system.

Despite the advantages of ontologies and their applications, there is a high cost in implementing a well-designed ontology. Also, a poorly designed ontology might have a negative effect on another part of a system such as disambiguation processes.

Querying ontologies could be very costly too. This is because of the high cost of calculations, especially in querying large ontologies. Also, ontologies are stored in different formats and there is not a unified format for storing ontologies. This leads to problems while querying them.

However, more views on ontologies and their application and in general, their usefulness, will be presented in the conclusion chapter, when other studies of this research have been concluded and a realization of ontology effectiveness and use in practice has been shaped.

2.4.1 Related Works

This section discusses studies similar to the one performed in this chapter. There were very few research papers that had specifically reviewed ontologies or application of ontologies in neuroscience. However, in order to have a better benchmark to compare this study with, research papers on this topic in biomedical science were taken into account too.

Bodenreider (2008) is a paper that has discussed application of ontologies in biomedical science. It argued the biomedical ontologies application in detail and discussed the use of ontologies in knowledge management, data integration and exchange and finally, decision support in biomedical science.

However, it does not consider any differences between ontologies, terminologies or controlled vocabularies. Moreover, as mentioned, it concerns the domain of biomedical science and not neuroscience. This paper limited its search criteria to recent journal papers only.

Larson and Martone (2009) talked about ontologies, along with advantages and disadvantages and applications of ontologies in information management. They saw ontologies as an answer to the problem of heterogeneity and dispersity of neuroscience data. Challenges ahead, such as data integration and some other uses of ontologies were discussed in this paper too.

Another research is the work presented in Rubin and Napel (2010). This research focused on images rather than ontology. It proposed ontologies as a new means of describing radiology reports and images. It also limited its scope to articles in PubMed only.

It is worth mentioning that the above research namely focused on applications such as data dictionaries, structured representation of image content and image retrieval for decision support.

Horrocks (2013), who presented a well written but brief paper on ontologies, reviewed ontologies in general. The paper also discussed their advantages and disadvantages, and then gave examples of applications that ontology can be useful in.

A study that incorporated a systematic literature review is Mohamad Hazawawi et al. (2015). It tried to explore how neuroscience reflects in information system studies and other related fields.

However, it only discusses primary studies from a few specific journals in a limited five-year period using a simple search method, and its results are more towards a systematic mapping study.

Hoehndorf, Gkoutos and Schofield (2016) present another related work, which offers a review on datamining methods using ontologies. According to this research, applications of ontologies include using ontologies as graph structures, finding semantic similarities, rule-mining or finding patterns and correlations, improving performance of clustering algorithms, text mining and using them as formalized theories.

All in all, according to the above information, it can be said that there have been some shortcomings regarding the number of reviews, methods of review studies and limitations in the scope and focus of the previous research in the field.

Rarely have detailed research papers been performed on the relation of ontologies and neuroscience. Furthermore, few major research papers used a systematic review protocol towards reviewing ontologies in biomedical science or neuroscience. Therefore, the field needed vast, rigorous research to cover and show applications of ontologies in neuroscience. Table 2-1 shows a comparison of this study versus similar existing studies.

Research paper	Field of study	Main Focus	Used systematic method or not	Protocol described clearly	Period Covered	Resources covered
Bodenreider (2008)	Biomedical science	Ontology Applications	×	×	Recent papers	Journal papers
Larson and Martone (2009)	Neuroscience	Data Management	×	×	- (not available)	-
Rubin and Napel (2010)	Neuroscience	Neuroimages	×	×	-	PubMed articles
Horrocks (2013)	General	Ontology Applications	×	×	-	-
Mohamad Hazawawi et al. (2015)	Neuroscience	Information Systems	✓	✓	5 years	Few journals

Hoehndorf, Gkoutos and Schofield (2016	Biology and Biomedicine	Ontology applications in datamining	×	×	-	-
This study	Neuroscienc e	Ontology Applications	✓	✓	All papers until 2016 (updated to 2019)	All papers in 8 Major databases , famous journals

Table 2-1- A comparison of this study and other existing studies

2.4.2 The Future of Ontologies in Neuroscience

Until now, useful information regarding ontologies and their applications have been presented and discussed. Based on the information given so far, this section seeks to speculate a potential future overview of ontologies in neuroscience.

The usefulness of ontologies has been pointed out as early as the start of 17^{th} century (Lorhard and Uckelman, 1606). In fact, neuroscientists have been using them for a long time now. This research was interested in computational ontologies, which are expressed in a way so they can be used by machines (computers), which in return allows computer automation.

These types of ontologies in biology and medical science were recognized by the famous work of Ashburner et al. (2000) which introduced the Gene Ontology (GO). Therefore, despite the process of searching for documents in this research did not envisaging any limitations on the year of the primary studies, the starting point was the year 2000.

Figure 2-3 demonstrates the number of ontology applications in neuroscience per year. As can be seen from the bar chart, the application of computational ontologies started from year 2000, but major studies using ontologies were rarely performed in neuroscience.

There are few research papers prior to the year 2006 that use ontologies for different purposes. However, it was in 2006 that the game really heated up, as ontologies became popular and were used multiple times in neuroscience for different goals, such as integration and information retrieval.

As can be seen from the bar chart, ontology application has had a steady increase in neuroscience over time. This was interesting and somewhat opposite to the common perception, including of the author of this research.

In fact, based on the trends shown by the research papers, and according to the instructions proposed by research organizations such as the International Neuroscience Coordination Framework (INCF) that encourages the application of ontologies, it seems the domain of ontology applications is going to become broader.

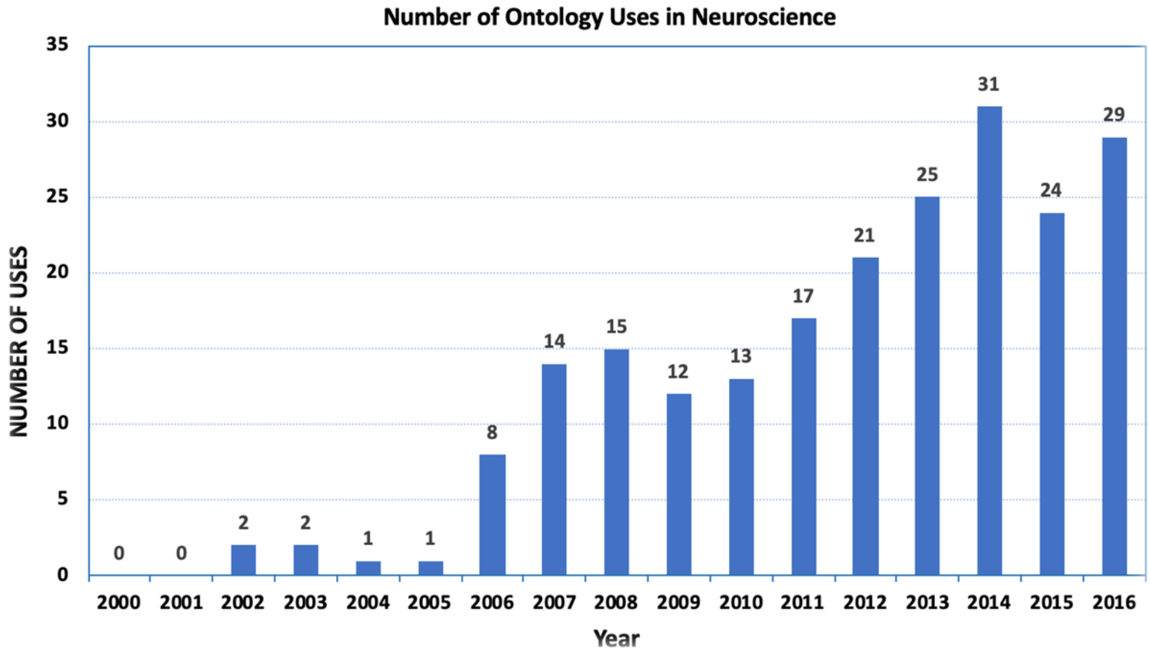

Figure 2-3- Number of applications per year

Ontology is, in fact, one of the symbols of the semantic era in computer science. It was introduced as a means of the semantic web and semantic computing, and gained popularity as a method of adding extra metadata and information to the data.

It changed the unstructured data to structured data in a way that data became self-explanatory and made sense by itself. This way, machines could understand the data and perform sophisticated automated tasks on it.

Ontologies have been an important aspect of the semantic computer science, and the case is not any different for neuroinformatics. As discussed earlier, ontologies that rose from philosophy have been used by different fields in the past, including neuroscience.

Nonetheless, the trend starting with the gene ontology was possibly a turning point for neuroinformatics. Therefore, “What lies in the path of ontologies in neuroinformatics from now on?”, “How would or should they evolve?” are some of the questions that come to mind.

It is a very efficient and effective tool for neuroscientists to infer information organizations, such as classifications or to perform data analysis. The gene ontology is still the most popular in information capturing among neuroscientists. Gene ontology dedicated almost one-third of all ontology applications to itself. The applications of the gene ontology were discussed in the result section.

In order to have semantic applications in neuroinformatics, the approaches should be semantic and automated as well. This means that either they should be created based on semantic tools such as ontologies; or techniques used in them, such as machine learning, should adopt a similar tool.

Although the starting point of the semantic era in neuroinformatics was marked by the gene ontology, and the application of ontologies in neuroinformatics has grown significantly in the recent years, more effort is needed to achieve the full potential of ontologies in neuroscience.

Ontologies will be of more value and assist scientists more when, in addition to providing a comprehensive definition of the domain concepts and their relations, they can also provide information on details of processes in neuroscience.

In other words, it can be claimed that ontologies will reach their full potential once the applications of ontologies such as the cognitive atlas ontology in performing tasks, become as popular as ontologies used in disambiguation and classification such as gene ontology.

Only when ontologies such as the cognitive atlas (Poldrack *et al.*, 2011) become widely useful and popular and get as many citations as ontologies such as gene ontology, it can be said that systems in neuroscience are truly ‘semantic’.

2.4.3 Future Work

Here, the potential future directions of this study are discussed. Keeping this study updated can be a future direction, since this study is a very detailed and comprehensive investigation both in the number of reviewed research papers and the approach towards reviewing them.

To the best of my knowledge, this is the most comprehensive study on the applications of ontologies in neuroscience. In fact, mentioning all future directions of this research can be a small separate chapter on its own because of the huge number of potential research directions this study has provided. Moreover, a great number of theories, applications and attempts were discussed in this study. Many of them can be a starting point for a researcher in neuroscience or neuroinformatics to start a new study. This is especially because, this study has listed important research papers in each category. These research papers can be a significant starting point. A researcher can start by first reading the studies listed for each application, and then continue with the research journey.

For example, using this study, scientists can pick single or multiple applications of ontologies in which they are interested and initiate a new investigation. This is what the author of this study has done in the next two studies (chapters) of this research (thesis). The next two studies in this research used this approach. They can be seen as future directions for this study, because ontology applications discussed in this study will be used in them.

These applications include classification and categorization, disambiguation and data representation in the second (next) study. Furthermore, the third study in this research will use all applications mentioned for the second study alongside the question answering application of ontologies.

Reviewing research papers in biomedical science regarding the application of ontologies, it was emphasised that neuroscience has close boundaries with biomedical science. Therefore, results found in this study should be valid at least in biomedical science, if not in other related domains. However, making sure of this might require further investigations.

Another future direction for this research could be to study the applications of ontologies in neuroscience knowledge management, since a review of the Results section demonstrates that ontologies can assist in almost every aspect of a knowledge management framework.

That is because some research such as BrainFrame (Barnes and Shaw, 2009), Rubin and Napel (2010), NeuroVIISAS (Schmitt and Eipert, 2012), SchizConnect (Wang *et al.*, 2016) and others did not limit their use of ontologies to only one application and benefited from them on multiple levels. Brain Architectural knowledge Management System (BAMS) provides an ontology that considers knowledge management.

Bratsas et al. (2009) discusses knowledge management goals in neuroscience as a framework that can capture and model all aspects of neuroscience research, such as knowledge representation, knowledge retrieval, knowledge sharing and knowledge discovery.

Ontologies can assist in achieving all the above goals. Therefore, it seems that knowledge management, along with the application of ontologies in neuroscience, can be a fruitful future direction. Such a study will be beneficial for many projects in neuroscience.

2.4.4 Limitations of the Study

Despite being the most comprehensive study as mentioned above, this study has some limitations. Most limitations were discussed in detail in the methods section. However, here we summarize the limitations again.

The resources of this study have included several reputable databases, such as two medical databases, three informatics databases and two multidomain databases. Moreover, three high impact and reputable journals were investigated. More databases and journals can be added to the resource pool of this study in future.

No time limitations were applied on the databases. This means that their records were searched from their beginnings up to the time of creating this document. However, the journals were searched after 2010, as this was when their entries were more up to date.

The other limitation that was discussed in the methods too, is the risk of bias with this study. Because all stages of the study were performed by a single researcher, there might be a risk of bias.

However, this risk was minimized by constantly consulting different experts from related disciplines including computer scientists, a neuroscientist and a librarian for setting up and organizing the study methods and performing it.

2.5 Summary of the Study

This study was designed to answer the first research question of this thesis and to investigate the applications of ontologies in neuroscience. In order to answer this question, first a review was done on ontologies and their history, mathematical foundations, characterizations and capabilities.

Also, the reasons for using ontologies in neuroscience and why they suit this domain were discussed. After that, systematic literature review protocols were introduced, and a rigorous study was designed based on them to find out ontology applications in neuroscience.

As a result of this study, eight major application groups were pointed out which were accompanied by multiple other applications. These applications included a diverse list of search and data retrieval, data capture and representation, integration, collaboration, classification and categorization, disambiguation, resource management and data analysis.

In the Discussion section, results were further explained. Studies similar to this study were mentioned and compared with the study performed in this chapter. After that, future work and limitations of the study were pointed out.

3 Analysing and Representing Neuroscience Questions and Classifying Them Using Ontologies

As discussed in the previous study, ontologies can be used for data representation and classification because of their characteristics. In this study, the aim is to further discuss these uses, as well as represent and classify neuroscience data - more specifically, questions using ontologies.

While tackling different questions in neuroscience, an automated system or even a researcher should be able to distinguish between different types of questions in order to be able to suggest appropriate answers for them; tackle questions in a systematic manner; and in some instances, avoid working on repetitive questions due to misinterpretation.

The reason for the above statement is because each question has different parts and neuroscientists ask questions at many levels from simple queries to complex, multi-word questions, using different terminologies and vocabularies. Therefore, considering the range of questions as a spectrum, the overall complexity will increase from one side to the other, and this calls for a system capable of distinguishing between different types of questions.

According to the above discussion, a representation and classification of questions seems necessary. Consequently, the second research question, which is "How can questions in neuroscience be represented and classified using ontologies?" is proposed.

Resolving this question is not only valuable by itself, since it enables scientists to represent and categorise questions, but it also can lead to proposing answer types and answers for questions; designing interfaces for searching; and systems designated for resolving questions, also known as question answering systems.

Ontologies are one of the many tools and techniques that can be used to develop and enhance question classification. While tackling questions, a system can benefit from ontologies at different levels, and up to different scales. As Baharudin et al. (2010) and Tenenboim et al. (2008) discuss, ontologies can assist question classification systems.

Being able to represent questions is very important, since the quality of breaking down and representing questions has a direct effect on question classification and therefore, resolving them. Moreover, by doing this, questions would be simplified which is important for finding answers for them (Chali, Hasan and Imam, 2012; Roberts *et al.*, 2014).

Question classification is a subcategory of short-text classification, which itself is a subcategory of text classification. It is important that there are systems solely designed for tackling and performing question classification.

Question classification is a vital part of the question analysis and processing module of systems for resolving questions, since it reveals information about the question and its answer types by assigning categories (classes) to them, and paves the way for other modules of systems for resolving questions (Athenikos and Han, 2010; Tomás and Vicedo, 2012; Tsatsaronis *et al.*, 2015).

If the use of ontologies in tackling questions is thought of as a spectrum, minimal usage would be to use ontologies as a terminology in checking synonyms, and maximal usage would be to use ontologies in question classification, sophisticated query expansion and pattern matching (McGuinness, 2004). This study will discuss some additional usages of ontologies such as lemmatization.

A complete review of multiple ontology uses, which include reasoning, question answering, resource organization, semantic annotation, classification, communication platform and several other uses can be found in the literature review in chapter 2.

To achieve the objective of this study, three different tasks have been performed- a question analysis approach for finding question dimensions that can be used to represent questions; and two approaches in classifying questions, including ontology-based and statistical approaches.

First, a question analysis process is performed. This is to analyse the neuroscience questions and find appropriate dimensions for them. Dimensions (attributes or features) can represent the questions. The reasons for using them will be discussed later in the Methods section and further, in the Discussion section alongside the implications of using them.

Then, an ontology-based question classification approach in neuroscience is presented, which uses some modules of a neuroscience ontology (NIFSTD) to enhance the accuracy of the classification. This part of the experiment demonstrates the extent to which neuroscience ontologies can assist in neuroscience question classification, and how ontologies can assist question classification with a minimal dataset (questions).

A question classification based on supervised statistical techniques (machine learning techniques) is presented as the third task in this chapter towards question classification. This is done as an additional task, only to get an idea of how ontology-based classification performs. Three different sets of experiments, using four different classifiers, have been executed for this reason.

Performing different classification approaches and using multiple techniques allows this research to depict a profound view of question classification in neuroscience using ontologies, and how ontologies can assist question classification systems to accomplish better results.

In the remainder of this chapter, first a background on question classification is given, including related works to this study. Then, methods will discuss how each of the tasks in this study will be performed in order to achieve the objectives of the study. The result section will discuss the output of each task, and finally, the discussion section will analyse the findings of this chapter and what they say about question classification. It will also draw some conclusions based on the analysis of the results.

3.1 Background

In this section, first an overview of the concepts discussed in this study, such as question classification and related works, are presented. In addition, the role of ontologies in question classification is discussed to further demonstrate the link between the classification and ontologies.

Before starting the main discussion, it is a good idea to define the difference between query and question, both of which are used in this chapter. There is a slight difference between these two terms.

A query usually implies a precise probing of a dataset, like *"Retrieve the value of field X."*, while a question is more general and unstructured such as *"What is the relationship between Alzheimer's disease and insomnia?"*.

Text classification is to assign one or more categories (classes) to the text, based on the content (Sanchez-Pi, Martí and Garcia, 2015), and it can be used for different purposes, such as question classification (Tomás and Vicedo, 2012), which is the focus of this study; spam filtering (Delany, Buckley and Greene, 2012); web page classification (Song *et al.*, 2005); personalized news selection (Tenenboim, Shapira and Shoval, 2008); information filtering (Sriram *et al.*, 2010); and other purposes.

As mentioned in the introduction of the chapter, before starting to classify questions, it is better to analyse, break down (decompose) and represent questions in a way that is simpler and can be defined by certain rules (Roberts *et al.*, 2014). In fact, this is so important that it is a part of the question classification process (Moreda *et al.*, 2011).

Question classification is the task of assigning categories to questions (Yu, Sable and Zhu, 2005; Tomás and Vicedo, 2012). It is very important since as discussed above, question classification is a vital part of systems for resolving questions, and directly affects the performance of these systems.

Question classification frameworks have many benefits, such as improving efficiency and effectiveness of search systems, ensuring that search systems; are progressing and not just working repetitively on the same type of the questions; assisting scientists with asking valid and valuable questions; and simplifying the identification of proper answers (Ely *et al.*, 2002; Yu, Sable and Zhu, 2005), since they enable systems resolving questions to narrow down the answer and remove irrelevant results (Yu *et al.*, 2010).

Early works in using ontologies for improving text classification mostly used WordNet (Sanchez-Pi, Martí and Garcia, 2015). Wordnet (Fellbaum, 2010) is like an encyclopaedia and includes information about the brain, but the information is not very comprehensive, and it cannot replace domain ontologies. Recently, ontology-based classification is being applied to many domains, and domain ontologies are used more often.

Recently, ontologies have been used in classification studies as a source of well-organized information, which provides adequate semantics between the concepts. In addition, neuroscientists have created many detailed and vast ontologies covering different sections of the neuroscience domain.

Two important ontologies in the domain of neuroscience are the Neuroanatomical domain of the Foundational Model of Anatomy, known as NeuroFMA (Nichols *et al.*, 2014) and the Neuroscience Information Framework Standard ontology, known as NIFSTD (Bug *et al.*, 2008; Fahim T. Imam *et al.*, 2011).

The NeuroFMA is a sub-category of a larger ontology called Foundational Model of Anatomy (FMA) that provides information about anatomy. NIFSTD is a very large ontology, which itself adapted or imported many other ontologies and taxonomies, such as the Sub-Cellular Anatomy Ontology, known as SAO (Larson *et al.*, 2007) and Birnlex (Bug *et al.*, 2008).

However, these are not the only ontologies in the neuroscience domain. There are other ontologies in both application and domain ontologies (described in the first study) too. For example, one of the ontologies that used modules or subsections of different ontologies for a specific cause is FMA-Radlex (Mejino, Rubin and Brinkley, 2008), which is an application ontology of radiological anatomy.

Three steps of creating a question classification system are question representation, classifier construction and classifier evaluation. As will be seen later, these steps are the foundation of the three stages for each of the two question classification approaches discussed in the Methods section.

As discussed in the first study of this research (thesis), one of the ontology applications is their use in classification. They can assist with text and question classification too. Some issues and shortcomings in statistical classification methods, which will be discussed later, can be taken care of by using ontologies (Song *et al.*, 2005).

Generally, classification can be performed in one of the three ways- supervised, semi-supervised and unsupervised. Supervised techniques include Rocchio’s algorithm, K-nearest neighbour (KNNs), support vector machines (SVMs), decision tree, neural networks, latent semantic analysis, fuzzy correlation, genetic algorithms and Bayesian algorithms (Korde and Mahender, 2012; Safae, Habib and Abderrahim, 2018). Some of these techniques are used in this study, and their description will be given later in the section related to Statistical Approach.

Machine learning techniques have some shortcomings in text classification (Sanchez-Pi, Martí and Garcia, 2015; Safae, Habib and Abderrahim, 2018). Training data must be large in order for the statistical techniques to be able to train the classifier; most traditional techniques used in statistical analysis have not considered the semantic relationships between the words, making it harder to improve the accuracy of these techniques; and finally, there are issues arising when language changes.

Studies such as (Sriram *et al.*, 2010; Yu *et al.*, 2010; Tomás and Vicedo, 2012; Sanchez-Pi, Martí and Garcia, 2015) demonstrate the importance and effectiveness of using domain taxonomy or ontology in text classification.

An ontology-based classifier increases the accuracy by using the domain knowledge that may be contained in each ontology (Sanchez-Pi, Martí and Garcia, 2015); and as Tomás and Vicedo, (2012) discuss, in question classification, semantic information is more useful than syntactic information.

Two other problems that occur in statistical techniques are polysemy and synonymy. Polysemy occurs when a word has multiple meanings and synonymy happens when multiple words have the same meaning (Safae, Habib and Abderrahim, 2018).

These two problems can be avoided to a great extent by using domain ontologies. Ontologies provide different synonyms of the words that can be used in resolving the synonymy; also, while using the domain ontology, words are restricted to that specific domain. Therefore, words are resolved and interpreted into their domain-specific meaning.

In this study, the NIFSTD is used for inferring information in question processing and classifying them in the ontology-based approach as will be described further in the following section. Possible uses of NeuroFMA are also discussed.

Figure 3-1 demonstrates the focus and scope of this study, and how ontologies are applied into the study, which help in understanding the overall view of the domain of this research.

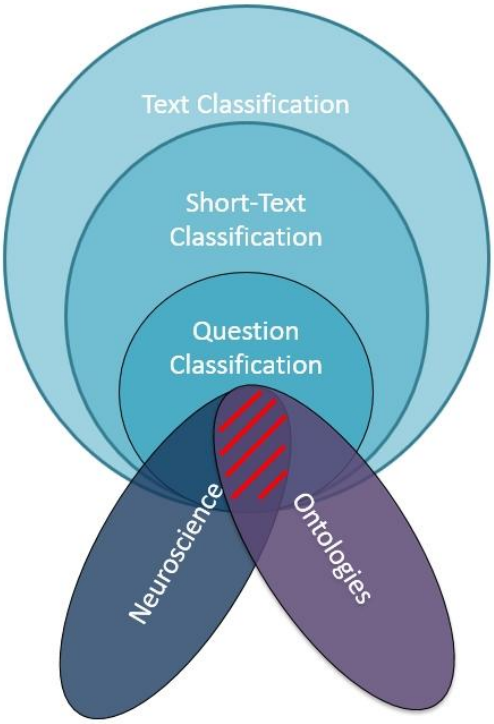

Figure 3-1- An overview of the domain of this study

3.2 Methods

This section describes how different tasks in this study are shaped in order to achieve a representation and classification of neuroscience questions. Processes involved in creating each of these are described in detail in order to set a strong foundation for the Results section and tests performed in that section.

Methods of this study consist of a set of tasks for gathering data (data collection) and a set of tasks for data analysis. The data collection explains why and how the questions were selected, and describes different approaches used in question representation and classification. Data analysis consists of three different methods, including one related to finding question dimensions and two others related to question classification approaches.

The first part of the method involves gathering a combination of different questions from two different resources and then performing an analysis on them. Questions posed by domain experts were used as the first resource of this study since in neuroscience, users of computer-based systems are mostly scientists and experts. Domain experts were selected based on their expertise and their eagerness to participate in this research. Two of the domain experts were members of the research panel.

In the question set, questions 5 to 13 and 26 to 32 are from the first team of domain experts from the University of Melbourne; questions 15 to 25 are from the second team of domain experts from Monash University. It was asked from the experts to propose questions that are as different as possible.

As the second resource, questions were gathered from literature published by high standard research teams and credible publishers, as these types of questions are of interest to neuroscientists, important, up-to-date, and worth studying. Questions 1 and 14 are taken from Imam *et al.* (2012); questions 2 to 4 are taken from Turner *et al.* (2010). After creating the question set, all questions were validated by an expert. Please see a list of questions below:

1. What are the synonyms of hippocampus?
2. Which parts of the precentral gyrus are active in schizophrenia and healthy volunteers in these data?
3. Which parts of Brodmann Area 6 are activated in Schizophrenia and Healthy Volunteers?
4. Do the schizophrenia subjects show more or less co-activation than healthy subjects do, in cortical regions that are connected by the superior longitudinal fasciculus?
5. What is the white matter volume of the regions of the inferior parietal lobule in the data?
6. What is the average anatomical region of the precentral gyrus in the brain in different ontologies?
7. What is the precentral gyrus?
8. What is the relationship of telencephalon and diencephalon?
9. Is the left amygdala part of the amygdala?
10. Which regions overlap with the precentral gyrus?
11. How many subjects gained extra volume in a region of the diencephalon in the patient's data?
12. Which annotated structural region has changed volume in patients?
13. What are the positive effects of aspirin on the brains of elderly people?
14. What is the anatomical region of the premotor cortex?
15. What is the thickness of inferior parietal lobule in the data?
16. What are the brain regions with atrophy in Huntington's disease (HD).
17. Which sub-cortical region has the largest atrophy in HD?
18. Which cortical region has the highest amount of cortical thinning?
19. Which regions are functionally connected to the most atrophied region in HD?
20. Which regions are structurally connected to the most atrophied region in HD?
21. Which cortical regions show changes in functional activity in HD?
22. Which regions show early atrophy in pre manifest HD?
23. Which cortical regions show thinning in pre manifest HD?
24. Which regions show thinning/atrophy over 12 months in HD?
25. Does any region show increased activity in pre manifest HD?
26. How many regions are in the parietal cortex?
27. Is the thalamus atrophic in the Multiple Sclerosis patients?
28. How does connectivity between the thalamus and motor cortex differ between patients and controls?
29. Find all the subjects with an average cortical thickness less than 5mm.
30. What is the activation level in the anterior cingulate gyrus?
31. Is the degree of change in volume of the thalamus over 1 year in patients greater than the change in controls?
32. Is cortical thickness correlated across all cortical regions in this subject group?

As pointed out before, tasks regarding the data analysis in this study are performed using three different methods. One is the question representation and analysis. From the other two methods, which are approaches toward question classification, one is ontology-based and the other one uses statistical methods (machine learning).

Please note that throughout this study where multiple tasks were needed, tasks and approaches have been broken down and divided into smaller segments called stages. Correspondingly, each stage might consist of a number of steps.

The first method, which performs question analysis suggests a comprehensive set of dimensions for question representation in neuroscience, which leads to a question hierarchy. The question hierarchy is created based on the number, formation and plausibility of different dimensions in a question. This is because each dimension is represented by a Resource Description Framework (RDF) triple. Each RDF triple adds to the complexity of the data that is being retrieved; therefore, levels are created based on the number of dimensions.

Moreover, the way a valid question (Ely *et al.,* 2002) is posed might not allow some dimensions, such as the conditional ones, to be present without other dimensions. In other words, such a question is not plausible. With regard to the formation and plausibility, the hierarchy was informed by Bloom's Taxonomy (Anderson, Krathwohl and Bloom, 2001), Web Ontology Language (OWL) definition and classes, and to some extents, Aristotelian categories and relationships explained in (Ryle, 1937; Studtmann, 2018). Bloom's Taxonomy and valid questions are discussed later in this chapter; OWL is discussed in Appendix 2 of this thesis.

The question hierarchy developed using these methods was corroborated by eight experts on 21/3/2014, in a confirmation seminar at the University of Melbourne; also presented as a poster (Eshghishargh *et al.*, 2015) in a Frontiers in Neuroinformatics event back in 2015; and in a paper (Eshghishargh *et al.*, 2018), which won the best student paper prize of the 11^{th} Australasian conference in 2018.

Dimensions were also used in question classification approaches, and in the next study. Different types of question classification approaches and reasons for them were covered in the Background section. Here, an ontology-based classification (categorization) was done, following a set of statistical approaches. This was in order to investigate the advantages and disadvantages of each of the approaches in the presence of neuroscience domain information in the shape of ontology and curated question representation.

Moreover, as pointed out and explained before, a major goal of this research is to test the application of domain ontologies in neuroscience. In this study, the effect of ontology on question classification is another aspect of the investigation.

The next reason was to examine neuroscience ontologies themselves, and to determine their coverage of the field, and seek future augmentations to ontologies in this domain, as well as how to expand terminologies and ontologies in order to make them more useful in neuroscience question classification. Figure 3-2 demonstrates a high-level view of methods and approaches, along with their different stages.

Ontologies were used in a semi-automatic manner in this study. Information was fed to codes and scripts, answers were retrieved and then, if required, they were fed manually to the other processes in other stages. The overall process has the potential to become fully automated by connecting the processes, but this was not necessary for the proof of concept that this thesis took.

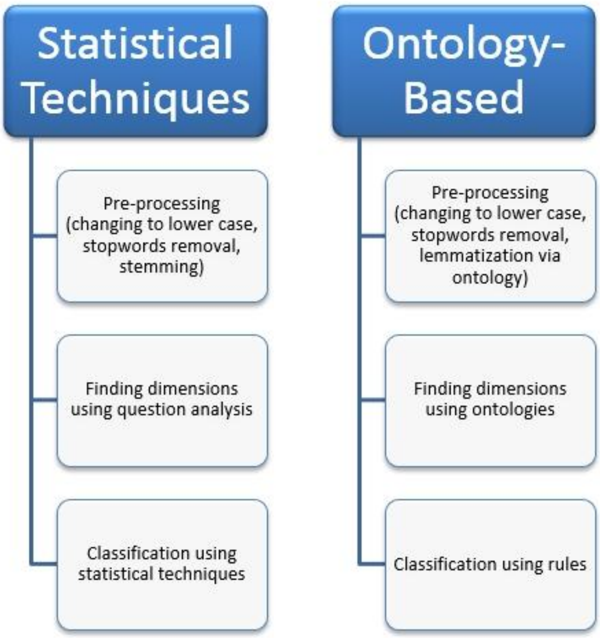

Figure 3-2- The approaches toward question classification

3.2.1 Question Analysis

The first method of the analysis in the method section involves gathering questions and analysing them in order to figure a representation for questions. For this means first, 32 questions were gathered and then analysed as follows.

In order to perform question representation and finding dimensions, half of the questions were parsed manually and individually deconstructed to their language building blocks (tokens). During tokenization, which is to deconstruct the text into tokens, words that were meaningful together, were left in one token.

Only the first half of the question set was used in order to avoid any biased dimensions selection. If the whole set had been used while testing dimensions for coverage ability, one could have argued that dimensions were comprehensive in representing questions, since they have been shaped based on the whole dataset. Thus, results might have become biased, and they might not be applicable to a different set of questions.

In addition, words were changed to their primary form. For example, both 'activation' and 'activated' changed to 'active'. This is called stemming. Consequently, stop-words (accessible from http://www.ranks.nl/stopwords), such as 'what', 'is' and 'the', were removed. Stop-words are prepositions and conjunctions that are seen in texts without being related to a specific topic.

Then, tokens were visually inspected and placed into separate clusters according to their role in questions. If a new token did not match any previous cluster, a new cluster was assigned to it. This process continued until no other token remained to be fitted in a cluster, a state known as theoretical saturation (Bowen, 2008).

After that, clusters were named according to their tokens and were referred to as dimensions (features). The results will be covered in Question Dimensions and Hierarchy part of the Results section.

These dimensions were then used to represent questions and were used in question classification approaches later. As will be seen further on, they have been used as coarse-grained (Li and Roth, 2006) categories (classes). In the ontology-based approach, ontology terms, which are fine-grained or detailed categories, were mapped to dimensions. Because, as Dridan and Baldwin (2007) and Sanchez-Pi, Martí and Garcia (2015) state, using fine-grained classes and categories will have a negative effect on the accuracy. This will be discussed in detail in the Discussion section.

3.2.2 Ontology-Based Approach

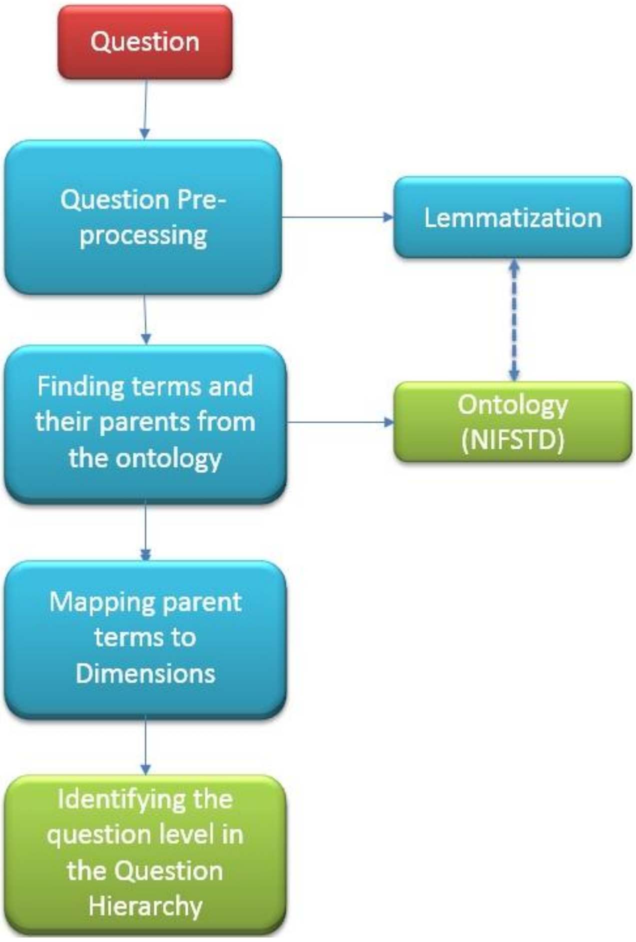

Figure 3-3- Ontology-Based classification diagram

The first approach towards classification and categorization of the questions in this study was the ontology-based classification. In this approach, the focus was to use neuroscience domain ontologies with a partially rule-based (Haris and Omar, 2012; Asghar *et al.*, 2017) mechanism in question classification and finding dimensions in questions as demonstrated in Figure 3-3.

As stated before, modules of the NIFSTD have been used as the domain ontology. This ontology was searched and used for finding appropriate dimensions using the National Centre for Biomedical Ontology (NCBO) BioPortal (Salvadores and Horridge, 2012).

First, questions were tokenized and stopwords were removed. Then the lemmatization was done according to the NIFSTD ontology. Lemmatization (Sanchez-Pi, Martí and Garcia, 2015) is to map the term to its original form, which is called lemma based on a vocabulary or dictionary. If needed, a neuroscientist was consulted to resolve potential ambiguates.

Despite stemming and lemmatization appearing to be the same, they are different in the way they operate. As demonstrated before, stemming usually cuts the end of the words to get to their original form. Therefore, there is a chance that it does not create a valid term.

For example, the word ‘saw’ might be returned ‘s’ by some stemmers; whereas, lemmatization will return ‘see’ or ‘saw‘, depending on the role of the word (verb or noun). Therefore, lemmatization is usually more accurate.

Second, an ontology-based search for terms in the question and their parent nodes was performed. According to Baharudin et al. (2010), a well-designed question classification approach, considers the domain. The approach in this section follows this indication, as it uses ontologies in performing a classification.

Two different approaches could be used for searching the ontology, and both of them could work. Either ontologies could be parsed first to select the appropriate terms and store them in separate files to later be matched against questions, or ontologies could be parsed and searched at the time of matching tokens with terms.

Here, the latter was used. In this method, and in order to create an ontology-based method for classifying the question, modules of the NIFSTD ontology were searched, and if there was a match, meaning that $token(i) = term(t)$, then the ancestor node of the *term (t)* was found.

Please note that here, the ancestor node of a term, meant the highest parent node in the branch of ontology related to that term. The parent node is the term right above the term in the ontology. As an example, Figure 3-4 demonstrates the branch related to the *superior longitudinal fasciculus*, a term used in question number 4 in the NIFSTD ontology. Here, n*erve fasciculus* is the parent node and *anatomical entity*, the ancestor node. It is worth mentioning that where coding and queries was needed in this chapter, SPARQL was used.

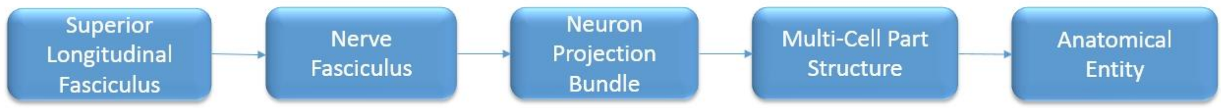

Figure 3-4- The hierarchy of Superior Longitudinal Fasciculus in NIFSTD ontology

Sometimes more than one match was found for a term. In these cases, the suitable term was chosen by using the information provided by the ontology; and if that did not work, then an expert was consulted. Some rules were also devised for a few repetitive words and dimensions.

In the next stages, ontology terms were mapped to dimensions in a quasi-logical (Crosswhite, 2010; Svačinová, 2019) manner, and the level of each question was identified in the question hierarchy, according to its dimensions. The percentage of questions that were classified correctly was then calculated as the overall ontology-based classification result.

3.2.3 Statistical Approach

The statistical approach used four different supervised statistical techniques, including Naïve-Bayes, SVM, KNN and Random Forrest, introduced in the Background section, and the approach was performed on a dataset of 16 and 32 questions, except the upper bound tests, which were performed on a dimensions set instead of questions based on its definition, which will be provided later in this section.

Naïve-Bayes is a technique that applies the Bayes theorem to each term individually in order to find the category for that term. SVMs are based on a hypothesis that tries to find the lowest true error. KNN is used to determine the classification based on the similarity of two categories. Decision trees, such as Random Forrest, try to rebuild the manual categorization by building true-false trees.

For each test, all named statistical techniques were used. The Weka (Frank *et al.*, 2005) application has been used as the classification platform. Weka provides a collection of machine learning algorithms and data processing tools. The classifiers used in Weka were respectively Naïve-Bayes, SMO (SVM), Lazy-IBK (KNN) and Random Forrest.

Document representation (Safae, Habib and Abderrahim, 2018) is one of the ways of reducing the complexity of documents. There are too many words in a set of questions. Therefore, it is better to minimize the number of words.

Dimensionality reduction can assist in achieving this goal. Dimensionality reduction (Safae, Habib and Abderrahim, 2018), which consists of feature extraction and selection, can assist in creating an efficient system. The following two stages are dedicated to feature extraction and selection, and as mentioned, each stage might consist of several different steps.

The first stage described feature extraction. Feature extraction consists of tokenization, lower case conversion, stemming and lemmatization. Tokenization has been defined before in the Question Analysis section.

Stemming (Uysal and Gunal, 2014) is used to get the different variations of a word into its root (stem) form. For example, applying the Snowball stemmer on 'volume', 'subparts' and 'precentral' will result in 'volum', 'subpart' and 'precentr'. As can be seen in 'volum' and 'precentr', stemming might affect some words. This will be discussed later in more detail.

The first step was to feed the questions to Weka in its own format (arff). Then, lower case conversion, tokenization and stemming were done automatically using Weka tools. StringToWordVector filter was used for tokenization and the Snowball stemmer was used for the stemming process.

The second stage involves feature selection (Feng *et al.*, 2012), which is the task of selecting appropriate dimensions from the training dataset and using them in text classification. Feature selection techniques can be divided into wrappers, filters and embedded methods (Uysal and Gunal, 2012).

Wrappers use learning algorithms along with search algorithms that feature evaluation functions, while filters use scoring frameworks which are independent of learning algorithms. Wrappers are usually costly and consider feature dependencies, whistle filters are the opposite. Embedded methods introduce the feature selection in the training phase.

Weka has a tool for finding features, but as will be seen, features suggested by the tool did not result in classifications with high accuracy. Therefore, in this stage, the dimensions found in the Question Dimensions and Hierarchy section, which were a result of an accurate question analysis and representation of questions, were used.

The third stage which was the core of the statistical experiment, consisted of performing all four listed classification techniques, using the dimensions inferred in the Question Dimensions and Hierarchy section as a result of question analysis.

Leave-one-out cross-validation is a subtype of (K-fold) cross-validation that uses one instance from the dataset as the validation set (test dataset) and other instances as the training dataset. In this technique, the classifier runs on the dataset for '$n-1$' number of times, where 'n' is the number of instances in the dataset. Therefore, all questions will be used as the training dataset and test dataset automatically in turns so there is no need to identify training and test datasets.

As discussed before, and will be seen later, the ontology-based approach does not need a big dataset to be effective in classifying tasks. However, in statistical techniques, usually, a bigger dataset results in better results. The leave-one-out cross-validation, along with other delicate details such as pre-processing and using suitable tools, namely suitable stemmers, assist in compensating for a larger dataset.

The classifier uses the training dataset to learn how to classify, and it uses the test dataset to try to classify based on the method it has learnt. While performing cross-validation, there is no need to identify the training and test datasets manually, as the classifier automatically divides the dataset into two subsets of train and test.

At this stage, the two values of baseline and upper-bound were calculated using the Weka application. At the first step, questions were classified using no dimension. This was considered as the baseline, as no strategy has been taken towards using dimensions. The baseline can be seen as random classification, and is the simplest possible prediction with no additional information or data manipulation. It provides a basic result, which can be checked later against results of other algorithms and approaches to see if techniques used in them made any improvements or not.

The second step included a classification with no questions. The classification was done only on the dimensions resulting from the first stage. This resulted in an upper bound for classifications; since instead of questions, dimensions that were the representation of the questions were used.

In the third step, classification was done on half of the dataset, and in the final step, it was performed on the whole dataset. The reason behind this was to demonstrate the effects of the size of the dataset on statistical approaches outcomes.

3.3 Results

In this section, first the question set described as the data collection phase of the method is presented in the Set of Neuroscience Questions section. Then in Question Dimensions, and Hierarchy the question analysis that led to dimensions is discussed.

After that, the results of the ontology-based categorization are presented. Next, different approaches of the statistical classification are performed, and their results are presented in the Statistical Approach section.

3.3.1 Question Analysis

As discussed in the Methods section, the first thing to do was to gather questions (data collection). As explained in the methods section, questions were gathered by experts and via scanning neuroscience papers from respected publications.

All questions were examined by one of the domain experts to assure that they were scientifically correct and could be validated by a neuroscientist. Questions such as Q13 might be a general question to a neuroscientist. However, it is a plausible question when the integration of these data for all types of users including health specialists in other fields, is considered.

Please note that logical operands such as ‘and’ and ‘or’ were divided into separate sub-questions, where they only glued two different questions together. For example, the question “What are the synonyms for amygdala and precentral gyrus?” can be seen as two separate questions: “What is the synonym for amygdala?” and “What is the synonym for precentral gyrus?”. On the other hand, phrases that add a condition to the question that might affect the result have been considered, especially because they add to the complexity of questions.

It has been assumed that the user poses valid questions. Furthermore, the difference between a valid and poor (bad) question has been discussed in Ely *et al.* (2002). For instance, “What is causing her anemia?” is an example of a poor question.

As the first step towards finding dimensions in questions, 14 questions were parsed manually and deconstructed to tokens as described in the Question Analysis, stopwords were omitted, and words were changed to their initial form. A result of this phase for the sample question list would be:

1. synonym/ hippocampus
2. parts/ precentral gyrus/ active/ schizophrenia/ healthy volunteers/ data
3. parts/ brodmann area 6/ activate/ schizophrenia/ healthy volunteers
4. schizophrenia/ less/ more/ activate/ healthy volunteers/ cortical regions/ connect/superior longitudinal fascicules
5. white matter/ volume/ subpart/ inferior parietal lobule/ data
6. most/ common/ anatomical location/ precentral gyrus/ ontology
7. description/ precentral gyrus
8. relationship/ telencephalon/ diencephalon
9. left amygdala/ part/ amygdala
10. regions/ overlap/ precentral gyrus
11. how many/ subject/ extra/ volume/ subpart/diencephalon/ data
12. structural data/ changed/ patient
13. positive/ effect/ aspirin/ brains/ elderly/ people
14. subclass/ premotor cortex

In the second step, tokens were placed into clusters based on their nature and similarities until theoretical saturation (Bowen, 2008) was reached. This means that no token was left without being placed in a cluster. Tokens were assigned to clusters and when faced with a new type of token that did not fall under existing clusters, a new cluster was assigned (as discussed earlier).

For example, while parsing Q2, 'data' was one of the tokens that could not be characterized under existing clusters from Q1. Therefore, a new cluster would be assigned to it. Parsing other questions, other data references such as 'schizophrenia' and 'healthy volunteers' would be placed under this new cluster. Repeating tokens in individual clusters were ignored. A summary of clusters is as follows:

- Cluster 1: synonyms, volume, description
- Cluster 2: hippocampus, part, precentral gyrus, brodmann area 6, cortical region, superior longitudinal fasciculus, white matter, sub-parts, inferior parietal lobule, telencephalon, diencephalon, left amygdala, amygdala, aspirin, brain, premotor.
- Cluster 3: activated, anatomical location, relationship, part of, overlap, effects
- Cluster 4: schizophrenia, healthy volunteers, data, ontologies, subjects, patients, people, structural data
- Cluster 5: less, more, most common, how many, change, positive
- Cluster 6: connected, inside, elderly

Furthermore, similar clusters were merged. For example, cluster 1 and 3 were both related to entities. Appropriate names were given to clusters, and subsequently referred to as dimensions (features).

Table 3-1 shows the list of dimensions which includes ***entities*** involved, such as hippocampus and precentral gyrus; ***domain-specific phrases*** that specify scientific processes or attributes such as curvature, activation, volume and thickness; ***aggregation or statistical phrases*** such as summary and total number; ***data references*** such as ontologies or data-files; and finally ***conditional phrases*** such as elderly and connected.

Dimension	Example
Entities	hippocampus, precentral gyrus
Domain-Specific Phrases	activation, curvature
Aggregation/ Statistical Phrase	summary, total
Data References	data, ontology
Conditional Phrases	elderly, connected

Table 3-1- Question Dimensions

Tokens such as 'changed', 'atrophy' or temporal tokens discussed changes in the data and multiple inputs or readings were needed to resolve them, possibly over a fraction of time. Therefore, they were grouped under the 'aggregation or statistical phrases' dimension. However, they could be grouped under a separate dimension to avoid confusion and complications.

As demonstrated above, dimensions have been used to represent questions in this study. Therefore, a hierarchy of questions was created based on the number of dimensions involved in a question, which in turn increases the complexity of the question.

Please also keep in mind that later in this research, dimensions will be mapped to concepts from the ontology, which are stored as RDF triples, and SPARQL will be used to query them similar to the process performed in studies such as (Lopez, Pasin and Motta, 2005; Lopez *et al.*, 2007). Moreover, as Pérez et al. (2009) reveals, SPARQL is a PSPACE-complete language; therefore, the ranking of dimensions in the question hierarchy does not make differences on querying complexity. The question hierarchy is as follows:

- Level 0: Questions with only entities in them.
- Level 1: Level 0 plus domain-specific phrases.
- Level 2: Level 1 plus references to data.
- Level 3: Level 2 plus aggregation/ statistical phrases.
- Level 4: level 3 plus conditions, comparisons or changes.

Table 3-2 demonstrates which questions belongs to which level:

Table 2-2 Questions and their levels in question hierarchy

Question	**Level in Hierarchy**
7. What is the precentral gyrus?	0

1. What are the synonyms of hippocampus? 8. What is the relationship of telencephalon and diencephalon? 9. Is the left amygdala part of the amygdala? 10. Which regions overlap with the precentral gyrus? 14. What is the anatomical region of the premotor cortex? 30. What is the activation level in the anterior cingulate gyrus?	1
2. Which parts of the precentral gyrus are active in schizophrenia and healthy volunteers in these data? 3. Which parts of Brodmann Area 6 are activated in Schizophrenia and Healthy Volunteers? 5. What is the white matter volume of the regions of the inferior parietal lobule in the data? 15. What is the thickness of inferior parietal lobule in the data? 16. What are the brain regions with atrophy in Huntington's disease (HD).	2
17. Which sub-cortical region has the largest atrophy in HD? 18. Which cortical region has the highest amount of cortical thinning? 19. Which regions are functionally connected to the most atrophied region in HD? 20. Which regions are structurally connected to the most atrophied region in HD? 26. How many regions are in the parietal cortex? 29. Find all the subjects with an average cortical thickness less than 5mm.	3
4. Do the schizophrenia subjects show more or less co-activation than healthy subjects do, in cortical regions that are connected by the superior longitudinal fasciculus? 6. What is the average anatomical region of the precentral gyrus in the brain in different ontologies? 11. How many subjects gained extra volume in a region of the diencephalon in the patient's data? 12. Which annotated structural region has changed volume in patients? 13. What are the positive effects of aspirin on the brains of elderly people? 21. Which cortical regions show changes in functional activity in HD? 22. Which regions show early atrophy in pre manifest HD? 23. Which cortical regions show thinning in pre manifest HD? 24. Which regions show thinning/atrophy over 12 months in HD? 25. Does any region show increased activity in pre manifest HD? 27. Is the thalamus atrophic in the Multiple Sclerosis patients? 28. How does connectivity between the thalamus and motor cortex differ between patients and controls? 31. Is the degree of change in volume of the thalamus over 1 year in patients greater than the change in controls? 32. Is cortical thickness correlated across all cortical regions in this subject group?	4

3.3.2 Ontology-Based Approach

In this part of the study, the ontology-based question classification approach is performed. The focus is on creating a classification of questions and categorizing them according to the question hierarchy, using the NIFSTD ontology.

The pre-processing step, including the lemmatization, was performed on questions. Also, ambiguities were resolved through information from the NIFSTD ontology and consulting a neuroscience expert.

The NIFSTD ontology comprises different modules. Therefore, different modules from the NIF ontology, such as NIF Dysfunction, which contains diseases information; NIF Anatomy, which comprises brain segments; and NIF Function, which contains functions of the brain were used in this study.

At the beginning of the classification process, pre-processing was performed; questions were tokenized, and lemmatization was done using the domain ontology (NIFSTD in here). The process of searching for terms in ontology and tagging them using dimensions was then started. The process for finding parents of a term (node) in the ontology can be represented via an algorithm or code for automated implementation as below.

```
SELECT ?directSub ?super

 WHERE { ?directSub rdfs:subClassOf ?super .

       FILTER NOT EXISTS {

         ?directSub rdfs:subClassOf ?otherSub .

         FILTER (?otherSub != ?directSub)

  }}
```

Some of the terms in the questions were multiple word terms. There were many ways to search for them in the ontology. One way was to consider white spaces as separation of tokens. While searching and in order to avoid any incorrect findings, whenever a token was found, its following tokens were then also added to it and searched for as a second search.

If the two were in the ontology as one term, then they were counted as a single token. For example, during the tokenization, the ‘superior longitudinal fasciculus’ in question 4 was broken into three tokens of ‘superior’, ‘longitudinal’ and ‘fasciculus’.

When ‘superior’ is found in the ontology, ‘longitudinal’ is added to it and searched for. Again, this was found in the ontology; therefore, the next token, which was ‘fasciculus’, was added to the combination and tested again. Since these three were in the ontology as a single term, the single term was used.

There are different methods available for searching terms in the ontology. The method above, is also known as string matching. String matching could be performed using different methods such as using personalized code, SPARQL string matching functions and third party applications such as Lucene (McCandless, Hatcher and Gospodnetic, 2010). This study used the second method.

Top-level terms or ancestry nodes in the ontology were selected as the dimensions for their sub-class terms, using the code from the last page. This way, questions could be matched with the dimensions and as a result, question hierarchy. The results of mapping for the sample question, “Which part of the precentral gyrus is active in these data?”, to ontology terms, based on the procedures described until now, can be seen in Figure 3-5.

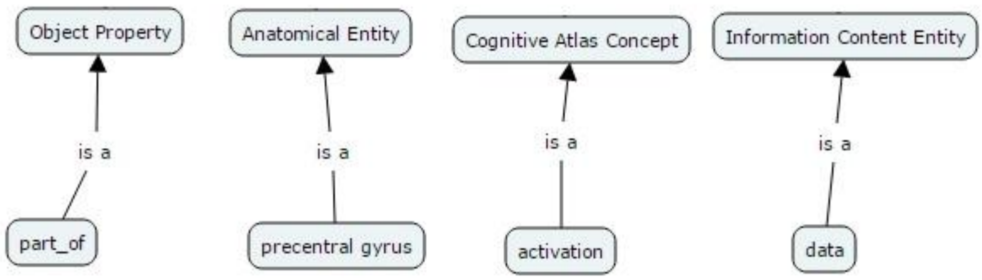

Figure 3-5- Classification of the sample question according to the NIFSTD ontology

Having ontology terms mapped to dimensions, avoids the accuracy issue discussed in the Methods section and allows categorizing questions according to the question hierarchy. As will be seen in the next study, the latter assists with resolving questions too.

Therefore, as discussed in the Method section, ontology terms were mapped to dimensions in a quasi-logical manner. For example, anatomical entities, material entities and spatio-temporal regions from the ontology were mapped to the *entity* dimension; quality and object property were mapped to the *domain-specific* dimension; processual and increased object quality were mapped to *change* and *aggregation* dimensions, and information content and disease mapped to *data* dimension.

Then, and based on the dimensions and the question hierarchy, the levels of the questions were identified. For example, the sample question used above, was a level 2 question. The categorization using the NIFSTD ontology, resulted in the classifier achieving 93% correctness, which as will be seen later, is better than all statistical based techniques. Of course, it is worth mentioning that the ontology-based classification, needed expert consultation for devising rules more than the statistical based methods, which almost did not need expert consultation at all.

3.3.3 Statistical Approach

In this approach, the aim was to create statistical based classifiers and measure their classification accuracy. The information regarding the classifiers was given in the Background and Methods sections.

The first stage was the feature extraction stage. At the first step, questions were fed to the Weka in its specific format, along with their dimensions and hierarchy levels inferred in the previous section. Questions were then tokenized using the 'StringToWordVector' filter in the Weka. Also, tasks such as tokenization and stop-word removal were done via Weka automatic tools. Stemming was done using the Snowball stemmer.

After the feature extraction, there is a need to select features. Weka provides a tool for selecting features, and using them resulted mostly in very low classification percentages. Therefore, the dimensions resulting from the question analysis were used.

In the next step, each classifier was applied on the dataset without using the dimensions to calculate a baseline. As can be seen from Table 3-3, Naïve-Bayes and SVM classifiers scored the highest percentages in classifications with 43.75% correct classification.

During the second step, classification was done without the questions and only on the dimensions as models of questions, which resulted in 96% correct classification. This means if the questions were supposed to be ranked manually, 96% of the classifications were correct and there is an upper bound of 96%. While classifying with the dimensions only, there is no need to use all four techniques, since they all result in the same percentage.

The third step consisted of running different classifiers on half of the questions with the dimensions found in the feature selection phase, which resulted in a significant shift in correct classifications with the Random Forrest technique, with respectively 52.32% and 77.27% correct classifications.

There are three different parameters for each classification including precision, recall and f-measure (Powers, 2011). Precision is to see how many instances are relevant; recall is to examine how many relevant instances were selected; and f-measure, which is a combination of both previous parameters, is calculated through the following formula: $f = 2.(precision + recall)/(precision * recall)$.

In the fourth step, classifications were done on the whole dataset. The Naïve-Bayes classifier scored 0.69 on precision, 0.55 on recall and 0.57 on f-measure. The SVM scored 0.72 on precision, 0.66 on recall and 0.66 on f-measure. KNN scored 0.76 on precision, 0.55 on recall and 0.52 on f-measure. Random Forrest scored 0.54 on precision, 0.62 on recall and 0.56 on f-measure.

As a summary, in this section, questions were analysed and dimensions (features) were inferred as representative of neuroscience questions, and a question hierarchy was formed based on their complexity.

Four different statistical techniques from different groups were tested for classifying neuroscience questions, and it was shown that the Random Forrest technique, which is a sub-type of decision trees had the best performance among statistical ones in classifying the question (data) set.

The statistical method was presented and its classification accuracy was calculated, which was 77.27% in the Random Forrest technique. As seen in the ontology-based classification, that categorization resulted in 93%, which exceeded the statistical technique by 15.73%.

	Naïve-Bayes (%)	SVM (%)	KNN (%)	R. Forrest (%)
No Dimensions (Baseline)	**43.75**	**43.75**	25	31.25
With Dimensions(16 Questions)	46.15	46.15	46.15	52.32
With Dimensions (32 Questions)	54.54	72.72	68.18	77.27
Upper Bound		**96**		

Table 3-3- Results of different QC experiments

3.4 Discussion

In this section, results of the experiments performed in this study are further discussed, and the goals described at the beginning of this chapter are revisited to inspect if experiments were able to achieve those goals, and to what extent they have been successful. Limitations and Future work of the study are also discussed.

The study began by bringing together a list of diverse questions from two reputable resources, including experts and high standard literature. The list of questions was then validated consulting a neuroscientist.

The first investigation of the study was the question analysis and representation of dimensions. As demonstrated, first, questions were analysed and organized as a list of dimensions.

Then, it was tested if dimensions were able to cover the questions and correctly represent them. After it was demonstrated that the dimensions could represent questions, the experiments regarding the classification of questions was initiated.

As for the first part of the approach, an ontology-based approach towards classification of the questions was implemented. Questions were parsed and tokenized, lemmatization was done, and the ontology entities were matched with the dimensions.

The ontology-based classification was then performed, and achieved 93% correct classification of questions, which is a high percentage. Therefore, one outcome of the ontology-based classification could be that ontologies can be very useful in creating classifiers.

After the ontology-based question classification, statistical classification was performed to be able to draw a better conclusion on the effectiveness of ontology-based classifications through comparing results of both methods of classification.

The statistical techniques results demonstrated that the ontology-based approach provided better results. As can be seen from Table 3-3, the statistical techniques despite being successful, could not achieve a better result than the ontology based one.

The best statistical technique result has been achieved while using the Random Forrest technique. Using this technique, the best result on question classification has been 77.27%. However, as mentioned, the ontology-based approach reached 93% correct classification.

It is worth mentioning again that experiments proved dimensions were beneficial for the classifications. As can be seen from Table 3-3, using them resulted in a better classification rate of at least 8.57% and up to 28.97%.

According to the information provided in the table, while dimensions were absent from the experiments, the best classification rate was 43.75% even while using Weka and its suggestions; whereas, in the presence of dimensions, classification rate reached up to 72.72%.

As demonstrated in the Results section, no training data were used in the ontology-based approach, and the NIFSTD ontology compensated the training process. Therefore, ontologies allow semi-supervised classifications to some extents.

However, using ontology-based classification has some disadvantages too. For example, term and ambiguity resolution might be needed, as it was done in this study by lemmatization techniques via the ontology.

It can be concluded from the results that domain ontologies are useful in lemmatization and term disambiguation, especially where comprehensive and rigorous ontologies had extensively described concepts and processes of field. Neuroscience would be a good example; however, it still lacks fully functional ontologies discussing neuroscientific processes.

One of the problems of ontology-based classification has been said to be decreased accuracy (Sanchez-Pi, Martí and Garcia, 2015); especially, when there are far too many classes and entities available as a result of using a large ontology, known as being too fine-grained (Van Zaanen, Pizzato and Moll??, 2005). Studies such as Dridan and Baldwin (2007) suggest that using ontology increases the accuracy on main categories, but decreases the overall accuracy.

However, as experiments in this chapter showed, at least in this study and with the introduction of dimensions alongside the ontology, this was not the case and use of ontology increased the accuracy of classification. Issues regarding the accuracy was avoided by matching the fine-grained ontology terms with the coarse-grained or abstract dimensions, inferred from question analysis.

Therefore, benefits of having a domain ontology that covers detailed information and leads to correct classification is preserved, while a decrease in classification accuracy is avoided. So, it can be said that ontology-based question classification and its effect on classification, heavily depends on the quality of the ontology and the method used for incorporating it.

Despite being very useful, using ontologies can lead to an increase of the computational cost without using techniques such as modularizing or as Shaw *et al.* (2011) explain, generating specific views from them. Generating views from ontologies is basically extracting desired parts from them. Moreover, designing an ontology of high quality is costly and time consuming. If the ontology is not designed properly, it will have a negative effect on the question classification.

Furthermore, despite being powerful in classifying questions, the process of using ontologies is very time consuming and costly due to the reasons described earlier and demanding domain expert consultations. On the other hand, statistical techniques are easy, since most of them are predesigned and ready to use, and do not need on going expert consultation. They are much faster too.

Therefore, using an ontology-based approach for classification of questions in neuroscience depends on the time and cost the researcher is willing to spend on the project. The researcher should choose between accuracy and higher percentage of correct classification or speed and ease of use, and decide if results of using domain ontologies justify the costs.

3.4.1 Related Works

In this section, a review on related research is given. Works in this area can be viewed and classified from different aspects including question classification, domain ontology and taxonomy application, question complexity and formation and their role in representing questions. Here we briefly review these three aspects.

From the question classification aspect, different research has been performed, including (Hacioglu and Ward, 2003; Zhang and Lee, 2003; Li and Roth, 2006; Magnini, Speranza and Kumar, 2009; Yu *et al.*, 2010; Haris and Omar, 2012; Tomás and Vicedo, 2012). These are discussed in the following sections according to their focus, approach and domain.

In the case of taxonomy and ontology application, using taxonomies for text, short text and question classification can be seen in several research papers such as (Hermjakob, 2001; Pasca and Harabagiu, 2001; Hacioglu and Ward, 2003; Yu, Sable and Zhu, 2005; Li and Roth, 2006; Dridan and Baldwin, 2007; Z. Liu *et al.*, 2010; Yu *et al.*, 2010; Chen, Jin and Shen, 2011; Yahya and Osman, 2011; Haris and Omar, 2012; Tomás and Vicedo, 2012), while a few classification research incorporated ontologies including (Song *et al.*, 2005; Fang *et al.*, 2007; Janik and Kochut, 2008; Tenenboim, Shapira and Shoval, 2008; Magnini, Speranza and Kumar, 2009; Sanchez-Pi, Martí and Garcia, 2015). From those ontology-based studies, two have focused on general question and text classification and the other four have used ontologies for classification of web documents, news and operational (occupational) health and safety. Hence, a majority of studies are related to general text and question classification, while a few are related to specific domains and on subjects other than question classification. It can be said that not much research has been done on question classification using ontologies in a specific domain.

In the case of domain related research in neuroscience, few papers are related to health and medicine text or question classification including (Yu, Sable and Zhu, 2005; Ribadas *et al.*, 2013). Almost no study focused on classifying neuroscience questions among the literature reviewed.

Another aspect was the question complexity. Almost no research was found to classify questions based on their complexity in a restricted domain. Although, a few systems for resolving questions, have discussed the complexity of questions including (Stevens *et al.*, 2003; Lopez *et al.*, 2007; Yahya and Osman, 2011; Haris and Omar, 2012; Veeraraghavan and Miller, 2014). However, the above research was not presented with much detail. For example, Stevens et al. (2003) had mentioned 'complex query formulation', but did not address the complexity of queries. Another research paper that dealt with complexity, used Resource Description Framework (RDF) triples as complexity elements, and focused on computational complexity is Aqualog (Lopez, Pasin and Motta, 2005; Lopez *et al.*, 2007). Nevertheless, that study did not focus on question classification.

Research such as (Yahya and Osman, 2011; Haris and Omar, 2012) used Bloom's taxonomy to represent questions that can be used to measure the level of cognitive capabilities among students via a rule-based approach. This taxonomy divides questions into six different categories of remember, understand, apply, analyse, evaluate and create. This taxonomy can be seen as a nice way for classifying questions; however, it does not seem to fit very well with neuroscience questions, since most questions in neuroscience can be mapped to level one and two of this taxonomy and other levels are very abstract. Still, studying the use of Bloom's taxonomy and its implications, can be a future direction of this study.

A research paper that categorized the complexity of the queries into four groups of simple, compound, specialized and complex, is Veeraraghavan & Miller (2014), where a simple query is about an 'individual' structure; a compound query refers to 'group of structures'; a specialized query is a refined version of simple or compound queries; and finally, a complex query is a combination of previous types. This approach is one of the first that categorizes the questions with respect to complexity, and tries to classify questions, but it does not necessarily achieve a detailed and comprehensive question classification for the purpose of using in automated question answering and other uses. It is mostly a way of categorizing questions for human users, while it does not seem to have a high level of accuracy either.

Yan and Tourangeau (2008) studies question response times from the cognitive psychology point of view. They considered the complexity of questions as a factor in response times and used number of clauses, number of words and question types as factors playing a role in question complexity.

Bu *et al.* (2010) suggests a taxonomy for representing general questions. The taxonomy presented by them provides six different categories for general questions. They then use an approach based on Markov logic networks (Richardson and Domingos, 2006) in order to assign questions to the taxonomy. This taxonomy is not suitable for our purpose since it is designed for general questions. However, as stated before, studying the implications of using this taxonomy and others, against the question hierarchy presented in this study, can be a future direction.

Another research paper that studies questions is Roberts *et al.*, (2014), which discusses annotating general medical questions and decomposing them. This research annotates questions using a set of arbitrary terms alongside UMLS (Unified Medical Language System).

Two other research papers that worked on representing questions, are (Chali, Hasan and Imam, 2012) which studied complex question decomposition; and (Saquete *et al.*, 2004), which tackled temporal questions. Both these systems worked with the question in the general domain.

Despite some activities on question classification, reviewing related work shows that a few question classification studies have incorporated ontologies and even fewer have used domain ontologies. Some studies have performed text and question classification in domains such as medicine and biomedical research, but no one focused on neuroscience.

The study in this chapter is unique, based on the wide and different approaches it used, the domain it was performed in, and the size of the ontology it took advantage of. The approach used in this paper, has an effect on most parts of a question answering system. In this research, question classification is performed based on a question hierarchy (question categorization) because of two reasons. First, as Ambert & Cohen (2012) states, one of the questions that neuroinformatics is seeking an answer for is to make sure that neuroscientists do not repeat themselves and are not answering the same sorts of questions again and again. Second, as Tsatsaronis & Balikas (2015) argues, the quality of an answer depends on the difficulty of the question. A hierarchy of questions can help addressing these issues.

3.4.2 Limitations and Future Work

Now that the results have been discussed to some extent, in this part, the limitations of the study, along with ways for rectifying them, will be explained. Furthermore, directions will be pointed out.

The dataset for this study was not very large. This was because different types and shapes of questions were more important for the purpose of this research than the number of questions, which might be important for a study solely focused on achieving better results using statistical techniques. Nonetheless, it might still be argued this had an effect on statistical methods performance in this study. However, the study compensated for the number of training dataset and maximized the accuracy of statistical classification techniques by several tactics, including using the leave-one-out cross-validation (K-fold) technique; using the stemmer, which resulted in the best outcome; and other optimization tasks available from Weka.

Moreover, different approaches in classification are emerging every day. However, testing all of them against the ontology-based approach presented here was outside the scope of this research, as most of them are towards advanced statistical natural language processing techniques.

Therefore, adding more questions to the question set through consulting experts and searching the literature, and then performing a follow up study with any type of supervised, semi-supervised or unsupervised techniques, or even hybrid techniques involving ontologies and statistical techniques, can be a future direction for this study.

The question hierarchy represented in this research was static. Therefore, there might be questions that do not fit into the descriptions of this question hierarchy and hence, remain unclassified. As will be seen in the next study, there are ways for overcoming this problem. However, a potential method for rectifying this issue is discussed here.

This issue might be resolved through assigning weights to dimensions and as a result to a question. Investigation of a question hierarchy based on weighted dimensions can be a future direction alongside a comparison of this technique, the current question hierarchy and using taxonomies discussed in the Related Works section.

Classification based on other factors can be another future direction. Sometimes implementing a good classification can help improve output of a system. Moreover, as described in chapter two, some research questions in neuroscience have been answered just by utilizing classifications. For example, one other useful category for question classification might be to classify questions based on the type of the user or scientist that poses questions. By understanding who is posing the question, tailored answers can be provided for each user. This is because it is not only the neuroscientists who pose neuroscience questions; other experts in health informatics, such as biologists or carers, might pose neuroscience questions too, and each of them uses their own terminologies. Even inside the neuroscience domain, a cognitive psychologist calls a part of the brain 'brodmann area IV', while a behavioural neuroscientist calls it 'primary motor cortex'. Therefore, having a classification that assigns questions to different types of users can help a lot in designing systems, such as web portals or question answering systems.

Another example for a useful classification might be to classify questions based on the resource they need to be answered with, because there are many different types of resources in neuroscience; including different modalities of neuro-images, text documents, lab experiment descriptions, trials of experiments, publications and other types of digital resources.

Therefore, a system can benefit from a classification system that assigns types of resources to a question by decreasing cost and time. Here, again, a well-designed ontology can be very beneficial by providing information regarding different resources. Samples of this type of ontology have been pointed out in the previous study, where ontology applications regarding resource management and data capturing were discussed.

As demonstrated, the question hierarchy presented in this study listed questions based on their elements in certain levels or categories. Since each level contains similar questions, one can use them in implementing an automated system that resolves questions according to these levels.

The dimension set discussed in this study can be seen as a terminology for annotating neuroscience questions. It can potentially be connected with or mapped to other ontologies too. However, this study did not explore this matter. Therefore, it did not claim a question terminology. A future direction can be to expand dimensions into other domains and investigate if it can work as a question terminology or not.

Despite NIFSTD being a major ontology in neuroscience, there are other ontologies in this field too. Ontologies such as FMA, which its neuroscience section called NeuroFMA, dedicated a massive body of information to neuroscience. Using these other ontologies in classification studies can be another future direction.

Creating lemmatization algorithms and protocols based on neuroscience ontologies can be another possible future direction for this thesis. As discussed, manual lemmatization is very time consuming and almost impossible for scientists without a neuroscience background. Therefore, an algorithm or protocol with the capability of performing automated lemmatization for the neuroscience domain and based on neuroscience ontologies could be very useful.

1.1 Summary of the Study

This study set out to answer the second research question and figure out how ontologies in neuroscience can assist in classifying questions. First, a background on classification and question representation was given.

In the method section, information regarding tasks and experiments in this study was given. The information included detailed descriptions of approaches towards question analysis, ontology-based classification and statistical classification.

In the Result section, first a question set was gathered. After that, the question analysis which led to the formation of question hierarchy, was performed. Following this step, both ontology-based and statistical-based approaches towards question classification in neuroscience were performed.

In the discussion section, circumstances of the outcomes and results of the experiments were discussed. Furthermore, limitations of the study and future work were outlined in detail.

4 Resolving Neuroscience Questions Using an Ontology-Based Method

In this study, the aim was to answer the third research question: "How can questions in neuroscience be answered using ontologies?" For this reason, several tasks were planned and performed. Outcomes of the previous chapters were used in this chapter, including techniques and research discussed regarding use of ontologies in answering questions in chapter 2, and the question representation and hierarchy proposed in chapter 3.

The importance of ontologies in neuroscience has been pointed out in chapter 2, and in chapter 3 it has been shown how domain ontologies in neuroscience can be used to classify and characterize neuroscience questions. Therefore, as a part of the overarching theme of this research, which is enhancement and improvements in tackling neuroscience questions by using ontologies, this study will focus on resolving neuroscience questions.

In this part of the research, multiple tasks and modules shape a multi-task process model (platform), which aims to answer questions in neuroscience. The resources for answering questions have been limited to neuroimages based on their importance which was mentioned in chapter 2. It will also be advised how to expand resources to other types of documents in the Discussion section.

Domain ontologies used in this chapter are the Neuroscience part of the Foundational Model of Anatomy (NeuroFMA) (Nichols *et al.*, 2014) ontology, Foundational Model of Neuroanatomy (FMN) (Nichols et al. 2011) ontology and Neuroscience Information Framework Standard (NIFSTD) (Bug *et al.*, 2008; Fahim T. Imam *et al.*, 2011) ontology. These resources will help in shaping the resource integration module, which will be discussed later. NIFSTD will be used to address question resolution, while NeuroFMA and FMN will be used for mapping the NIFSTD with Freesurfer (Fischl, 2012), which is a collection of tools for analysing neuroimaging data.

In the remainder of this chapter, a review of current approaches is given in the Background section, then the Methods section describes how this research was performed, and Results shows the outputs of this research. In the discussion, results will be further analysed, and goals and achievements of this research will be discussed.

4.1 Background

Multiple, diverse ontology uses were described in the literature review chapter. In this study a few of those were used including disambiguation, query translation and ontology application in resolving questions. Therefore, there is a need to expand on these specific uses here.

Ontology applications in resolving questions in the neuroscience domain can be studied from different aspects. In fact, question resolving approaches that use ontologies in neuroscience benefit from them in many ways.

Two of the approaches are prominent among others, including ones that use the ontology for integration of resources so that questions can be answered more accurately and comprehensively (integration-oriented approaches), and ones that use ontologies for resolving and answering questions (question-oriented approaches).

It is worth mentioning that borders between these two approaches are not solid and they have a close and almost bilateral relation. This is because while enhancing the semantics and information of the domain by integrating ontologies, a solid resource containing detailed information is being created. Therefore, questions can be answered with more detail and more accurately.

In integration-oriented approaches, the ontology can be used as a resource organizer in order to integrate disparate resources, or to map the pre-existing data to ontology and by doing this, letting the system to be viewed, reached and used by other services.

For example, as Ascoli (2012) states, a problem with finding answers for neuroscience questions is that simple questions might need several time consuming searches on different search systems. Therefore, finding a way to integrate these searches can be of great value.

Ontologies can assist in creating frameworks that bring resources together in order to answer the above issue. For example, NIF search (Gupta *et al.*, 2008; Marenco *et al.*, 2008; Jeffrey *et al.*, 2014) was created to integrate neuroscience resources and answer questions.

NIF performs searches on different types of documents and provides answers for both neuroscientists and biomedical researchers. Furthermore, any other integration initiative or database to ontology mapping and integration systems such as the system defined by Gupta et al. (2010) correspond to this type of approach.

In the question-oriented approach, the disambiguation role of ontologies discussed in chapter 2 can be used as a tool to disambiguate neuroscience terms including terms, related to the nervous system, experiments regarding them and even knowledge management in this field.

In this approach, some ontology-based question resolving systems such as AquaLog (Lopez, Pasin and Motta, 2005; Lopez *et al.*, 2007), use ontologies mostly for disambiguation or query expansion alongside integration for adding to question semantics.

These systems mostly use ontologies, lexicons or controlled vocabularies such as NIFSTD, NeuroLex (Fahim T. Imam *et al.*, 2011) and NeuroNames (Bowden *et al.*, 2012) to disambiguate the terms in a phrase and hence, resolve the query.

The resolution is usually done by matching ontological terms against keywords. While using ontologies, query expansion is applied to a part of a hierarchy of the ontology, and the result would be used for querying the data from single or multiple data sources (Gupta, Condit and Qian, 2010).

For example, the study presented by Ashish and Toga (2015) used ontologies for rewriting queries in a project cantered around Alzheimer's disease. The work described by Samwald *et al.*, (2009) presents a query platform for querying neurons using an ontology repository.

The BioDB system described by Gupta, Condit and Qian, (2010) that tried to answer queries in the biomedical field through bringing ontologies together is another fine example of ontology being used for integration and query answering.

There are some other research papers that while not completely related to this study, facilitate some sort of query answering or search and resolution in neuroscience. For example, Mautner, Novotný and Prokop (2015) use an ontology for bringing data together; sharing among scientists and labs; and searching the data.

Textpresso (Müller, Kenny and Sternberg, 2004; Müller *et al.*, 2008, 2018) which is one of the oldest and most updated projects, is a text mining system that performs searches on research papers using an ontology. A project similar to Textpresso is the work presented in (Li *et al.*, 2007), which trawls through research papers and indexes them according to its ontology.

WhiteText (French *et al.*, 2015) uses text-mining techniques alongside ontologies for the purposes of query expansion, receiving user inputs and finding synonyms. Finding synonyms and query expansion is done using the NIF gross anatomy ontology, while receiving user information is done using NeuroLex.

The other project worth mentioning is the (Malhotra *et al.*, 2014). In this project, the focus has been on creating an ontology for tackling Alzheimer's disease. However, the authors also demonstrated that their ontology supports semantic search.

RegenBase (Callahan *et al.*, 2016) is another project that has tried to bring together information regarding spinal cord injuries. It has used ontologies to represent and standardize experiments and reports related to this injury. It has demonstrated as an outcome of its efforts, the RegenBase can be used in answering queries too, although not in an automated manner.

Textrous (Chen *et al.*, 2013) is a web-based framework designed to extract biomedical semantic meaning based on user inputs. It uses some controlled vocabularies alongside natural language processing techniques.

The approaches discussed in this section can be extended to other domains too. Other research papers that present an overview and some ideas on the wider biomedical and web query federation include research papers such as (C., 2008; Cheung, Frost, *et al.*, 2009; Cheung, Lim, *et al.*, 2009; Marenco, Wang and Nadkarni, 2009).

Regardless of the orientation, every system that answers and resolves questions needs to understand or translate the questions it receives. For automated and computer-based systems, questions have to be translated to a form that machine can understand and process.

Please note that although this discussion is a part of the system interface, the focus is from the point of having the question onwards, not query or question formulation. Creating a complete interface for the model presented in this system either by free-text or keywords will be addressed in the discussion section. Again, here the focus is on processes that translate a question to machine language, sometimes known as question translation.

There are differences in how systems tackle this issue. While using ontologies, some try to shape a resource description framework (RDF) triple, which represent the questions and solve them based on that; this is also known as the template-based approach.

Some even receive the question as RDF triples from the user to make it easier. There are others that use natural language processing techniques in order to translate the questions into code instead of ontology application.

For example, there are some research papers such as (Abacha and Zweigenbaum, 2012), which worked on answering medical questions. They depend on the expressive power of SPARQL in defining different questions and try to translate a free-form question to SPARQL. As demonstrated in this paper SPARQL can cover translating a large scale of questions.

AskHermes (Cao *et al.*, 2011) is a system that is focused on resolving clinical questions. It categorizes resources such as abstracts, full text documents, guidelines and Wikipedia articles, and uses ontologies in resolving them.

Some research papers try to reduce questions to templates in a textual entailment process (Kouylekov *et al.*, 2006) and then resolve questions. An example of this type of research is the work performed in Ou et al. (2008).

Articles such as (Dietze and Schroeder, 2009; Athenikos and Han, 2010; Asiaee *et al.*, 2013) provide further information on different types of approaches towards resolving questions in a semantic way, and compare those methods.

This approach is beneficial since potentially many questions can be reduced to a certain form. For example, “Where is amygdala located?”, “Which part of the brain contains the amygdala?”, “Where in the brain is the amygdala?” and “Amygdala is part of which anatomic entity?” are essentially asking the same thing and can be represented by or reduced to the same query.

Therefore, as can be seen, the way a system models or translates a question can be very important and has its own advantages and disadvantages. Moreover, there are many different methods for doing so from purely using domain related or domain-independent ontologies to using natural language processing techniques.

4.2 Methods

This section discusses the methods regarding the design and implementation of different parts of this study. It explains how templates, integrated platform and query expansion unit (the main contributions of this study) were shaped. After discussing the process of building those parts, it explains how they come together, along with question classification from the previous study to make a question processing model in neuroscience.

A template-based approach (Dwivedi and Singh, 2013) is one that reduces and represents a question as a template. Templates mirror the internal structure of questions (Unger and Bühmann, 2012). They include some missing elements called slots, which are filled according to the question. After filling the slots, the template is queried against the data.

This study exploited the domain knowledge which is presented in the form of neuroscience ontologies. It benefited from the expressive power of ontologies to resolve questions according to a template-based process.

In the Introduction chapter, the importance of neuroimages in neuroscience studies was discussed. This study used them in its approach for resolving questions. This was done through a neuroimaging application called Freesurfer and using the NeuroFMA ontology.

As mentioned above, the tasks and processes created in this study shape a question processing model. These tasks include three major groups, including question preparation module, query expansion module and resource integration module. The question processing model is nothing but the organization of different modules and processes. It describes how modules are connected and how they operate. In other words, it puts different parts of this study in a package.

In the remainder of this section, the templates and how they were shaped are discussed; the chapter continues with a discussion regarding the query expansion module; then the resource integration platform (module) is discussed. Following the resource integration, it is explained how all modules come together to create the question processing model.

4.2.1 Templates

The initial thing to do was to change questions to a machine-understandable form. This study offered an efficient and flexible method by creating templates for translating a user question to a machine-understandable question.

This section shows how the model tackled the transition from question to a template. Templates were made according to the question hierarchy from the previous study. Therefore, it is better to revisit the questions hierarchy. It contained five different levels as below.

- Level 0: Questions with only entities in them.
- Level 1: Level 0 plus domain-specific phrases
- Level 2: Level 1 plus references to data
- Level 3: Level 2 plus aggregation/ statistical phrases
- Level 4: level 3 plus conditions, changes or comparison

In the previous study, it was demonstrated that question hierarchy covers questions. Therefore, building templates according to it should turn out fine in resolving questions. A template is created for each level. Furthermore, for level 4, that includes changes, comparisons and conditional questions, three different templates should have been created.

SPARQL and Python Language were used for creating templates; this was because SPARQL is a suitable language for querying ontologies, graphs and linked data. Python was used to add more flexibility to the SPARQL codes. Moreover, because SPARQL does not have the means to allow communication with the user. Also, some Python libraries like Pandas were used to manage the information from the Freesurfer which contained the subject neuroimaging data.

Please note that some formations, such as questions consisting of two domain-specific phrases and one entity, were not taken onto account. This was because as discussed in the previous study, they were seen as two different questions.

4.2.2 Query Expansion

While resolving questions, it was important to remove any sort of ambiguities first. Query expansion seeks to remove ambiguities. Moreover, if the user wants to receive more information, query expansion can help in increasing the recall, which is the number of results returned.

In the approach of this study, it was envisioned to give the user the option to select query expansion or not. This is a similar process to some systems where it is asked if the user wants to also seek for related results.

Then, upon having the query expansion selected, terms (dimensions) in the question were expanded from several aspects. These aspects include synonyms, parent and child nodes of the dimension.

However, as standard, while trying to resolve a question, the query expansion was performed under a certain condition- that was if the question asked for subclasses or subparts of an entity. All query expansions were performed using ontologies. For example, if a question had ‘parts of gyrus’ in it, then first the subparts of gyrus were found from the ontology via query expansion. Then they were placed in the question instead of the gyrus.

4.2.3 Resource Integration

As discussed in the introduction of this study, the focus of the study was on using ontologies alongside neuroimages for resolving neuroscience questions. One key factor for answering a question is the data resource upon which the question is being queried.

Therefore, to provide a solid foundation for the questions to be answered, an integrated answering module was shaped through integrating domain ontologies with neuroimage-based data resources. The results were calculated by posing templates against integrated resources. For creating the integrated resources, ontologies used in this study including the NIFSTD and NeuroFMA were integrated according to the directions of research papers such as (Nichols, Mejino Jr and Brinkley, 2011; Nichols *et al.*, 2014).

These research papers laid foundations for this integration. NIFSTD and NeruoFMA were discussed briefly in previous studies and the full amount of information can be found in Appendix B. Module B of Figure 4-2 shows how these resources have been shaped.

Then, the ontologies were integrated with the Freesurfer (Fischl, 2012) software. Freesurfer is an application for processing and analysing Magnetic Resonance Imaging (MRI). Detailed information regarding Freesurfer can be also found in Appendix B.

The reason for using NeuroFMA is that this ontology includes information about the human brain. It acts as a medium for mapping different ontologies and applications too. For example, it includes links to the NIFSTD entities, which contain information about the brain and nervous system in general, regardless of the species type.

As an example, Figure 4-1 demonstrates how NeuroFMA maps brain parts into Freesurfer data points. This process is done via IDs that have been manually mapped from NeuroFMA to applications such as the Freesurfer, and ontologies such as the NIFSTD.

For example, the NeuroFMA subject with the URI “http://.../ONARD_Instance_1360104” has the FMA_ID number 272454. This ID has been mapped to the Freesurfer_ID “1006”. The Freesurfer_ID is in turn pointing to the Freesurfer subject “ctx_lh_entorhinal”. This is how the mapping between NeuroFMA and Freesurfer is performed. The mapping between NeuroFMA and NIFSTD is achieved using the same method.

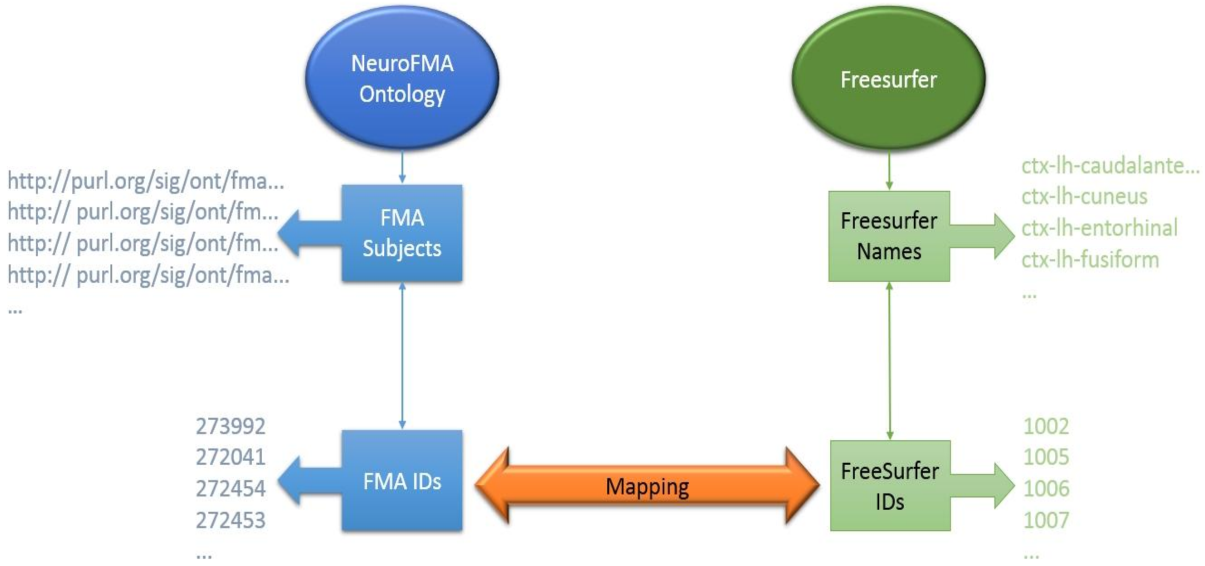

Figure 4-1- NeuroFMA to Freesurfer mapping

4.2.4 Question Processing Model

After creating templates and query expansion unit, also integrating the resources, all three came together as modules, to create the question processing model. This model describes how a question is processed from the beginning to the end, and where each module is located along this process. Figure 4-2 demonstrates how different modules are connected.

The first module concerned question-related tasks and handled question preparation processes. This module of the study received and translated questions with the help of ontologies according to tasks described in the Templates section, Query Expansion section and partially the work in the previous chapter in ontology-based question classification.

Since most of the tasks regarding pre-processing were covered in the previous study, there is no need to discuss them in detail here. However, the following paragraph gives a short summary of the pre-processing steps.

Upon tackling each question, first, the question was parsed and tokenized, stopwords were removed, dimensions pointed out in the previous study were searched for and tokens were then tagged with those dimensions. It is worth reminding the reader that dimensions included entities, domain-specific phrases, aggregation and statistical phrases, data references and conditional phrases. So, for example, the token ‘hippocampus’ was tagged as an *entity.*

Please note that query expansion was used for ambiguity resolution by tasks such as searching terms entered by the user in the ontology and finding their synonyms. However, if needed, user interaction was used for resolving ambiguous words that could not be found from ontologies.

Sometimes it was necessary to perform string matching to select and find different terms in a question from the ontology. String matching can be performed using various methods and tools such as programming languages, libraries, APIs and even standalone applications.

SPARQL supports string matching to some extents through its string matching functions such as ‘contains’ and ‘regex’. Moreover, some SPARQL query tools use third-party tools such as Lucene (McCandless, Hatcher and Gospodnetic, 2010) for this purpose. Lucene can be used for string matching by itself too.

With the string-matching capabilities of SPARQL, even when incomplete parts of the information are in hand, the full details of the information can be found. A sample SPARQL code for finding information regarding a keyword is as follows. An argument can be passed to the code snippet either by the program or user.

```
SELECT DISTINCT ?dim ?properties ?values

WHERE{

 ?dim ?properties ?values .

  FILTER (

    CONTAINS (lcase(str(?dim)), "?argument") )

}
```

After this, the information was passed to the templates to be queried against the second module, which was the integration platform module. The resources were used as discussed in the Resource Integration section.

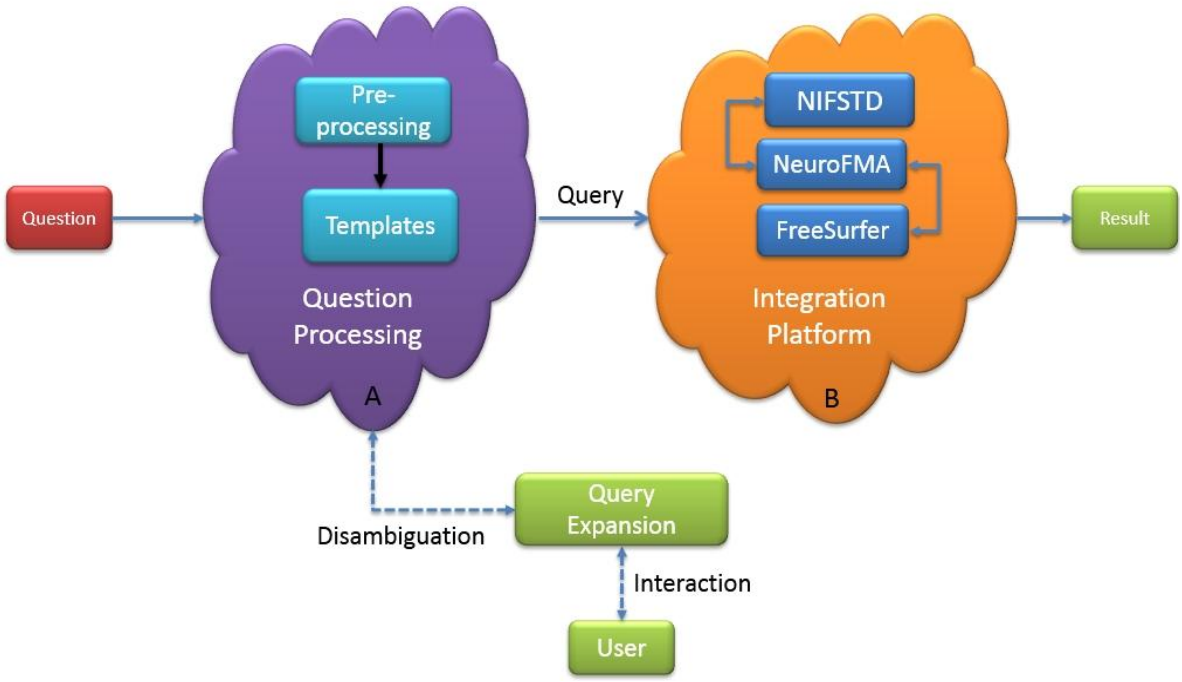

Figure 4-2- The Question Answering Process Model

4.3 Results

In the methods section, it was discussed how tasks in this study were going to be performed. During the current section, it is shown how these tasks are implemented and presented according to the Methods section.

First, experts reinspected the data set to match it to MRI images. The outcome was a slightly different dataset from the dataset in the previous study. The new one is presented in the following section.

Then, templates were created to cover as many questions as possible, and it was demonstrated how questions can be resolved and translated into templates. As described in the Methods section, each template was created based on a level from the question hierarchy.

Then the integration of resources was done according to the ways discussed in the Methods section. Next was the question processing model. It was demonstrated how this model works through an example.

4.3.1 Question Set

In this section, the question set used for this study is discussed. Neuroimage resources provided by experts for this research were MRI neuroimages and their annotated data. However, the set of questions we had in the previous study included some questions that needed functional magnetic resonance imaging (fMRI) to be answered.

Functional magnetic resonance imaging is blood oxygen level dependant (BOLD). This means special experiments must be performed so their outcome can be used to resolve questions. Therefore, some questions have been changed slightly by an expert to be resolvable using the MRI images. Also, some questions have been added and replaced in the new set. The set of questions is as follows:

1. What are the synonyms of hippocampus?
2. Which parts of precentral gyrus have volume in the data?
3. What is the thickness of the gyrus rectus in the data?
4. Do the schizophrenia subjects show more or less ventricular enlargement than healthy subjects do, in cortical regions that are connected by the superior longitudinal fasciculus?
5. What is the white matter volume of the inferior parietal lobule in the data?
6. What is the anatomical region of the lingual gyrus?
7. What is the amygdala?
8. What is the anatomical relationship of telencephalon and diencephalon?
9. What are amygdala parts?
10. Which regions overlap with the precentral gyrus?
11. How many subjects gained extra volume in a region of the diencephalon in the patient's data?
12. Which annotated structural region has changed volume in patients?
13. What are the branches of the seventh thoracic nerve?
14. What is the anatomical region of the premotor cortex?
15. What is the thickness of the inferior parietal lobule in the data?
16. What is the grey matter volume of the short insular gyrus in the data?
17. Which part of the cerebral cortex has the highest atrophy in HD?
18. Which parts of the cortical region has highest amount of cortical thinning?
19. What are the adjacent regions to the most atrophied region in HD?
20. Which regions are connected to the most atrophied region in HD?
21. Which cortical regions show curvature in HD?
22. What are the parts of the superior temporal gyrus?
23. Which cortical regions show thinning in pre manifest HD?
24. Which regions show thinning/atrophy over 12 months in HD?
25. Does any region show increased volume in pre manifest HD?
26. How many regions are in the parietal cortex?
27. Is the thalamus atrophic in these Multiple Sclerosis patients?
28. What are the regional parts of the posterior short insular gyrus?
29. What are the subjects with an average cortical thickness less than 5mm?
30. What is the thickness of the anterior cingulate gyrus?
31. Is the degree of change in volume of the thalamus over 1 year in patients greater than the change in controls?
32. Are cortical thicknesses correlated across all cortical regions in this subject group?

The tasks described and related to this study were produced based on the above questions. Of course, as will be demonstrated in the following sections, the model represented in this study theoretically has the potential to answer fMRI questions upon access to those images and incorporate information related to them.

4.3.2 Templates

This section is dedicated to creating and discussing templates, which are a major contribution of this chapter and the centre of focus in this study. They are built according to the question hierarchy from the previous study.

As a reminder, in the previous chapter, neuroscience terms were found from the ontology first. They were then mapped to the dimensions such as entity; domain-specific (attribute); and statistical, data reference and conditional phrases. Finally, based on the number and formation of dimensions, each question was mapped to the appropriate level of the question hierarchy.

For example, the question "What is the amygdala?" consisted of the *amygdala*, which is an *entity* and was mapped to level 0. Similarly, the question "What are the synonyms of hippocampus?" consists of the hippocampus, which is an *entity* and synonym which is a *domain-specific (attribute)* term matched to level 1.

Furthermore, the question "What are the regions with atrophy in Huntington Disease?" included regions as an *entity*, atrophy as a *domain-specific* and Huntington disease as a *data reference*. Therefore, it matches level 2. Table 4-1 reviews what level each question was corresponding to, based on the previous study presented in the third chapter.

Table 4-1- Questions, their dimensions and level

Question Number	Question Contains	Level in Hierarchy
7	Only one dimension	0
1, 6, 8, 9, 10, 13, 14, 22, 28,30	Entity + Domain-Specific (attribute) Phrase	1
2, 3, 5, 15, 16, 21	Entity + Domain-Specific Phrase + Data-Reference	2
17, 18, 19, 20, 26, 29	Previous formations + Aggregation Phrases	3
4, 11, 12, 23, 24, 25, 27, 31, 32	Previous formations + Conditions / Changes/ Comparisons	4

The next step was to create templates based on question hierarchy levels. In order to account for all forms of questions while changing the abstract form of the question hierarchy levels, a greater number of templates were created than the initial number of levels.

This was because as can be seen, a question in level 4 can have conditions, change or comparisons and each of them needs its own code. Moreover, in all levels, templates should have been created that accounted for different types of results and dimensions wanted by the user.

Please note that to save space, codes and files regarding questions could not be fully demonstrated here and in cases that the codes were lengthy, only pseudocodes were used. Please find more complete codes here: https://github.com/Aref-cs.

A level 0 question will have one template since based on definitions in question hierarchy levels, a question is located under this group whenever there is only one dimension in it and the result for such a question would be the information regarding that dimension. The general pseudocode for this template is as follows:

```
SELECT distinct (?dim as ?Part) (Str(?label) as ?Name) (?def as ?Definition) (?com as ?
Additional_Info)

  WHERE {

    GRAPH <http://localhost:3030/myDataset/data/NeuroFMA> {

        dim rdfs:label ?label .

        optional {?dim rdfs:comment ?com}

        ?dim fma:definition (or skos:definition for nif) ?def

  }
```

The general pseudocode is a comprehensive form of the query containing code that can be queried against ontologies used in this study. It is worth mentioning there is more than one method to implement these codes. For simplicity reasons, sometimes codes are queried against only one ontology.

The question “What is the amygdala?” is an example that used this template, since it had only one dimension and falls under level 0. The *amygdala* is the keyword used in this question. Based on this, and since amygdala is an entity, the code for resolving this question used the above template with *amygdala* being passed to the query as the dimension which its value is searched for.

```
SELECT DISTINCT (?term as ?PartURI) (?def as ?Definition)

  WHERE

{ ?term rdfs:label "Amygdala"^^xsd:string .

    ?term skos:definition ?def

}
```

Level 1 consists of two dimensions. It has an entity and a domain-specific dimension. The template for this level is straight forward like the question template for level 0. Here is the general pseudocode for this level:

```
SELECT DISTINCT (?dim as ?Result) {
```

```
 {VALUES ?type {ontology1 namespace (nif) :related predicates}.

   ?subject ?type ?dim.

 }

  UNION

  {

 VALUES ?type {ontology2 namespace (fma):related predicates}.

      ?subject ?type ?dim.

  }

      }
```

A question which is relevant to level 1 is “What are the synonyms of hippocampus?”. This question contains an entity and a *domain-specific* phrase. The dimension in question or the dimension which is being queried, is a synonym; therefore, its code would be:

```
SELECT distinct (?syn as ?Synonyms)

WHERE{

{?s rdfs:label "Hippocampus";

    nif:synonym ?syn . }

UNION

{?s rdfs:label "Hippocampus";

    fma:synonym ?syn . }

}

ORDER BY ?synonyms
```

Next is level 2. Level 2 is a little bit different from the previous two levels since it is working with outputs of neuroimaging applications such as the Freesurfer. For this question, which has a reference to the data, a Python library called Pandas (McKinney, 2012; McKinney, W. and Team, 2015) is used as mentioned before.

Figure 4-3 is a snapshot of a Freesurfer data file. Freesurfer produces multiple files per MRI image process and each of the files contain information regarding the brain. More data on Freesurfer is given in the Appendix B.

Index	SegId	NVoxels	Volume_mm3	StructName	normMean	normStdDev	normMin	normMax	normRange
1	3001	3109	3109	wm-lh-bankssts	102.9081	6.1054	81	120	39
2	3002	2278	2278	wm-lh-caudalanteriorcingulate	104.9004	6.422	73	118	45
3	3003	6107	6107	wm-lh-caudalmiddlefrontal	101.8753	6.2081	80	116	36
4	3005	2402	2402	wm-lh-cuneus	102.7179	8.545	79	120	41
5	3006	868	868	wm-lh-entorhinal	90.8382	6.6576	71	109	38
6	3007	5712	5712	wm-lh-fusiform	96.2881	7.8874	63	116	53
7	3008	8976	8976	wm-lh-inferiorparietal	107.3302	7.0717	81	125	44
8	3009	5630	5630	wm-lh-inferiortemporal	94.5687	9.2353	66	116	50
9	3010	3332	3332	wm-lh-isthmuscingulate	110.9344	7.7504	48	130	82
10	3011	7880	7880	wm-lh-lateraloccipital	101.1066	7.6086	76	124	48
11	3012	6454	6454	wm-lh-lateralorbitofrontal	98.7598	8.1996	72	121	49
12	3013	4715	4715	wm-lh-lingual	96.6863	8.3226	59	118	59
13	3014	2659	2659	wm-lh-medialorbitofrontal	96.223	8.1062	63	116	53
14	3015	5175	5175	wm-lh-middletemporal	94.929	8.4992	70	114	44
15	3016	1471	1471	wm-lh-parahippocampal	97.8883	7.6657	65	118	53
16	3017	3264	3264	wm-lh-paracentral	102.5959	6.4102	78	118	40
17	3018	3244	3244	wm-lh-parsopercularis	99.2119	7.4122	78	115	37
18	3019	840	840	wm-lh-parsorbitalis	90.8217	7.5229	71	107	36
19	3020	2420	2420	wm-lh-parstriangularis	98.1689	7.1232	76	112	36
20	3021	3488	3488	wm-lh-pericalcarine	98.7263	9.9965	19	118	99

Figure 4-3- A sample of Freesurfer data

In this level, first the information was located from the ontology. Then the identification number for the part and its related data was also found from the ontology and was based on the general code mentioned in the Resource Integration section.

The identification number, referred to as *id* in the code, was then used for calculating the information needed by the user. These data include surface area, mean curvature, average thickness and other information regarding the brain. Figure 4-4 shows the general code for accessing and working with the Freesurfer data.

```
import pandas as pd

df = pd.read_excel(r'freesurfer data file')

#Reading the data based on the Freesurfer ID
print(df[df.SegId == idFromOntology].'required data cell'.values[0])
print(df[df.SegId == idFromOntology].'required data cell'.values[0])
 .
 .
 .

#Reading secondary data in case of tracking change or comparison
norm_a = df[df.SegId == idFromOntology1].'required data cell'.values[0]
norm_b = df[df.SegId == idFromOntology2].'required data cell'.values[0]
 .
 .
 .

Calculate calculations
print("Result of calculations: ", calculations)
```

Figure 4-4- Pandas general form for working with data (Freesurfer)

The question “What is the volume of the insula in the data?” is a sample question for level 2. This question contains insula as the *entity* and volume as a *domain-specific* phrase and *data* as a data reference.

For resolving this question, the *insula* is found in Fressurfer using the mapping provided by the NeuroFMA ontology. The code for finding the Fressurfer IDs for these parts based on the general pseudocode provided earlier, is as follows.

```
SELECT DISTINCT ?fma_uri ?neurolex_id ?fs_id
 WHERE
  {
    ?fma_uri fma:Preferred_name "Insula"^^xsd:string ;
                   fma:Freesurfer_ID ?fs_id.
```

Then, the information regarding the Freesurfer IDs is passed to the Python and using the Pandas library, the final result of this question is calculated and returned to the user.

The next level on the list is level 3. Level 3 codes are based on and similar to level 2 codes. Level 3 accepts questions that SPARQL could implement conventionally with its statistical functions. This limits level 3 questions to those that could be implemented with commands such as COUNT, which is used for counting instances of a concept in the ontology; MIN, which calculates the minimum amount of all instances; MAX, for calculating the maximum of all instances; AVG, which calculates the average of all instances and finally SUM, which is used for calculating the summary of all instances. Therefore, questions of this level included statistical phrases in addition to level 2. The general pseudocode for this level is as follows.

```
SELECT ?num (aggregation command

    (DISTINCT ?dim) as ?appropriate name) {

 {VALUES ?type {ontology namespace :related predicates}.

   ?subject ?type ?num.

 }

     } GROUP BY ?num
```

The final level of the question hierarchy that a template was built for, was level 4. Level 4 contained different subgroups including questions containing changes, comparisons and conditions.

Since changes in the data usually happen across different experiments or readings, more than one dataset was needed while resolving such questions. However, the foundations of the codes are similar to previous ones.

A routine sample question associated with level 4 is "What region of the premotor cortex shows atrophy?". Atrophy can be measured only by looking at different datasets from different times or groups of patients.

Therefore, the user identifies the first two datasets for temporal questions. For groups, two groups of patients are identified. Then, according to each of these, after the query expansion, the Python code with the Pandas library will get the information regarding the sizes of the parts in a loop using the pseudocode below.

```
#Load Pandas as shown above (pd)


# Read data from related files 'file1.csv', 'file2.csv'

data1 = pd.read_csv("filename.csv")

#one file per subject data (data2 for the second one)

```

```
count = 0
while (count < 'number of loops needed for reading the data'):
    count = count + 1
    read the data from the file, as shown in Figure 4-4

#compare and calculate the result

#print out the result
```

Then the results of the loops were compared together. Please note that each loop can consist of several available datasets. Also, it is obvious that all these loading and calculations add to the cost of calculations.

Another type of level 4 questions are the comparison questions. Templates of these questions were similar to questions containing change. The only difference was where the comparison templates wanted to demonstrate the result.

The last template for questions containing conditions was the most complex, and a little bit different to the other level 4 templates. This was because, in questions containing conditions, the condition had to be resolved first. After all, the rest of the question was basically a question from other levels, queried on the result of the condition.

An example of a conditional question is “Which subjects show more atrophy in cortical regions that are connected by the superior longitudinal fasciculus?” In these types of questions, first, the condition should be satisfied.

Therefore, regions that are connected by superior longitudinal fasciculus had to be found as a separate query and stored as a graph first. This graph was created using the ‘construct’ command in SPARQL and will be a subset of the ontology that the condition was true inside it. ‘Part’ variable was passed to the query. Here, ‘part’ value equalled ‘superior longitudinal fasciculus’.

```
CONSTRUCT ?regions FMA:connected_to ?part
  WHERE {
     regions FMA:connected_to ?part
}
```

The rest of the conditional template is not very different from the other ones. From here, according to the remaining parts of the question, one of the other templates was used. Please note that there are other ways of tackling conditions too.

It is worth mentioning that altogether, about 10 main templates and several extra templates were made to cover and resolve different types of questions. Main templates included four templates for levels 0 to 3, and six different templates for level 4 questions containing comparison, change and conditions.

4.3.3 Resource Integration

This section describes the resource integration module of the model described in this study. It brings together data resources that the questions are asked against and assist in calculating the answers.

Since this study used different resources such as ontologies and MRI images for answering questions, there was a need to find a way for mapping and connecting them. This enabled users to ask more complex questions with a wider scope.

The task was to connect NeuroFMA ontology, NIFSTD ontology and Freesurfer. There were a couple of ways to resolve the mappings between different ontologies and applications. One was to extend one of the terminologies to cover the terminology of the application (Freesurfer in this case) and the other ontology. The problem with this method is that it is very time-consuming.

The more convenient method was to use third party resources such as UMLS and NeuroNames terminologies to resolve the problem of mapping. In this method, the third-party resource is used as a bridge between the two main resources through its vocabulary.

Fortunately, in the case of the resource integration module designed for this model, the Foundational Model of Neuroanatomy Ontology (FMN) (Nichols, Mejino Jr and Brinkley, 2011), has been created which helps us to map the FMA to the Freesurfer.

FMN is created to provide a framework for mapping disparate views of neuroanatomy. In other words, it uses the second approach for mapping different semantic resources described above.

FMN works as an aligner (Acampora, Loia and Salerno, 2012) for NeuroFMA ontology and Freesurfer. This means that it tries to match ontologies that are discussing the same topic but using different terminologies based on their curators and human subjectivity.

Also, there was a need for mapping the NeuroFMA to the NIFSTD. This can be done using coding. For example, the following code maps NeuroFMA (FMA) concepts to NIFSTD terms, using NeuroLex, a terminology that is linked to NIFSTD. The general template for using this mapping which in fact creates the underpinning for the integrated resource is as follows.

```
SELECT DISTINCT ?fmaS  ?fmaName ?neurolexId ?nlxName

      WHERE {? fmaS   fma:preferred_name ?fmaName;

                     fma:Neurolex ?neurolexId .

             OPTIONAL {?nlxAnt owl:annotatedTarget ?neurolexId;

                                      fma:name ?nlxName .}

            }
```

Where 'fmaS' is the uniform resource identifier of the FMA subject (part), 'fmaName' is the name of the subject, 'neurolexId' is the identification for the subject in NeuroLex (Fahim T Imam *et al.*, 2011) and 'nlxName' is the name for that part.

Having resources mapped and connected together as the resource integration module of the model, the foundations for answering different neuroscience questions is almost ready and the model is one step closer to resolving questions.

4.3.4 Question Processing Model

So far, it has been demonstrated how different parts such as templates, query expansion and resource integration were implemented. Furthermore, examples were provided for each part.

In this section, it is discussed how all parts came together as modules to create the question processing model which is shown in Figure 4-2. Here, it is demonstrated how each module works through presenting and solving a sample question. The sample question is as follows:

- What is the volume of the inferior parietal lobule in the data?

In resolving this question, each of the tasks discussed so far, played a role, including the question processing module, query expansion module and resource integration module. As mentioned in the Methods section, the question processing module is created by merging the ontology-based question classification from the previous study and templates.

In the beginning, as discussed before and as a part of a question processing module, the pre-processing procedure was performed and all stopwords were removed from the question; moreover, all dimensions were tagged. The outcome was *volume*, *inferior parietal lobule* and *data*.

According to the discussion regarding dimensions in the previous study, *volume* was a domain-specific phrase and *inferior parietal lobule* was an entity in this question. The query expansion was performed to have a richer set of results.

Therefore, the code below was posed on the ontology to find children nodes of the inferior parietal lobule. The reason for this action, will be discussed later. The *id* variable will be filled with the name of the entity. Please note that parts of the code such as the part that defined namespaces are not mentioned to save space. This query returned a set of inferior parietal lobule sub-parts including left inferior parietal lobule and right inferior parietal lobule.

```
SELECT DISTINCT (?p as ListOfSubClasses) (str(?label) as ?Name)

{

 ?s fma:FMAID \(id) ^^xsd:string.

 ?p rdfs:subClassOf ?s.

 ?p rdfs:label ?label.

}
```

Having the sub-parts, the second step was to search for these entities alongside their parent in the neuroimage related data resource, which was a Freesurfer data file. This file is an annotated data representation of neuroimages created by the Freesurfer. A sample of one of the Freesurfer output files can be seen from Figure 4-3. For this means, the resource integration module was used via the code mentioned in its section. Figure 4-1 shows how this code works.

For the final step, this task of resolving the question was completed by returning to the Freesurfer data file and finding identification numbers of two subparts of the inferior parietal lobule. Then their volumes were calculated separately and added together to calculate the volume of the white matter of the inferior parietal lobule, which was equal to 19,432 mm^3.

As could be seen from the results, using the query expansion assisted in finding results for this question that could not be found in any other way; this was because of the resolution of MRI images, that results in not providing any volume for the inferior parietal lobule itself. Detailed information regarding how query expansion assists in achieving better outcomes will be discussed in the Role of Ontologies from the Discussion section.

The model presented in this study could resolve 78 per cent of questions in total. Table 4-2 demonstrates a review of which questions could, and which ones could not be resolved using the model. For example, the question “Is cortical thickness correlated across all cortical regions in this subject group?” could not be resolved, because ‘correlated’ could not be resolved using the ontologies. Of course, it could have been resolved using user-interaction or creating a new template but creating templates for individual questions was against the automated nature of the model. Similarly, “Which regions are functionally connected to the most atrophied region in HD?” could not be resolved because detecting if a region is ‘functionally connected’, needs information that is not available in the datasets used. It needs additional databases, such as functional MRI databases to be integrated into the model. Another example is “Which cortical regions show thinning in pre-manifest HD?”, in which, the term ‘pre-manifest’ could not be resolved. In the question “Is the degree of change in volume of the thalamus over 1 year in patients greater than the change in controls?”, resolving the condition turned out to be a problem and in question “What are the positive effects of aspirin on the brains of elderly people?”, terms ‘positive’ and ‘elderly’ could not be resolved due to their relative and non-definitive nature. In some questions like “What are the adjacent regions to the most atrophied region in HD?” the ontology did not contain the information needed for resolving the question.

Therefore, in many cases, the question was not resolved, not because of the model, but because the ontologies or the image did not contain the information needed for these questions. Potential ways of resolving these questions, including using multimodal imaging and fuzzy logics, will be discussed further in the Discussion section as future work.

It is obvious that the model will be able to achieve even better results if modules or parts of them, such as the question classification are improved. This is because the overall success of the model is dependent on the overall performance of each module.

For example, one important outcome of this study was its success in resolving different types of questions to templates and issues arising during this process. This was because the templates were a major contribution of this study and play a significant role in resolving questions. Therefore, if more questions can be represented by templates, more questions can be resolved. This has been discussed further in the Discussion section.

Despite this, it is worth mentioning that the model is working fine as it is, because the goal of this study has been to cover different questions and its goal was different with systems that try to achieve a high number of recall and precision only. This has been discussed further in the discussion section at the Critical Reflection on the Results section.

Question	Resolved or not
What are the synonyms of hippocampus?	✓
Which parts of precentral gyrus have volume in the data?	✓

What is the thickness of the gyrus rectus in the data?	✓
Do the schizophrenia subjects show more or less ventricular enlargement than healthy subjects do, in cortical regions that are connected by the superior longitudinal fasciculus?	×
What is the white matter volume of the inferior parietal lobule in the data?	✓
What is the anatomical region of the lingual gyrus?	✓
What is the amygdala?	✓
What is the anatomical relationship of telencephalon and diencephalon?	✓
What are amygdala parts?	✓
Which regions overlap with the precentral gyrus?	✓
How many subjects gained extra volume in a region of the diencephalon in the patient's data?	✓
Which annotated structural region has changed volume in patients?	✓
What are the branches of the seventh thoracic nerve?	✓
What is the anatomical region of the premotor cortex?	✓
What is the thickness of the inferior parietal lobule in the data?	✓
What is the grey matter volume of the short insular gyrus in the data?	✓
Which part of the cerebral cortex has the highest atrophy in HD?	✓
What are the positive effects of aspirin on the brains of elderly people?	×
Which parts of the cortical region has highest amount of cortical thinning?	✓
What are the adjacent regions to the most atrophied region in HD?	×
Which regions are connected to the most atrophied region in HD?	✓
Which regions are functionally connected to the most atrophied region in HD?	×
Which cortical regions show curvature in HD?	✓
What are the parts of the superior temporal gyrus?	✓
Which cortical regions show thinning in pre-manifest HD?	×
Which regions show thinning/atrophy over 12 months in HD?	✓
Does any region show increased volume in pre manifest HD?	✓
How many regions are in the parietal cortex?	✓
Is the thalamus atrophic in these Multiple Sclerosis patients?	✓
What are the regional parts of the posterior short insular gyrus?	✓
What are the subjects with an average cortical thickness less than 5mm?	✓
What is the thickness of the anterior cingulate gyrus?	✓

Is the degree of change in volume of the thalamus over 1 year in patients greater than the change in controls?	×
Are cortical thicknesses correlated across all cortical regions in this subject group?	×

Table 4-2- Outcome of the model for question-set

4.3.5 Implementation Issues

Many additional challenges had to be overcome during this research on top of the challenges solely focused on answering the research questions. This is because sometimes strategic and technical obstacles in the way of implementation process are harder, or as hard as the implementation process . Working with ontologies in neuroscience is a good example of this.

This was mainly due to the large size of ontologies. For example, only the NIFSTD ontology has about 3.6 million triples and still is not considered as a very large ontology. Therefore, one of the major challenges in this study was in querying domain ontologies in neuroscience such as NIFSTD and NeuroFMA that consume huge amounts of memory.

Different tools were available, including different libraries and APIs from different programming languages; applications designed for working with ontologies and SPARQL query endpoints via different methods, such as ontology modularization on different machines and virtual machines with different settings. All of these were tested to find an efficient and viable way of tackling questions.

Java and Python both provide different libraries for working with ontologies. Namely, Java provides the Jena (Reynolds, 2004) and OWL (Matthew Horridge, 2011) libraries and Python provides RDFLib and OWLReady2 libraries. All of these libraries had some shortcomings since this study needed a library that can work with OWL2 (Sattler *et al.*, 2008) and support SPARQL.

Protégé (Gennari *et al.*, 2003; Musen, 2015) is able to demonstrate a hierarchy of concepts inside ontologies and is used for curating them too. More information on Protégé can be found in the appendices.

Protégé also offered an API for use in programming languages. Unfortunately, protégé versions up to 3.x supported SPARQL did not support OWL2, while the more recent versions from 4.0 and later support OWL2 but did not support SPARQL according to its Wiki.

Next were the Blazegraph and Bioportal web-based portals and query endpoint for working with ontologies. Blazegraph can load external ontologies using a local machine. Both Blazegraph and Bioportal provide APIs for working with the ontologies.

However, as mentioned, the major problem in using large neuroscience ontologies is the amount of memory needed. For example, both Protégé and Blazegraph had problems on a machine with 8GB of memory and could not load the NIFSTD ontology.

Therefore, modularization of the ontologies was tested using a method in which researchers generate a view or extraction from FMA ontology (Shaw et al., 2008). This method is usually used for extracting application ontologies from foundation ontologies.

This procedure was performed using vSPARQL (Shaw *et al.*, 2011). vSPARQL is an extension to SPARQL which enables the creation of application ontologies from reference ontologies. This is done by changing the ARQ, which is an open-source query processor for SPARQL (Shaw *et al.*, 2008). However, this did not work either, even while using the original code described by the vSPARQL research paper.

Finally, the 8GB memory machine was replaced with a 48GB memory machine on a cloud computing service, Linux-based provider to allow Blazegraph and other coding experiments to be executed. In the meanwhile, a new approach with Apache Jena Fuseki showed great results even on the machines with a regular amount of memory.

However, as mentioned, each method had its benefits. Therefore, a mixed setting was used, which turned out to be successful. This mixed setting was created by using Fuseki server, Python and RDFLib. Tools were selected according to their capabilities and where they were more suitable.

Possibly one of the drawbacks of using ontologies lies in the fact that most platforms designed for working with them are hard to set up and the amount of memory they need is usually considerable, which in return, sometimes results in slow speeds of calculations.

Nonetheless, the memory issue could probably be taken cared by using light platforms like Fuseki server or using approaches such as allocating a separate process to each of the tasks or parallel computing. Moreover, once the environment is set up, working with ontologies gets much easier.

4.4 Discussion

In this section, several issues including how this study helped resolve questions, the role of ontologies in this study, limitations and shortages of the study and future directions are discussed. Furthermore, problems faced during this study and the set up for implementing tests in the study are addressed.

A noticeable issue worth discussing regarding the model is its flexibility, ability to be extended and the comprehensiveness of it. It has a modular design that makes applying changes or adding to the model easier.

Furthermore, it reduces questions through its modules, specifically the question processing module. It also brings together different resources in the resource integration module, which results in answering more questions.

Moreover, this model tackled a variety of different questions posed by different experts and resources, which demonstrated how comprehensive it was. These factors can potentially make this model a more comprehensive one than most current template-based approaches towards resolving questions in neuroscience.

One objective of the current chapter was to study results of using ontologies in answering questions in neuroscience. Ontologies worked as primary sources of data, complementary data and means of disambiguation in this study. They played a significant role in resolving almost all sample questions. Following sections discuss these items in more detail.

4.4.1 A Critical Reflection on Results

It is a good idea to discuss some issues that occurred during resolving questions and experiments of this study. One is that the approach discussed in this study does not answer some questions, while potentially it can answer them.

The reason is that in this study the focus has been on resolving questions in an automated process. Therefore, even though the model can theoretically answer the majority of questions, resolving questions that cannot be performed automatically for the most part or require individually tailored templates have been intentionally discarded.

For example, the question "What are the positive effects of aspirin on elderly people?" might be resolved through simple or sophisticated methods. Simple methods include user interaction, defining *positive* and *elderly* in advance, and sophisticated methods include using knowledge graphs according to methods investigated by Cimiano (2017), in addition to domain ontologies.

In general, this question is a very abstract one with indefinite terms that can have extensive value margins with no certain answer, even with using methods mentioned above. However, investigating methods of resolving very abstract questions like this one can be a future direction for this study.

The next issue was with questions containing terms that were partially resolvable through ontologies. For example, in the questions "What is the average anatomical location of the precentral gyrus?" the term '*average'* is not resolvable. However, the same term could be resolved via the use of ontologies in a question asking about the average volume of a specific part of the brain.

Something that is worth discussing is the resolution of questions containing temporal phrases. In this study, they were resolved using a mixture of Python and SPARQL. However, there are ontologies that contain information, which can assist in resolving such questions.

These ontologies include the Digital Imaging and Communications in Medicine (DICOM) ontology, which contains information regarding the patient and imaging procedures; OBO relation ontology (OBO-RO), which contains information regarding relations in biomedical ontologies and RadLex ontology (Langlotz, 2006). Radlex ontology includes information regarding neuroimages.

They can be added to the resource integration module of the model discussed in this study as a future direction. Especially, since Mejino *et al.* (2010) already established a way to connect RadLex ontology to the FMA ontology (NeuroFMA) and as a result, created an application ontology.

According to the issue explained in the Result section regarding the templates, a future direction of this research can be to re-design the question hierarchy according to its initial assumptions and conditions, plus findings of this study. Doing this will result in a comprehensive and practical question hierarchy in neuroscience.

The current question hierarchy was created according to the types of dimensions and phrases that shaped the questions. Despite the current hierarchy not considering RDF triples as a factor in question complexity while constructing its levels, it was solely designed with good question coverage in mind and based on the questions.

After implementing the templates and going through experiments in resolving questions, it was noted that it is better to factor the issues regarding the implementation of templates into question hierarchy as a feedback mechanism. It will be better if more weight is given to the structure of the questions and RDF triples in shaping the next version of the question hierarchy.

4.4.2 Related Works

In this section, related works to this study are discussed. This study mentioned two major uses for ontology in resolving questions and described them as integration-oriented and question-oriented approaches. However, others provided different categorizations too.

Another fine categorization based on the role of ontologies was presented by Gupta et al. (2010) and consists of four major groups. This research indicates the application of ontologies as a major attempt towards resource integration and by this, answering questions.

It divides applications of ontologies in resolving questions into four groups: F-logic based systems, database to ontology mapping systems, integration systems and query expansion systems. As Gupta et al. (2010) state themselves, out of these four categories, the F-logic based systems are out-dated to a great extent.

AquaLog (Lopez, Pasin and Motta, 2005; Lopez *et al.*, 2007) and later, PowerAqua (Lopez *et al.*, 2012) and Freebase (Yao and Van Durme, 2014; Cui, Xiao and Wang, 2016) are systems designed for tackling queries on the semantic web. They are among the first and best systems designed using ontologies.

They are different from the work presented in this chapter since they are designed to work in the general domain. Furthermore, apart from having different scope from this study, they have some shortcomings while tackling questions.

For example, these systems form a triple to model the question. These triples can have problems, which is the gap between a RDF triple to actual SPARQL code that can be processed by a machine. The study presented in this chapter tried to remove this problem by using templates that cover more than just a triple.

One other system that is related to this study is NIF search (Imam *et al.*, 2012). Of course, the NIF search system has been described in a set of research papers including (Akil, Martone and Van Essen, 2011; Fahim T. Imam *et al.*, 2011; Jeffrey *et al.*, 2014).

It is worth mentioning that NIFSTD, the ontology used in this study, was created as a part of NIF (neuroscience information framework) attempts. Despite being fundamental for neuroscience, queries that were demonstrated to be resolved and processed by NIF were basically simple queries.

Queries included searching for synonyms; some conjunctive queries; searching hierarchies and seeking for collection of terms. Whereas, the model introduced in this chapter tackles a variety of questions instead of queries, which are much more complex to resolve by a machine.

The approach discussed in this study has the ability to resolve questions containing synonyms, term hierarchies such as parents and children of a term, comparisons, patient related data, changes in the data, conditions, description of terms and aggregation phrases. Therefore, the range of questions that can be resolved using this approach is certainly wider than other approaches.

It is worth mentioning that due to the use of ontologies, this approach can be used to resolve questions regarding various parts and functions of the brain. These include cell and subcellular entities, functions, anatomy and structure and lastly, experiments.

The other difference between the system in this study and other systems is that most other systems provide a single or combination of terms search. Whereas, the model described here provides a free text search that can become much more efficient and advanced with a few improvements in the question processing module. These improvements are discussed in the Model and Its Design section as future works.

There are other related studies outside the scope of neuroscience such as the TAMBIS system (Stevens *et al.*, 2003) which provides a transparent access to some bioinformatics resources or KA-SB (Roldán-García *et al.*, 2009) that integrates resources and performs reasoning in biology.

4.4.3 Model and Its Design

The modular design of the model is a positive point since it increases flexibility while allowing for a lower cost of applying changes to the model or replacing parts of it. For example, if the part responsible for pre-processing questions is to be changed, then it can easily be changed without affecting other parts of the model, especially if its output stays the same.

Furthermore, adding new parts to the model would be easier. In order to make the model more comprehensive, new ontologies can be added to the model and make it more powerful in resolving questions or increase the types of questions that can be answered via this model.

For example, in this study the main focus of the model was on image-driven data resources and ontologies regarding the brain itself. However, other sorts of neuroscience documents could be used or integrated with other resources such as text-based resources or ontologies discussing tools regarding implementation of neuroscience tests, and labs could be added, as well allowing questions about the test methods.

Another benefit is in mapping and connecting the model to other systems in order to make them cooperate. For example, adding the model to a search system that can search for articles such as Textpresso (Müller, Kenny and Sternberg, 2004; Müller *et al.*, 2008, 2018).

The overall results of the model show that a template-based model that uses ontologies can be very successful in resolving questions. Such a system does not seem to need any sophisticated technique or mechanism in order to achieve better results.

One reason for this could be that this process model is designed and used in a domain such as neuroscience. As discussed in the first study, neuroscience is a suitable domain for applying ontologies.

Also, neuroscience is a restricted domain and there is no need to incorporate sophisticated natural language processing techniques to resolve questions. Furthermore, disambiguating terms is not very hard in such a domain.

However, something that was also observed while running the experiments was that the question- processing module was the bottleneck in the process model. The problem with this module was that some questions that could be answered using templates could not be mapped to the correct template via this module. For example, while the question classification part was doing its job in categorizing questions just fine, it had a hard time mapping questions containing conditions to the correct template. This was because as the definition shows, questions containing conditions basically consist of two separate queries that form a single question. Finding which part of the question is the conditional one and which one is the rest of the question was a problem for the question processing module.

Of course, resolving this problem was outside the scope of this research and towards natural language processing. Resolving dependencies was not an aim of this research. However, a future direction of this study can be to improve the question processing module, since doing so will improve the overall results of the model too.

One way to resolve this issue is through pattern recognition methods such as sequence labelling (Sun, Luo and Chen, 2017). Sequence labelling can be performed in multiple ways. Using conditional random fields (CRFs)(Sutton and Machine, 2012) or BiLSTM-CRF (Lample *et al.*, 2016) seems a reasonable candidates for this study.

4.4.4 Role of Templates

Another aspect of results that is worth discussing is the approach of this research towards creating SPARQL codes through a template-based method. Templates played an important role in this study by providing a general method of resolving questions for each level of them.

Templates added multiple advantages to the model. For example, being based on the question hierarchy, templates inherited the hierarchical form of the question hierarchy. This allowed templates in higher levels to be created on the foundations of lower-level templates and added to the efficiency of creating templates and ease of debugging them.

They played a role in automating the process of resolving questions. This is because by using them, there was no need for a developer to write the codes; and the model, especially its code translation module, can work somewhat similar to pipelines (Bienia and Li, 2012). Pipelines are processes in which codes and tasks run simultaneously in an organized way. Therefore, templates added to the overall efficiency.

The other advantage of templates is that if there is a problem or a need to change how a template works, only the relevant template has to be changed. Therefore, the debugging processes become much easier.

Since templates were based on the question hierarchy and it had defined specification of each level, there is not a chance of problems such as overlap occurrence among them. However, this means that if the question hierarchy changes, templates have to change accordingly, which might have an effect on cost.

The question preparation module, which as discussed, can reduce the results' capability of resolving a wider spectrum of questions. However, reducing questions might lead to a negative side effect, and that is the risk of creating an answer different to what the user or scientist might have been looking for due to incorrect interpretation of the question by the model.

4.4.5 Role of Ontologies

A question might arise on how the model has helped in resolving the questions and what was the benefit of using ontologies in various parts of the model. Therefore, a walkthrough of the model and explanation of how the use of ontologies and the model itself assisted in resolving questions seems necessary.

Therefore, Figure 4-5 demonstrates a high-level view on how the model and ontologies assisted in finding answers for questions. It demonstrates how modules such as the query expansion and disambiguation (module) assisted in finding and expanding the information that was queried for.

The figure also shows what sorts of problems would have been faced if ontologies were not present, and how ontologies were used for resolving the question as a part of the resource integration framework. Furthermore, it discusses why it was not possible to answer certain questions without the model and ontologies.

If ontologies were to be omitted from this study and as a result, the model, the question and the Freesurfer data file would be the only resources for answering questions. This would have left no way for connecting the question to the Freesurfer data file if the user had used a different terminology for creating the questions. Also, it might have led to incomplete answers.

The importance of ontologies in the model is demonstrated through going back to the question that was used as an example at the end of the Results section, "What are the subparts of the white matter of the inferior parietal lobule?" The flow of the process model starts from the red rectangle showing the question.

As can be seen from the image, if after processing the question, the entity was directly asked from the neuroimage data file (the annotated file created by the Freesurfer); then there was a problem, and the information regarding the entity could have not have been found.

This was because the entity with that specific name was not present in that data file. The reason was that MRI images have a certain degree of resolution which matches *left inferior parietal lobule* and *right inferior parietal lobule*, but not *inferior parietal lobule*.

Therefore, using the NeuroFMA, the inferior parietal lobule parts are found and then matched against the neuroimage data. Thus, it can be said that NeuroFMA works as a medium for breaking the inferior parietal lobule into its subparts, which are visible from the data file created by the Freesurfer.

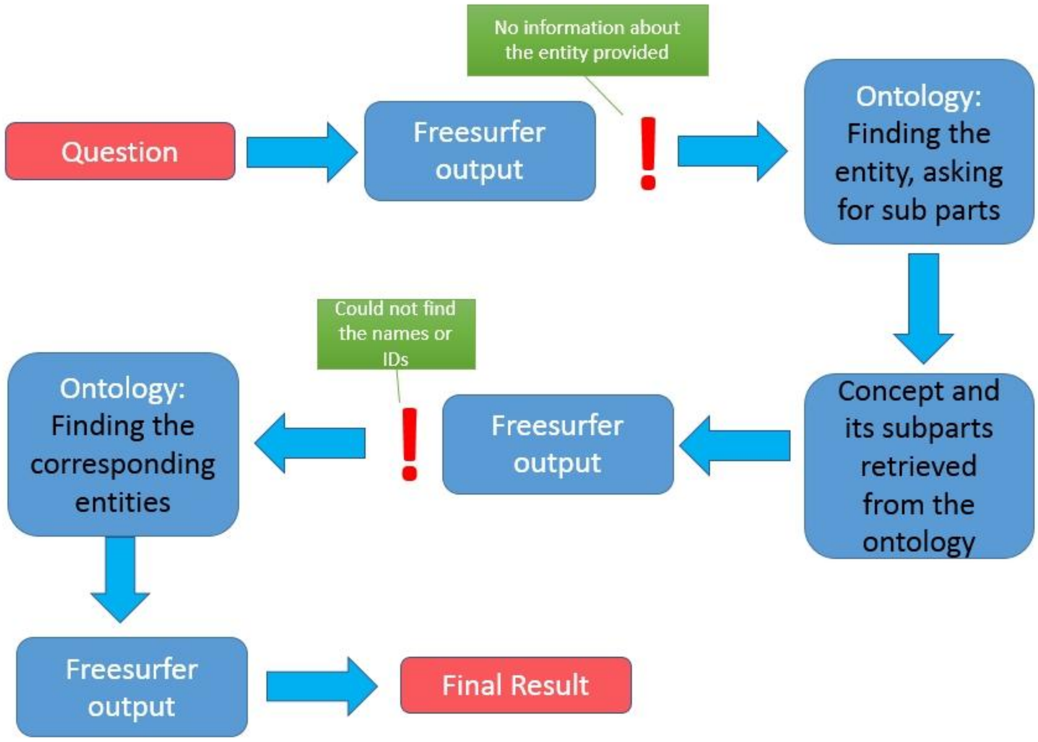

Figure 4-5- Role of ontology in query expansion and resolving ambiguities

Even after having the name of the correct parts, there was a problem and those parts could not be found in the Freesurfer data file easily. This time, the NeuroFMA helped as a medium between the parts in the data file from the Freesurfer and parts found in the other data by having them linked according to their identification numbers.

The above example and explanations demonstrated ontologies being used as a means of disambiguation. It described how ontologies assisted the model in resolving questions and how the lack of ontologies would have made the model unable to do so properly. However, other examples could have been given regarding other uses of ontologies in the model, including the use for finding synonyms and other related disambiguation uses.

In the presented mode, ontologies can be used in order to expand terms and by this, make the search for answers more comprehensive and ensure searching for all possible answers. For example, while searching for a question containing *frontal lobe*, using the ontologies, its synonyms including *precentral gyrus, frontal gyrus, cerebral cortex, cerebral mantle, frontal cortex, cortex and pallium* can be found and used as complementary terms in the search.

The other method was to break down a term into its subparts and then using those subparts in the search in order to have a more comprehensive search. For example, breaking amygdala into left amygdala and right amygdala.

That been said, there were still some problems regarding the scope and comprehensiveness of ontologies, which has to be addressed in the future. For example, some domain-specific words such as *volume* are easy to find from ontology and calculate.

Meanwhile, there are some domain-specific dimensions that do not have a clear definition in the ontology and need user interaction in order to be resolved. Some functional terms such as *activation* should be defined prior to use. Of course, it is worth mentioning that there are attempts by other research teams in to resolve this issue.

Introducing other image modalities, such as fMRI, to this module and taking it towards a multi-modal approach can be another future direction. This way, a much wider area of questions can be posed, and scientists can enrich their results and research via this model.

4.4.6 Role of Standards

What has been experienced during this study was that the power of resolving questions with the aid of ontologies depended on the quality and expressive power of the ontology before anything else.

However, this is only one side of the equation. The other side is that applications cannot be built on ontologies and flourish unless there is a standard way of accessing ontologies. For the general ontologies, there are rules defined by the resource description framework schema (RDFS) and web ontology language (OWL). But there should be some standardizations for neuroscience ontologies too.

The open biomedical ontologies (OBO) foundry (Smith *et al.*, 2007) is an initiation that seeks to standardize biomedical ontologies as well as remove redundancies in these ontologies. It tries to shape some common principles for ontologies.

There have been some efforts to implement OBO instructions in ontologies. OBO relation ontology (OBO-RO) (Smith *et al.*, 2005) is an ontology that has been created in order to enhance relations in biomedical ontologies.

However, OBO instructions and as a result, standardization of ontologies, is progressing slowly. For example, the NeuroFMA ontology has been changed to follow OBO directions, but still have not implemented all relations defined in OBO-RO, such as '*derives_from*', '*has_participant*' and '*adjacent_to*'.

Therefore, this limitation with the NeuroFMA ontology ends in a limitation in querying it via a standardized method that follows OBO instructions. This is while NeuroFMA at least follows OBO partially. The NIFSTD ontology does not implement OBO instructions relations yet. In addition, there are some inconsistencies with the OBO ontologies, such as the owlInObo, which provides OBO to OWL mappings.

It is known that all popular and widely used standards have a standard for naming, and types and values of variables. However, there is not a standard syntax for OBO identification numbers (IDs) or in the ontology property names. Furthermore, while most of its properties use underscore for joining keyword names, only one of them use a hyphen.

In conclusion, defining a set of standards and building ontologies upon them will improve the ease of access and use both in creating and querying the data. Once initiations such as OBO reach their goals, ontologies will become more effective and useful.

4.4.7 Interface Design

Interface and discussions regarding it in a system designed for resolving questions does not only cover how the user input page should look like. It should also be discussing how the questions are inserted into the system. The interface can determine how questions are received not only from the user, but also from other applications.

Systems that resolve questions can either use a keyword-based or text-based interface. In a keyword-based search method, the user builds up the question based on predefined terms commonly derived from a terminology or an ontology.

Some systems that have been discussed or referred to along this research, including NIF (Bug *et al.*, 2008) and TAMBIS (Goble *et al.*, 2001; Stevens *et al.*, 2003), use this method for creating queries. Even hierarchical faceted search interfaces (Hearst, 2006) can be seen as improved keyword-based interfaces.

In text-based methods, the question is received as a text phrase and interactive disambiguation processes in addition to natural language processing techniques often underpin the deconstruction of the question.

Each approach has advantages and disadvantages; for instance, a keyword-based method has high recall, but low precision. However, the text-based method has high precision and low recall (Müller, Kenny and Sternberg, 2004).

The approach the model presented in this study towards receiving questions, is close to a hybrid approach. This means that it received questions as a text-based system and then tried to reduce the questions to a keyword-based input through questions groups.

As mentioned before, the focus of the study described in this chapter was resolving neuroscience questions through the template-based model. Therefore, despite designing and implementing an input mechanism in the model, this study did not perform a very detailed and complex discussion regarding the interface.

However, the model has the capability to be implemented via either approaches (keyword or text-based input). The modular design of the model that was explained before, lets the model to be altered in the future.

Therefore, one of the future directions of this research could be to implement a fully text-based interface, including calculating parameters like estimated answer types (EATs) automatically and through automated techniques such as machine learning.

4.5 Summary of the Chapter

In this chapter, first a background of current research focused on tackling questions in neuroscience, especially via a template-based approach, was given, alongside related information such as ways of translating questions.

Then methods were described in a detailed discussion and based on them, a process of tasks was performed, creating a multi-layered template-based process model for answering questions in neuroscience.

A question hierarchy from the previous study was mapped to groups in this study, and pseudo-codes were created based on them. It was shown how to break down questions into several smaller queries, and then solve them by processing questions via an integrated process model.

Then in the discussion section, the outcomes of the model, its advantages and disadvantages and possible future directions, including connecting the model to other systems, designing interface for the systems and some other directions were discussed.

Several issues including those issues about the integration part of the model that handles mappings, disambiguation processes, and problems regarding creating and executing the model were explored in the Discussion section too.

5 Conclusion

Now that ontologies and their applications have been explored in detail in the first study and two further studies have been performed on both the classification and resolving neuroscience questions using ontologies, a conclusion can be drawn regarding this research.

Therefore, this chapter discusses conclusions and possible future directions of this research. It reviews the aims and research questions stated at the beginning of the research in Chapter 1, and discusses how those questions were answered.

For each study, first a review of research aims discussed in the introduction will be presented. Then, attempts towards achieving those aims and results of those attempts during the course of this thesis will be discussed.

Furthermore, while describing the results, limitations of the research will be discussed. Moreover, during each study, multiple potential future directions were identified. Those directions will be pointed out too.

Please note that this chapter briefly summarizes each study and then discusses their conclusions. Detailed discussions, limitations and future directions of each study are explained in more detail in the discussion section of each study.

At the end of the separate conclusions of each study, an overall conclusion shaped from performing three different but related studies to the ontologies and their application in neuroscience will be presented as an answer for the general aim of the research stated in the introduction chapter.

In the remainder of this chapter, first a conclusion is given on the first study. This is followed with the conclusion of the second study and then, the third study. After that, an overall conclusion is given and in the future of ontologies section, a potential insight to the future of ontologies in neuroscience is discussed.

5.1 Conclusions of the First Study

The first study set out to investigate the versatility of ontology applications in neuroscience. It did so by answering the first research question "What are the applications of ontologies in neuroscience?" This was done by presenting a background of ontologies, alongside their characteristics, significance, their place among similar concepts such as terminologies and taxonomies, followed by mentioning some tools and methods to work with them. Furthermore, it was explained why ontologies are suitable for the neuroscience domain.

Systematic literature reviews were introduced and used as the primary method for this study. Notions related to systematic literature review and systematic review protocol were introduced and discussed too.

In the Results section, information from eight major databases and three reputable journals was investigated through an intense search. Approximately 1000 primary research papers were reviewed, and finally 208 results were selected as the outcome of full-text analysis.

Different applications of ontologies in neuroscience were pointed out based on the 208 results from the systematic search and selection process described earlier. Then, the results were analysed and homogenous applications were grouped together as a result of a detailed and accurate process.

Results pointed out several applications for ontologies in neuroscience, including search and data retrieval, data capture and representation, integration, collaboration, classification and categorization, disambiguation, knowledge management and data analysis. Outcomes of the study on the versatility of ontology applications in neuroscience exceeded the expectations and speculations.

There were several popular and well-known applications for ontologies before this study, including mainly disambiguation and integration. However, this study shed light on applications of ontologies that were not very popular, such as resource organization in neuroscience.

In addition, even those popular applications of ontologies are much more detailed and can be easily divided into different subbranches. Research papers that used them, executed their studies by implementing and using different methods which shows their different understanding of some concepts such as integration using ontologies.

However, it is worth mentioning that being unpopular does not mean they are totally new applications. It means the use of these successful applications has been very limited in neuroscience so far and using them in future research might open new possibilities in this field. Therefore, there are research areas in the application of ontologies that still need further examination. Ontologies can be applied to areas that have seen little presence of them to provide better results.

It was discussed that the information provided in the Result section can be used for initiating other studies because of containing different directions in the field in addition to a list of quality work in each domain. Scientists can use this information for setting up their own studies.

It is worth mentioning that this research has followed one of the future directions discussed here; because as discussed before, outcomes of this study were used as a starting point for the next two studies described in this research.

A limitation of this study included the number of databases and journals investigated in this study. Despite the large number of reputable resources, there are still some available that can assist in achieving a better understanding of application ontologies in neuroscience.

Therefore, expanding the resources of this study can be seen as a future orientation in examining application of ontologies in neuroscience by adding new databases to the resources.

Performing a study like the one presented in this research in the biomedical field is also seen as a possible future direction. This is due to the similarity of both fields and the large scope of ontologies application and assistance in neuroscience.

Risk of bias was another limitation of this study. However, efforts were made to avoid this risk. These efforts included consulting librarians regarding the methods of performing the search, consulting computer scientists to check the technical matters regarding the ontologies and consulting a neuroscientist to double check the results.

Therefore, setting up a team for the next updates on this study can be seen as a future direction. However, this direction is more toward the methods of practice rather than a concept or area of interest and study.

5.1.1 Update to the First Study

Since from the time of the systematic literature review to the time of submitting this thesis new research papers were published, so an update of the literature seemed necessary. It was performed before finalizing the research in order to check if the results of the study are still accurate, correct and applicable to the field.

The criteria and protocol for searching for research papers followed the ones explained in the first study discussed in the second chapter. However, results were limited to the PubMed database. The date of this second search was started from when the initial search was stopped which was September 2016 and covered all papers up to October 2019.

This search resulted in 351 new research papers. After the initial review of this papers, 34 of them were selected as relevant research papers. All the relevant research papers were studied thoroughly.

Analysing research papers showed that ontologies are getting more popular even in areas outside neuroscience and its related fields. For example, Dedhia *et al.*, (2019) have introduced high school students to the gene ontology classification system.

The data analysis aspect of ontologies is still present in many research including (Babenko *et al.*, 2016), which used ontologies in comparison of enhanced aggressiveness and tolerant behaviour between lab animals; Bordi *et al.*, (2016), which used ontologies in evaluating the autophagy process in Alzheimer disease, Miyata *et al.*, (2016), which used ontologies in a study of depression during menopause; the work of E Hirbec, Noristani and Perrin (2017), which used ontology classification to compare activated pathways in different neuropathological conditions; Davis *et al.*, (2018) which used ontologies for finding phenotype similarities and differences in brain of mouse strains and Morgan *et al.*, (2019), which used the ontologies in studying schizophrenia,

The same way, research such as (Koskimäki *et al.*, 2019) used ontologies to find information regarding diseases. This specific research tries to identify major genes, functions and neural functions that play a role in pathobiology of cerebral cavernous malformation.

On a larger scale, Swanson (2018) has created a great foundation for other research by releasing the last version of Brain Map which contains the structure of the rat brain. The information presented in this knowledge base, forms an ontology for knowledge management systems that deal with neural circuitry.

Data representation benefits of ontologies is also shown in works such as the Smart Protocols Ontology (SMO) (Giraldo *et al.*, 2017). This ontology is used for representing experimental protocols. Moreover, the work presented by O'Reilly, Iavarone and Hill (2017) applies ontologies in an annotation process, while creating a framework for collaborative curation of neuroscientific literature.

Timón, Rincón and Martínez-Tomás (2017) try to improve data sharing and integration by extending a platform called XNAT (Herrick *et al.*, 2016) which is an open source neuroimaging platform. Koopmans *et al.* (2017) uses ontologies while creating a data repository which is enabled with data mining tools called AHCODA-DB. Using ontologies in this study leads to better data export and as a result, sharing.

5.2 Conclusions of the Second Study

The second aim of this study was to examine how neuroscience questions can be classified using ontologies and how well ontology-based classification suits neuroscience; in other words, an answer to the research question 'How can questions in neuroscience be classified using ontologies?'

For this means, first questions were gathered from two reputable resources including experts from two different research teams and publications. Then, an expert from one of the teams inspected the questions for validation purposes.

Then questions were analysed and similar elements among them were outlined. These elements were called dimensions and were used to represent and model questions. Further to this, a question category or hierarchy was created using these dimensions. The hierarchy was later used for assigning categories to questions in question classification and in the third study for creating templates.

Afterwards, attempts towards classification were initiated by using the ontology-based approach. Questions were parsed, tokenized, and lemmatization was done. The ontology-based classification was performed and it achieved a 93 percent correct classification.

Next, multiple statistical classification techniques were performed to draw a better conclusion on the effectiveness of the ontology-based classification. Statistical (machine learning) classification techniques included SVM, Naive-Bayes, Random Forrest and KNN and they respectively achieved 73 percent, 54.5 percent, 77 percent and 68 percent correct classification.

Apart from showing that ontology-based classification was effective, results achieved while using statistical techniques demonstrated that using dimensions are useful in these techniques. This was because using dimensions as features while classifying questions using statistical techniques, improved classification percentages from at least 8.5 percent up to 30 percent.

Performing both ontology-based and statistical approaches towards question classification shows that ontology-based classification in neuroscience is powerful and can achieve a good classification.

Furthermore, it does not need a training set, unlike popular statistical techniques such as the supervised ones. It can be said that ontology-based classification works like a semi-supervised technique.

However, designing and implementing ontology-based classification is hard, very time consuming, costly from the computational point of view and is very dependent on the quality of the ontology from which it is created upon.

One limitation for statistical techniques was that the dataset was not very big- this might result in statistical techniques not being able to perform at their best. However, improving results of the statistical techniques was not in the scope of this study and research.

It tried to compensate for this shortage using several methods including the leave-one-out (K-fold) cross-validation technique. However, gathering more questions and experimenting with an expanded set of statistical techniques can be seen as a future direction of this study.

Using a static question hierarchy might result in being left with some questions that cannot be grouped in any category. However, there are methods for overcoming and bypassing this problem. These methods were discussed in the third study where question hierarchy was used for creating templates.

Nevertheless, investigating methods for overcoming this shortage can be seen as a future direction for this thesis. These methods can be using weights for dimensions to which numbers are assigned or just using dimensions without the question hierarchy alongside machine learning techniques in classification.

The other future direction can be to perform the classification based on other factors since it can be of use and assistance. Classification should not be viewed as a simple concept or tool. Sometimes a classification based on a critical and important factor can improve user experience in a system or even answer a research question as described in the Result section of the first study.

For example, different scientists interact with information and systems in the neuroscience domain including biologists, neurosurgeons, carers and other people working in health informatics. They all use their own terminologies for accessing the neuroscience data.

Even within the neuroscience domain a part of the brain which is called 'Brodmann Area IV' by a cognitive psychologist, is called 'Primary Motor Cortex' by a behavioural neuroscientist. Therefore, having a system that classifies questions based on the user in such an environment can contribute in achieving a rich and tailored result.

Another example would be to classify questions according to the resource they need to be answered with. This classification will help with cutting the computational cost and time, since there are various resources in neuroscience including neuroimages, text documents, lab experiment descriptions, publications, trials of experiments and other types of digital resources.

Therefore, preselecting the resource with the aid of classification will cut the cost by several degrees. Ontologies that can assist with such approaches have been discussed in the Result section of the first study of this research.

Anther future direction for this study can be to use the dimension set introduced in this study as a question terminology via mapping to other ontologies. This terminology can then be used for annotating neuroscience questions.

As explained before, there are other ontologies in neuroscience apart from the NIFSTD used in this study. Those other ontologies, such as the NeuroFMA, can be used to expand this study and investigate the question terminology described in the previous paragraph.

It was described that lemmatization is achievable using ontologies. However, it is a time consuming and costly process. Therefore, creating algorithms and protocols for performing automated lemmatization through ontologies in neuroscience can be a future direction.

The last future direction of this study was to use the question hierarchy in a system that resolves questions. This can be done since questions can be placed into this question hierarchy. Furthermore, each level of this hierarchy has a structure that can be translated to code. This is a direction that was further investigated in the third study of this thesis.

5.3 Conclusions of the Third Study

This study was undertaken in order to answer the third research question, which was 'How can questions in neuroscience be answered using ontologies?'. It set out to examine the power of ontologies in answering neuroscience questions.

For this means, the study used the question hierarchy introduced in the previous study and expanded it to some templates that answered questions by utilizing ontologies. After that, a question process model was created around the templates by connecting different sections of the study including question processing, templates and the integration module.

Templates that were in the form of pseudocodes and translatable to SPARQL codes, which was the coding (querying) language in this study, were created according to the question hierarchy introduced in the previous study.

Templates played a role in automating the process of resolving questions. They did this by translating the question to machine understandable format. Therefore, they assisted in automation and moreover, they freed scientists from learning a coding or querying language and let them focus on their research.

The other benefit of using templates lies in flexibility; because if a change was needed as a result to a problem or improvement, changing templates would probably fix the issue and there is no need for redoing an extensive group of tasks. For example, if there is improvement with the SPARQL, changing the templates would be enough.

However, this advantage can backfire in some certain cases. For instance, if there is a need in the future to add a new type of question and this change leads to an overlap among templates, then the templates that overlap have to be changed.

Ontologies also played a significant role in answering questions in this chapter. After all, this study was designed to investigate their role in answering neuroscience questions. They have been used in various parts of the approach discussed in this study, including for disambiguation attempts and creating an integration platform.

They have been used for disambiguation purposes by being a means for resolving or expanding the neuroscience terms. Furthermore, ontologies acted as mediators between different data resources including neuroimage segmentation applications and other ontologies. Moreover, they acted as translators between data resources and terms used in questions.

Therefore, multiple uses of ontologies and advancements achieved through their use demonstrates this study was not possible without them. This was not just because part of process in this study was based on ontologies or the way this study was designed, but because certain tasks performed in this study were not achievable without using ontologies.

The third study was different from the second study since there were multiple ways for performing each task present in the second study, but certain capabilities of the third study was solely based on ontologies.

However, using ontologies in answering neuroscience questions has some burdens too. These burdens include the cost of using them. The third study demonstrated that ontologies need powerful machines to work properly. Even while using powerful systems, querying them is tedious and time consuming.

The other issue with ontologies is the availability and completeness of information. Despite providing huge amounts of information, some ontologies are still being completed. Therefore, a scientist designing a system based on ontologies might hit some barriers while using them.

A natural progression of this study is to expand it using other ontologies. This can be done by adding to ontologies of the integration module of the model. This will allow scientists to pose a wider scope of questions and reach more expressive results.

Moreover, the study can be advanced further by adding other neuroimage modalities. Adding other neuroimage modalities such as fMRI and taking this study towards multi-modal research opens more possibilities for scientists to achieve far more enriched results by asking questions that can only be answered by those specific modalities.

A remaining concept that was introduced during this study was the process model. The process model was shaped by connecting different parts discussed in this research, which included the question processing and integration modules.

The presented model can be used not only by the neuroscientist, but scientists in other branches such as neurosurgeons, biomedical scientists, care-givers and several other professions. It is capable of resolving different types of questions that were discussed in the Discussion section of chapter 4 in detail.

The model can tackle questions regarding various parts and functions of the brain including cell and subcellular entities, functions, anatomy and structure and experiments. Questions can contain synonyms, term hierarchies such as parents and children of a term, comparisons, patient related data, changes in the data, conditions, description of terms and aggregation phrases.

One advantage of the process model that allows changes to take effect easier is its modular design. This modular design leads to flexibility; this means that it allows the model to be improved by changing different modules instead of changing it as a whole and by this, reduces the cost of changes.

The question processing module of the model was essentially a module shaped around the templates. This module worked as a translator of questions into machine understandable format. As discussed, it was based on the question hierarchy from the second study. It was slightly changed and expanded in order to categorize questions into templates.

Further studies on improving this module using techniques such as sequence labelling or semantic similarity techniques can be a future direction. By using machine learning techniques, or semantic similarity techniques and matching questions to the ontologies directly, using templates could be minimized.

This minimization might be beneficial since as discussed above, altering templates can become very costly in case of an issue such as overlap between templates. However, all these suggestions need accurate further investigations before being implemented.

The process model introduced in this study demonstrated how a question can be processed and resolved from start to the end. Building a system or web portal with an interface for users based on the instructions given in the model introduced in this study can be a future direction. The information regarding the interface design was discussed in chapter 4. It is worth mentioning that ontologies can help with interface design too.

This model can be connected to other applications in neuroscience or biomedical science such as Textpresso or an ontology-based mining system for competitive intelligence in neuroscience (OMSCIN). This can be done using various techniques such as ontologies or even using web-based services such as simple object access protocol (SOAP) or web service definition language (WSDL).

The scope of this research was kept within the boundaries of ontology applications. However, a future direction for the whole approach in the third study could be to use ontologies as information containers alongside techniques such as machine learning.

However, further investigations are needed to decide which technique to use since it has to have the potential to be integrated and mixed with the ontology-based approach described in this research.

5.4 Overall Research Conclusions

In previous sections, a short summary alongside the conclusion of each study, their limitations and future directions were discussed. In this section, the whole research (thesis) will be reviewed to reach an overall conclusion.

In the first study, applications of ontologies in neuroscience were investigated. It was shown that ontologies are very useful in this field. In the second study, the power of ontologies in presentations and classification were discussed and in the third study resolving neuroscience questions with the help of ontologies was discussed.

To continue this research further, some tasks can be done as overall future directions. Namely, redesigning the question hierarchy according to the outcomes from the third study and adding a sequence labelling method to the question processing module.

However, while performing the above studies, some points were raised that might lead to new horizons and some different, but related, directions for this research. Here, some of them are pointed out briefly.

When questions like “What is the volume of amygdala in elderly people?” or “What are the negative effects of Alzheimer’s Diseases on patients?” are asked, it is not easy for an automated system to find their answer.

The reason is that terms such as ‘negative’ and ‘elderly’ are not definitive terms, but they are relative. The process model described in this research is able to resolve these words through user interaction and asking an expert to resolve them. However, then the process will not be automated and questions have to be resolved case by case.

Therefore, obtaining more information or heuristic techniques are needed for these terms in order to enable an automated system to establish answers for questions containing them. Finding such information or techniques can be the subject of future investigations. One potential method for clarifying these terms can be the utilization of fuzzy logics.

The other issue with ontologies in neuroscience is standardization. Despite the effort by the world wide web consortium (W3C), such as introducing OWL2 or bodies such as OBO foundry (Smith *et al.*, 2007), standardization still needs a large amount of time and effort to become popular and beneficial.

Considering the OBO as an example, some ontologies like NIFSTD have not yet implemented their standards, while some like FMA (NeuroFMA) have partially implemented OBO standards. Furthermore, there are some inconsistencies with their standards.

This leads to difficulties while experimenting with ontologies or implementing models and systems based on them. This is because there has to be some level of curation and specification for each ontology in such a system so that the user can see it as a useful and consistent system.

Many times in this research it was demonstrated that certain calculations or processes could have been performed easier and with better precision if their foundations or definitions were provided in the ontology.

This is why contributing to the development of ontologies that discuss tasks and processes in neuroscience is very important; especially in research regarding tackling questions or any other automated task.

It is clear from results of this thesis that it could benefit a lot from such ontologies. Taking advantage of terminologies and ontologies such as the Cognitive Atlas (Essen *et al.*, 2013) once they are ready for use will enhance results of process models like the one introduced in this research.

In general, according to discussions until now, ontologies can be very powerful both in creating and enhancing a variety of systems and in different applications. However, benefiting from the full potential of ontologies has some prerequisites.

Ontologies will reach their full potential once new ontologies that describe processes such as the above example are completed. Moreover, ontologies and standards describing detailed information in neuroscience and biomedical ontologies such as OBO-RO and OBO are completely accepted and implemented by all ontologies across the field.

The other direction for this research is to investigate effects of its implementation or even implement it in other fields of study. This research did not examine using its proposed solutions and models in other domains.

Doing so seems possible with a little bit of work. It does not appear to be very costly since the modular design of this research might assist the overall process. What basically needs to be done is to replace neuroscience data with data from the destination domain. Of course, there will definitely be some obstacles and compatibility issues to be taken care of.

In general, it seems the approach and methods used in this research can be used in fields outside neuroscience or even biomedical science. Example of other fields that can benefit from this research are geography studies, cartography and urban planning.

This is because just like there are different versions of brain segmentations, there are different versions of maps. Furthermore, there are some older maps that have to be mapped or integrated with newer versions. Ontologies and therefore, this research, are good choices for undertaking such studies.

Representing, classifying or resolving neuroscience questions could have been performed via various tools and techniques. Some of those, such as strictly statistical techniques, fuzzy logics and pattern recognition techniques, have been discussed in the Discussion sections related to each study and in this chapter. However, the reason that this research used ontologies instead or alongside some of those techniques, was because of multiple reasons.

These reasons included the benefits of using ontologies, which were discussed in Chapter 2 of this research, such as the expressive power of using domain knowledge, especially in restricted domains like neuroscience; if and how ontologies can assist goals such as question analysis and resolution; moreover, as mentioned in the preface, this research was a part of a grant from Australian Research Council (ARC). It was designed as a component to a data management platform and the project required to use terminologies or ontologies. This research was set out to uncover the role and effectiveness of ontologies in neuroscience and their application in processing neuroscience questions using ontologies.

The outcomes of this research are significant, not only because of the reasons that were given in the first chapter, the Introduction and Discussion section of each chapter; but also, because of other reasons. These reasons include providing an understanding of the brain for different groups of scientists and users; demonstrating various applications of the ontologies in neuroscience and possibly other fields; also, opening the restricted and specialized domain of neuroscience to general science by using ontologies.

This research provides an understanding of the brain by processing neuroscience questions for various types of scientist. Neuroscientists, neurosurgeons, neurologists, neurophysiologists, psychiatrists, healthcare workers and even the general public can be among the users of the system presented in this research.

In addition, other applications and systems, such as data management systems can benefit from this system in the form of a question resolution module too. Specifically, since this research provided a summary of ontology uses and showed it can be used successfully in certain parts of the neuroscience research.

Furthermore, this research can be used in conjunction with web-based services and other huge brain atlases, databases, or general systems like Google. This is because ontologies are very close to concepts such as knowledge-bases or knowledge graphs and linked data. Therefore, this research and its outcomes can be connected to, or shared with services, data storages and systems mentioned above. This characteristic of the research opens up a whole new variety of applications.

5.4.1 A Brief Discussion on the Different Views of Ontologies

Here and after experimenting and discussing the ontologies in detail in this research, I want to point out some views about ontologies alongside my own view, which was shaped after performing this in-depth research on the ontologies.

Some researchers view ontologies as just another trivial tool. They might use ontologies in small sections of their work, like the bag-of-words in a machine learning technique. This is so even while the title of their research or platform states ‘ontology application’.

The reason for this is mostly because of incorrect or insufficient understanding of ontologies. This misunderstanding sometimes even leads them to mistake concepts similar to ontologies (such as taxonomies) for ontologies.

On the other hand, some researchers think ontologies are some of the most powerful tools and answer many problems in computer systems, especially in the semantic domain of biomedical science or neuroscience.

One example is in-depth and strong research papers such as (Chandrasekaran, Josephson and Benjamins, 1999). Despite possibly being old for the fast progressing field of computer science, this research paper is still valid and accurate. It puts ontologies and basically all content theories on top and ahead of different techniques and approaches.

After reviewing many research papers on the topic of ontologies and their applications during this research, and in the hindsight of performing different experiments with ontologies in this research, I think it is safe to say both these perspectives on ontologies are correct to some extent.

In fact, as I explained before, ontologies can be very useful and can have great effects on a system or they can just be very trivial tools that are rarely used in answering and resolving serious issues. It all depends on the quality of the design of the system or ontology.

If an ontology is designed carefully and well, it can be effective. If not, then the ontology would be of little benefit to a system. Similarly, a very detailed and carefully curated ontology cannot be effective if it is applied to a system with a poor design.

The best result would be achieved when a carefully curated ontology is used in a very carefully designed system. It would be even better if the system benefits from a fast and advanced mechanism or approach such as new machine learning techniques or fuzzy logic systems as well as ontologies.

This brings me to an enhanced definition of ontologies that I think best suits them. I have to point out that I do not see myself in a place to offer a new definition, neither do I believe that I have a better understanding from the ontologies than those elite researchers who already gave definitions for ontology in computer science.

However, I think a more explicit definition for ontologies might help in understanding their role better. In my opinion, a more detailed definition will demystify ontologies and assist scientists in using them more often and according to what ontology delivers. Moreover, I think this will help scientists with a native language other than English.

As I explained in the second study, there are different definitions for ontologies used in the computer science field. From Gruber's definition which is 'a specification of a conceptualization' (Gruber, 1993), to the enhanced version proposed later by Studer et al. (1998) that says 'a formal, explicit specification of a conceptualization' or 'ontology is a formal specification of a shared conceptualization', to the definition from Chandrasekaran (1999) described as 'content theories about the sorts of objects'.

My enhanced definition of ontologies that resulted from studying them during this research is 'ontologies are containers of a formal specification of shared conceptualization that can act as additional sources of data and maps for their affiliated domains'. I believe this definition should be valid across the field of computer science.

Furthermore, I have to state again that this is not a totally new definition and as the term 'enhanced' shows, my definition is based on those formulated by the great scientists I mentioned above.

5.5 The Story of Ontologies in Neuroscience

The three different studies explained in this document depict a vivid picture of ontologies in neuroscience. Therefore, and based on the information given so far, this section seeks to speculate on a potential future overview of ontologies in neuroscience.

Ontologies have been used in science for a long time. In neuroscience, they have been used for no more than two decades, with the first major instance being the gene ontology (Ashburner et al., 2000).

From 2000 to 2006, gene ontology was basically the only major real ontology in neuroscience. It was only after 2006 that ontologies became popular and used widely in neuroscience for different purposes, including integration and information retrieval.

From that time on, ontology applications have been on a steady rise. They have been used in various researches and proved to be beneficial. They have also changed the landscape of the data and research.

Ontologies are in fact one of the symbols of semantic computing. After being introduced as means of the semantic web and semantic computing, they gained popularity as a method of adding extra metadata and information to the data.

They changed the form of the data from unstructured to structured, in a way that data became self-explanatory and made sense by itself. This way, machines could understand the data and perform sophisticated automated tasks on it.

Studying the current state of the ontologies can assist in drawing a view for their future, both for their evolution and application. At the moment, the gene ontology is still the most popular ontology.

Semantic approaches are the core of any semantic system. Achieving semantic approaches is possible through using tools such as ontologies because of their specifications generally discussed in this research and specifically in the above lines.

However, in order to make ontologies a better choice for application in semantic approaches, some improvements have to be made. These improvements can be achieved via expanding ontologies to different aspects of the neuroscience research, such as explaining how tasks, calculations and processes should be done.

1 Appendix A: Database Search for Chapter 2

This appendix is related to chapter 2 (study1). Chapter 2 was trying to find out an answer for the following research question: “what are the uses of ontologies in neuroscience knowledge management?”. This appendix discusses detailed search approach which was performed for finding related documents.

As described in the Method section of Chapter 2, the initial search performed using neuroscien* and ontolog* keywords. The asterisk (*) wildcard was used as truncation in order to retrieve all different variations of keywords. Eight databases including PubMed, Medline, IEEE, ACM Digital Library, Compendex, Scopus, Science Direct and Web of Science were searched. Search methods were updated as the research progressed. The detail of each database search is as follows.

1.1 PubMed

Search was performed using the two keywords (neuroscine* and ontolog*) and the time of articles from earliest possible to 01 September 2016. Overall, a total of 334 documents were retrieved as the result of this search. The advanced query for this search is shown below.

> Query phrase: *(neuroscien[All Fields] OR neuroscienc[All Fields] OR neurosciencc[All Fields] OR neuroscience[All Fields] OR neuroscience'[All Fields] OR neuroscience's[All Fields] OR neuroscience1[All Fields] OR neuroscience162[All Fields] OR neuroscience6[All Fields] OR neuroscienceaging[All Fields] OR neuroscienceamsterdam[All Fields] OR neuroscienceand[All Fields] OR neuroscienceandnorms[All Fields] OR neurosciencebdepartment[All Fields] OR neurosciencebeer[All Fields] OR neuroscienceberlin[All Fields] OR neuroscienceblueprint[All Fields] OR neurosciencebmrc[All Fields] OR neuroscienceboston[All Fields] OR neurosciencebunteweg[All Fields] OR neurosciencecentro[All Fields] OR neurosciencechu[All Fields] OR neurosciencedalhousie[All Fields] OR neurosciencedepartment[All Fields] OR neurosciencedublinireland[All Fields] OR neuroscienceemory[All Fields] OR neuroscienceeuronmaastricht[All Fields] OR neurosciencefaculty[All Fields] OR neurosciencefunctional[All Fields] OR neurosciencegeriatric[All Fields] OR neurosciencegoettingen[All Fields] OR neurosciencegottingen[All Fields] OR neuroscienceicahn[All Fields] OR neuroscienceina[All Fields] OR neuroscienceinstitute[All Fields] OR neurosciencekarolinska[All Fields] OR neuroscienceking's[All Fields] OR neurosciencekings[All Fields] OR neurosciencelondon[All Fields] OR neurosciencem[All Fields] OR neurosciencemaastricht[All Fields] OR neurosciencemax[All Fields] OR neurosciencemcknight[All Fields] OR neurosciencemilan[All Fields] OR neurosciencemilano[All Fields] OR neurosciencemunich[All Fields] OR neurosciencemunster[All Fields] OR neurosciencenational[All Fields] OR neurosciencenewcastle[All Fields] OR neurosciencenortheastern[All Fields] OR*

neurosciencepadova[All Fields] OR neurosciencepadua[All Fields] OR neurosciencephysical[All Fields] OR neurosciencequeen[All Fields] OR neuroscienceriken[All Fields] OR neurosciencerose[All Fields] OR neurosciences[All Fields] OR neurosciences'[All Fields] OR neurosciencesbangalore[All Fields] OR neurosciencescambridge[All Fields] OR neurosciencescience[All Fields] OR neurosciencesclinical[All Fields] OR neurosciencesdaggerjanos[All Fields] OR neurosciencesdepartment[All Fields] OR neuroscienceshenzhen[All Fields] OR neurosciencesinstitute[All Fields] OR neurosciencesku[All Fields] OR neurosciencesmcmaster[All Fields] OR neurosciencesmedical[All Fields] OR neurosciencesmrc[All Fields] OR neurosciencesophie[All Fields] OR neurosciencespadova[All Fields] OR neurosciencesparis[All Fields] OR neurosciencesradboud[All Fields] OR neurosciencesroyal[All Fields] OR neurosciencessemmelweis[All Fields] OR neurosciencestel[All Fields] OR neurosciencestephanstr[All Fields] OR neurosciencesuniversity[All Fields] OR neuroscienceszamzuri[All Fields] OR neurosciencetel[All Fields] OR neurosciencethe[All Fields] OR neurosciencetokyo[All Fields] OR neurosciencetraining[All Fields] OR neurosciencetrieste[All Fields] OR neurosciencetrondheimnorway[All Fields] OR neurosciencetubingen[All Fields] OR neuroscienceucl[All Fields] OR neuroscienceuniversity[All Fields] OR neuroscienceuppsala[All Fields] OR neuroscienceverona[All Fields] OR neuroscienceweill[All Fields] OR neurosciencewest[All Fields] OR neurosciencezamudio[All Fields] OR neurosciencia[All Fields] OR neurosciencias[All Fields] OR neurosciencies[All Fields] OR neurosciencists[All Fields] OR neurosciencs[All Fields] OR neurosciencve[All Fields] OR neurosciene[All Fields] OR neuroscienec[All Fields] OR neuroscienee[All Fields] OR neurosciences[All Fields] OR neurosciense[All Fields] OR neuroscientfic[All Fields] OR neuroscienti[All Fields] OR neuroscientic[All Fields] OR neuroscientific[All Fields] OR neuroscientifically[All Fields] OR neuroscientifique[All Fields] OR neuroscientifiques[All Fields] OR neuroscientist[All Fields] OR neuroscientist's[All Fields] OR neuroscientists[All Fields] OR neuroscientists'[All Fields] OR neuroscienza[All Fields] OR neuroscienze[All Fields] OR neuroscienzes[All Fields]) AND (ontolog[All Fields] OR ontologia[All Fields] OR ontologiajarol[All Fields] OR ontologias[All Fields] OR ontologic[All Fields] OR ontologica[All Fields] OR ontological[All Fields] OR ontological'[All Fields] OR ontologically[All Fields] OR ontologicas[All Fields] OR ontologici[All Fields] OR ontologicial[All Fields] OR ontologico[All Fields] OR ontologicos[All Fields] OR ontologie[All Fields] OR ontologiek[All Fields] OR ontologien[All Fields] OR ontologies[All Fields] OR ontologies'[All Fields] OR ontologique[All Fields] OR ontologiques[All Fields] OR ontologische[All Fields] OR ontologischen[All Fields] OR ontologist[All Fields] OR ontologist's[All Fields] OR ontologistic[All Fields] OR ontologists[All Fields] OR ontologists'[All Fields] OR ontologizacion[All Fields] OR ontologization[All Fields] OR ontologize[All Fields] OR ontologized[All Fields] OR *ontologizer[All Fields] OR ontologizing[All Fields] OR ontology[All Fields] OR ontology'[All Fields] OR ontology's[All Fields] OR ontologyand[All Fields] OR ontologybased[All Fields] OR ontologybiological[All Fields] OR ontologyentry[All Fields] OR ontologyentry'[All Fields] OR ontologyfingerprint[All Fields] OR ontologyindex[All Fields] OR ontologyinformation[All Fields] OR ontologyplot[All Fields] OR ontologysimilarity[All Fields] OR ontologytrade[All Fields] OR ontologytraverser[All Fields] OR ontologywidget[All Fields] OR ontologyworks[All Fields] OR ontologyx[All Fields]) AND ("0001/01/01"[PDAT] : "2016/09/01"[PDAT])*

1.2 Medline

The search for Medline was done on its web of science interface. The initial search using keywords from year 1950 up to the day of the search retrieved 116 results. Then, the results were filtered by Mesh headings based on previous searches and results related to neuroscience were selected including neuroscience, brain, brain mapping, models neurological, neurons, magnetic resonance imaging, neuroanatomy, biological ontologies, nerve net, image processing computer assisted, diagnostic imaging, schizophrenia, neuronal plasticity, neuroimaging, natural language processing, mental processes, mental disorders, memory, medical informatics, gene expression profiling, emotions, behaviour, social behaviour, unified medical language system, neurophsycology, neurology, neurobiology, neural pathways, nervous systems, cerebral cortex, anatomy and aging.

In addition, no time span for documents were defined and the language was limited to English and then, the final results came to 67 documents. The advanced query phrase can be seen below.

Query phrase: *TOPIC: (neuroscien* ontolog*)*

Refined by: *MeSH HEADINGS: (DIAGNOSTIC IMAGING OR NEUROPSYCHOLOGY OR NEUROSCIENCES OR SCHIZOPHRENIA OR NEUROLOGY OR NEUROBIOLOGY OR BRAIN OR NEURAL PATHWAYS OR NERVOUS SYSTEM OR BRAIN MAPPING OR NEURONAL PLASTICITY OR NATURAL LANGUAGE PROCESSING OR MENTAL PROCESSES OR MENTAL DISORDERS OR MEMORY OR MODELS NEUROLOGICAL OR MEDICAL INFORMATICS OR CEREBRAL CORTEX OR GENE EXPRESSION PROFILING OR NEURONS OR EMOTIONS OR MAGNETIC RESONANCE IMAGING OR BEHAVIOR OR UNIFIED MEDICAL LANGUAGE SYSTEM OR ANATOMY OR NEUROANATOMY OR AGING OR BIOLOGICAL ONTOLOGIES OR SOCIAL BEHAVIOR OR NEUROIMAGING OR NERVE NET OR IMAGE PROCESSING COMPUTER ASSISTED) AND LANGUAGES: (ENGLISH)*

Timespan: *All years.* Indexes: *MEDLINE*

1.3 Science Direct

The science direct database was accessed on 1 November 2016 and no limitations was posed on the year; therefore, articles were retrieved from 1823 till the data of access. Resources were limited to journals, all subscribed publications and open access articles. In addition, the field was limited to neuroscience and filters were applied on topics such as brain, schizophrenia, Alzheimer disease and Alzheimer. The result of this search was 153 documents. The advanced search query is as follows.

> Query phrase: *ALL(neuroscien* and ontolog*) AND LIMIT-TO(contenttype, "JL,BS,RW","Journal,Reference Work") AND LIMIT-TO(topics, "brain,schizophrenia,alzheimer disease,alzheimer").*

1.4 Compendex

This database was accessed through engineering village on 2 November 2016. No date limit was enforced; therefore, the search period was from the year 1884 to the present day. Keywords were searched in fields such as subject, title, abstract of the documents. The search language was limited to English. The final result of this search included 67 results and the query phrase is as follows.

> Search phrase: *((((((neuroscien*) WN KY) AND ((ontolog*) WN KY)))) AND ({english} WN LA))*

1.5 Scopus

Scopus was the next database which was accessed on 12 November 2016. No time span was considered for the search, therefore documents were searched from 1960 till the access date.

Search keywords including neuroscien* and ontolog* were applied on subject, title, abstract and field. The field was limited to sciences which includes neuroscience. The number of results equalled 132 documents at this stage.

Then, subject areas other than neuroscience were filtered as life science was much broader and included other fields too. Articles, conference papers, book chapters, reviews, articles in press and short surveys were among the documents selected. The final number of results included 92 documents. The advanced search query for this search comes below.

> The Final search query: *(TITLE-ABS-KEY (neuroscien*) AND TITLE-ABS-KEY (ontolog*)) AND SUBJAREA (mult OR agri OR bioc OR immu OR neur OR phar) AND PUBYEAR < 2017 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "re") OR LIMIT-TO (DOCTYPE , "ch") OR LIMIT-TO (DOCTYPE , "sh") OR LIMIT-TO (DOCTYPE , "cp") OR LIMIT-TO (DOCTYPE , "ip")) AND (LIMIT-TO (SUBJAREA , "NEUR"))*

1.6 IEEE

IEEE (Institute of Electrical and Electronics Engineers) was another database searched during this study. The database was accessed on 30 October 2016 and no time limits were defined. The keywords were applied on the search engine of the database and resulted in only 17 answers. The search phrase is as follows.

> Search phrase in IEEE: ((neuroscien*) AND ontolog*)

1.7 ACM Digital Library

The ACM Digital Library was accessed on 2 November 2016 and no time limit was enforced in the search. Applying search keywords via two different queries on specific fields such as title and abstract resulted in two sets of results of 76 and 74 documents (total of 150 search results). It worth mentioning that two searches had many results in common. The two search phrases are as follows.

> First search query which resulted in 76 results: *acmdlTitle:(ontology ontologies ontological onto neuro neuroscience neurosciences neuroscientific) AND recordAbstract:(neuro neuroscience neurosciences neuroscientific)*
> Second search query which resulted in 74 results: *recordAbstract:(ontology ontologies ontological onto neuro neuroscience neurosciences neuroscientific) AND acmdlTitle:(neuro neuroscience neurosciences neuroscientific)*

1.8 Web of Science

The web of science was accessed on 12 November 2016 and no year limits were defined. Therefore, documents from the year 1900 to present day were searched. A search that applied keywords on topic and title of the documents resulted in 255 results. Then documents in English language were selected, neuroscience selected as the field and book chapters, articles, reviews and proceedings papers were selected as document types. This resulted in 52 results. The search phrase is as follows:

> The search query phrase: *TOPIC: (neuroscien* ontolog*) OR Title: (neuroscien* ontolog*)*
> Refined by: *WEB OF SCIENCE CATEGORIES: (NEUROSCIENCES) AND LANGUAGES: (ENGLISH) AND DOCUMENT TYPES: (ARTICLE OR PROCEEDINGS PAPER OR REVIEW OR BOOK CHAPTER)*
> Timespan: *All years*
> Indexes: *SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC.*

2 Appendix B: Resources Used in this Research

In this appendix, the information regarding platforms, resources and data types used in the thesis is discussed.

2.1 OWL and its Usage

Several languages are used for defining and expressing ontologies. OWL (Web Ontology Language) is a semantic, web-based ontology language, proposed by W3C (World Wide Web Consortium) and can be used in academic, commercial and medical fields. It had been purposely built based on the DL and by doing so, benefits from the DL functionalities (Baader, Horrocks and Sattler, 2007).

The OWL using the RDF syntax, adds to the level of expressiveness. Using OWL also enriches the property relations (ObjectProperties) which increases the reasoning power (Bug *et al.*, 2008). OWL is a means of storing logical statements about the relationships among classes by using object properties (relations) and logical quantifiers. It also stores logical characteristics of relations and entailments between them. OWL provides many constructs in defining classes but it is limited in defining relations (Osumi-Sutherland *et al.*, 2012).

OWL is a part of semantic web, in which the information is understandable by machines and has exact meaning. It is designed to process the contents and not to display them and is XML-based. It has three sublanguages: OWL Lite, OWL DL and OWL Full (each version covers the previous one plus some extra features).

An Ontology could be compared to a DL TBox plus the role hierarchy (Baader, Horrocks and Sattler, 2007). By representing an ontology in OWL, we have made its context available to the OWL reasoners and RDF query engines (Larson *et al.*, 2007). It has an enriched and at the same time generalized expressing power. OWL has the ability to declare hierarchies and to infer them based on specific properties (Bug *et al.*, 2008). OWL DL can handle up to 10^5 concepts (Haarslev and Möller, 2001).

OWL helps scientists to technically and semantically improve their information in several aspects. It helps them in imposing restrictions and by doing so, preventing inconsistencies and inferring classifications. Imposing “restrictions” on classes means that compulsory and adequate conditions have to be met for that specific class (Larson *et al.*, 2007).

OWL avoids inconsistencies in classes by applying restrictions and additional rules (Larson and Martone, 2009). Actually, the importance of OWL could be shown through the inconsistencies that could be found after transferring large amount of information into ontologies and then being checked by the OWL (Baader, Horrocks and Sattler, 2007).

Restrictions also allow OWL-reasoners to infer classifications about the classes (Larson and Martone, 2009). OWL has the ability to identify the class of unknown instance and then locate it under a known class. Also, it will accommodate for future changes to this classification (Larson *et al.*, 2007). OWL 2 provides better features in case of ontology handling such as defining property chain rules (Imam *et al.*, 2012).

2.1.1 Using OWL for Neuroscience Ontologies

Most of the ontologies which we studied in the neuroscience use OBO principle of single inheritance. This results in a '*is a'* hierarchy which is a simple tree and is called *flat* ontology and results in limited classification ability. Of course, OWL enables us to describe the unique characteristics of entities, also relations among then. But computer scientists manage to make new and multiple hierarchies through using the '*part of'* relationships and cross-cutting the '*is a'* hierarchy.

In an ontology which has been represented by the OWL, we connect the *subclasses* to *superclasses* via "is a": neuron *is a* nerve cell. The classes are connected to the properties by using "has a": cell *has a* nucleus. Also we assign properties to other properties through symmetric, transitive or inverse (opposite of transitive) relationships (Larson *et al.*, 2007).

2.1.2 Limitations of OWL

All properties in OWL language are first class entities. This means that they are independent of the classes that they describe. But this leads to the structural properties and relationships between the classes not being distinguishable (Larson *et al.*, 2007).

Despite all the powerful aspects of OWL, its current version cannot cover the vast number of instances (e.g. gene expressions) and is limited (Bug *et al.*, 2008). Also, using RDF syntax, which is lengthy and not easily understandable by humans could be counted as a disadvantage for the OWL (Horrocks, 2013).

2.2 SPARQL

SPARQL is a recursive acronym for SPARQL Protocol and RDF Query Language. It is an RDF query language which is a semantic query language for databases, able to retrieve and manipulate data stored in Resource Description Framework (RDF) format.

RDF is a flexible and extensible way to represent information about semantic web resources. It is used to represent, among other things, personal information, social networks, metadata about digital artefacts, as well as provide a means of integration over disparate sources of information.

A standardized query language for RDF data with multiple implementations offers developers and end users a way to write and to consume the results of queries across this wide range of information. Used with a common protocol, applications can access and combine information from across the Web.

2.3 OBO Foundry

The OBO Foundry is a collective of ontology developers that are committed to collaboration and adherence to shared principles. The mission of the OBO Foundry is to develop a family of interoperable ontologies that are both logically well-formed and scientifically accurate. To achieve this, OBO Foundry participants voluntarily adhere to and contribute to the development of an evolving set of principles including open use, collaborative development, non-overlapping and strictly-scoped content, and common syntax and relations, based on ontology models that work well, such as the Gene Ontology (GO).

The OBO Foundry is overseen by an Operations Committee with Editorial, Technical and Outreach working groups. The processes of the Editorial working group are modelled on the journal refereeing process.

2.4 BioPortal

Image Biomedical ontologies provide essential domain knowledge to drive data integration, information retrieval, data annotation, natural-language processing and decision support. BioPortal (http://bioportal.bioontology.org) is an open repository of biomedical ontologies that provides access via Web services and Web browsers to ontologies developed in OWL, RDF, OBO format and Protégé frames.

BioPortal functionality includes the ability to browse, search and visualize ontologies. The Web interface also facilitates community-based participation in the evaluation and evolution of ontology content by providing features to add notes to ontology terms, mappings between terms and ontology reviews based on criteria such as usability, domain coverage, quality of content, and documentation and support.

BioPortal also enables integrated search of biomedical data resources such as the Gene Expression Omnibus (GEO), ClinicalTrials.gov, and ArrayExpress, through the annotation and indexing of these resources with ontologies in BioPortal. Thus, BioPortal not only provides investigators, clinicians, and developers straight access to programmatically use biomedical ontologies, but also provides support to integrate data from a variety of biomedical resources.

2.5 Image

Neuroimaging or brain imaging is the use of various techniques to either directly or indirectly image the structure, function/pharmacology of the nervous system. It is a relatively new discipline within medicine, neuroscience, and psychology. Physicians who specialize in the performance and interpretation of neuroimaging in the clinical setting are neuroradiologists.

Neuroimaging falls into two broad categories: Structural imaging, which deals with the structure of the nervous system and the diagnosis of gross (large scale) intracranial disease (such as tumour) and injury; Functional imaging, which is used to diagnose metabolic diseases and lesions on a finer scale (such as Alzheimer's disease) and also for neurological and cognitive psychology research and building brain-computer interfaces.

Functional imaging enables, for example, the processing of information by centres in the brain to be visualized directly. Such processing causes the involved area of the brain to increase metabolism and "light up" on the scan.

2.5.1 Digital Imaging and Communications in Medicine (DICOM)

DICOM or Digital Imaging and Communications in Medicine is a hugely influential and broad standard for managing (storing, printing and etc.) and transmitting information in medical imaging, originated by eSDI (Electronic Source Data Interchange). It can hold any kind of file types and is extendible via having standard and non-standard (private) tags (odd numbered groups).

Unfortunately, it is complex and its documentation is very lengthy (its 2008 version is over 3600 pages). Its model is also limited since it cannot manage multiple subjects per project, the status of subjects and the method used in the experiment cannot be stored (Lohrey, Killeen and Egan, 2009). Another problem is that by using DICOM’s messaging protocol for transferring the images, we cannot be sure that the images are received unchanged at the other end. Last but not least is the problem with the de-identification process. DICOM adds lots of additional data to the image which have to be cleared before being used in a trial (El-Ghatta, Cladé and Snyder, 2010). All in all it can be said that its model is a primary and comprehensive one, but hard to use and adopt.

2.5.2 FreeSurfer

FreeSurfer is a brain imaging software package developed by the Athinoula A. Martinos Centre for Biomedical Imaging at Massachusetts General Hospital for analyzing magnetic resonance imaging (MRI) scan data. It is an important tool in functional brain mapping and facilitates the visualization of the functional regions of the highly folded cerebral cortex. It contains tools to conduct both volume based and surface based analysis, which primarily use the white matter surface. FreeSurfer includes tools for the reconstruction of topologically correct and geometrically accurate models of both the grey/white and pial surfaces, for measuring cortical thickness, surface area and folding, and for computing inter-subject registration based on the pattern of cortical folds. In addition, an automated labelling of 35 non-cortical regions is included in the package.

2.6 Ontologies Used in the Thesis

The following ontologies were used in this thesis for different means in study 2 and study 3.

2.6.1 NIFSTD

NIFSTD is an ontology which can be seen as an example for using of ontology for reasoning, query answering, data integration and many other uses. It is from Biomedical Informatics Research Network (BIRN) (available from http://www.birncommunity.org/). However, the search system around it, designed by the NIF is still far from answering abstract and high-level queries.

Structrully, BIRNLex was created based on NeuroNames, UMLS and some other community ontologies and later, NIFSTD was created based on the BIRNLex. It is designed on top of the BIRNLex which is a lexicon (language) about the clinical neuroimaging concepts (Bug *et al.*, 2008) and has modular design. When ontology has modular design, the user does not have to import all parts of the ontology for being able to use it. He can just import s specific part. BIRNLex has imported many entities from the Simple Knowledge Organization System (SKOS) and Dublin Core Metadata.

NIFSTD is not just following the OBO foundry; they also followed the National Center for Biomedical Ontology (NCBO) guidelines. NIFSTD also used OBO-RO which helped them in separation of the definition and representation also confines the relation types from increasing. This is essential for automated relation parsing (Bug *et al.*, 2008).

OBO-RO is a set of standard relations which are to be used in ontology standardization across the OBO foundry ontologies (Smith *et al.*, 2005) and helps us by avoiding the mixture of relations such as 'is-a' and 'part-of' because mixing them in a single hierarchical graph ends up in poor computational power (Smith, Williams and Schulze-Kremer, 2003).

NIFSTD has created its ontology based on two different approaches: in some cases the NIFSTD adapted an existing ontology (OBO Cell ontology) with its own design and in some cases imported it (Sub-cellular Anatomy ontology); and in some instances, they used two of the above approaches together (e.g. Protein Ontology) (Bug *et al.*, 2008). NIFSTD is not dedicated to a specific specie. That is why it does not import the FMA ontology concepts; since FMA is specifically designed for human anatomy (Imam *et al.*, 2012).

2.6.2 Foundational Model of Anatomy

The Foundational Model of Anatomy (FMA), initially developed as an enhancement of the anatomical content of UMLS, is a domain ontology of the concepts and relationships that pertain to the structural organization of the human body.

It encompasses the material objects from the molecular to the macroscopic levels that constitute the body and associates with them non-material entities (spaces, surfaces, lines, and points) required for describing structural relationships. The disciplined modeling approach employed for the development of the FMA relies on a set of declared principles, high level schemes, Aristotelian definitions and a frame-based authoring environment.

FMA was proposed as a reference ontology in biomedical informatics for correlating different views of anatomy, aligning existing and emerging ontologies in bioinformatics ontologies and providing a structure-based template for representing biological functions

2.7 Protégé

Protégé is a platform for knowledge-based systems development and research. Protégé is not an expert system itself and does not create them directly but is a tool for building tools that assist expert systems in acquisition of data. Expert system *shell* is an interface engine that is reusable and knowledge engineers use it in attendance with diverse knowledge bases to accommodate building expert systems. Opal knowledge acquisition tool and Oncocin knowledge base are the ancestors of Protégé. The versions of Protégé are: Protégé-I, Protégé-II, Protégé/Win and Protégé-2000.

Protégé's general methodology is to obtain the information in stages and the information in each stage would be the 'meta-knowledge' for the next step. A generic structure for all fields is present in the Protégé.

Protégé-II has separated problem solving methods (PSMs) from knowledge bases and this is good, since PSMs are very unlikely to change, whereas knowledge bases are subject to change. Protégé includes a tool for editing ontologies graphically which is called Maître.

Protégé/win allowed modular ontologies by using an ontology inclusion mechanism. It improved the task of custom-tailoring the knowledge-acquisition tool. This is done via defining some ontologies which are common. The users will use these pre-defined ontologies to come up with a new larger one or just use it for their specific purpose. Many knowledge bases in medical domain share drugs and laboratory tests could be the example for this.

Protege-2000 tried to do two things: first, to do a complete check up on the previous knowledge model of the Protégé in order to make it able to interoperate with other knowledge-based formalism (united and 'agreed on' ontology). Second, it unified all applications of the Protégé package in a single application and third, it developed to be a plug-in and supported by java programming language. Its OntoViz accessory can visualize the knowledge bases as diagrams.

Protoge-2000 supports more file types and can read and write to RDF format. Its default was to use a flat file for I/O (I guess they meant text file when they used flat file). It can also store information in RDBMS.

There are some counterparts for Protégé: Ontology editing and knowledge base construction tools such as Ontolingua Editor, Ontosaurus system built on LOOM and GKB Editor. Protégé is open source under the Mozilla public licence (Gennari *et al.*, 2003).

References

Abacha, A. Ben and Zweigenbaum, P. (2012) 'Medical question answering: Translating medical questions into SPARQL queries', *IHI'12 - Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, pp. 41–49. doi: 10.1145/2110363.2110372.

Acampora, G., Loia, V. and Salerno, S. (2012) 'A hybrid evolutionary approach for solving the ontology alignment problem', *International Journal of*. Available at: http://onlinelibrary.wiley.com/doi/10.1002/int.20517/full (Accessed: 13 August 2016).

Akil, H., Martone, M. E. and Van Essen, D. C. (2011) 'Challenges and opportunities in mining neuroscience data.', *Science (New York, N.Y.)*, 331(6018), pp. 708–712. doi: 10.1126/science.1199305.

Alexaki, S. and Christophides, V. (2001) 'The ICS-FORTH RDFSuite: Managing Voluminous RDF Description Bases.', *SemWeb*. Available at: http://ceur-ws.org/Vol-40/alexakietal.pdf (Accessed: 20 August 2014).

Ambert, K. and Cohen, A. (2012) 'Text-mining and neuroscience', *Int. Rev. Neurobiol*. Available at: https://books.google.com.au/books?hl=en&lr=&id=cLP_3lWFC6gC&oi=fnd&pg=PA109&dq=text+mining+and+neuroscience&ots=rBgvsPknbF&sig=DjTo16auPa0QAhBp5lK_Eq0lp1U (Accessed: 3 March 2016).

Amunts, K. *et al.* (2016) 'The Human Brain Project: Creating a European Research Infrastructure to Decode the Human Brain', *Neuron*, 92(3), pp. 574–581. doi: http://dx.doi.org/10.1016/j.neuron.2016.10.046.

Anderson, L., Krathwohl, D. and Bloom, B. (2001) 'A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives'. Available at: https://scholar.google.com.au/scholar?hl=en&q=bloom+taxonomy&btnG=&as_sdt=1%2C5&as_sdtp=#0 (Accessed: 11 February 2016).

Arbib, M. A. *et al.* (2014) 'A neuroinformatics of brain modeling and its implementation in the brain operation database BODB', *Neuroinformatics*, 12(1), pp. 5–26. doi: 10.1007/s12021-013-9209-y.

Ascoli, G. A. (2013) 'Global neuroscience: Distributing the management of brain knowledge worldwide', *Neuroinformatics*, 11(1), pp. 1–3. doi: 10.1007/s12021-012-9173-y.

Ascoli, G. A. G. (2012) 'Twenty questions for neuroscience metadata', *Neuroinformatics*, 10(2), pp. 115–117. doi: 10.1007/s12021-012-9143-4.

Asghar, M. Z. *et al.* (2017) 'Sentence-Level Emotion Detection Framework Using Rule-Based Classification', *Cognitive Computation*, 9(6), pp. 868–894. doi: 10.1007/s12559-017-9503-3.

Ashburner, M. *et al.* (2000) 'Gene Ontology: tool for the unification of biology', *Nature genetics*, 25(1), pp. 25–29.

Ashish, N. and Toga, A. W. (2015) 'Medical data transformation using rewriting', *Frontiers in Neuroinformatics*, 9. doi: 10.3389/fninf.2015.00001.

Asiaee, A. H. *et al.* (2013) 'From Questions to Effective Answers: On the Utility of Knowledge-Driven Querying Systems for Life Sciences Data', *Proceedings of the 9th International Conference on Data Integration in the Life Sciences (DILS'13)*, pp. 1–15. doi: 10.1007/978-3-642-39437-9.

Athenikos, S. J. S. and Han, H. (2010) 'Biomedical question answering: A survey', *Computer Methods and Programs in Biomedicine*. Elsevier Ireland Ltd, 99(1), pp. 1–24. doi: 10.1016/j.cmpb.2009.10.003.

Baader, F., Baader, U. and Sattler (2001) 'An overview of tableau algorithms for description logics', *Studia logica*, 69(1), pp. 5–40. Available at: http://sfx.unimelb.hosted.exlibrisgroup.com/sfxlcl41?sid=google&auinit=F&aulast=Baader&atitle=An overview of tableau algorithms for description logics&id=doi%3A10.1023%2FA%3A1013882326814&title=Studia logica&volume=69&issue=1&date=2001&spage=5&issn=0039-3.

Baader, F., Brandt, S. and Lutz, C. (2005) 'Pushing the EL envelope', in *International Joint Conference on Artificial Intelligence*. LAWRENCE ERLBAUM ASSOCIATES LTD, p. 364.

Baader, F., Horrocks, I. and Sattler, U. (2007) 'Description Logics', in *Handbook of knowledge representation*. Elsevier Science.

Babenko, V. N. *et al.* (2016) 'Analysis of differential gene expression by RNA-seq data in brain areas of laboratory animals.', *Journal of integrative bioinformatics*, 13(4), p. 292. doi: 10.2390/biecoll-jib-2016-292.

Bakker, R. and Wachtler, T. (2012) 'CoCoMac 2.0 and the future of tract-tracing databases', *in neuroinformatics*. Available at: http://journal.frontiersin.org/article/10.3389/fninf.2012.00030/full (Accessed: 15 November 2016).

Barillot, C. *et al.* (2006) 'Federating distributed and heterogeneous information sources in neuroimaging: the NeuroBase Project', *Stud Health Technol Inform*. Netherlands, 120, pp. 3–13. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16823118%5Cnhttp://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16823118.

Barnes, S. J. and Shaw, C. D. (2009) 'BrainFrame: A knowledge visualization system for the neurosciences', in *Proceedings of SPIE - The International Society for Optical Engineering*. School of Interactive Arts and Technology, Simon Fraser University, 250-13450 102 Avenue, Surrey, BC V3T 0A3, CanadaDept. of Psychology, University of British Columbia, 2136 West Mall, Vancouver, BC V6T 1Z4, Canada: SPIE (Proceedings of SPIE - The International Society for Optical Engineering), p. 72430F. doi: 10.1117/12.812290.

Batrancourt, B. *et al.* (2014) 'A Multilayer Ontology of Instruments for Neurological, Behavioral and Cognitive Assessments', *Neuroinformatics*, 13(1), pp. 93–110. doi: 10.1007/s12021-014-9244-3.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The semantic web', *Scientific american*, 284(5), pp. 28–37.

Bezgin, G. *et al.* (2009) 'Matching spatial with ontological brain regions using java tools for visualization, database access, and integrated data analysis', *Neuroinformatics*. Department of Cognitive Neuroscience, Nijmegen Medical Centre, Radboud University, P. O. B. 9101, 6500 HB Nijmegen, Netherlands, 7(1), pp. 7–22. doi: 10.1007/s12021-008-9039-5.

Bienia, C. and Li, K. (2012) 'Characteristics of workloads using the pipeline programming model', in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, pp. 161–171. doi: 10.1007/978-3-642-24322-6_14.

Birgiolas, J. *et al.* (2015) 'Ontology-assisted Keyword Search for NeuroML Models', in *Proceedings of the 27th International Conference on Scientific and Statistical Database Management - SSDBM '15*. New York, NY, USA: ACM (ACM International Conference Proceeding Series), pp. 37:1--37:6. doi: 10.1145/2791347.2791360.

Bodenreider, O. (2008) 'Biomedical ontologies in action: role in knowledge management, data integration and decision support.', *Yearbook of medical informatics*, 47, pp. 67–79. doi: 10.1055/s-0038-1638585.

Bonino, D. *et al.* (2004) 'Ontology Driven Semantic Search', *WSEAS Transactions on Information Science and Applications*, 1(6), pp. 1597–1605. Available at: http://www.wseas.us/e-library/conferences/venice2004/papers/472-334.pdf (Accessed: 13 January 2017).

Bordi, M. *et al.* (2016) 'Autophagy flux in CA1 neurons of Alzheimer hippocampus: Increased induction overburdens failing lysosomes to propel neuritic dystrophy.', *Autophagy*, 12(12), pp. 2467–2483. doi: 10.1080/15548627.2016.1239003.

Bota, M. (2008) 'BAMS neuroanatomical ontology: design and implementation', *Frontiers in Neuroinformatics*. Switzerland, 2, p. 2. doi: 10.3389/neuro.11.002.2008.

Bowden, D. M. *et al.* (2012) 'NeuroNames: An ontology for the braininfo portal to neuroscience on the web', *Neuroinformatics*, 10(1), pp. 97–114. doi: 10.1007/s12021-011-9128-8.

Bowden, D. M. and Dubach, M. F. (2003) 'NeuroNames 2002', *Neuroinformatics*. Department of Behavioral Sciences, Washington Natl. Primate Res. Center, University of Washington, Seattle, WA, United States, 1(1), pp. 43–59. Available at: https://www.scopus.com/inward/record.uri?eid=2-s2.0-0038068998&partnerID=40&md5=35237705899c7600b982ae8c743c20bf.

Bowen, G. a. (2008) 'Naturalistic inquiry and the saturation concept: a research note', *Qualitative Research*, 8(1), pp. 137–152. doi: 10.1177/1468794107085301.

Bratsas, C. *et al.* (2009) 'Towards a semantic framework for an integrative description of neuroscience patterns and studies: A case for emotion-related data', *Studies in Health Technology and Informatics*. Netherlands, 150, pp. 322–326. doi: 10.3233/978-1-60750-044-5-322.

Broekstra, J. and Kampman, A. (2003) 'SeRQL: a second generation RDF query language', *Proc. SWAD-Europe Workshop on Semantic Web*. Available at: http://www.w3.org/2001/sw/Europe/events/20031113-storage/positions/aduna.pdf (Accessed: 13 January 2017).

Bu, F. *et al.* (2010) 'Function-based question classification for general QA', in *EMNLP 2010 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1119–1128. Available at: https://dl.acm.org/citation.cfm?id=1870767 (Accessed: 9 October 2019).

Bug, W. J. *et al.* (2008) 'The NIFSTD and BIRNLex vocabularies: Building comprehensive ontologies for neuroscience', *Neuroinformatics*. Humana Press Inc, 6(3), pp. 175–194. doi: 10.1007/s12021-008-9032-z.

Burns, G. A. P. C. and Turner, J. A. (2013) 'Modeling functional Magnetic Resonance Imaging (fMRI) experimental variables in the Ontology of Experimental Variables and Values (OoEVV)', *NeuroImage*, 82, pp. 662–270. doi: 10.1016/j.neuroimage.2013.05.024.

C., P. (2008) 'Biological data integration using Semantic Web technologies', *Biochimie*, 90(4), pp. 584–594. Available at: http://www.sciencedirect.com/science/article/pii/S0300908408000412.

Calabrese, E. *et al.* (2013) 'A quantitative magnetic resonance histology atlas of postnatal rat brain development with regional estimates of growth and variability', *NeuroImage*, 71, pp. 196–206. doi: 10.1016/j.neuroimage.2013.01.017.

Calabrese, E. *et al.* (2015) 'A diffusion tensor MRI atlas of the postmortem rhesus macaque brain', *NeuroImage*, 117, pp. 408–416. doi: 10.1016/j.neuroimage.2015.05.072.

Callahan, A. *et al.* (2016) 'RegenBase: A knowledge base of spinal cord injury biology for translational research', *Database*, 2016. doi: 10.1093/database/baw040.

Calvanese, D., De Giacomo, G. and Lenzerini, M. (1999) 'Reasoning in expressive description logics with fixpoints based on automata on infinite trees', in *IJCAI*, pp. 84–89.

Cao, Y. G. *et al.* (2011) 'AskHERMES: An online question answering system for complex clinical questions', *Journal of Biomedical Informatics*, 44(2), pp. 277–288. doi: 10.1016/j.jbi.2011.01.004.

Chakrabarti, C. *et al.* (2014) 'Statistical algorithms for ontology-based annotation of scientific literature.', *Journal of biomedical semantics*, 5(Suppl 1 Proceedings of the Bio-Ontologies Spec Interest G), p. S2. doi: 10.1186/2041-1480-5-S1-S2.

Chali, Y., Hasan, S. A. and Imam, K. (2012) 'Learning Good Decompositions of Complex Questions', in, pp. 104–115. doi: 10.1007/978-3-642-31178-9_10.

Chan, M. (2010) 'Mental health: central to human development', *Mental Health and Development: Targeting People with Mental Health Conditions as a Vulnerable Group*, pp. 1–6. Available at: http://wwwlive.who.int/entity/mental_health/policy/development/mh_devel_targeting_summary_2010_en.pdf (Accessed: 16 November 2015).

Chandrasekaran, B., Josephson, J. R. and Benjamins, V. R. (1999) ‘What Are Ontolgoies, and Why Do We Need them?’, *IEEE Intelligent Systems and their applications*, 14(1), pp. 20–26. doi: 1094-7167/99/$10.0.

Chen, H. *et al.* (2013) ‘Textrous!: Extracting Semantic Textual Meaning from Gene Sets’, *PLoS ONE*, 8(4), p. e62665. doi: 10.1371/journal.pone.0062665.

Chen, L. *et al.* (2006) ‘OntoQuest: exploring ontological data made easy’, *Proceedings of the 32nd international conference on Very large data bases*. Seoul, Korea: VLDB Endowment, pp. 1183–1186.

Chen, M., Jin, X. and Shen, D. (2011) ‘Short text classification improved by learning multi-granularity topics’, *IJCAI*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.208.4202&rep=rep1&type=pdf (Accessed: 1 January 2016).

Cheung, K.-H., Frost, H. R., *et al.* (2009) ‘A journey to Semantic Web query federation in the life sciences’, *BMC Bioinformatics*. BioMed Central, 10(S10), p. S10. doi: 10.1186/1471-2105-10-S10-S10.

Cheung, K.-H., Lim, E., *et al.* (2009) ‘Approaches to neuroscience data integration’, *Briefings in bioinformatics*. England, 10(4), pp. 345–353. doi: 10.1093/bib/bbp029.

Ciccarese, P., Ocana, M. and Clark, T. (2012) ‘Open semantic annotation of scientific publications using DOMEO.’, *Journal of biomedical semantics*, 3 Suppl 1(Suppl 1), p. S1. doi: 10.1186/2041-1480-3-S1-S1.

Cimiano, P. (2017) ‘Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods’, *Semantic Web*, 8(3), pp. 489–508. Available at: https://content.iospress.com/articles/semantic-web/sw218 (Accessed: 14 July 2019).

Courtot, M. *et al.* (2011) ‘MIREOT: The minimum information to reference an external ontology term’, *Applied Ontology*, 6(1), pp. 23–33. doi: 10.3233/AO-2011-0087.

Crosswhite, J. (2010) 'The new rhetoric project', in *Philosophy and Rhetoric*. Dordrecht: Springer Netherlands, pp. 301–307. doi: 10.1353/par.2010.0001.

Cui, W., Xiao, Y. and Wang, W. (2016) 'KBQA: An online template based question answering system over freebase', in *IJCAI International Joint Conference on Artificial Intelligence*, pp. 4240–4241. Available at: https://www.ijcai.org/Proceedings/16/Papers/640.pdf (Accessed: 2 July 2019).

Das, S. *et al.* (2014) 'Pain Research Forum: application of scientific social media frameworks in neuroscience.', *Frontiers in neuroinformatics*, 8, p. 21. doi: 10.3389/fninf.2014.00021.

Das, S. *et al.* (2016) 'The MNI data-sharing and processing ecosystem', *NeuroImage*, 124, pp. 1188–1195. doi: 10.1016/j.neuroimage.2015.08.076.

Davis, R. G. *et al.* (2018) 'Top-Down Proteomics Enables Comparative Analysis of Brain Proteoforms Between Mouse Strains.', *Analytical chemistry*, 90(6), pp. 3802–3810. doi: 10.1021/acs.analchem.7b04108.

Dedhia, M. *et al.* (2019) 'Introducing high school students to the Gene Ontology classification system.', *F1000Research*, 8, p. 241. doi: 10.12688/f1000research.18061.4.

Delany, S., Buckley, M. and Greene, D. (2012) 'SMS spam filtering: methods and data', *Expert Systems with Applications*. Available at: http://www.sciencedirect.com/science/article/pii/S0957417412002977 (Accessed: 31 December 2015).

Dictionary, M. (2006) *The Merriam-Webster Dictionary*. Available at: https://www.merriam-webster.com/dictionary/ontology (Accessed: 26 November 2013).

Dietze, H. and Schroeder, M. (2009) 'GoWeb: a semantic search engine for the life science web', *BMC Bioinformatics*, 10(S10), p. S7. doi: 10.1186/1471-2105-10-S10-S7.

Djamanakova, A. *et al.* (2014) ‘Tools for multiple granularity analysis of brain MRI data for individualized image analysis’, *NeuroImage*, 101, pp. 168–176. doi: 10.1016/j.neuroimage.2014.06.046.

Dridan, R. and Baldwin, T. (2007) ‘What to Classify and How: Experiments in question classification for Japanese’, *Proceedings of the 10th Conference of the Pacific Association for Computational Linguistics*, pp. 333–341. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.6142&rep=rep1&type=pdf (Accessed: 21 February 2016).

Dwivedi, S. K. and Singh, V. (2013) ‘Research and Reviews in Question Answering System’, *Procedia Technology*, 10, pp. 417–424. doi: 10.1016/j.protcy.2013.12.378.

E Hirbec, H., Noristani, H. N. and Perrin, F. E. (2017) ‘Microglia Responses in Acute and Chronic Neurological Diseases: What Microglia-Specific Transcriptomic Studies Taught (and did Not Teach) Us.’, *Frontiers in aging neuroscience*, 9, p. 227. doi: 10.3389/fnagi.2017.00227.

Edmunds, R. C. *et al.* (2016) ‘Phenoscape: Identifying candidate genes for evolutionary phenotypes’, *Molecular Biology and Evolution*, 33(1), pp. 13–24. doi: 10.1093/molbev/msv223.

El-Ghatta, S., Cladé, T. and Snyder, J. (2010) ‘Integrating Clinical Trial Imaging Data Resources Using Service-Oriented Architecture and Grid Computing’, *Neuroinformatics*. Humana Press Inc., 8(4), pp. 251–259. doi: 10.1007/s12021-010-9072-z.

Ely, J. W. *et al.* (2002) ‘Obstacles to answering doctors’ questions about patient care with evidence: qualitative study’, *BMJ*, 324(7339), pp. 710–710. doi: 10.1136/bmj.324.7339.710.

Eshghishargh, A. *et al.* (2015) ‘An Ontology-Based Semantic Question Complexity Model and Its Applications in Neuroinformatics’, *Frontiers in Neuroscience*, 9. doi: 10.3389/conf.fnins.2015.91.00015.

Eshghishargh, A. *et al.* (2018) 'A semantic system for answering questions in neuroinformatics', in *ACM International Conference Proceeding Series*. New York, New York, USA: ACM Press, pp. 1–5. doi: 10.1145/3167918.3167960.

Essen, D. Van *et al.* (2013) 'The WU-Minn human connectome project: an overview', *Neuroimage*. Available at: http://www.sciencedirect.com/science/article/pii/S1053811913005351 (Accessed: 11 August 2016).

Fang, J. *et al.* (2007) 'Ontology-based automatic classification and ranking for web documents', *Proceedings - Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007*, 3, pp. 627–631. doi: 10.1109/FSKD.2007.432.

Fellbaum, C. (2010) 'WordNet', in *Theory and applications of ontology: computer applications*. Springer, pp. 231–243.

Feng, G. *et al.* (2012) 'A Bayesian feature selection paradigm for text classification', *Information processing & management*. Available at: http://www.sciencedirect.com/science/article/pii/S0306457311000811 (Accessed: 31 December 2015).

Fensel, D. *et al.* (2001) 'OIL: An Ontology Infrastructure for the Semantic Web', *IEEE Intelligent Systems*, 16(2), pp. 38–45. doi: 10.1109/5254.920598.

Fischl, B. (2012) 'FreeSurfer', *Neuroimage*. Available at: http://www.sciencedirect.com/science/article/pii/S1053811912000389 (Accessed: 20 July 2016).

Fong, L. L. *et al.* (2007) 'An ontology-driven knowledge environment for subcellular neuroanatomy', in *CEUR Workshop Proceedings*. Innsbruck, Austria.

Francken, J. C. and Slors, M. (2014) 'From commonsense to science, and back: The use of cognitive concepts in neuroscience', *Consciousness and Cognition*, 29, pp. 248–258. doi: 10.1016/j.concog.2014.08.019.

Frank, E. *et al.* (2005) 'Weka', *Data Mining and Knowledge Discovery Handbook*, pp. 1305–1314. doi: 10.1007/0-387-25465-X_62.

French, L. *et al.* (2015) 'Text mining for neuroanatomy using WhiteText with an updated corpus and a new web application', *Frontiers in Neuroinformatics*, 9(May), p. 13. doi: 10.3389/fninf.2015.00013.

Gangemi, A. *et al.* (2002) 'Sweetening Ontologies with DOLCE', in Gómez-Pérez, A. and Benjamins, V. R. (eds) *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*. Springer Berlin Heidelberg, pp. 166–181. doi: 10.1007/3-540-45810-7_18.

Gardner, D. (2008) 'The neuroscience information framework: a data and knowledge environment for neuroscience', *Neuroinformatics*, 6(3), p. 149. Available at: http://sfx.unimelb.hosted.exlibrisgroup.com/sfxlcl41?sid=google&auinit=D&aulast=Gardner&atitle=The neuroscience information framework: a data and knowledge environment for neuroscience&id=doi:10.1007/s12021-008-9024-z&title=Neuroinformatics&volume=6&issue.

Gennari, J. H. *et al.* (2003) 'The evolution of Protégé: An environment for knowledge-based systems development', *International Journal of Human Computer Studies*, 58(1), pp. 89–123. doi: 10.1016/S1071-5819(02)00127-1.

Gertz, M. *et al.* (2002) 'Annotating scientific images: A concept-based approach BT - 14th International Conference on Scientific and Statistical Database Management, SSDBM 2002, July 24, 2002 - July 26, 2002', in. Department of Computer Science, University of California, Davis; CA, United StatesDepartment of Pathology and Internal Medicine, University of California, Davis; CA, United StatesDepartment of Neurology, University of California, Davis; CA, United StatesC: IEEE Computer Society (Proceedings of the International Conference on Scientific and Statistical Database Management, SSDBM), pp. 59–68. doi: 10.1109/SSDM.2002.1029706.

Gibaud, B. *et al.* (2011) 'NeuroLOG: sharing neuroimaging data using an ontology-based federated approach.', *AMIA Annual Symposium Proceedings/ AMIA Symposium. AMIA Symposium*, 2011, pp. 472–480. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243145&tool=pmcentrez&rendertype=abstract (Accessed: 14 September 2016).

Giraldo, O. *et al.* (2017) 'Using semantics for representing experimental protocols.', *Journal of biomedical semantics*, 8(1), p. 52. doi: 10.1186/s13326-017-0160-y.

Gleeson, P. *et al.* (2010) 'NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail', *PLoS Comput*. Available at: http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1000815 (Accessed: 13 January 2017).

Goble, C. A. *et al.* (2001) 'Transparent access to multiple bioinformatics information sources', *IBM Systems Journal*, 40(2), pp. 532–551. doi: 10.1147/sj.402.0532.

Grant, M. and Booth, A. (2009) 'A typology of reviews: an analysis of 14 review types and associated methodologies', *Health Information & Libraries Journal*. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.1471-1842.2009.00848.x/full (Accessed: 13 September 2016).

Grewe, J., Wachtler, T. and Benda, J. (2011) 'A Bottom-up Approach to Data Annotation in Neurophysiology.', *Frontiers in neuroinformatics*, 5, p. 16. doi: 10.3389/fninf.2011.00016.

Groza, T. *et al.* (2015) 'The Human Phenotype Ontology: Semantic Unification of Common and Rare Disease', *American Journal of Human Genetics*, 97(1), pp. 111–124. doi: 10.1016/j.ajhg.2015.05.020.

Gruber, T. R. (1993) 'A translation approach to portable ontology specifications', *Knowledge acquisition*, 5(2), pp. 199–220.

Gruber, T. R. (1995) 'Toward principles for the design of ontologies used for knowledge sharing', *International journal of human-computer studies*, 43(5–6), p. 907. Available at: http://sfx.unimelb.hosted.exlibrisgroup.com/sfxlcl41?sid=google&auinit=TR&aulast=Gruber&atitle=Toward principles for the design of ontologies used for knowledge sharing&id=doi:10.1006/ijhc.1995.1081&title=International journal of human-computer studies&vo.

Gupta, A. *et al.* (2003) 'Towards a formalization of disease-specific ontologies for neuroinformatics', *Neural Networks*. San Diego Supercomputer Center, University of California San Diego, 9500 Gilman Drive, San Diego, CA 92093-0505, United States, 16(9), pp. 1277–1292. doi: 10.1016/j.neunet.2003.07.008.

Gupta, A. *et al.* (2008) 'Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF)', *Neuroinformatics*, 6(3), pp. 205–217. doi: 10.1007/s12021-008-9033-y.

Gupta, A., Condit, C. and Qian, X. (2010) 'BioDB: An ontology-enhanced information system for heterogeneous biological information', *Data and Knowledge Engineering*. Elsevier B.V., 69(11), pp. 1084–1102. doi: 10.1016/j.datak.2010.07.003.

Haarslev, V. and Möller, R. (2001) 'High performance reasoning with very large knowledge bases: A practical case study', in *IJCAI*, pp. 161–168.

Haarslev, V. and Müller, R. (2001) 'RACER system description', in *Automated Reasoning*. Springer, pp. 701–705.

Haase, P. *et al.* (2008) 'The neon ontology engineering toolkit', *Www*, (April), pp. 4–6. doi: 10.1.1.141.4163.

Hacioglu, K. and Ward, W. (2003) 'Question classification with support vector machines and error correcting codes', in *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology*. ACL, pp. 28–30. doi: 10.3115/1073483.1073493.

Hamilton, D. J. *et al.* (2012) 'An ontological approach to describing neurons and their relationships.', *Frontiers in neuroinformatics*, 6(APRIL 2012), p. 15. doi: 10.3389/fninf.2012.00015.

Haris, S. S. and Omar, N. (2012) 'A rule-based approach in Bloom's Taxonomy question classification through natural language processing', *Proceedings - 2012 7th International Conference on Computing and Convergence Technology (ICCIT, ICEI and ICACT), ICCCT 2012*. IEEE, pp. 410–414. Available at: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6530368 (Accessed: 11 February 2016).

Hastings, J. *et al.* (2012) 'Wanting what we don't want to want: Representing addiction in interoperable bio-ontologies', in *CEUR Workshop Proceedings*. Cheminformatics and Metabolism, European Bioinformatics Institute, Cambridge, United KingdomSwiss Center for Affective Sciences, University of Geneva, SwitzerlandComputational Systems Neurobiology, European Bioinformatics Institute, Cambridge, United King: Sun SITE Central Europe CEUR-WS (CEUR Workshop Proceedings).

Hawrylycz, M. *et al.* (2014) 'The allen brain atlas', *Springer Handbook of Bio-/Neuroinformatics*, pp. 1111–1126. doi: 10.1007/978-3-642-30574-0_62.

Hearst, M. (2006) 'Design recommendations for hierarchical faceted search interfaces', *ACM SIGIR workshop on faceted search*, pp. 5–9. Available at: http://flamenco.sims.berkeley.edu/papers/faceted-workshop06.pdf (Accessed: 11 September 2014).

Hermjakob, U. (2001) 'Parsing and question classification for question answering', in *Proceedings of the ACL 2001 Workshop on Open-Domain Question Answering*. Morristown, NJ, USA: Association for Computational Linguistics, pp. 1–6. doi: 10.3115/1117856.1117859.

Herrick, R. *et al.* (2016) 'XNAT Central: Open sourcing imaging research data', *NeuroImage*, 124, pp. 1093–1096. doi: 10.1016/j.neuroimage.2015.06.076.

Heymans, S. *et al.* (2008) 'Ontology reasoning with large data repositories', in *Ontology Management*. Springer, pp. 89–128.

Higgins, J. and Green, S. (2008) *Cochrane handbook for systematic reviews of interventions*. Available at: http://onlinelibrary.wiley.com/doi/10.1002/9780470712184.fmatter/summary (Accessed: 27 October 2016).

Hoehndorf, R. *et al.* (2015) 'Aber-OWL: A framework for ontology-based data access in biology', *BMC Bioinformatics*, 16(1). Available at: http://dx.doi.org/10.1186/s12859-015-0456-9 (Accessed: 16 December 2016).

Hoehndorf, R., Gkoutos, G. V and Schofield, P. N. (2016) 'Datamining with Ontologies.', *Methods in molecular biology (Clifton, N.J.)*, 1415, pp. 385–97. doi: 10.1007/978-1-4939-3572-7_19.

Hollink, L. *et al.* (2003) 'Semantic annotation of image collections', *Knowledge capture*, 2. Available at: http://www.cs.vu.nl/~laurah/papers/Hollink_saic.pdf (Accessed: 11 September 2014).

Horrocks, I. (2013) 'What are ontologies good for?', *Evolution of Semantic Systems*, pp. 175–188. doi: 10.1007/978-3-642-34997-3.

Horvat, M., Bogunović, N. and Ćosić, K. (2014) 'STIMONT: a core ontology for multimedia stimuli description', *Multimedia Tools and Applications*, 73(3), pp. 1103–1127. doi: 10.1007/s11042-013-1624-4.

Huang, F. *et al.* (2015) 'Investigation of gene regulatory networks associated with autism spectrum disorder based on MiRNA expression in China', *PLoS ONE*. United States, 10(6), p. e0129052. doi: 10.1371/journal.pone.0129052.

Hustadt, U. and Schmidt, R. A. (2002) 'Using resolution for testing modal satisfiability and building models', *Journal of Automated Reasoning*, 28(2), pp. 205–232.

Imam, Fahim T. *et al.* (2011) ‘NIFSTD and neurolex: A comprehensive neuroscience ontology development based on multiple biomedical ontologies and community involvement’, in *CEUR Workshop Proceedings*. Neuroscience Information Framework, Center for Research in Biological Systems, University of California, San Diego, CA, United States: ICBO (CEUR Workshop Proceedings), pp. 349–356.

Imam, Fahim T *et al.* (2011) ‘NIFSTD and neurolex: A comprehensive neuroscience ontology development based on multiple biomedical ontologies and community involvement BT - 2nd International Conference on Biomedical Ontology, ICBO 2011, July 26, 2011 - July 30, 2011’, in. Neuroscience Information Framework, Center for Research in Biological Systems, University of California, San Diego, CA, United States: Sun SITE Central Europe CEUR-WS (CEUR Workshop Proceedings), pp. 349–356.

Imam, F. T. *et al.* (2012) ‘Development and use of ontologies inside the neuroscience information framework: A practical approach’, *Frontiers in Genetics*, 3(JUN), p. 111. doi: 10.3389/fgene.2012.00111.

Jakoniene, V. and Lambrix, P. (2005) ‘Ontology-based integration for bioinformatics’, in *Proc VLDB Workshop on Ontologies-based techniques for Databases and Information Systems (ODBIS)*, pp. 55–58.

Janik, M. and Kochut, K. (2008) ‘Training-less ontology-based text categorization’, *workshop on Exploiting Semantic Annotations ….* Available at: https://getd.libs.uga.edu/pdfs/janik_maciej_g_200808_phd.pdf (Accessed: 4 February 2016).

Jayapandian, C. *et al.* (2015) ‘A scalable neuroinformatics data flow for electrophysiological signals using MapReduce’, *Frontiers in Neuroinformatics*, 9(MAR), p. 4. doi: 10.3389/fninf.2015.00004.

Jeffrey, G. *et al.* (2014) ‘The Neuroscience Information Framework (NIF): A Unified Semantic Framework and Associated Tools for Discovery, Integration, and Utilization of Biomedical Data and Resources on the Web’, *Frontiers in Neuroinformatics*, 8, p. 53. doi: 10.3389/conf.fninf.2014.08.00115.

Jernigan, T. L. *et al.* (2016) ‘The Pediatric Imaging, Neurocognition, and Genetics (PING) Data Repository’, *NeuroImage*, 124, pp. 1149–1154. doi: 10.1016/j.neuroimage.2015.04.057.

Ježek, P. and Mou, R. (2011) ‘Semantic Web in EEG / ERP Portal Ontology development and NIF registration’, in *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)*. IEEE, pp. 2058–2062. doi: 10.1109/BMEI.2011.6098757.

Jezek, P. and Moucek, R. (2011) ‘Semantic web in EEG/ERP portal: Ontology development and NIF registration BT - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011, October 15, 2011 - October 17, 2011’, in. Department of Computer Science and Engineering, University of West Bohemia, Pilsen, Czech Republic: IEEE Computer Society (Proceedings - 2011 4th International Conference on Biomedical Engineering and Informatics, BMEI 2011), pp. 2058–2062. doi: 10.1109/BMEI.2011.6098757.

Johnstone, P. L. (2004) ‘Mixed methods, mixed methodology health services research in practice’, *Qualitative Health Research*, 14(2), pp. 259–271. doi: 10.1177/1049732303260610.

Kalyanpur, a *et al.* (2006) ‘Swoop: A Web Ontology Editing Browser’, *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(2), pp. 144–153. doi: 10.1016/j.websem.2005.10.001.

Kasabov, N., Jain, V. and Benuskova, L. (2008) ‘Integrating evolving brain-gene ontology and connectionist-based system for modeling and knowledge discovery’, *Neural Networks*, 21(2–3), pp. 266–275. doi: 10.1016/j.neunet.2007.12.041.

Kassahun, Y. *et al.* (2014) ‘Automatic classification of epilepsy types using ontology-based and genetics-based machine learning’, *Artificial Intelligence in Medicine*, 61(2), pp. 79–88. Available at: http://dx.doi.org/10.1016/j.artmed.2014.03.001.

Katifori, A., Vassilakis, C. and Dix, A. (2010) ‘Ontologies and the brain: Using spreading activation through ontologies to support personal interaction’, *Cognitive Systems Research*. Available at: http://www.sciencedirect.com/science/article/pii/S1389041709000138 (Accessed: 15 January 2017).

Keator, D. B. *et al.* (2013) ‘Towards structured sharing of raw and derived neuroimaging data across existing resources’, *NeuroImage*, 82, pp. 647–661. doi: 10.1016/j.neuroimage.2013.05.094.

Kennedy, D. N. *et al.* (2012) ‘The internet brain volume database: A public resource for storage and retrieval of volumetric data’, *Neuroinformatics*, 10(2), pp. 129–140. doi: 10.1007/s12021-011-9130-1.

Kitchenham, B. *et al.* (2009) ‘Systematic literature reviews in software engineering - A systematic literature review’, *Information and Software Technology*. Butterworth-Heinemann, pp. 7–15. doi: 10.1016/j.infsof.2008.09.009.

Kitchenham, B. and Charters, S. (2007) ‘Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3’, *Engineering*, 45(4ve), p. 1051. doi: 10.1145/1134285.1134500.

Köhler, J., Philippi, S. and Lange, M. (2003) ‘SEMEDA: Ontology based semantic integration of biological databases’, *Bioinformatics*, 19(18), pp. 2420–2427. doi: 10.1093/bioinformatics/btg340.

Köhler, S. *et al.* (2012) ‘Ontological phenotype standards for neurogenetics’, *Human Mutation*. Wiley Subscription Services, Inc., A Wiley Company, 33(9), pp. 1333–1339. doi: 10.1002/humu.22112.

Köhler, S. *et al.* (2013) ‘Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research.’, *F1000Research*. England, 2(0), p. 30. doi: 10.12688/f1000research.2-30.v2.

Kollia, I., Glimm, B. and Horrocks, I. (2011) ‘SPARQL Query Answering over OWL Ontologies’, in *The Semantic Web: Research and Applications*. Springer, pp. 382–396. Available at: http://link.springer.com/10.1007%252F978-3-642-21034-1_26 (Accessed: 28 August 2016).

Koopmans, B. *et al.* (2017) ‘AHCODA-DB: a data repository with web-based mining tools for the analysis of automated high-content mouse phenomics data.’, *BMC bioinformatics*, 18(1), p. 200. doi: 10.1186/s12859-017-1612-1.

Korde, V. and Mahender, C. N. (2012) ‘Text Classification and Classifiers: A Survey’, *International Journal of Artificial Intelligence & Applications*, 3(2), pp. 85–99. doi: 10.5121/ijaia.2012.3208.

Koskimäki, J. *et al.* (2019) ‘Comprehensive transcriptome analysis of cerebral cavernous malformation across multiple species and genotypes.’, *JCI insight*, 4(3). doi: 10.1172/jci.insight.126167.

Kouylekov, M. *et al.* (2006) ‘Towards entailment-based Question Answering: ITC-irst at CLEF 2006’, in *CEUR Workshop Proceedings*. Available at: http://link.springer.com/content/pdf/10.1007/978-3-540-74999-8.pdf#page=547 (Accessed: 3 July 2017).

Laird, A. R. *et al.* (2015) ‘Neural architecture underlying classification of face perception paradigms.’, *NeuroImage*, 119, pp. 70–80. doi: 10.1016/j.neuroimage.2015.06.044.

Lample, G. *et al.* (2016) ‘Neural Architectures for Named Entity Recognition’. Available at: https://github.com/ (Accessed: 15 July 2019).

Landhuis, E. (2017) ‘Neuroscience: Big brain, big data’. Available at: https://www.nature.com/articles/541559a (Accessed: 28 October 2019).

Langlotz, C. P. (2006) ‘RadLex: A New Method for Indexing Online Educational Materials’, *RadioGraphics*, 26(6), pp. 1595–1597. doi: 10.1148/rg.266065168.

Larson, S. D. *et al.* (2007) ‘A formal ontology of subcellular neuroanatomy’, *Frontiers in Neuroinformatics*, 1(NOV). doi: 10.3389/neuro.11/003.2007.

Larson, S. D. and Martone, M. E. (2007) 'Rule-based reasoning with a multi-scale neuroanatomical ontology', in *CEUR Workshop Proceedings*, pp. 73–1613.

Larson, S. D. and Martone, M. E. (2009) 'Ontologies for neuroscience: What are they and what are they good for?', *Frontiers in Neuroscience*. Switzerland, 3(MAY), p. 60. doi: 10.3389/neuro.01.007.2009.

Larson, S. D. and Martone, M. E. (2013) 'NeuroLex.org: an online framework for neuroscience knowledge.', *Frontiers in neuroinformatics*, 7, p. 18. doi: 10.3389/fninf.2013.00018.

Lee, E. S. *et al.* (2009) 'Incorporating collaboratory concepts into informatics in support of translational interdisciplinary biomedical research', *International Journal of Medical Informatics*. Elsevier, 78(1), pp. 10–21. doi: 10.1016/J.IJMEDINF.2008.06.011.

Lenartowicz, A. *et al.* (2010) 'Towards an Ontology of Cognitive Control', *Topics in Cognitive Science*, 2(4), pp. 678–692. doi: 10.1111/j.1756-8765.2010.01100.x.

Li, J. *et al.* (2007) 'An ontology-based mining system for competitive intelligence in neuroscience', in *Web Intelligence Meets Brain Informatics*. Beijing, China: Springer Verlag (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)), pp. 291–304. doi: 10.1007/978-3-540-77028-2.

Li, X. and Roth, D. (2006) 'Learning question classifiers: the role of semantic information', *Natural Language Engineering*, 12(03), p. 229. doi: 10.1017/S1351324905003955.

Liggins, C. *et al.* (2014) 'International Alzheimer's Disease Research Portfolio (IADRP) aims to capture global Alzheimer's disease research funding', *Alzheimer's and Dementia*, 10(3), pp. 405–408. doi: 10.1016/j.jalz.2013.12.013.

Liu, H. *et al.* (2010) ‘Ontology-based mining of brainwaves: A sequence similarity technique for mapping alternative features in event-related potentials (ERP) data’, in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Hyderabad, India: Springer Verlag (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)), pp. 43–54. doi: 10.1007/978-3-642-13672-6_5.

Liu, S. *et al.* (2015) ‘Multimodal neuroimaging computing: a review of the applications in neuropsychiatric disorders’, *Brain Informatics*. Available at: http://link.springer.com/article/10.1007/s40708-015-0019-x (Accessed: 19 November 2015).

Liu, Z. *et al.* (2010) ‘Short text feature selection for micro-blog mining’, *2010 International Conference on Computational Intelligence and Software Engineering, CiSE 2010*. doi: 10.1109/CISE.2010.5677015.

Lloyd, C., Halstead, M. and Nielsen, P. (2004) ‘CellML: its future, present and past’, *Progress in biophysics and*. Available at: http://www.sciencedirect.com/science/article/pii/S007961070400015X (Accessed: 13 January 2017).

Lohrey, J. M., Killeen, N. E. B. and Egan, G. F. (2009) ‘An integrated object model and method framework for subject-centric e-Research applications’, *Frontiers in neuroinformatics*, 3. doi: 10.3389/neuro.11.019.2009.

Lopez, V. *et al.* (2007) ‘AquaLog: An ontology-driven question answering system for organizational semantic intranets’, *Web Semantics*, 5(2), pp. 72–105. doi: 10.1016/j.websem.2007.03.003.

Lopez, V. *et al.* (2012) ‘PowerAqua: Supporting users in querying and exploring the Semantic Web’, *Semantic Web*, 3(3), pp. 249–265. doi: 10.3233/SW-2011-0030.

Lopez, V., Pasin, M. and Motta, E. (2005) ‘AquaLog: An Ontology-Portable Question Answering System for the Semantic Web’, in Gómez-Pérez, A. and Euzenat, J. (eds) *The Semantic Web: Research and Applications*. Springer Berlin Heidelberg, pp. 546–562. doi: 10.1007/11431053_37.

Lorhard, J. and Uckelman, S. L. (1606) 'Diagraph of Metaphysic or Ontology'.

Magnini, B., Speranza, M. and Kumar, V. (2009) 'Towards interactive question answering: an ontology-based approach', in *Semantic Computing, 2009. ICSC'09. IEEE International Conference on*. IEEE, pp. 612–617.

Majka, P. *et al.* (2012) 'Common atlas format and 3D brain atlas reconstructor: Infrastructure for Constructing 3D brain atlases', *Neuroinformatics*, 10(2), pp. 181–197. doi: 10.1007/s12021-011-9138-6.

Malhotra, A. *et al.* (2014) 'ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease', *Alzheimer's and Dementia*, 10(2), pp. 238–246. doi: 10.1016/j.jalz.2013.02.009.

Marenco, L. *et al.* (2007) 'Interoperability across neuroscience databases.', *Methods in molecular biology (Clifton, N.J.)*. United States, 401, pp. 23–36. doi: 10.1007/978-1-59745-520-6_2.

Marenco, L. *et al.* (2008) 'Issues in the design of a pilot concept-based query interface for the Neuroinformatics Information Framework', *Neuroinformatics*, 6(3), pp. 229–239. doi: 10.1007/s12021-008-9035-9.

Marenco, L. *et al.* (2010) 'The NIF DISCO framework: Facilitating automated integration of neuroscience content on the web', *Neuroinformatics*, 8(2), pp. 101–112. doi: 10.1007/s12021-010-9068-8.

Marenco, L., Wang, R. and Nadkarni, P. (2009) 'Automated Database Mediation Using Ontological Metadata Mappings', *Journal of the American Medical Informatics Association*, 16(5), pp. 723–737. doi: 10.1197/jamia.M3031.

Martone, M. E. *et al.* (2008) 'The cell centered database project: an update on building community resources for managing and sharing 3D imaging data', *Journal of structural biology*, 161(3), pp. 220–231.

Matthew Horridge, S. B. (2011) 'The OWL API : A Java API for OWL Ontologies', *Semantic Web journal*, pp. 1–11. Available at: https://content.iospress.com/articles/semantic-web/sw025 (Accessed: 23 June 2019).

Maynard, S. M. *et al.* (2013) 'A knowledge based approach to matching human neurodegenerative disease and animal models.', *Frontiers in neuroinformatics*, 7(May), p. 7. doi: 10.3389/fninf.2013.00007.

McCandless, M., Hatcher, E. and Gospodnetic, O. (2010) *Lucene in action: covers Apache Lucene 3.0*. Available at: https://dl.acm.org/citation.cfm?id=1893016 (Accessed: 1 July 2019).

McCray, A. T., Trevvett, P. and Frost, H. R. (2014) 'Modeling the autism spectrum disorder phenotype', *Neuroinformatics*, 12(2), pp. 291–305. doi: 10.1007/s12021-013-9211-4.

McGuinness, D. L. *et al.* (2003) 'DAML-ONT: An Ontology Language for the Semantic Web', *Spinning the Semantic Web: Bringing the World Wide Web to Its Full Potential*, pp. 65–94. Available at: http://ksl.stanford.edu/pub/KSL_Reports/KSL-01-08.htm.gz (Accessed: 11 September 2014).

McGuinness, D. L. (2004) 'Question answering on the semantic Web', *Intelligent Systems, IEEE*, 19(1), pp. 82–85. doi: 10.1109/mis.2004.1265890.

McGuinness, D. L. and Van Harmelen, F. (2004) 'OWL web ontology language overview', *W3C recommendation*, 10(2004–03), p. 10.

McKinney, W. and Team, P. D. (2015) 'Pandas—powerful Python data analysis toolkit', *pandas.pydata.org*. Available at: http://pandas.pydata.org/pandas-docs/version/0.7.3/pandas.pdf (Accessed: 5 July 2019).

McKinney, W. (2012) *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. Available at: https://books.google.com.au/books?hl=en&lr=&id=v3n4_AK8vu0C&oi=fnd&pg=PR3&dq=pandas+excel+python&ots=rgJG6iCzsy&sig=5i_GYJmtlwuj8UH6rFFViM5FVlA (Accessed: 5 July 2019).

Meehan, T. F. *et al.* (2011) 'Logical Development of the Cell Ontology', *BMC Bioinformatics*, 12(1), p. 6. doi: 10.1186/1471-2105-12-6.

Mejino, J. L. V., Rubin, D. L. and Brinkley, J. F. (2008) 'FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology.', in *AMIA Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*. American Medical Informatics Association, pp. 465–469.

Mejino, J. L. V *et al.* (2010) 'Enabling RadLex with the Foundational Model of Anatomy Ontology to Organize and Integrate Neuro-imaging Data', *AMIA 2010 Symposium Proceedings*, p. 1171.

Merlo, L. *et al.* (2014) 'Alteration in synaptic junction proteins following traumatic brain injury', *Journal of Neurotrauma*, 31(16), pp. 1375–1385. doi: 10.1089/neu.2014.3385.

Miller, L., Seaborne, A. and Reggiori, A. (2002) 'Three implementations of SquishQL, a simple RDF query language', *International Semantic Web Conference*. Available at: http://link.springer.com/chapter/10.1007/3-540-48005-6_36 (Accessed: 13 January 2017).

Milyaev, N. *et al.* (2012) 'The virtual fly brain browser and query interface', *Bioinformatics*, 28(3), pp. 411–415. doi: 10.1093/bioinformatics/btr677.

Mirbel, I. and Crescenzo, P. (2009) 'Improving collaborations in neuroscientist community', in *Proceedings - 2009 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IAT Workshops 2009*. IEEE, pp. 567–570. doi: 10.1109/WI-IAT.2009.351.

Miyata, S. *et al.* (2016) ‘Gene expression alterations in the medial prefrontal cortex and blood cells in a mouse model of depression during menopause.’, *Heliyon*, 2(12), p. e00219. doi: 10.1016/j.heliyon.2016.e00222.

Mohamad Hazawawi, N. A. *et al.* (2015) ‘A structured literature review on applied neuroscience in Information Systems (neuroIS) BT - 2nd International Symposium on Technology Management and Emerging Technologies, ISTMET 2015, August 25, 2015 - August 27, 2015’, in *2nd International Symposium on Technology Management and Emerging Technologies, ISTMET 2015 - Proceeding*. Comm, Multimedia, Network and Sys Research Group, Universiti Teknikal Malaysia Melaka, Durian Tunggal Melaka, MalaysiaMicrowave Research Group, Universiti Teknikal Malaysia Melaka, Durian Tunggal Melaka, Malaysia: Institute of Electrical and Electronics Engineers Inc. (2nd International Symposium on Technology Management and Emerging Technologies, ISTMET 2015 - Proceeding), pp. 20–24. doi: 10.1109/ISTMET.2015.7358993.

Moreda, P. *et al.* (2011) ‘Combining semantic information in question answering systems’, *Information Processing and Management*. Elsevier Ltd, 47(6), pp. 870–885. doi: 10.1016/j.ipm.2010.03.008.

Morgan, S. E. *et al.* (2019) ‘Cortical patterning of abnormal morphometric similarity in psychosis is associated with brain expression of schizophrenia-related genes.’, *Proceedings of the National Academy of Sciences of the United States of America*, 116(19), pp. 9604–9609. doi: 10.1073/pnas.1820754116.

MouÄ⍰ek, R. (2014) ‘Corrigendum: Software and hardware infrastructure for research in electrophysiology’, *Frontiers in Neuroinformatics*, 8(March), p. 20. doi: 10.3389/fninf.2014.00075.

Müller, H.-M. *et al.* (2018) ‘Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature’, *BMC Bioinformatics*, 19(1), p. 94. doi: 10.1186/s12859-018-2103-8.

Müller, H.-M., Kenny, E. E. and Sternberg, P. W. (2004) ‘Textpresso: an ontology-based information retrieval and extraction system for biological literature’, *PLoS biology*, 2(11), p. e309.

Müller, H. M. *et al.* (2008) 'Textpresso for neuroscience: Searching the full text of thousands of neuroscience research papers', *Neuroinformatics*. Division of Biology, Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA, United States, 6(3), pp. 195–204. doi: 10.1007/s12021-008-9031-0.

Musen, M. (2015) 'The protégé project: a look back and a look forward', *AI Matters*. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4883684/ (Accessed: 19 May 2019).

Nichols, B. N. *et al.* (2014) 'Neuroanatomical domain of the foundational model of anatomy ontology.', *Journal of biomedical semantics*, 5(1), p. 1. doi: 10.1186/2041-1480-5-1.

Nichols, B. N., Mejino Jr, J. L. V and Brinkley, J. F. (2011) 'The Foundational Model of Neuroanatomy Ontology: An Ontology Framework to Support Neuroanatomical Data Integration', in *ICBO*.

Nielsen, F. Å. (2014) 'Brede tools and federating online neuroinformatics databases', *Neuroinformatics*, 12(1), pp. 27–37. doi: 10.1007/s12021-013-9183-4.

Nowke, C. *et al.* (2015) 'Integrating Visualizations into Modeling NEST Simulations', *Frontiers in Neuroinformatics*, 9(October), pp. 1–20. doi: 10.3389/fninf.2015.00029.

Noy, N. F. *et al.* (2009) 'BioPortal: ontologies and integrated data resources at the click of a mouse.', *Nucleic acids research*. Oxford University Press, 37(Web Server issue), pp. W170-3. doi: 10.1093/nar/gkp440.

O'Reilly, C., Iavarone, E. and Hill, S. L. (2017) 'A Framework for Collaborative Curation of Neuroscientific Literature.', *Frontiers in neuroinformatics*, 11, p. 27. doi: 10.3389/fninf.2017.00027.

Olesen, J. *et al.* (2012) 'The economic cost of brain disorders in Europe', *European Journal of Neurology*, 19(1), pp. 155–162. doi: 10.1111/j.1468-1331.2011.03590.x.

Onwuegbuzie, A. and Johnson, R. (2006) 'The validity issue in mixed research', *Research in the Schools*, 13(1), pp. 48–63. Available at: http://www.msera.org/docs/rits-v13n1-complete.pdf#page=55 (Accessed: 28 October 2019).

Osumi-Sutherland, D. *et al.* (2012) 'A strategy for building neuroanatomy ontologies', *Bioinformatics*, 28(9), pp. 1262–1269. doi: 10.1093/bioinformatics/bts113.

Ou, S. *et al.* (2008) 'Automatic question pattern generation for ontology-based question answering', *Proceedings of the 21st International Florida Artificial Intelligence Research Society Conference.*, pp. 183–188. doi: 10.1.1.477.1018.

Ozyurt, I. *et al.* (2010) 'Federated web-accessible clinical data management within an extensible neuroimaging database', *Neuroinformatics*. Humana Press Inc., 8(4), pp. 231–249. doi: 10.1007/s12021-010-9078-6.

Pasca, M. and Harabagiu, S. (2001) 'High performance question/answering', in *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM. Available at: http://dl.acm.org/citation.cfm?id=384025 (Accessed: 4 February 2016).

Patel, S. *et al.* (2016) 'Gene Prioritization for Imaging Genetics Studies Using Gene Ontology and a Stratified False Discovery Rate Approach.', *Frontiers in neuroinformatics*, 10, p. 14. doi: 10.3389/fninf.2016.00014.

Pérez, J., Arenas, M. and Gutierrez, C. (2006) 'Semantics and complexity of SPARQL', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4273 LNCS, pp. 30–43. doi: 10.1007/11926078_3.

Petr, B. *et al.* (2013) 'The Ontology for Experimental Neurophysiology: a first step toward semantic annotations of neurophysiology data and metadata.', *Frontiers in Neuroinformatics*, 7(August), pp. 37–38. doi: 10.3389/conf.fninf.2013.09.00026.

Poldrack, R. A. *et al.* (2011) 'The cognitive atlas: toward a knowledge foundation for cognitive neuroscience.', *Frontiers in neuroinformatics*, 5(September), pp. 1–11. doi: 10.3389/fninf.2011.00017.

Poliakov, A. V *et al.* (2007) 'Unobtrusive integration of data management with fMRI analysis', *Neuroinformatics*, 5(1), pp. 3–10. doi: NI:5:1:3 [pii].

Poline, J.-B. *et al.* (2012) 'Data sharing in neuroimaging research', *Frontiers in Neuroinformatics*, 6(April), p. 9. doi: 10.3389/fninf.2012.00009.

Portales-Casamar, E. *et al.* (2013) 'Neurocarta: aggregating and sharing disease-gene relations for the neurosciences.', *BMC genomics*, 14(1), p. 129. doi: 10.1186/1471-2164-14-129.

Powers, D. (2011) 'Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation'. Available at: http://dspace2.flinders.edu.au/xmlui/handle/2328/27165 (Accessed: 5 March 2016).

Prodanov, D. (2011) 'Data Ontology and an Information System Realization for Web-Based Management of Image Measurements', *Frontiers in Neuroinformatics*, 5(November), pp. 1–18. doi: 10.3389/fninf.2011.00025.

Provenzano, G. *et al.* (2016) 'Comparative Gene Expression Analysis of Two Mouse Models of Autism: Transcriptome Profiling of the BTBR and En2−/− Hippocampus', *Frontiers in Neuroscience*. Switzerland, 10, p. 396. doi: 10.3389/fnins.2016.00396.

Prud, E. and Seaborne, A. (2006) 'Sparql query language for rdf', *W3C Recommendation, http://www. w3. org/TR/rdf-sparql-query*. Available at: https://scholar.google.com.au/scholar?hl=en&q=sparql&btnG=&as_sdt=1%252C5&as_sdtp=#0 (Accessed: 25 November 2015).

Raikov, I. and De Schutter, E. (2012) 'The Promise and Shortcomings of XML as an Interchange Format for Computational Models of Biology', *Neuroinformatics*, 10(1), pp. 1–3. doi: 10.1007/s12021-011-9136-8.

Ranjan, R. *et al.* (2011) 'Channelpedia: an integrative and interactive database for ion channels.', *Frontiers in neuroinformatics*, 5(December), p. 36. doi: 10.3389/fninf.2011.00036.

Refolo, L. M. *et al.* (2012) 'Common Alzheimer's Disease Research Ontology: National Institute on Aging and Alzheimer's Association collaborative project.', *Alzheimer's & dementia : the journal of the Alzheimer's Association*, 8(4), pp. 372–5. doi: 10.1016/j.jalz.2012.05.2115.

Reynolds, D. (2004) 'Jena 2 inference support', *Online manual at http://jena. sourceforge. net/inference/index. html*.

Ribadas, F. J. *et al.* (2013) 'Two hierarchical text categorization approaches for BioASQ semantic indexing challenge', *CEUR Workshop Proceedings*, 1094. Available at: http://www.grupolys.org/biblioteca/RibCamDarRom2013a.pdf (Accessed: 1 January 2016).

Riboni, D., Civitarese, G. and Bettini, C. (2016) 'Analysis of long-term abnormal behaviors for early detection of cognitive decline', in *2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, pp. 1–6. doi: 10.1109/PERCOMW.2016.7457139.

Richardson, M. and Domingos, P. (2006) 'Markov logic networks', in *Machine Learning*, pp. 107–136. doi: 10.1007/s10994-006-5833-1.

Roberts, K. *et al.* (2014) 'Annotating question decomposition on complex medical questions', in *Proceedings of the 9th International Conference on Language Resources and Evaluation, LREC 2014*, pp. 2598–2602. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.675.3142&rep=rep1&type=pdf (Accessed: 9 October 2019).

Roldán-García, M. del M. *et al.* (2009) 'KA-SB: From data integration to large scale reasoning', *BMC Bioinformatics*, 10(SUPPL. 10), p. S5. doi: 10.1186/1471-2105-10-S10-S5.

Rosse, C. and Mejino, J. L. V. (2003) 'A reference ontology for biomedical informatics: The Foundational Model of Anatomy', *Journal of Biomedical Informatics*, 36(6), pp. 478–500. doi: 10.1016/j.jbi.2003.11.007.

Rubin, D. L. *et al.* (2009) 'Computational neuroanatomy: ontology-based representation of neural components and connectivity.', *BMC bioinformatics*. England, 10 Suppl 2, p. S3. doi: 10.1186/1471-2105-10-S2-S3.

Rubin, D. L., Greenspan, H. and Brinkley, J. F. (2014) 'Biomedical imaging informatics', *Biomedical Informatics: Computer Applications in Health Care and Biomedicine: Fourth Edition*, pp. 285–327. doi: 10.1007/978-1-4471-4474-8_9.

Rubin, D. L. and Napel, S. (2010) 'Imaging informatics: toward capturing and processing semantic information in radiology images.', *Yearbook of medical informatics*, 19(01), pp. 34–42. doi: 10.1055/s-0038-1638686.

Ryle, G. (1937) 'Categories', *JSTOR*, 38, pp. 189–206. Available at: https://www.jstor.org/stable/4544305.

Safae, L., Habib, B. El and Abderrahim, T. (2018) 'A Review of Machine Learning Algorithms for Web Page Classification', *Colloquium in Information Science and Technology, CIST*, 2018-Octob, pp. 220–226. doi: 10.1109/CIST.2018.8596420.

Sahoo, S., Lhatoo, S. and Gupta, D. (2014) 'Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care', *Journal of the*. Available at: http://jamia.oxfordjournals.org/content/21/1/82.short (Accessed: 14 December 2016).

Sahoo, S. S. *et al.* (2012) 'OPIC: Ontology-driven Patient Information Capturing system for epilepsy.', *AMIA Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, pp. 799–808. Available at: http://www.ncbi.nlm.nih.gov/pubmed/23304354 (Accessed: 14 December 2016).

Sahoo, S. S. *et al.* (2016) 'NeuroPigPen: A scalable toolkit for processing electrophysiological signal data in neuroscience applications using apache pig', *Frontiers in Neuroinformatics*, 10(JUNE). doi: 10.3389/fninf.2016.00018.

Salvadores, M. and Horridge, M. (2012) 'Using sparql to query bioportal ontologies and metadata', *The Semantic Web– ….* Available at: http://link.springer.com/chapter/10.1007/978-3-642-35173-0_12 (Accessed: 11 September 2014).

Samwald, M. *et al.* (2009) 'Entrez neuron RDFa: A pragmatic Semantic Web application for data integration in neuroscience research', *Studies in Health Technology and Informatics*. Netherlands, 150, pp. 317–321. doi: 10.3233/978-1-60750-044-5-317.

Samwald, M. *et al.* (2010) 'Semantic SenseLab: Implementing the vision of the Semantic Web in neuroscience', *Artificial Intelligence in Medicine*, 48(1), pp. 21–28. doi: 10.1016/j.artmed.2009.11.003.

Samwald, M. and Adlassnig, K.-P. (2005) 'Bringing neuroscience to the semantic web: The semantic synapse project BT - International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2005 and International Conference on Intelligent Agents, Web Technologies and', in. Section on Medical Expert and Knowledge-Based Systems, Core Unit for Medical Statistics and Informatics, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, AustriaLudwig Boltzmann Institute for Expert Systems and Quality Management in Medicine, : Inst. of Elec. and Elec. Eng. Computer Society (Proceedings - International Conference on Computational Intelligence for Modelling, Control and Automation, CIMCA 2005 and International Conference on Intelligent Agents, Web Technologies and Internet), pp. 145–769525040.

Sanchez-Pi, N., Martí, L. and Garcia, A. (2015) 'Improving ontology-based text classification: An occupational health and security application', *Journal of Applied Logic*. Available at: http://www.sciencedirect.com/science/article/pii/S1570868315000774 (Accessed: 22 December 2015).

Saquete, E. *et al.* (2004) ‘Splitting complex temporal questions for question answering systems’, in *dl.acm.org*, pp. 566-es. doi: 10.3115/1218955.1219027.

Sattler, U. *et al.* (2008) ‘OWL2: The Next Step for OWL’, *SSRN Electronic Journal*, (November). doi: 10.2139/ssrn.3199412.

Schmitt, O. and Eipert, P. (2012) ‘NeuroVIISAS: Approaching multiscale simulation of the rat connectome’, *Neuroinformatics*, 10(3), pp. 243–267. doi: 10.1007/s12021-012-9141-6.

Schofield, P. N. *et al.* (2013) ‘The mouse pathology ontology, MPATH; structure and applications.’, *Journal of biomedical semantics*, 4(1), p. 18. doi: 10.1186/2041-1480-4-18.

Schwartz, Y., Thirion, B. and Varoquaux, G. (2013) ‘Mapping cognitive ontologies to and from the brain BT - 27th Annual Conference on Neural Information Processing Systems, NIPS 2013, December 5, 2013 - December 10, 2013’, in. Parietal Team, Inria Saclay Ile-de-France, Saclay, France: Neural information processing systems foundation (Advances in Neural Information Processing Systems), p. Air Force Office of Scientific Research (AFOSR); A.

Setchi, R. and Tang, Q. (2007) ‘Concept indexing using ontology and supervised machine learning’, *World Academy of Science, Engineering and Technology*. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.122.9574&rep=rep1&type=pdf (Accessed: 13 January 2017).

Shang, Y. *et al.* (2013) ‘Enhancing entity annotation using web service and ontology hierarchy in biomedical domains BT - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013, December 18, 2013 - December 21, 2013’, in *2013 IEEE International Conference on Bioinformatics and Biomedicine*. College of Computing and Informatics, Drexel University, Philadelphia, PA, United States: IEEE Computer Society (Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013), pp. 465–468. doi: 10.1109/BIBM.2013.6732537.

Shaw, M. *et al.* (2008) ‘Generating application ontologies from reference ontologies’, in *AMIA Annual Symposium Proceedings*. American Medical Informatics Association, p. 672.

Shaw, M. *et al.* (2011) ‘vSPARQL: A view definition language for the semantic web’, *Journal of Biomedical Informatics*, 44(1), pp. 102–117. doi: http://dx.doi.org/10.1016/j.jbi.2010.08.008.

Shearer, R., Motik, B. and Horrocks, I. (2008) ‘HermiT: A Highly-Efficient OWL Reasoner’, in *OWLED*.

Sirin, E. *et al.* (2007) ‘Pellet: A practical OWL-DL reasoner’, *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), pp. 51–53. doi: http://dx.doi.org/10.1016/j.websem.2007.03.004.

Slater, L. *et al.* (2015) ‘Using Aber-OWL for fast and scalable reasoning over BioPortal ontologies’, in *CEUR Workshop Proceedings*. Lisbon, Portugal, p. 49. doi: 10.1186/s13326-016-0090-0.

Smith, B. *et al.* (2005) ‘Relations in biomedical ontologies.’, *Genome biology*, 6(5), p. R46. doi: 10.1186/gb-2005-6-5-r46.

Smith, B. *et al.* (2007) ‘The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration’, *Nature Biotechnology*. United States: Nature America Publishing : New York Ny, 25(11), pp. 1251–1255. doi: 10.1038/nbt1346.

Smith, B. and Grenon, P. (2002) ‘Basic formal ontology’, *Draft. Downloadable at http://ontology. buffalo. edu/bfo*.

Smith, B., Williams, J. and Schulze-Kremer, S. (2003) ‘The ontology of the gene ontology.’, *AMIA Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2003, pp. 609–613. Available at: http://sfx.unimelb.hosted.exlibrisgroup.com/sfxlcl41?sid=google&auinit=B&aulast=Smith&atitle=The ontology of the gene ontology&title=AMIA Annual Symposium proceedings&volume=2003&date=2003&spage=609&issn=1559-4076.

Song, M. H. *et al.* (2005) 'Automatic classification of web pages based on the concept of domain ontology', *Proceedings - Asia-Pacific Software Engineering Conference, APSEC*, 2005, pp. 645–651. doi: 10.1109/APSEC.2005.46.

Sriram, B. *et al.* (2010) 'Short text classification in twitter to improve information filtering', *Proceedings of the 33rd international ACM SIGIR conference on Research\r\nand development in information retrieval*, pp. 841–842. Available at: http://dl.acm.org/citation.cfm?id=1835643 (Accessed: 24 December 2015).

Staab, S. and Studer, R. (2009) *Handbook on Ontologies*, *Handbook on Ontologies*. Springer. doi: 10.1007/978-3-540-92673-3.

Stevens, R. *et al.* (2003) 'Complex query formulation over diverse information sources in TAMBIS', *Bioinformatics: Managing Scientific Data. Morgan Kaufmann*.

Studer, R., Benjamins, V. R. and Fensel, D. (1998) 'Knowledge engineering: principles and methods', *Data & Knowledge Engineering*, 25(1), pp. 161–197.

Studtmann, P. (2018) *Aristotle's Categories*, *The Stanford Encyclopedia of Philosophy (Fall 2018 Edition)*. Available at: https://plato.stanford.edu/archives/fall2018/entries/aristotle-categories/.

Subbarao, J. *et al.* (2010) 'Towards an ontology for psychosis Action editor : Ning Zhong', *Cognitive Systems Research*, 11(1), pp. 42–52. doi: 10.1016/j.cogsys.2008.08.005.

Sun, S., Luo, C. and Chen, J. (2017) 'A review of natural language processing techniques for opinion mining systems', *Information Fusion*, 36, pp. 10–25. doi: 10.1016/j.inffus.2016.10.004.

Sutton, C. and Machine, A. M.- (2012) 'An Introduction to Conditional Random Fields for Relational Learning', in *Introduction to Statistical Relational Learning*. doi: 10.7551/mitpress/7432.003.0006.

Svačinová, I. (2019) ‘The Thesis of the Effectiveness of Quasi-logical Arguments’, *Argumentation*, 33(1), pp. 75–106. doi: 10.1007/s10503-018-9464-z.

Swanson, L. W. (2018) ‘Brain maps 4.0—Structure of the rat brain: An open access atlas with global nervous system nomenclature ontology and flatmaps’, *Journal of Comparative Neurology*, 526(6), pp. 935–943. doi: 10.1002/cne.24381.

Tan, A. (1999) ‘Text mining: The state of the art and the challenges’, *Proceedings of the PAKDD 1999 Workshop on*. Available at: http://www.ntu.edu.sg/home/asahtan/papers/tm_pakdd99.pdf (Accessed: 13 January 2017).

Tashakkori, A. and Creswell, J. W. (2008) ‘Editorial: Mixed Methodology Across Disciplines’, *Journal of Mixed Methods Research*, 2(1), pp. 3–6. doi: 10.1177/1558689807309913.

Temal, L. *et al.* (2008) ‘Towards an ontology for sharing medical images and regions of interest in neuroimaging’, *Journal of Biomedical Informatics*, 41(5), pp. 766–778. doi: 10.1016/j.jbi.2008.03.002.

Tenenboim, L., Shapira, B. and Shoval, P. (2008) ‘Ontology-based classification of news in an electronic newspaper’. Available at: http://sci-gems.math.bas.bg/jspui/handle/10525/1035 (Accessed: 31 December 2015).

Terrell, S. (2012) ‘Mixed-methods research methodologies’, *The Qualitative Report*, 17(1), pp. 254–280. Available at: http://nsuworks.nova.edu/cgi/viewcontent.cgi?article=1819&context=tqr/ (Accessed: 6 June 2019).

Tiesinga, P. *et al.* (2015) ‘Feeding the human brain model’, *Current Opinion in Neurobiology*, 32, pp. 107–114. doi: 10.1016/j.conb.2015.02.003.

Timón, S., Rincón, M. and Martínez-Tomás, R. (2017) ‘Extending XNAT Platform with an Incremental Semantic Framework.’, *Frontiers in neuroinformatics*, 11, p. 57. doi: 10.3389/fninf.2017.00057.

Tomás, D. and Vicedo, J. L. (2012) 'Minimally supervised question classification on fine-grained taxonomies', *Knowledge and Information Systems*, 36(2), pp. 303–334. doi: 10.1007/s10115-012-0557-y.

Tran, T. *et al.* (2007) 'Ontology-based interpretation of keywords for semantic search', *The Semantic Web*. Available at: http://link.springer.com/chapter/10.1007/978-3-540-76298-0_38 (Accessed: 13 January 2017).

Tsarkov, D. and Horrocks, I. (2006) 'FaCT++ description logic reasoner: System description', in *Automated reasoning*. Springer, pp. 292–297.

Tsatsaronis, G. *et al.* (2015) 'An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition', *BMC Bioinformatics*, 16(1), p. 138. doi: 10.1186/s12859-015-0564-6.

Turner, J. A. and Laird, A. R. (2012) 'The cognitive paradigm ontology: Design and application', *Neuroinformatics*, 10(1), pp. 57–66. doi: 10.1007/s12021-011-9126-x.

Turner, J. a *et al.* (2010) 'Application of neuroanatomical ontologies for neuroimaging data annotation.', *Frontiers in neuroinformatics*, 4(June), pp. 1–12. doi: 10.3389/fninf.2010.00010.

Turner, M. D. *et al.* (2013) 'Automated annotation of functional imaging experiments via multi-label classification', *Frontiers in Neuroscience*, 7(7 DEC), p. 240. doi: 10.3389/fnins.2013.00240.

Unger, C. and Bühmann, L. (2012) 'Template-based question answering over RDF data', *Proceedings of the 21st international conference on World Wide Web*, pp. 639–648. doi: 10.1145/2187836.2187923.

Uren, V. *et al.* (2006) 'Semantic annotation for knowledge management: Requirements and a survey of the state of the art', *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(1), pp. 14–28. doi: http://dx.doi.org/10.1016/j.websem.2005.10.002.

Uysal, A. and Gunal, S. (2012) ‘A novel probabilistic feature selection method for text classification’, *Knowledge-Based Systems*. Available at: http://www.sciencedirect.com/science/article/pii/S0950705112001761 (Accessed: 31 December 2015).

Uysal, A. K. and Gunal, S. (2014) ‘The impact of preprocessing on text classification’, *Information Processing & Management*, 50(1), pp. 104–112. doi: 10.1016/j.ipm.2013.08.006.

Veeraraghavan, H. and Miller, J. V. (2014) ‘Faceted visualization of three dimensional neuroanatomy by combining ontology with faceted search’, *Neuroinformatics*, 12(2), pp. 245–259. doi: 10.1007/s12021-013-9202-5.

Verbeeck, R. *et al.* (2010) ‘Relational to RDF mapping using D2R for translational research in neuroscience BT - 4th International Workshop on Formal Biomedical Knowledge Representation: Semantic Applications in Life Sciences, KR-MED 2010, July 9, 2010 - July 10, 2010’, in. Johnson and Johnson Pharmaceutical Research and Development, Turnhoutseweg 30, Beerse, BelgiumJohnson and Johnson Pharmaceutical Research and Development, Welch and McKean Roads, Spring House PA, United StatesJohnson and Johnson Pharmaceutical Research an: Sun SITE Central Europe CEUR-WS (CEUR Workshop Proceedings).

Wan, Y. *et al.* (2015) ‘BlastNeuron for Automated Comparison, Retrieval and Clustering of 3D Neuron Morphologies’, *Neuroinformatics*, 13(4), pp. 487–499. doi: 10.1007/s12021-015-9272-7.

Wang, L. *et al.* (2013) ‘Northwestern University Schizophrenia Data and Software Tool (NUSDAST).’, *Frontiers in neuroinformatics*, 7(November), p. 25. doi: 10.3389/fninf.2013.00025.

Wang, L. *et al.* (2016) ‘SchizConnect: Mediating neuroimaging databases on schizophrenia and related disorders for large-scale integration’, *NeuroImage*, 124, pp. 1155–1167. doi: 10.1016/j.neuroimage.2015.06.065.

Wiese, C. *et al.* (2012) 'A genome-wide screen to identify transcription factors expressed in pelvic ganglia of the lower urinary tract', *in neuroscience*. Available at: http://journal.frontiersin.org/article/10.3389/fnins.2012.00130 (Accessed: 26 November 2016).

Wu, D. *et al.* (2016) 'Resource atlases for multi-atlas brain segmentations with multiple ontology levels based on T1-weighted MRI', *NeuroImage*, 125, pp. 120–130. doi: 10.1016/j.neuroimage.2015.10.042.

Yahya, A. A. and Osman, A. (2011) 'Automatic Classification of Questions Into Bloom's Classification', *Proceedings of the International Arab Conference on Information Technology*, pp. 1–6. Available at: www.nauss.edu.sa/acit/PDFs/f2914.pdf (Accessed: 11 February 2016).

Yan, T. and Tourangeau, R. (2008) 'Fast times and easy questions: the effects of age, experience and question complexity on web survey response times', *Applied Cognitive Psychology*, 22(1), pp. 51–68. doi: 10.1002/acp.1331.

Yao, X. and Van Durme, B. (2014) 'Information extraction over structured data: Question answering with freebase', *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, 1, pp. 956–966. doi: 10.3115/v1/p14-1090.

Yu, H., Sable, C. and Zhu, H. R. (2005) 'Classifying medical questions based on an evidence taxonomy', *AAAI Workshop - Technical Report*, WS-05-10, pp. 27–35. Available at: http://www.aaai.org/Papers/Workshops/2005/WS-05-10/WS05-10-005.pdf (Accessed: 1 October 2014).

Yu, Z. *et al.* (2010) 'Question classification based on co-training style semi-supervised learning', *Pattern Recognition Letters*, 31(13), pp. 1975–1980. doi: 10.1016/j.patrec.2010.06.010.

Van Zaanen, M., Pizzato, L. A. and Moll??, D. (2005) 'Classifying sentences using induced structure', *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 3772 LNCS, pp. 139–150. doi: 10.1007/11575832_15.

Zaldivar, A. and Krichmar, J. L. (2014) 'Allen Brain Atlas-Driven Visualizations: a web-based gene expression energy visualization tool.', *Frontiers in neuroinformatics*, 8(May), p. 51. doi: 10.3389/fninf.2014.00051.

Zao, J. K. *et al.* (2014) 'Pervasive brain monitoring and data sharing based on multi-tier distributed computing and linked data technology.', *Frontiers in human neuroscience*, 8(JUNE), p. 370. doi: 10.3389/fnhum.2014.00370.

Zaveri, A. *et al.* (2010) 'Achieving high research reporting quality through the use of computational ontologies', *Neuroinformatics*, 8(4), pp. 261–271. doi: 10.1007/s12021-010-9079-5.

Zhang, D. and Lee, W. S. (2003) 'Question Classification using Support Vector Machines', *SIGIR Forum (ACM Special Interest Group on Information Retrieval)*, (SPEC. ISS.), pp. 26–32. doi: 10.1145/860442.860443.