A MEDICAL ULTRASOUND REPORTING SYSTEM BASED ON DOMAIN ONTOLOGY

A THESIS SUBMITTED TO THE UNIVERSITY OF SALFORD
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
IN THE SCHOOL OF COMPUTING, SCIENCE AND ENGINEERING

2017

By
Nur Zareen Zulkarnain

Contents

List of Tables

List of Figures

Abstract

Ultrasound reports are produced in different ways by radiologists. These variations in reporting style could impact on the value of the report and the way it is interpreted, which in turn may have implications for patients' management and decision making. As the images produced will not give the whole view of the examination, it is vital that a high quality and standardised ultrasound report is produced. In addition to their medical value, ultrasound reports contain a lot of important information that can be very useful in research and education. Reports can contain a variety of terms or heterogeneous terminologies used for describing similar findings. This research project aims to develop a medical ultrasound reporting system that uses domain ontology as its knowledge base to support the generation of standardised reports as well as Rhetorical Structure Theory (RST) to transform free text reports to the preferred structured and standardised format. The domain ontology will specifically focus on abdominal ultrasound scanning which includes both the anatomy and pathology of the organs in the abdominal area. The ontology was developed using an ontology reuse methodology where terms from the sample reports were mapped to existing biomedical ontologies. It is anticipated that a standardised report based on domain ontology will improve the quality of ultrasound reports and encourage its implementation.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Acknowledgements

I would like to thank everyone who has helped in making the completion of this thesis possible. My gratitude goes to my supervisor Professor Farid Meziane for his dedication, constant support and patience, also for his supervision and constructive criticisms.

I would also like to thank my co-supervisor, Dr Gillian Crofts for her invaluable assistance and expertise in the ultrasound field and for helping me in obtaining the data for my research. My thanks to my colleagues in the Data Mining and Pattern Recognition Research Centre for making my postgraduate research journey a pleasant one.

Many thanks to my previous and present housemates who have made my stay in Manchester an enjoyable and memorable one. I am indebted to my best friend for putting up with me and my constant babble about my work. Special thanks to Manchester United for showing me that passion and perseverance are the keys to success.

Thank you to the Universiti Teknikal Malaysia Melaka (UTeM) for their financial assistance and the Malaysian Students Department for looking after my welfare in all my years in Manchester.

And most important of all, thank you to my parents and siblings for which their undivided love, unfailing support and constant encouragements have made it possible for me to come this far. Their love and sacrifices will be treasured forever. A special thank you to my father for the sacrifices he has made for the family. And also to my mother whom I have always looked up to in my journey to becoming an academician. This PhD degree is dedicated to the both of you.

Publications

The following are conference and journal articles that have been submitted and published based on intermediate result of this research:

1. Zulkarnain, N. Z., Meziane, F., & Crofts, G. (2016). A Methodology for Ontology Reuse: Application to the Biomedical Domain. Manuscript pre-selected and submitted to the Data and Knowledge Engineering Journal, Elsevier.

2. Zulkarnain, N. Z., Meziane, F., & Crofts, G. (2016, June). A Methodology for Biomedical Ontology Reuse. In International Conference on Applications of Natural Language to Information Systems (pp. 3-14). Springer International Publishing.

3. Zulkarnain, N. Z., Crofts, G. S., & Meziane, F. (2015). An architecture to support ultrasound report generation and standardisation. In Proceedings of the 8th International Conference on Health Informatics. International Conference on Health Informatics (HEALTHINF).

Several posters have also been displayed and presented as follows:

1. Zulkarnain, N. Z., Crofts, G. S., & Meziane, F. (2016, Dec). Automatic classification of abdominal ultrasound findings as normal, abnormal or inconclusive. Poster session presented at the British Medical Ultrasound Annual Scientific Meeting & Exhibition, York, UK.

2. Zulkarnain, N. Z., Crofts, G. S., & Meziane, F. (2015, June). A Proposed Model for the Standardisation of Ultrasound Report. Poster session presented at the UK Radiological Congress, ACC Liverpool, UK.

3. Zulkarnain, N. Z., Meziane, F., & Crofts, G. (2015, May). An Architecture for Ultrasound Report Generation and Standardisation. Poster session presented at

the College Of Science And Technology Deans's Annual Research Showcase, University of Salford, UK.

4. Zulkarnain, N. Z., Meziane, F., & Crofts, G. (2014, June). A Medical Ultrasound Reporting System Based On Domain Ontology. Poster session presented at the College Of Science And Technology Deans's Annual Research Showcase, University of Salford, UK.

Chapter 1

Introduction

1.1 Introduction and Motivation

Knowledge and information sharing nowadays plays an important role in research and education. The availability of online journals, articles and other knowledge sources over the World Wide Web allows people to share and receive numerous amount of information in their fields across the world. In order for everyone to understand and make full use of the information, it is important that the knowledge and information are being communicated clearly. This is even more crucial in fields where misinformation may affect the lives of a human. However, in the field of medical ultrasound, variations in reporting styles are seen as the main reason why knowledge and information are being communicated poorly. This then causes the limitation in knowledge and information sharing and at the same time affect patient diagnosis.

> "From a referring physician's point of view, a radiology department's ultimate product is the radiologist's report." - (Boland, 2007 [9])

In an ultrasound examination, medical ultrasound reports serve as the main tool for communicating results to a referring clinician. The images produced, though being generated with exceptional quality, provide relatively little value compared to the report [9] since the images do not give the whole view of the examination. Important information such as tissues characterisation alongside quantitative measurements, are features typically reported on during the scan [84] and these information are often not seen clearly in the images produced. The amount of data therefore obtained during the examination is huge and should be reported clearly. There are vast amounts of information that can be extracted in a single radiology report and even more when a huge amount of reports are analysed. This however is not an easy task because of the problem of variations in medical imaging reporting. These variations are often seen in the structure of the report, its reporting styles as well as the terminologies used.

Most reports are written in free-form format making it hard to extract important information [8, 29, 31]. As a result, these information might be overlooked making it unused and wasted. Some reports are written in a lengthy essay style where it might be hard to locate findings. Others are written in a very short and summarised form which caused some important information to be left out. These variations on the style and structure of the report may impact on the way a report is interpreted and in turn may affect the decision making process and the way a patient is managed. Whenever there are any medico-legal issues that arise, the radiology report will be the first document scrutinised [21]. Miscommunication resulting from variations in ultrasound reports could cause grave effects. This can be reflected by the number of malpractice lawsuits against radiologists where 80% of them were caused by miscommunication and misinterpretation of radiology reports [48].

Other than that, reports are also written using various terminologies that actually have the same meaning. These variations in terminologies become a paramount obstacle in the dissemination of knowledge and information retrieval because they may cause the search engine to fail in finding reports that are relevant to a specific keyword unless the search is repeated using different synonyms of the keyword. This problem is demonstrated in a study by Erinjeri et al. where they tested a tool they have developed to securely mine data from radiology reports [24]. From the testing, the tool failed to return any results for 15% of the total queries which are mostly caused by the improper search keywords. This shows that proper usage of terminologies is important in confirming that the correct result is returned when answering a query.

The problem of variations in medical imaging reports can be solved by standardising the structure of the reports and the terminologies used. Standardisation will allow for the full usage of information available in these reports and facilitate data and information retrieval as well as encourage audit and research for various purposes such as learning and teaching. The standardisation of these reports will also improve the organisation of information and promotes consistency. This enable the reports to be interpreted better and avoid miscommunication in conveying obeservations which in return increases diagnostic accuracy and reduces the risk for medico-legal problems. The immense benefits of auditing and analysing medical imaging reports, especially in ultrasound reports, have motivated us to find the best way to reduce the problem of variations by implementing standardisation in the structure of the report and the terminologies used. Report standardisation in the form of structured reporting has been widely promoted in the medical imaging field. However, there was limited use of structured reporting because of the rigidity of the approaches proposed as well as its interference with the imaging process. This has prompted us to design a standardised reporting system that adapts to the preferences of the radiologists with the intention to increase the implementation and use of structured reporting within the medical imaging field.

1.2 Research Aim and Objectives

This research aims to overcome the problem of variations in ultrasound reporting by developing a medical ultrasound reporting system that uses a domain ontology as its knowledge base to support the standardisation of terminologies being used in the report. This reporting system will take various reporting styles as an input and generates the report in a structured form using Rhetorical Structure Theory (RST) with the support of a domain ontology.

These are the objectives of this research in order to achieve the research aim:

- To determine the variations in ultrasound medical reporting styles.
- To identify the characteristics of a good quality ultrasound report.
- To develop and evaluate an abdominal ultrasound ontology using an ontology reuse methodology.
- To find out the effectiveness of using RST on medical ultrasound reports.
- To develop and evaluate the medical ultrasound reporting system that produces standardised ultrasound reports using an ontology and RST.
- To investigate the possibility of automatically classifying the findings in the ultrasound reports as normal, abnormal or inconclusive based only on the information available in the reports.

1.3 Research Contributions

This research is a multidisciplinary research project which combines the theories from the computer science field to be applied in the medical healthcare field. Thus, the contributions of this research are dedicated towards both fields. The subsections below will explain further on the research contributions in both fields.

1.3.1 Computer Science Field

In the field of computer science, this research will first contribute in proposing a novel methodology to reuse existing biomedical ontologies. Since ontology reuse is fairly new in computer science, there has been no standard way to conduct it. This research contributes by outlining an approach to reuse the required and relevant concepts of large existing biomedical ontologies for a specific application with the help of various off the shelf supporting tools. The approach presented in this research is superior to the existing approaches in terms of concept matching and domain coverage because it selects and merges best matches from more than one existing biomedical ontology. The usage of this approach in developing a small domain specific ontology for the use of a specific application also reduces the development time and redundancy as it avoids the unnecessary process of "reinventing the wheel".

Another contribution of this research in the computer science field, which is also considered as the main contribution of this research is the implementation of RST in analysing medical ultrasound reports. As far as we are concerned, the implementation and usage of RST in analysing ultrasound reports has never been attempted before. This research contributes by defining a set of rules that is able to segment paragraphs into text spans and recognise the rhetorical relations that exist between them. The definition of the rules for each relation was inspired by the classic RST relations defined by Mann and Thompson [52]. In classic RST, segmentation and identification of rhetorical relation were performed using discourse markers and part-of-speech tags (POS tags). However, in our approach, the combination of discourse markers together with classes from an ontology in the abdominal ultrasound domain were used to segment the paragraphs and sentences in an ultrasound report into elementary discourse units (EDUs) or text spans. This approach allows for the sentences in a free form ultrasound report, which are usually not grammatically complete, to be segmented reconstructed and displayed in a structured form.

1.3.2 Medical Healthcare Field

As for the medical healthcare field, this research will contribute by proposing a solution to the problem of various report writing styles by developing a novel medical ultrasound reporting system that will provide flexibility for radiologists and sonographers in writing reports. This reporting system will allow for standardisation in report format as well as terminologies used despite having reports with various styles and heterogeneous terms as its input. Even though there have been several efforts in implementing structured reporting in standardising ultrasound reports, the overly constrained systems have deterred the radiologists from using them. The novelty of this system is that it does not restrict radiologists from writing their reports in any style and structure they prefer yet still having a structured version as the final output. This will hopefully encourage further usage of standardised and structured reporting among radiologists and in return will result in a better interpretation of the report and allows for better decision making and patient management. The ability of the system to transform free-form reports to structured form allows for research and studies to be conducted not only on the current reports but also on the old reports that were written in free-form format. This brings a lot of benefit in acquiring data that dates way back and allows for the discovery of new knowledge which currently may be restricted because of the format of the old reports.

1.4 The Research Approach

The main purpose of this research is to develop a medical ultrasound reporting system that will be able to solve the problem of variations in ultrasound reporting. Accordingly, the research has been divided into five phases encompassing the data collection and analysis process up until the development of the medical ultrasound reporting system. Figure 1.1 summarises the research approach.

In **the first phase** of this research, the aim is to further understand the problems faced by radiologists and referring clinicians in writing and interpreting various ultrasound reports. In doing that, a sample of 100 anonymised clinical abdominal ultrasound reports were reviewed. These reports were obtained through the Directorate of Radiology of the University of Salford. The purpose of the review was to undertake an audit style study of the reports in order to understand the variations in writing them. The format, content, terminologies used and reporting style of each report were among the factors being compared in the review. A comparison was performed against a set of

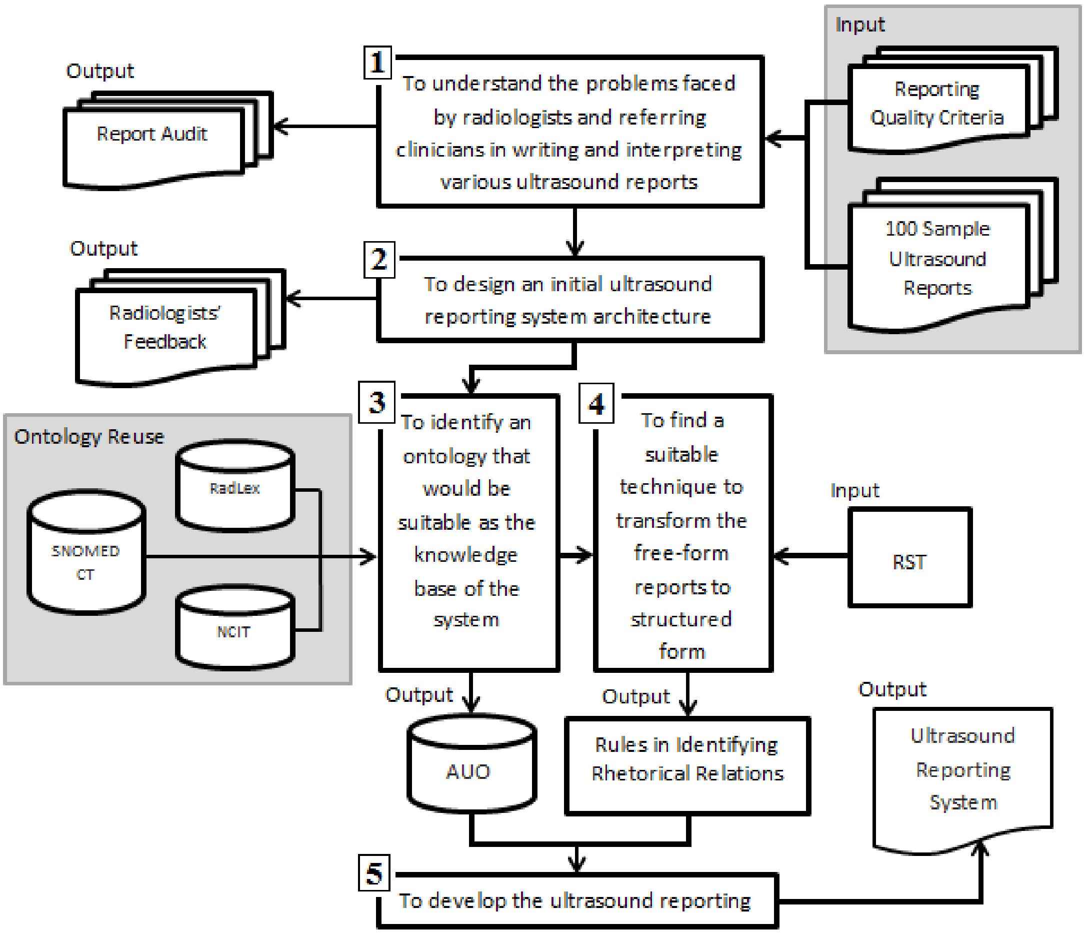


Figure 1.1: The five phases in the research approach

quality criteria acquired from the literature and each report was categorised according to its quality. From the audit, it was found that the sample ultrasound reports were of poor quality where most of them have grammatical errors, missing punctuations and obvious spelling mistakes. The reports were also written mostly in free-form format and some were written using terminologies that are vague. Therefore, the audit style study has proven that there is a need to improve the quality of ultrasound reports.

This has then led to **the second phase** of the research which is to design an initial medical ultrasound reporting system architecture that would address all the problems found during the analysis. The architecture took into account not only the problem of variations but also the preferences of the radiologists in writing the reports. Based on the system architecture, the user interface of the system was also designed with the focus being directed to the structured reporting page which is the essence of the system. The interface design of the structured reporting page takes into consideration the quality criteria that have been gathered from the literature review in the first phase of this research. Both the system architecture and the interface design of the structured reporting page were presented at the 2015 Health Informatics Conference as well as the 2015 UK Radiological Congress where initial feedback was obtained. A survey was also conducted (see Appendix B) during the 2015 UK Radiological Congress and the results are discussed in Section 3.4 of this thesis. The feedback from both conferences showed that the problems identified in the research are indeed real and validated the purpose of the research. Thus, the next step would be to find a suitable method to properly develop the system.

Comprehending the need to have a knowledge base to standardise the terminologies in the system, the usage of an ontology has been considered because of its ability to provide relations between classes and to include various annotations on each classes. The vast availability of ontologies in the biomedical domain was also seen as an advantage. Therefore, **the third phase** of the research was to identify an ontology that would be suitable for use as a knowledge base for the system. Several existing biomedical ontologies that were deemed as suitable for the ultrasound reporting system have been identified and analysed. From the analysis, it was found that these existing biomedical ontologies were too big to be used in a small application. Developing a new domain ontology specific to the needs of a system on the other hand will consume a lot of time and cause redundancy. Hence, there is a need for ontology reuse. Since ontology reuse has no standard methodology, a new methodology consisting of four steps was developed based on the regular ontology reuse practices observed in the literature review. The development of the ontology was executed in two phases where the first phase uses the terminologies extracted from 49 sample reports while the second phase uses the terminologies extracted from all 100 sample reports including the 49 sample reports. The purpose of this was to investigate whether the methodology allows for easy incorporation of new classes and information in the ontology based on new requirements from additional ultrasound reports. It is important that the methodology allows the ontology to be adjustable as more sample reports are acquired. In both phases, terminologies from the reports were extracted and used to generate a list of ontology recommendation based on scores provided by BioPortal. Ontology mapping was then initiated and a complete abdominal ultrasound ontology was developed and evaluated by domain experts.

With the knowledge base resolved, **the fourth phase** of the research was to find a suitable technique to transform the free-form ultrasound reports into the standard structured form. This has led to the choosing of RST as the suitable candidate as it promotes coherency in a text and is able to recognise rhetorical relations between text spans. It is important that the relations between text spans are recognised so that when a report is being transformed, no information will be lost. The traditional technique of segmenting texts and recognising relations in RST was by using discourse markers and POS tags. However, this technique does not seem to be effective in the 100 sample ultrasound reports that have been collected mainly because of the way the reports were written which is not grammatically complete. Therefore, a new way of implementing RST must be realised. This was achieved by analysing 60 out of the 100 sample reports to recognise their patterns and how the abdominal ultrasound ontology can be applied to recognise rhetorical relations between the text spans in the reports instead of using POS tags. A set of new rules was defined to identify rhetorical relations from the ontology annotated text spans. Evaluation was performed using the remaining set of the sample reports (40 reports) where the selected manual parsing, completed by three natural language processing experts, was compared to the automatic parsing performed by the system. The evaluation has resulted in a high accuracy rate for both the segmentation and the rhetorical relation identification process which indicates that the implementation of RST with the combination of an ontology is possible in parsing medical data especially in ultrasound reports.

Now that the knowledge base and the transformation system to generate structured reports have been developed, **the fifth and final phase** is to implement both of them in the medical ultrasound reporting system. First, an ontology in the abdominal ultrasound domain developed in the third phase was adopted to standardise the terminologies when processing the free-form reports. Then, the rhetorical relations defined in the previous phase will need to be reviewed together with the sample ultrasound reports to identify how it can be implemented in transforming the free-form reports to the structured form. Once this has been achieved, RST together with the ontology will be used to extract rhetorical relations between text spans in the report so that ultrasound findings can be presented in an itemised and structured format. Upon the completion of the medical ultrasound reporting system, it was evaluated using the manually standardised 100 sample ultrasound reports as a comparison to ensure that the reports were accurately transformed and no information was lost. A group of radiologists was also selected to verify that the information transfer from free-form to structured form was correct and complete. They have also given their opinion on the quality of the sample reports acquired and the medical ultrasound reporting system as a whole.

1.5 Research Scope and Limitations

The research will specifically focus on the abdominal ultrasound reports where the domain ontology will mainly consists of classes describing abdominal ultrasound scanning which includes both the anatomy and pathology of the abdominal area. It will also include technical terminologies that will assist in the reporting of the abdominal ultrasound scanning such as the word ultrasound itself as well as unit of measurements for instance meters and millilitres. Even though the research proposed a system architecture for a medical ultrasound reporting system, not all components in the architecture will be actualised as the focal point of this thesis will be the components that will allow for the transformation of free-form reports to structured form.

There have been some limitations encountered throughout the execution of the research as listed below:

1. There were difficulties in gathering the sample ultrasound reports from the radiology department which resulted in having only 100 sample reports. This limits the availability to perform a wider analysis of the variations in the reports as well as in conducting data training and testing. The limitation in the number of reports has also restricted the possibility of finding more patterns of rhetorical relations in the report which will enhance the rhetorical relation rules and in return increase the accuracy of the RST parser.

2. It was not possible for all components of the system architecture to be developed. The focus was given only to the main components of the architecture such as the free-form report page, the structured report page, the structured report generator and the knowledge base.

3. The number of reports acquired have also resulted in the implementation of RST using a rule-based approach only. It was not possible to apply machine learning with the time given. Currently there is no annotated corpus of radiology report and developing one was out of the scope of this research. However, this is one of the improvements that we look forward to in the near future.

4. A full implementation of the system for the usage of radiologists in their daily examination was also not possible in the given time. Therefore, evaluation can only be performed by comparing the manually standardised reports with the automatically standardised reports and by gathering feedback from radiologists.

1.6 Thesis Overview

This thesis has been structured into seven chapters and each one describes the major components of the research. The following is an overview of each of these chapters:

Chapter 1: Introduction. The first chapter introduces the research problem and the motivation which has initiated this research. It also presents the research aim and objectives as well as the contributions of this research in both the computer science and medical healthcare fields. It also describes the different phases underwent to achieve the aim of the research. Lastly, it details the scope and limitations of the research.

Chapter 2: Background and Literature Review. This chapter presents the background of the research and the literature review conducted on various topics pertaining to the research such as ultrasound medical reporting, structured reporting, ontology reuse and RST. It also presents related works on structured reporting that have been developed before and how this research is different or somewhat better than previous works. It also discusses previous works in ontology reuse and RST and how this research is unique compared to the others.

Chapter 3: The Design of a New Ultrsound Reporting System. Chapter 3 demonstrates the overview of the design of the ultrasound reporting system architecture and its components. It explains the operations of the system and the functions of each component in the system. This chapter also presents the result of the interface preferences survey undertaken during the 2015 UK Radiological Congress (UKRC). It then discusses several issues brought up during the survey and how changes have been made to reflect these issues.

Chapter 4: The Knowledge Base for the Standardised Reporting System. The fourth chapter of this thesis focuses on the development of the Abdominal Ultrasound Ontology (AUO) which serves as the knowledge base of the ultrasound reporting system. It first reviews existing biomedical ontologies and the suitability of using these ontologies as a whole in the ultrasound reporting system. It then presents the ontology reuse methodology and explains how ontology reuse can be carried out beginning from the term extraction process up until the ontology evaluation process. It also discusses the result of using the methodology in developing AUO including its statistics such as the number of classes it has and the coverage it achieved compared to existing ontologies.

Chapter 5: Structured Report Generation. Chapter 5 of this thesis discusses the RST which was used as the mechanism for the generation of the structured report. It first introduces the classic RST and the types of rhetorical relations it recognised. This chapter then introduces seven rhetorical relations together with a set of defined rules for each relation which was relevant to this research. It then demonstrates the approach in implementing these relations together with AUO in parsing medical ultrasound reports. Next, it discusses the result of implementing RST in segmenting and recognising rhetorical relations in these reports. Finally, it attempts to compare the work with an existing work and discusses arising issues and questions regarding the comparison.

Chapter 6: The Development of the Standardised Reporting System. This chapter presents the development of the whole medical ultrasound reporting system with focus being given to the structured report page. It explains how AUO and RST were implemented in the system to transform free-form reports to structured form. Feedback from the ultrasound specialists on the implementation was also recorded. This chapter also discusses the idea of automatically classifying findings as “normal”, “abnormal” and “inconclusive”. However, the evaluation of this part of the research by the specialists lead us to abandon it as it was not possible to validate it with the amount of information available in the reports only. This chapter presents the opinion of the ultrasound specialists regarding this matter.

Chapter 7: Conclusion and Future Directions. The final chapter concludes this thesis by restating the important findings and matters discussed in the previous chapters. It also revisits the research objectives and evaluates whether each objective has been met. Finally, it gives suggestions on future directions of the research so that the current work could be improved and enhanced.

Chapter 2

Background and Literature Review

2.1 Introduction

Ultrasound is a common medical diagnostic imaging technique which has been used for many years in clinical diagnosis as it generally gives most of the information in diagnosing a patient's problem [14]. It passes a pulse or continuous high frequency beam to the body and utilises echoes reflected from the biological tissues [57]. Ultrasound is often the first-line imaging modality used to evaluate the affected abdominal organs such as the kidneys, liver and abdominal aorta [78] in order to diagnose a variety of conditions such as abdominal aortic aneurysms, gallstones and acute cholecystitis. Hangiandreou [34] has stated three main reasons why ultrasound has been one of the most widely used medical diagnostic imaging techniques in clinical diagnosis:

1. Ultrasound images are acquired in real time allowing for minimally invasive procedures such as needle biopsies to be performed.

2. The ultrasound equipment is inexpensive and portable.

3. Ultrasound possesses no risk to patient. Though sometimes it may feel uncomfortable, ultrasound should not cause any pain to the patient.

To appreciate the importance of intelligible reports in ultrasound, we first need to understand the ultrasound examination process. There are at least two personnel involved in an ultrasound examination. One is the referring clinician and the other is the radiologist or sonographer as shown in Figure 2.1. First, a clinician will refer his patient for an ultrasound examination if he deems necessary. A radiologist or sonographer will then perform the ultrasound examination and make notes of the important

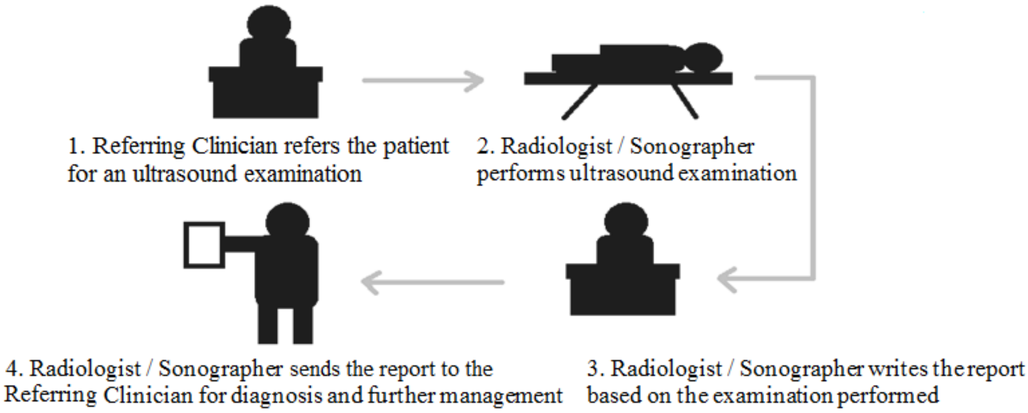


Figure 2.1: Process involved in an ultrasound examination

observations and findings to be included in the report. Once the examination is completed, the report should be written as soon as possible. The final report will then be sent to the referring clinician for diagnosis and the follow up of the patient. Here it can be understood that an ultrasound report does not only represent the observations and findings during the examination but it also serves as a communication tool between the two doctors.

The quality of an ultrasound examination is highly correlated to the skill and thoroughness of the radiologist or sonographer [14]. Since the images generated from an ultrasound examination do not give the whole view of the examination, the report produced by a radiologist can be considered as a vital part in diagnosing a patient because referring clinicians rely heavily on it [22]. Furthermore, ultrasound images are in real time which makes it impossible to produce images for the whole examination. Even though the quality of the ultrasound examination is prominent, if the report is poorly written, observations and findings during the examination might not be communicated properly. This will then affect proper diagnosis of the examination and could result in the skill and thoroughness of the radiologist or sonographer to be worthless. Therefore, it is important that the quality of the report to be at par with the quality of the ultrasound examination. The following sections will examine the reporting component of the ultrasound examination in more detail.

2.2 Ultrasound Medical Reporting

Reports generated by radiologists using real time images during an ultrasound examination have more value compared to the images produced, though being generated with exceptional quality [9]. The nature of the ultrasound examination process has made the ultrasound report an even more important document as it is used as the main tool for communicating the result of an examination between the sonographer and the referring clinician. Compared to other forms of radiology diagnostic tools, this role is much more significant in ultrasound. This is because the images produced during the ultrasound examination are only representative of the examination and are appropriately selected to assist subsequent discussion [3] and not the overview of the whole examination. This is different in other radiology diagnostic tools where images produced are the still images seen during the examination. It is now apparent that ultrasound reports are an essential element of an ultrasound examination. Therefore, to improve the quality of the reports, we first need to be familiarised with the different reporting styles which will be discussed next.

2.2.1 The Different Reporting Styles

There are various styles and formats that have been adopted in writing an ultrasound report. These styles and formats vary from hospital to hospital and sometimes even by different practitioners within the same hospital despite measures have been taken by the hospitals to standardise their reports. Generally, reports can be written either manually or using an editor within a computer system and in either free-form or structured form.

Traditional reports are written in free-form [65, 83, 77] whereby findings are described in an essay like document which makes it hard to locate specific findings without reading the whole report. Free-form reports are usually very detailed to ensure maximum information is transferred between the sonographer and the referring clinician. This information is important in assisting the referring clinicians in making a diagnosis on a patient as well as in discussing abnormal cases during meetings with other referring clinicians. However, Plumb, Grieve and Khan [69] warned that unnecessary information can cause distraction from the main message of the report and its key findings to the clinicians.

Ultrasound reports can also be written in a structured form whereby contents are grouped under specific headings. Findings are also itemised and written in a short and concise format so that it is easier to be read. This allows for not only a maximum information to be communicated to the referring clinicians [69] but also for specific information to be easily pinpointed. Cramer et al. have classified structured form reports into three tiers [18] as illustrated in Figure 2.2. The first tier of structured form reports is the most common one where contents are grouped under specific headings. The second tier is where contents are not just grouped under the same headings but are also itemised. The third tier of structured form reports is when the language of the report is also standardised using lists, checkboxes and buttons. The third tier is the most rigid and often hard to be accepted by radiologists.

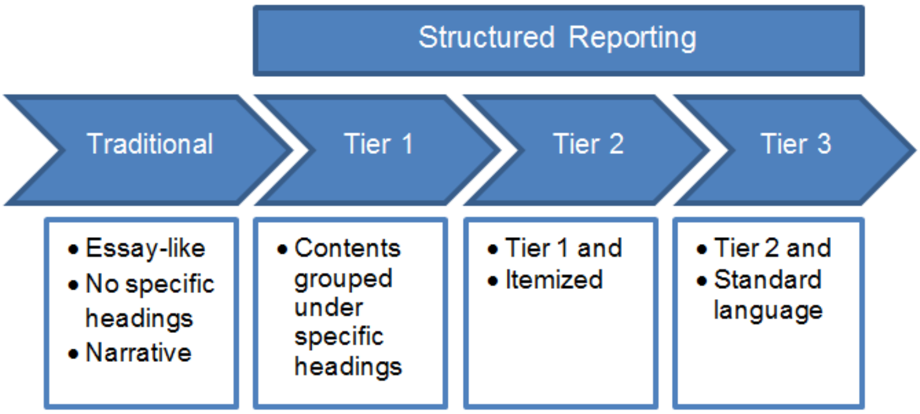


Figure 2.2: The three tiers of structured reporting as defined by Cramer et al. [18]

It is often believed that the structured form allows reports to be read in a shorter time which makes diagnosis better and faster compared to free-form reports. In order to verify this, Sistrom has constructed a study [82] using 12 cases consisting of both free-form and structured form reports where 16 senior medical students were asked to read the cases and answer 10 questions about them. It is hypothesised that less time is required to read and answer questions about the content of a structured report. Interestingly, the result showed no significant difference between both formats in terms of time, score and efficiency. However, almost all senior medical students in the study preferred the structured form compared to the free-form [82]. This is because of the benefits that it brings which among others include the improvement of communication between radiologists and referring clinicians as well as their satisfaction [47, 70]. The benefits and challenges of the structured form will be discussed further in Section 2.2.2 and 2.2.3.

Though different in format, both styles of ultrasound reports should contain key components which are the title of the examination, history, technique, comparison, findings and conclusion as mentioned by Wallis and McCoubrie [91]. The review of the 100 sample ultrasound reports that were acquired showed that all these reports have most of these key components, especially the title of the examination, findings and conclusion. The contents of all these reports were also grouped under different headings but most are still written in free-form. Therefore, in accordance with the three tiers of structured form reports previously defined by Cramer et al., it is safe to say that the 100 sample ultrasound reports that were gathered fall under the first tier of structured reporting.

Cramer et al. [18] have also mentioned that there are a lot of interests lately to standardise ultrasound reports by using structured reporting instead of the standard free-form. Considering that the reports we collected are already the first tier of structured reporting, it supports the claim that ultrasound reporting is moving towards a more standardised reporting. However, we believe that achieving the first tier of structured reporting is not enough to eradicate the problem of variations in ultrasound report. This problem has a higher chance to be eliminated if reports are written in accordance with the third tier of structured reporting. The problem now is, the third tier of structured reporting is the most rigid and most unlikely to be adopted by the radiologists. For that reason, to ensure that the reporting system is accepted by the radiologists, it is crucial to understand the rationale behind the radiologists' refusal in using the system.

2.2.2 The Many Benefits of Structured Reporting

The summary of the 2007 American College of Radiology Intersociety Conference as reported by Dunnick and Langlotz [21] stated that the majority of the participants recommended that the solution of the problem of variations in radiology reporting resides in structured reporting. This is especially when it also includes the support of lexicons or controlled vocabulary such as the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) and the Radiology Lexicon (RadLex) in order to standardise the terminologies used [45, 77]. In recent years, a lot of studies have been conducted to verify whether structured reporting brings positive changes to radiology reporting. Even though some studies do not show that structured reporting is better than free-form [36, 82], both radiologists and non-radiologists have shown their preference in using structured reporting [25, 60, 69, 77, 82]. Before we discuss further on the challenges of structured reporting which may contribute to the refusal of the radiologists in adopting the system, let us first explore its many benefits which have attracted the radiologists' preference towards it.

Production of a Standardised Report

The first advantage of structured reporting which we are most interested in is the production of a standardised radiology report. The nature of structured reporting imposes reports to be presented in a standard and structured way [76]. The strict format of structured reporting prompted for a more uniform report between radiologists [77] where it requires the radiologists to follow certain guidelines for each and every report produced. Ellis and Srigley [23] have included seven key parameters for quality reporting:

1. Timeliness
2. Accuracy
3. Completeness
4. Conformance with current agreed standards
5. Consistency
6. Clarity in communication

The adoption of structured reporting ensures at least two of these key parameters are met which are completeness and consistency [18, 60, 89]. An audit undertaken by Naik, Hanbidge and Wilson [60] on free-form reports confirmed that there are a lot of inconsistencies in the reports and quite a number of them were incomplete. Structured reporting provides separate fields for different elements in a report such as the conclusion, area examined and its observation. These fields help in reminding the radiologists on what needs to be included in a radiology report hence ensuring its completeness.

Consistent Terminology Usage

One important benefit of structured reporting is that it reduces variations by allowing consistent terminology [77] usage throughout the report. This upholds one of the key parameters for a quality report which is the conformance with current agreed standards. This is not only with regards to the structure and content of the report but also in the terminologies used. Structured reporting with the implementation of lexicons and controlled vocabulary improves the quality of the report [36] by removing ambiguity.

For example, a report that contains the word “calculi” should be associated with the mass found in the body and not the calculi used in reasoning. It should also be grouped together with reports that contained the word “stone” since it is the synonym of "calculi". Lexicons such as the National Cancer Institute Thesaurus (NCIT) have a preferred label annotation where it suggests which word is best used in describing a certain finding. Such confusion is reduced when standard lexicons similar to this are implemented because preferred terminologies will be used throughout all the reports. This at the same time promotes clarity in communication [30] between radiologists.

Improved Accuracy of the Reports

Ellis and Srigley [23] stressed that accuracy is the most important key parameter in determining the quality of a report. Therefore, another important benefit of structured reporting is that it has been proven to improve the accuracy of reports [13]. This is possible because of the usage of standardised terms [25, 89] resulting in a clear and accurate report which then leads to better diagnosis.

Structured reporting also allows for reports to be populated with measurements and key images corresponding to the examination [84]. The unit of measurements and qualifiers used are often standardised which contributes to the accuracy of the reports. As a result, it is easier for an estimation to be calculated and aids in decision making.

Furthermore, structured reporting gives the ability to easily asses the quality of individual elements [77] in the reports as it is written in an itemised format. Careful review of the diagnostic accuracy of the reports is also possible [41] by monitoring the diagnostic patterns in the reports. This will then indirectly improve the accuracy of the radiology reports because errors and flaws in the reports are recognised and corrected.

Increased Accessibility of Data for Research

Structured reporting also gives benefit by increasing the accessibility of data for research. Structured reporting nowadays are electronically archived which allows for the reports to be accessed easily and in short amount of time. Data in these reports are no longer only a historical review of a patient or a disease but could also be analysed in real time [23]. Compared to free-form text, data from structured reporting can be easily queried and populated because of the consistent terminology used in the reports. For example, a radiologist simply needs to query a keyword relevant to the case he is handling in order to find recent examples of similar cases. This speeds up the diagnosis and makes it easier for studies about the case to be conducted.

Report analysis for the purpose of research and education will also be more efficient [18] as standard terminologies often include other similar terminologies that have the same semantics in the form of metadata [25]. As a consequence, a query on a keyword will not only return reports containing the keyword but also other reports containing terminologies that have the same meaning as the keyword. Lexicons can be in the format of tabular, database and ontology [45]. In this research, we are more interested with the use of an ontology because of its capabilities in annotating knowledge details such as definitions and synonyms and in executing reasoning.

Higher Reproducibility of the Reports

The standardisation that structured reporting brings results in higher reproducibility of the reports [25]. Important information such as the details of the patients and radiologists can be electronically transferred from the database. This makes the production of radiology reports much faster and reduces the turnaround time, provided that the structured report is not too rigid.

The electronic nature of structured reporting also contributes to timeliness which is one of the key parameters for quality reporting. This is because every report produced will have several time stamps such as the time the report was written, the time of the examination and the time the radiologist signs off the report.

The usage of structured reporting also promotes the adherence to guidelines [18] which promotes good practices in reporting, allowing for reports to be produced in a timely manner. Structured reporting also allows for reports to be organised better and makes it easier for reports to be located and queried [89] for future data references and report reproductions.

Structured reporting allows for reports to comply with the seven key parameters of quality reporting as stated by Ellis and Srigley [23]. Five benefits of structured reporting that have been discussed are the production of a standardised report, consistent terminology usage, improved accuracy of the reports, increased accessibility of data and higher reproducibility of the reports. These benefits have motivated several researchers to design computerised tools that could take the most advantage of structured reporting. However, not all of these efforts succeeded and the reasons behind this will be discussed next.

2.2.3 The Challenges of Implementing Structured Reporting

Realising the effectiveness of conveying information in a structured form, several researchers have developed models and proposals to apply structured reporting as a method to create radiology and ultrasound reports [5, 43] as early as the 90's. Most of these early structured reporting systems were constructed using checkboxes and radio buttons to complete a report. One example of a structured reporting application that has been developed and used before is UltraSTAR (Ultrasound STructured Attribute Reporting) which is an application that allows sonographers to construct reports interactively by using checkboxes and radio buttons in a graphical user interface to select concepts from a hierarchical standardised vocabulary [5]. The concepts selected from the checkboxes and radio buttons were used to answer a set of predefined questions in order to construct the content of the report. Figure 2.3 illustrates an example of the interface of an early structured reporting application.

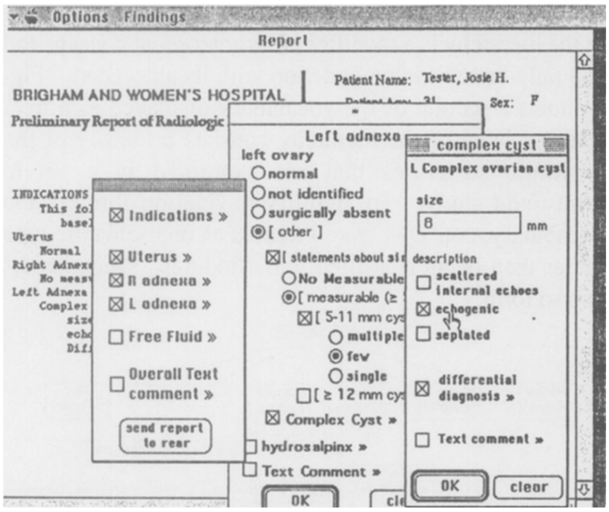


Figure 2.3: User interface of early structured reporting [5]

This method of designing a structured reporting application has formed the basis of most reporting applications. A more advanced implementation of structured reporting includes a predictive data entry feature [43] where it provides suggestions on concepts based on several letters that have been typed by the radiologist. Whilst this type of systems have been used in some Radiology Departments [4], there are several challenges which have limited their wider implementation. These challenges can be grouped into three which are the rigidity of the reporting applications, the technological constraints and the radiologists' resistance to change.

One of the biggest downfalls of structured reporting is the rigidity of the reporting applications developed [25, 70]. These applications often limit the inclusion of additional information [36, 46] that could be important for the case and would not allow the form to be submitted with certain fields being left blank. It is important that reporting applications allow for these options as sometimes reports may contain more information than the fields provided and in some instances, some information cannot even be obtained. This caused the radiologists to find that the implementation of structured reporting was time consuming [36, 46] as a lot of time was spent figuring out how to include all the necessary information in the structured form.

The effect caused by the rigidity of reporting applications was visible in the study conducted by Johnson et al. [36] where there was a huge decrease in completeness and slight decrease of accuracy in the reports written by its intervention group using standard reporting compared to its control group that used free-form report. As a result of this limitation, structured reporting was seen as better for normal findings compared to abnormal and complex cases [89, 92]. This contravenes with the benefit of using structured reporting which is to promote completeness and accuracy thus making the effort of implementing structured reporting ineffective.

Another challenge in expanding the implementation of structured reporting is the technological constraints faced by the radiologists. This is hugely contributed by the design of the reporting applications that requires a lot of mouse clicking, keyboard command and complicated hierarchical menu [33] which often interferes with the examination process. Other than that, Reiner and Siegel [71] stated that concerns over computer literacy among radiologists and the steep learning curve are among the reasons for the minimal adoption of structured reporting. These technological constraints are however not too worrying as technology improves over time. This will enable the reporting applications to be developed in a way which no longer frustrates the radiologists. Computer illiteracy and steep learning curve can also be resolved by providing trainings and consultations to the radiologists.

The major challenge however is the radiologists' resistance to change. A recent study undertaken by Tran, Wadhwa and Mann [89] demostrated evidence towards this concern. A departmental wide implementation of structured reporting has been conducted in their centre's ultrasound department where the implementation was executed in three phases. The first phase was the template development phase where reporting templates from the Radiological Society of North America (RSNA) and the American College of Radiology (ACR) was modified according to their needs. The second phase was the implementation of structured reporting where it was executed for seven months. Finally, the evaluation phase was conducted during the final month of the implementation where audit was completed and feedback was acquired.

During the implementation, the study saw a very small number of consistent usage of structured reporting. Some radiologists used structured reporting but not consistently while most did not even use structured reporting in constructing their reports. The study also saw higher usage of structured reporting between trainees as compared to staff radiologists with feedbacks expressing that it was difficult for them to change and adopt the new reporting style. This signals resistance towards change from the staff radiologists. The shift to structured reporting is indeed a proactive effort especially when the staff radiologists were so accustomed to the free-form format [71]. We view the radiologists' resistance to change as the biggest challenge in ensuring broader implementation of structured reporting as it is impossible to force people to accept something unfamiliar to them.

2.2.4 Improving the Implementation of Structured Reporting

The two challenges which are the rigidity of reporting applications and the radiologists' resistance to change need to be addressed so that the implementation of structured reporting can be improved. In order to reduce rigidity in structured reporting, a reporting application should include free-form text fields to enable radiologists to explain complex findings and include any related reflections [13, 25]. This is important because different cases would have different needs in reporting the findings. Thus, a one size fits all template would not be sufficient.

It is also important for a reporting application to allow for certain information to be left out but with some remarks explaining the reason [66], because not all information can always be obtained due to limitations often encountered during the scan process. Structured reporting would be better accepted by radiologists if it is more flexible whereby they would be able to choose what to include and what to leave out in the report depending on the case that is being reported. This is because if the reporting system is too rigid, the radiologist might get frustrated and ends up abandoning the system.

Even though radiologists resist the transition to structured reporting, they were never against it [89]. Therefore, in order for structured reporting to be accepted widely, it is important that it includes features that would appeal to them. Consequently, this research endeavours to provide a structured reporting system that would allow radiologists to take advantage of the benefits of structured reporting without needing to change the way they usually write their reports.

2.3 Ontology and Ontology Reuse

Standardisation of terminologies in a structured reporting system can be achieved using an ontology. Gruber [32] defined an ontology as "an explicit specification of a conceptualisation with the purpose of sharing knowledge and enable interoperation between applications based on shared conceptualisation". An ontology enables the representation of knowledge and allows it to be shared across domains. Noy and McGuinness [62] have specified five purposes of an ontology:

1. To share common understanding of the structure of information among people or software agents.
2. To enable the reuse of domain knowledge.
3. To make domain assumptions explicit.
4. To separate domain knowledge from the operational knowledge.
5. To analyse domain knowledge.

In this research, we are mostly interested in the first two purposes of an ontology which are to share a common understanding of the structure of information among people or software agents and to enable the reuse of domain knowledge. There are huge amounts of underlying information in medical reports and articles that can be beneficial in research and education. However, heterogeneous terms pointing to the same concept have deterred the maximum usage of this information. For example, a search for articles related to "ultrasound" will return results only with the keyword "ultrasound" while optimally it should also return results of other terms referring to ultrasound such as "ultrasonography". Terms that have various different meanings also cause unnecessary result to be displayed when being queried. Therefore, ontology has been favored for the purpose of sharing a common understanding of these terminologies so that they can be standardised and knowledge can be disseminated.

Smith [85] has defined ontology as a "dictionary of terms formulated in a canonical syntax and with commonly accepted definitions designed to yield a lexical or taxonomical framework for knowledge representation which can be shared by different information-systems communities". Ontology defines a common vocabulary of machine-interpretable definitions of basic concepts in the domain and the relations between them [62]. It is known that in medical and healthcare discipline, there are a huge number of terminologies being used. In addition to that, several terminologies are being used to describe the same concept.

For example, "carcinoma" and "cancer", "calculi" and "stone", as well as "neoplasm" and "tumour" are both different words that have the same meaning. This makes it hard to gather information in the form of reports or articles relating to a terminology without including all the terms that refer to it. This burdens the researchers as they will need to do multiple queries using different keywords in order to find articles relating to one concept. Some important reports or articles related to the concept might also be left out since computer might fail to return any result because of improper usage of keywords. The existence of ontologies makes it possible to annotate terms in websites, journals and articles published on the Internet to allow computers to query and extract relevant information pertaining to a concept even though it uses different terms.

Realising the needs for ontologies in the medical domain, several ontologies have been published to include various topics related to the medical domain such as anatomy, pathology as well as pharmacology. The usage of ontologies enables us to reuse the domain knowledge for the specific needs of our medical ultrasound reporting system instead of developing a new ontology from scratch. Reuse of these ontologies allows for common knowledge to be shared in the medical field which contributes to a standardised usage of terminologies. Examples of existing medical ontologies that have the potential to be reused will be described further in Subsection 2.3.2.

2.3.1 Ontologies in Medical Applications

Ontologies serve several purposes in the medical field. One of the main purposes of ontologies in the radiology domain is to annotate images and reports. Radiology departments produce thousands of images and reports concerning examinations performed on patients. The annotation of these images and reports makes it easier for automatic information searching and extraction allowing them to play a major role in integration with teaching and research.

RadiO, a prototype application by Marwede, Fielding and Kahn [55] is one example of annotating reports using ontologies. In this application, an ontology of imaging findings and their interpretation was used as a knowledge base in annotating features of anatomies found in images. Another example of a medical application using ontologies is the Interdisciplinary Prostate Ontology Project (IPOP) [68], which uses ontologies from OBO Foundry such as Phenotype Quality Ontology (PATO), Basic Formal Ontology (BFO) and the Foundational Model of Anatomy (FMA) to annotate clinical reports about prostate cancer.

Ontology also serves a purpose in report generation such as the one in the Medical Imaging and Advanced Knowledge Technologies (MIAKT) project [12]. In this project, reports were generated automatically from knowledge encoded in the domain ontology using Natural Language Generation Techniques (NLG). Semantic data such as patients' information and diagnosis were encoded in an ontology of the breast cancer domain. The role of NLG is to turn these data to textual description in order to generate complete reports. These reports were however in a free-form structure which contradicts with the aim of this research which is to develop an ultrasound reporting system that produces structured reports.

2.3.2 Existing Biomedical Ontologies

There are many existing biomedical ontologies available in various ontology libraries. In this section, we describe four examples of the most common and popular ontologies in the biomedical community which are the Foundational Model of Anatomy (FMA), the Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT), the National Cancer Institute Thesaurus (NCIT) and the Radiology Lexicon (RadLex). These four ontologies are in either the W3C Web Ontology Language (OWL) or the Open Biomedical Ontologies (OBO) format and can be downloaded for use depending on their license.

Overall, all four ontologies cover the domain that is of interest in this research namely the abdominal ultrasound domain. However, these four ontologies are very large in size and general in domain. Thus, it would be inefficient to use these ontologies as a whole in one specific application system as it will slow down the process and take a lot of storage space.

Foundational Model of Anatomy (FMA)

Foundational Model of Anatomy (FMA)[1] is one of the largest ontologies related to the medical domain developed and maintained by the Structural Informatics Group at the

[1] http://si.washington.edu/projects/fma

University of Washington. It documents the concepts and relationships concerning the structural organisation of the human body from its macrocellular to microscopic levels [73]. FMA is intended as a reusable and generalisable resource of deep anatomical knowledge. However, because of its size, it will not be practical to adopt the whole FMA into an application as it will take up space and time. Thus, a section of the FMA will need to be extracted to tailor to the needs of the application. FMA is open source and can be viewed online using ontology library such as BioPortal[2] .

SNOMED CT

Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT)[3] covers more domains than FMA where it also includes clinical findings, chemical substance scales and other miscellaneous health information [6]. It is owned and maintained by The International Health Terminology Standards Development Organisation (IHTSDO) where they ensure the quality of the ontology and clinically validates it. SNOMED CT is semantically rich and is widely used by more than 50 countries. It is also multilingual as it is available in languages other than English such as Spanish, Dutch and Swedish. Compared to FMA, SNOMED CT requires a licence to be used and it charges users a fee depending on the country of use.

National Cancer Institute Thesaurus (NCIT)

National Cancer Institute Thesaurus (NCIT)[4] is a biomedical reference terminology that covers both the clinical and biomedical domain especially carcinoma [44]. It is maintained by a team of multidisciplinary editors in the National Centre Institute's Centre for Bioinformatics. It is a large ontology which contains more than 100,000 terminologies with definitions, synonyms and other information. NCIT is updated monthly where it receives ad hoc criticism from users and reacts according to it. It is available in the OWL ontology format in BioPortal.

[2] http://bioportal.bioontology.org
[3] http://www.ihtsdo.org/snomed-ct
[4] http://ncit.nci.nih.gov/ncitbrowser/

Radiology Lexicon (RadLex)

Radiology Lexicon (RadLex)[5] is a smaller ontology compared to the other three ontologies recently presented, where it focuses on the radiology domain. It incorporates many complex radiology related domains from basic science to imaging technology [74]. Compared to FMA, SNOMED CT and NCIT, RadLex is more suited to this research as it focuses on radiology which is the parent domain of abdominal ultrasound. However, the ontology is still quite large for the usage of this research and the coverage is still not wide enough as some general concepts were still missing.

2.3.3 Ontology Reuse

General and established domains such as medicine, law and business usually have existing ontologies that cover the general concepts in the domain. These ontologies, however, are often too large to be manipulated or processed in a specific application. Furthermore, the generalisation of these ontologies requires substantial amounts of alteration before they can be applied in an application [10]. Thus, a domain specific ontology is needed to solve this problem. Building a new domain specific ontology from scratch would not be efficient since this will cause redundancy and takes a lot of time. Thus, ontology reuse has been potentially seen as a better method.

Ontology reuse can be defined as a process where a small portion of existing ontologies is taken as an input to build a new one [11] where parts of it are manipulated to meet the requirements of the application using the ontology [42]. The process of reusing parts of a large existing ontology allows for the ontology to be used in an application without slowing down the process. Ontology reuse also increases interoperability [81]. Indeed, when several other new ontologies reuse an ontology, interoperability between these ontologies can be achieved much easier since they share several features such as classes naming method and concept modelling.

Even though ontology reuse brings many benefits, there are currently no tools that provide adequate support for the ontology reuse process [51, 81] which becomes a hindrance for its implementation. This is with the exception of the Platform for the Reuse of Ontologies through Merging and Integration (PROMI)[6] which is a Java-based tool that guides users in reusing ontologies, developed by Bontas and her colleagues at the Freie Universität Berlin [10]. This tool, however, has its limitation which makes it

[5]http://www.rsna.org/RadLex.aspx
[6]http://promi.sourceforge.net/

inadequate in reusing biomedical ontologies. This will be elaborated further in Subsection 2.3.5.

Ontology reuse does not have one specific method agreed in conducting it. Even so, most practices in reusing ontologies in previous works [1, 10, 15, 16, 28, 75, 79, 81, 90] fall along the line of these four steps:

1. Ontology selection for reuse
2. Concept selection
3. Concept customisation
4. Ontology integration

The first step commonly adopted for ontology reuse is to select the ontology to be reused. Ontology selection is performed according to several criteria based on the needs of the new ontology, for example, the language of the ontology, its comprehensiveness and its reasoning capabilities. One or several ontologies can be selected for reuse depending on the needs of the new ontology. Once the ontology for reuse is chosen, the next step would be to select the concepts to reuse. The concepts selected are then translated into the same semantic language and then merged. Several different concepts can also be selected from several different ontologies which are then customised to suit the rest of the ontology. Finally, the ontology will be integrated into the system or application. The next section provides examples of how these steps were applied in previous works involving ontology reuse.

2.3.4 Previous Ontology Reuse Work

Ontology reuse has been practised in various ontology development works in order to optimise the usage of existing ontologies and reduce redundancy. In the ontology reuse system proposed by Alani [1], all the terms needed for the new ontology were first listed to determine which existing ontologies should be selected for reuse. The system then searches for relevant ontologies online using the terms that have been listed and from the obtained results, the ontologies will then be ranked and only the first few will be analysed and selected for ontology reuse. Then, the system will determine whether to reuse the whole ontology or only a segment of it. This step is the concept selection step where relevant concepts are being chosen. In his work, Alani's approach was to reuse several ontologies for one concept [1]. Therefore, each group of related concepts needs to be customised to ensure that they have a standardised semantic language so that they can be merged as one concept. Since the concept is reused from several ontologies, each concept contains different properties which resulted in additional knowledge representation. Finally, the ontology is automatically evaluated.

Another example of ontology reuse is in the development of the Oral-Systemic Health Cross-Domain Ontology (OSHCO) by Shah, Rabhi and Ray [79]. In developing OSCHO, the first step taken by the authors was to determine the scope of the ontology by recognising the domain of coverage, the intended use and the questions that OSCHO should be able to answer. Once the scope has been determined, they utilised a tool in Bioportal and submitted several domain related terms to assess the domain coverage of several ontologies that have the potential to be reused. This has resulted in SNOMED CT being the best candidate. Compared to Alani, they reused only one ontology; SNOMED CT where they selected the concepts needed and then added other relevant concepts not included in SNOMED CT. Since they reused only one ontology, the new ontology developed; OSCHO closely follows the model of SNOMED CT where the new concepts added are customised to follow the same model.

Russ et al. [75] in their work aimed to develop an aircraft ontology that can be used by several applications to ensure knowledge sharing. During the development of their work, they realised that there exist two ontologies that are related to the aircraft domain. Furthermore, some concepts exist in one ontology but not in the other. Thus, Russ et al. have decided to merge both ontologies to create a more complete one. The first step taken was to select ontologies for reuse which are the time ontology (which is publicly available) and the two existing aircraft ontologies. The time ontology was written in Ontolingua while the two aircraft ontologies were written in Loom. After both aircraft ontologies were merged, the time ontology is translated to Loom allowing its integration into the aircraft ontology. The same method was adopted by Caldarola, Picariello and Rinaldi [15] for developing an ontology in the food domain where metadata were manually translated to better understand the concepts.

Polionto [67] is another example of an ontology developed using ontology reuse. Ortiz in his work developed Polionto by reusing two ontologies in the political domain where the first ontology is in Portuguese and the other in English. These ontologies were selected based on their popularity, coverage and knowledge details. The author then selects relevant concepts to reuse from both ontologies and compares them to the corpus that has been formed using documents in the political domain in order to ensure its coverage. Those concepts are then translated and integrated to create a multilingual political ontology called Polionto.

These examples demonstrated that previous works involving ontology reuse loosely follows the four steps mentioned in Section 2.3.3. In this research, these four processes were also used as guidelines to develop a novel ontology reuse methodology for the biomedical domain. The development of an ontology reuse methodology is intended to increase the practice of ontology reuse among ontology developers. This methodology includes suggestions of sets of freely available tools that were utilised to aid the reuse process. There was a tool previously developed to assist in the process of ontology reuse. However, there are some limitations in using the tool in the biomedical domain. The next section will discuss further on this.

2.3.5 Platform for the Reuse of Ontologies through Merging and Integration (PROMI)

Comprehending the need for a tool that will assist the process of ontology reuse, Bontas and her colleague developed a tool named PROMI that is able to perform the steps required in an ontology reuse process [10]. Figure 2.4 illustrates the user interface of PROMI.

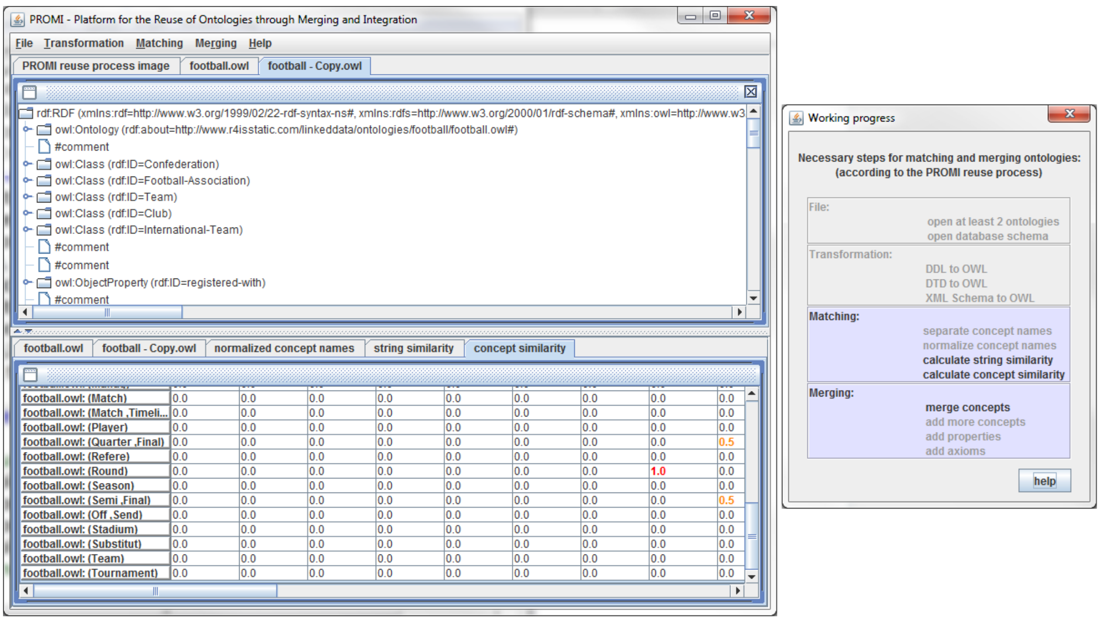


Figure 2.4: The user interface of PROMI

PROMI begins by prompting the users to upload at least two ontologies to be reused. Then, the language of the ontologies will be examined in order to decide whether the transformation to OWL language should be performed. Next, the matching of the concepts was conducted where PROMI first separates and normalises the concept names before calculating the string and concept similarity. Finally, the concepts will be merged and users are allowed to include more concepts, properties and axioms as necessary.

The development of PROMI is an immense step forward in assisting the process of ontology reuse. However, there are several limitations in using PROMI as a tool to assist ontology reuse. First, PROMI does not provide the facility to assess the suitability of an ontology to be reused for a specific domain. Instead, it begins by prompting the users to upload ontologies which they have selected beforehand. It also does not recommend any ontologies for reuse based on the domain which means that the users will need to use their own expertise in selecting suitable ontologies.

Second, the concept matching between the two ontologies depends heavily on the similarity measures chosen by the users. For example, the similarity measure for the word “tournament” and “competition” was 0 using the Euclidean Distance measure and 0.143 using the Hamming Distance measure. Therefore, the users need to have some knowledge about the differences between these measures for them to be able to select the most appropriate one.

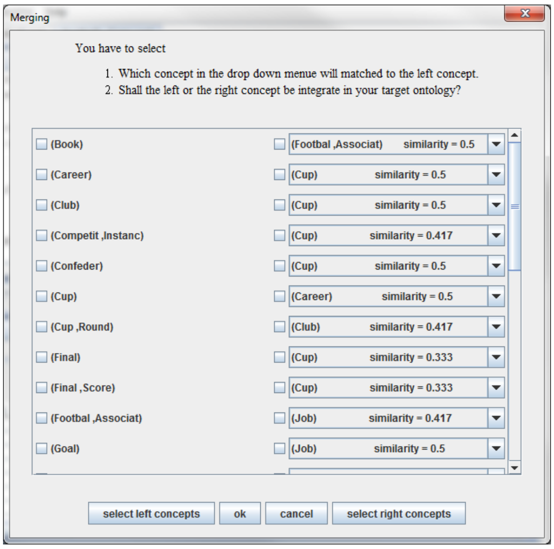


Figure 2.5: The process of merging concepts in PROMI

Finally, in order to merge concepts in PROMI, users will need to select from a list of concepts and its equivalent candidate concepts which have a certain degree of similarity. This is demonstrated in Figure 2.5.

The process of merging concepts in PROMI will not be efficient in this research as thousands of concepts in large biomedical ontologies will need to be assessed in order to select only the relevant ones. As a conclusion, PROMI could be a beneficial tool in assisting the process of ontology reuse. However, there are still several features that can be included and improved so that it can ease the ontology process even more and allows the usage on large ontologies to be more efficient.

2.4 Rhetorical Structure Theory

Reiner [72] mentioned that most technologies used to create structured medical reports were developed in a way that requires the radiologists to adapt to it, producing relatively static and inflexible data. This further adds to the resistance of the radiologists in adopting structured reporting in presenting their findings. Reiner [72] then hypothesised that if there is a way to create technologies that are able to adapt to the needs of the radiologists instead, there will be improvements in the productivity, workflow, diagnostic accuracy and clinical outcomes of the reports.

Hence, this research investigates the possibility of developing a medical ultrasound reporting system that adapts to the radiologists' preferences by allowing them to create reports in any way but still produces a structured version in the end. The generation of structured reports from free-forms can be performed by utilising the Rhetorical Structure Theory (RST) which was pioneered by Mann, Matthiessen and Thompson in the 1980's [87].

RST is a theory which describes the major features of natural text organisation [52]. It allows for the classification of a span of texts and the description of relations between two or more spans of texts that have independent functional integrity unit which are also known as Elementary Discourse Units (EDU) [53]. RST is insensitive to the size of the text [52] which makes it suitable to be applied in small text documents such as a memo as well as in large documents such as a journal article. In this research, RST has been chosen as the mechanism to transform free-form reports to structured form because of its ability to define clear relationships between texts [52]. This allows for text spans to be extracted from a whole paragraph or sentence without losing its relation with other text spans.

RST has the ability to present texts as coherent where it illustrates the hierarchical structure of each part of the texts as having a role to play in regards of each other [87]. These roles are defined by rhetorical relations between the texts and can be identified as either a nucleus or a satellite. Nucleus are parts of texts that can standalone without its satellite but satellite on the other hand losses its meaning without the nucleus [52]. A well written text is coherent where it should have relations between each of its text spans and no text span is isolated [26]. There are two types of rhetorical relations that can occur between the nucleus and the satellite which are mononuclear relation and multinuclear relation. These relations are defined as follow:

Mononuclear: RELATION(N,S)
Multinuclear: RELATION(N,N)

In a mononuclear relation, the nucleus has a higher importance compared to its satellite. On the contrary, a multinuclear relation puts the same importance on all of its nuclei. Mann and Thompson [52] have defined a total of 24 rhetorical relations which were known as the classical RST relations. Six more relations were then added [93] which makes it a total of 30 mononuclear and multinuclear relations where the mononuclear relations were further categorised into presentational relations and subject matter relations. Each relation is defined according to four fields; (i) constraints on nucleus, (ii) constraints on satellite, (iii) constraints on combination of nucleus and satellite, and (iv) the effect towards readers or the intention of the writer [52]. Table 2.1 shows some of these relations together with their nuclei and satellites description.

The first four relations in the table are mononuclear relations where the nucleus is seen as more important than the satellite as it is able to stand by itself. The other three relations are multinuclear relations where all nuclei have the same importance. Example of both types of relations can be seen in Figure 2.6 where there are two RST trees each representing a mononuclear relation, CONCESSION and multinuclear relations, CONTRAST adopted from Taboada [87]. In the CONCESSION relation, the text span "we shouldn't embrace every popular issue that comes along" is the nucleus of the text span. Without the text span "Tempting as it may be," which is the satellite, the nucleus still makes sense. However, it does not hold for the satellite. On the other hand, in the CONTRAST relation, both "Animals heal," and "but tress compartmentalize" text spans are equally important and can be understood on their own.

RST trees are used as a schema to visualise the structure of the text that has been parsed [52]. A curve arrow represents a mononuclear relation where the arrow starts

Relation Name	Nucleus	Satellite / Nucleus
Background	text whose understanding is being facilitated	text for facilitating understanding
Elaboration	basic information	additional information
Concession	situation affirmed by author	situation which is apparently inconsistent but also affirmed by author
Preparation	text to be presented	text which prepares the reader to expect and interpret the text to be presented
Contrast	one alternate	the other alternate
Joint	(unconstrained)	(unconstrained)
List	an item	a next item

Table 2.1: Example of rhetorical relations in RST [93]

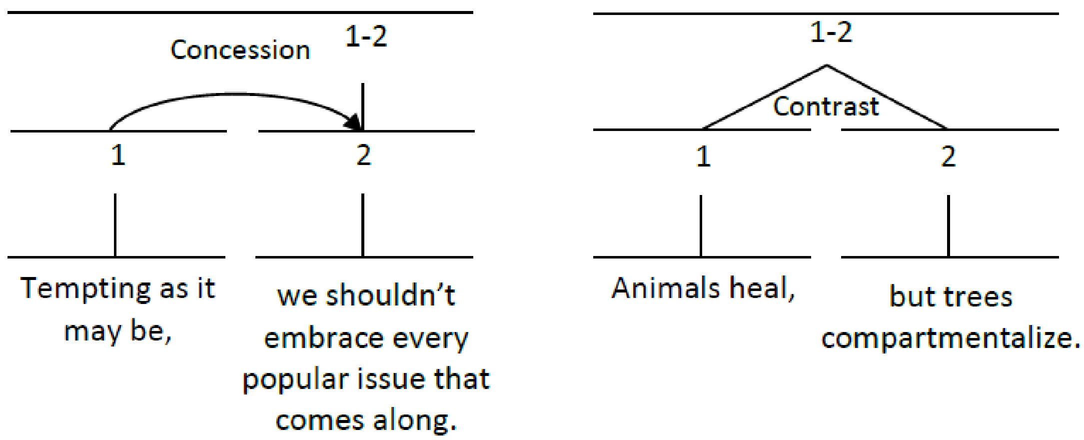


Figure 2.6: CONCESSION and CONTRAST relations [87]

from the satellite and points to the nucleus. On the other hand, straight lines are used to represent multinuclear relations. Both these subtrees can be seen in Figure 2.6 where each subtree represents a relation. The combination of these subtrees makes up a larger tree that contains several rhetorical relations for a given text.

2.4.1 Automatic RST Discourse Analysis

Discourse analysis using RST is performed by first segmenting the text into text spans based on discourse boundaries that have been set. These segmented text spans will then be parsed to identify rhetorical relations that exist between them. Discourse analysis was performed manually by Mann and Thompson [52]. Even though their work was successful, Marcu [53] argued that their work was too informal and not suitable to be automated. Thus, there were several works that have attempted to automate this process. This section explores some of the notable works that have been developed by Marcu [53, 54], Soricut and Marcu [86], Feng and Hirst [26, 27] as well as Joty et al. [38, 40, 39] in developing an automatic discourse parser.

Marcu (1997, 2000)

Daniel Marcu [53, 54] was one of the earliest to have developed an automatic discourse parser. He proposed a surface-based algorithm that is able to perform three tasks which are (i) identifying cue phrases and breaking sentences using these cue phrases, (ii) hypothesising rhetorical relation between text as well as (iii) producing an RST Tree. He developed what he called a shallow analyser that is able to determine elementary discourse unit of a text, the rhetorical relation that exist between these units as well as the importance of the unit whether it is a nucleus or a satellite. The term shallow analyser was introduced because it does not use the traditional parsing and tagging technique. Instead, Marcu argued that discourse parsing using RST can be performed automatically by relying only on the coherency and connectivity of the text. Assuming that texts are well formed, he stated that it is sufficient to segment text and hypothesising relations using discourse markers.

In his study, Marcu has compiled 450 discourse markers from several existing discourse markers list where for each marker, 17 text fragments have been selected from the Brown corpus[7] which was the first computer readable general corpora. These texts with discourse markers were then paired with a set of discourse related information such as a marker field that denotes the paragraph break of the text, a usage field which explain the role of the discourse marker on whether it is sentential, discourse or pragmatic, a position field which specifies whether the discourse marker is at the beginning, the middle or the end of the text as well as a rhetorical relation field which lists the name of the relation that is signalled by the discourse marker. From his study, Marcu found that discourse markers are sometimes ambiguous depending on the relation they are signalling as well as the size of the whole text. The more complex the text is, the more ambiguous it becomes to automatically detect rhetorical relations.

To solve this problem of ambiguity, Marcu took advantage of the text coherency

[7] http://www.hit.uib.no/icame/brown/bcm.html

where he looks for similarity measure of co-occurrences of words. If the measure exceeds a certain threshold, then a relation holds. When two text spans mention the same topics and share the same words, they will be assigned with the ELABORATION or BACKGROUND relation. However, if two text spans mention two different subjects, they will be assigned with a JOINT relation. To evaluate the accuracy of his parser, Marcu used the result from a manual discourse analysis completed by three judges with two third majority. Overall, Marcu's shallow analyser achieved 80.8% recall and 89.5% precision which is quite high compared to other discourse parsers at that time.

Soricut and Marcu (2003)

Soricut and Marcu [86] developed a sentence level discourse parser that uses a probabilistic model to identify EDUs and build a sentence level RST tree. Their work aims to show that it is possible to build an RST tree by exploiting the syntactic and lexical information of the sentence. They used a semantic and discourse annotated corpora called the RST Discourse Treebank (RST-DT)[8] which contains 385 Wall Street Journal articles from the Penn Treebank. Out of the 385 articles, RST-DT used 347 articles with 6132 sentences as training data while the other 38 articles with 991 sentences was used as testing data where each articles has its own corresponding manually built RST tree.

The limitation of Soricut and Marcu's parser is that it works only on the sentence level. They have assigned sets of tuples in the form of

$$<s, st(s), dt(s)>$$

where:

s = sentence

st(s) = syntactic tree for the sentence from the Penn Treebank

dt(s) = discourse tree for the sentence from RST-DT

The parser used a statistical approach to insert boundaries after each word if it has a probability score higher than 0.5. The segmenter then uses the boundaries to perform the segmentation. Soricut and Marcu's work uses a bottom-up algorithm. If there are two probability scores for a given EDU, the lower score will be removed. Overall,

[8] http://catalog.ldc.upenn.edu/LDC2002T07

their parser surpasses the performance of Marcu's parser [54]. They also claimed that the parser's accuracy can match near-human levels of performance if the segmentation was performed manually.

Feng and Hirst (2012, 2014)

Realising the limitations of sentence level parsers such as the one developed by Soricut and Marcu [86], Feng and Hirst [26, 27] have developed a text level RST discourse parser which imposed more constraints on global coherence. The first version of their parser [26] enhances the HILDA [35] discourse parser by including their own rich linguistic features such as cue phrases, production rules and contextual features combined with features from the work of Lin et al. [49] such as the dependency parse feature. Since HILDA has already achieved a high F-score of 93.8%, Feng and Hirst have focused on improving the RST tree building task instead. Adopting the same methodology as HILDA, Feng and Hirst used a greedy bottom-up approach with two Support Vector Machine (SVM) classifiers in their parser. The first classifier is the binary structure classifier which evaluates whether there is a relation that holds between two text spans and merges them into a new subtree. The other classifier, the multiclass relation classifier on the other hand, evaluates which relation should then be assigned.

In 2014, Feng and Hirst [27] aimed to improve the discourse parser developed by Joty et al. [40] using the Conditional Random Fields (CRF) as local classifiers. Joty et al.'s [40] parser which will be discussed further later, achieved an overall relation assignment accuracy of 55.73%. However, Feng and Hirst points out that Joty et al.'s parser is highly inefficient in practice as it takes hours for the parser to process texts since the Cocke-Kasami-Younger (CKY) parser they used searches all possible paths. Instead, Feng and Hirst uses a greedy bottom-up approach adopting four local models with two dimensions, (i) scope of the model, either intra-sentential or multi-sentential and (ii) the purpose of the model, either to determine structures or relations. Feng and Hirst [27] also introduced a novel feature in their recent parser which is a post-editing process that allows the RST tree to be modified based on top-down information such as the depth of the tree which can only be obtained after the tree has been fully built. This feature doubles the time required to parse the text. However, it is outweighed by the fact that it enhances the performance of the parser to close to 90% of human performance.

Joty et al. (2012, 2013, 2015)

Joty et al. [40] have developed a text level discourse parser similar to Feng and Hirst's [26] first version of a text level discourse parser but instead of using SVM, they have used two CRF model to build RST trees using optimal parsing algorithm. These two models were used, one for intra-sentential parsing which produces subtrees for each sentence, and one for multi-sentential parsing which combines these subtrees into a text level RST Tree. Joty et al. claimed that separating the parsing of both intra- and multi-sentential is much more effective compared to combining them. In parsing the text, Joty et al. have taken an approach which was non-greedy and optimal [38] compared to HILDA and Feng and Hirst's parser which used suboptimal and greedy approach. Their approach uses a probabilistic CKY-like bottom-up algorithm that resulted in a globally optimal RST Tree and received an overall of 55.73% relation assignment accuracy when tested on two types of corpus which were news articles (RST-DT corpus) and instructional how-to-do manual. Feng and Hirst [27] however argued that their approach is inefficient as it takes too much time.

In 2015, Joty et al. [39] have rebranded their discourse parser together with their discourse segmenter [38] as CODRA[9] which stands for "a complete probabilistic discriminative framework for performing rhetorical analysis in accordance to RST" and modifies their parsing algorithm to search for the k most probable RST trees for the text. This modification allows the parser to store and track k-best candidates simultaneously instead of storing just a single best parse. When tested on the RST-DT corpus, their k-best intra-sentential reranking parser improved the accuracy significantly by 13.45% (base accuracy is 79.77%) for 30-best. However, when tested on document level, their k-best parser did not give much improvement to the base accuracy (55.83%) with improvement as little as 1.91% for 30-best.

2.4.2 Ontology Usage in RST Discourse Analysis

Whilst existing works on automatic discourse analysis have achieved high accuracy, none that we know have been tested on biomedical corpus especially on ultrasound reports. We argue that because of the nature of most ultrasound reports where sentences are often not grammatically complete, the conventional method of parsing text would not achieve high accuracy. Thus, it has been proposed that an ontology be used to support the discourse parsing by using a rule-based approach. The usage of ontology

[9] http://alt.qcri.org/demos/Discourse˙Parser˙Demo/

in RST discourse analysis has not been explored much in previous works. One notable work would be the discourse parser developed by Bärenfänger et al. [2] which used two ontologies, a taxonomy of rhetorical relations (RRSet Ontology) and an ontology version of GermaNet[10] as a knowledge base for the discourse parser.

The RRSet ontology consists of 70 rhetorical relation types where 44 of them were basic types and the rest were subtypes. In their work, the ontologies were used to help determine the type of relation that exists between text spans. For example, in determining an ELABORATION subtype, ELABORATION-CONTINUATION where it is not indicated by any discourse markers, the ontology will be consulted to find synonymy or pertainymy between the two text spans. Their work is different compared to what is being carried out in this research. Instead of using an ontology to consult about the rhetorical relation types, our work uses the ontology to annotate the corpus with relevant classes. The annotated corpus will then be compared with the rules defined using classes from the same ontology to determine the type of rhetorical relations that exist.

2.5 Chapter Summary

The purpose of this chapter is to undertake a thorough study of the problem of variations in ultrasound reporting and explore some of the possible solutions to this problem. First, we study the process of an ultrasound examination in order to understand the importance of reports in ultrasound. We understood that because ultrasound examination involves at least two different personnel, the clarity of the reports produced is very important as they serve as a communication tool between the two personnel.

Following that, we explore the problems of variations in ultrasound reports where it can be divided into two major areas which are the reporting styles and format and the terminologies used. Ultrasound reports can be written in two different forms which are the free-form and the structured form.

Cramer et al. [18] further classified the structured form into three tiers. The first tier is where reports have contents grouped under a specific heading. The second tier on the other hand improves the first tier by having the content itemised. Finally the third tier further improves this by standardising the language used.

We learned from this that the 100 sample reports we collected belongs to the first tier of structured reporting. However, we observed that the first tier structured reports were not enough to reduce the problem of variations in ultrasound reports. From this,

[10]http://www.sfs.uni-tuebingen.de/GermaNet/

we decided that the medical ultrasound reporting system developed in this research should comply with Cramer et al.'s third tier of structured reporting.

To make sure that structured reporting is the best option to solve the problem of variations in ultrasound reporting, we then evaluate the potential benefits that structured reporting could bring. This leads us to five benefits of structured reporting which are the production of a standardised report, consistent terminology usage, improved accuracy of the reports, increased accessibility of data and higher reproducibility of the reports. We found that these benefits complied with Ellis and Srigley [23] seven key parameters of quality reporting namely, timeliness, accuracy, completeness, conferment with current organisation standard, consistency and clarity in communication.

Although we found many benefits of structured reporting, there are several reasons why wider implementation of it fails. Review of several literatures lead us to three main reasons which are the rigidity of the report, technological constraints and the radiologist's resistance to change. Out of this three challenges, the radiologist's resistance to change is seen as the biggest obstacle in implementing structured reporting.

Following this, we look at the possibility of our research improving this. First, the medical ultrasound reporting system that we developed should be able to standardise terminologies used in the reports. This has led us to the usage of ontology because of its ability to share common knowledge and enables reuse.

Our study found that there are several existing biomedical ontologies that can be used in our reporting system. However, these ontologies were too big to be implemented in a specific application. Following this, we decided to reuse classes from existing ontologies to develop an ontology that is more specific to our domain.

Finally, we investigated the possibility of giving flexibility to the radiologists in using the reporting system by allowing them to write their reports in free-form which will later be transformed to structured form. We found RST as the best mechanism to do the transformation because of its ability to recognise and retain relations between text spans. As a result of this, we investigated previous works that have implemented RST in discourse parsing whilst putting emphasis in the possibility of combining ontology together with RST.

The study in this chapter brings us to the conclusion that it is possible to reduce the problems of variation in ultrasound reporting by standardising the reporting style, format and terminologies used. It highlighted that the terminologies can be standardised using an ontology while the reporting style and format can be standardised using structured reports generated by RST.

Chapter 3

The Design of a New Ultrasound Reporting System

3.1 Introduction

Humans are known to sometimes being incapable of accepting changes and adapting to new things. For an ultrasound reporting system to be accepted by the radiologists, it is important that the system has a level of flexibility that would allow for reports to be produced in a structured form without causing too much inconveniences and frustrations to the radiologists. Studies developed by Bosmans et al. [13] and Danton [19] concluded that radiologists were found to have problems by not having more options in creating their report. This chapter presents a system architecture model that allows for these radiologists to have more options and flexibility in creating their report. It explains how the overall system works as well as how each component plays their part. The focal point of the system architecture model is the structured report page. This chapter will discuss in detail about the design of the page including the guidelines that were used as references. The overview of the system architecture model presented in this chapter has been published as a position paper in the 8th International Conference of Health Informatics [94]. The system architecture model was also presented as a poster in the 2015 UK Radiological Congress (UKRC) where a survey on the ultrasound reporting system was also carried out. This chapter presents the result of this survey and discusses the issues that were raised.

3.2 The Design Methodology

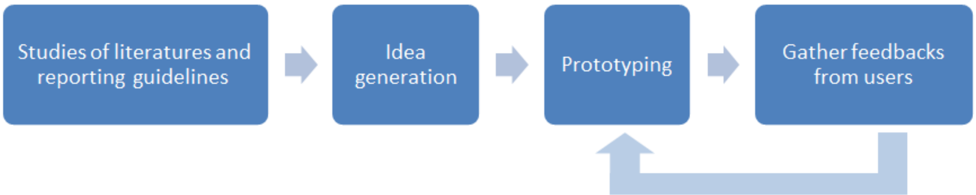


Figure 3.1: The design methodology

The design of a new ultrasound reporting system follows the methodology displayed in Figure 3.1. First, a study was conducted using various literatures and ultrasound reporting guidelines in order to compile the best practices in producing ultrasound reports. Accordingly, ideas on how these best pratices can be applied in the design of the architecture model were generated. A prototype of the design was then developed before feedback was sought from the end users during the 8th International Conference of Health Informatics. Based on the feedback, the protoype was redesigned and further feedback was gathered during the 2015 UK Radiological Congress (UKRC) in the form of a survey and discussions. Following this, the protoype was redeveloped before a final evaluation was performed with ultrasound specialists.

3.3 Components of the System Architecture Model

In the literature review, it was found that one of the biggest challenges in implementing structured reporting in ultrasound reports is the radiologists' resistance to change. This was clearly observed in the study undertaken by Tran et al. [89] where resistance was manifested by the senior radiologists compared to the trainees because they have been accustomed to writing in free-form. We believe that flexibility allows humans to slowly adapt to changes. With that in mind, a system architecture model that provides flexibility for the radiologists in writing their reports was designed. This enables the radiologists to choose whichever method they are comfortable with in writing the report but at the same time producing the same result which is a standardised ultrasound report. The model consists of a main page, an option to create or upload a report in either free-form or structured form and a mechanism to standardise all reports into structured form. Figure 3.2 depicts the components of the architecture model.

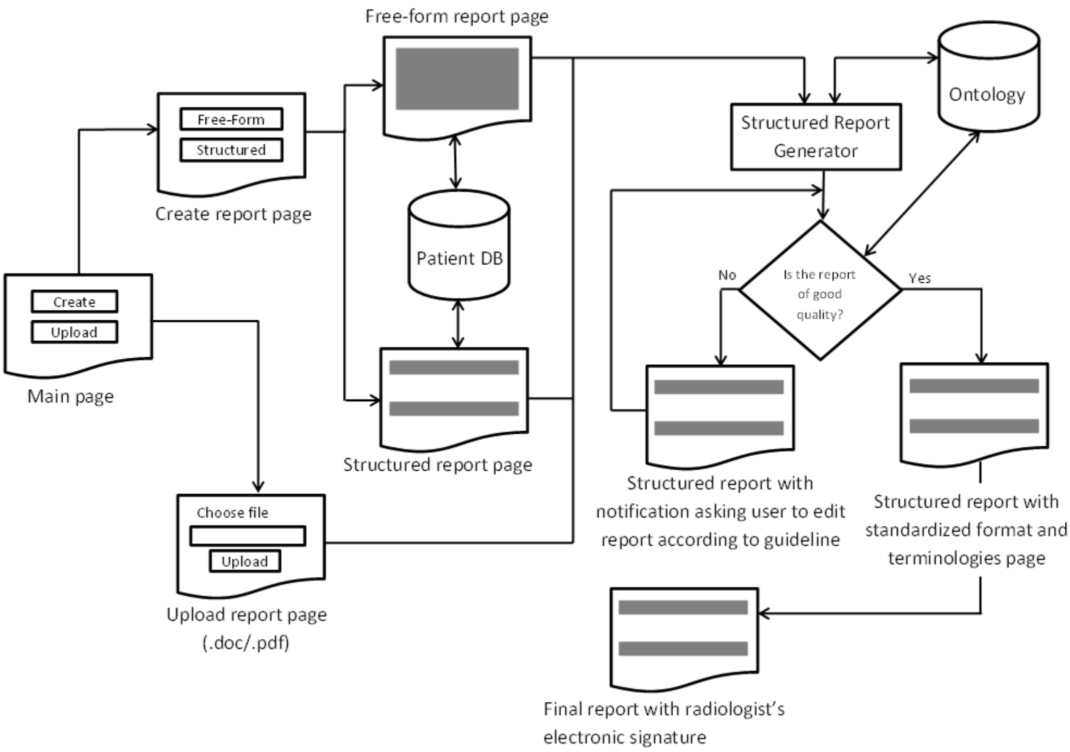


Figure 3.2: System architecture model of the reporting system

3.3.1 Main Page

The main page of the ultrasound reporting system consists of two options for the radiologists to choose from in creating their report. They could either choose to create the report from scratch or upload an existing report. If the radiologists want to generate a new report for an ultrasound image, they could choose the "create a report" option that will produce a structured form report without restricting the radiologists to write in a certain format. However, if they have existing reports that needs to be transformed to structured form, they can choose the "upload existing report" option. Figure 3.3 displays the user interface design of the main page.

3.3.2 Create Report Page

If the radiologists choose to create an ultrasound report from scratch using the system, they would have a further option in doing this. Most medical ultrasound reporting system uses a structured form for the radiologists to complete. This requires a lot of mouse control and clicking. A structured form is also often too rigid where it forces the radiologists to complete most parts of the form and not allowing them to submit

Figure 3.3: The user interface design of the main page

the form until all parts are completed. This could cause frustration to the radiologists and deter their interest in using the system. In our system architecture model, the radiologists would have an option whether to create their report using a guided free-form or by using our structured report form which has been designed to be less rigid compared to the existing structured report forms.

3.3.3 Free-form Report Page

Free-form report page allows the radiologists to create an ultrasound report by freely typing in their observation without the need to adhere to a certain structure. By creating a report using our system, the radiologists would not need to provide patient information because the system will automatically fetch the information from a database and displays it on the form. The free-form report page will consist of three questions related to the ultrasound examination as exhibited in Figure 3.4. These questions will act as a guide for the radiologists to write their report so that enough information is provided. Once the radiologists are finished with the report, they will be able to submit it directly and the system will then automatically transform it into the structured form.

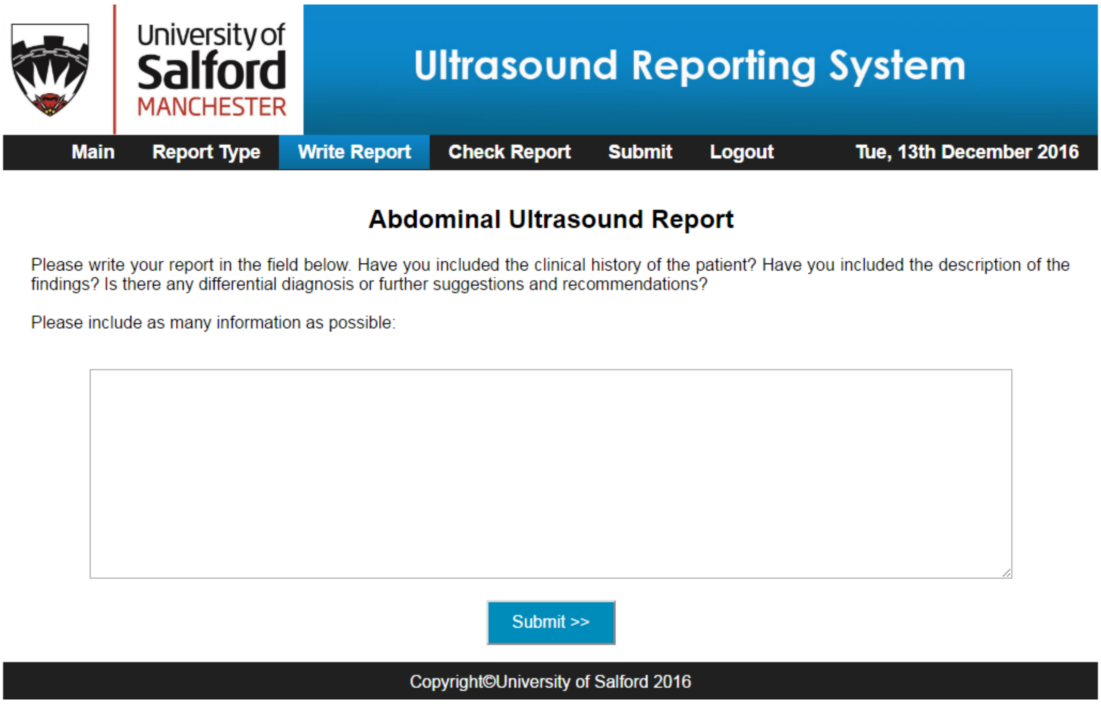


Figure 3.4: The user interface design of the free-form page

3.3.4 Structured Report Page

In our system architecture model, a structured report page similar to what have been developed by previous works was included. The main difference between the structured report page that we have developed and the ones used in previous ultrasound reporting systems is that ours was designed with much more flexibility and less rigidity. The structured report page was also designed according to the quality criteria guidelines as provided in Appendix A. These guidelines were developed based on suggestions from literatures and specifications created by the Society and College of Radiographers and British Medical Ultrasound Society [64] as well as the United Kingdom Association of Sonographers [66].

It is recommended that the ultrasound report be presented in a tabulated or itemised format [19, 22, 69] by separating the report with suitable headings. The report should include basic information about the patient such as his name, age and gender as well as the radiologist such as his name and status and the location of the examination [41, 66]. It should also include the patient's clinical history including the reason for referral if available [41]. Following this, several text fields and a radio button have been included to accommodate all of this information as seen in Figure 3.5.

Similar to the free-form report page, the radiologists would not need to provide

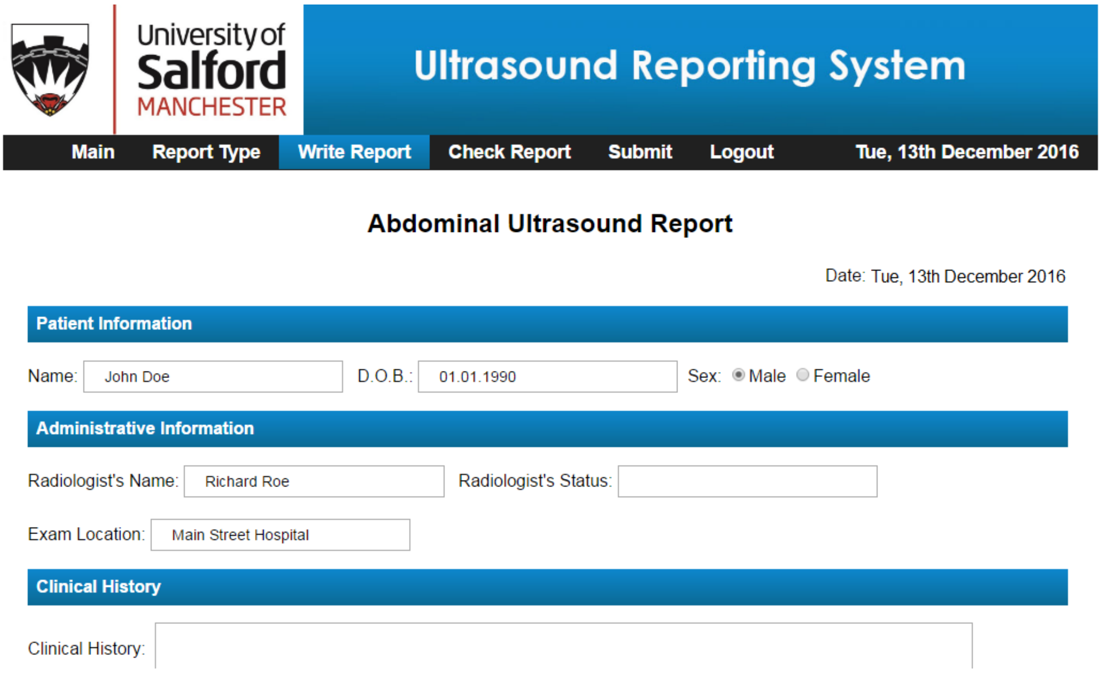


Figure 3.5: Basic information and clinical history fields

the patients' information as this will be included automatically by the system. This reduces the need for the radiologists to key in a lot of data and allow them to focus on the findings of the examination instead. The structured report page automatically records the time when the report is written and when the radiologists sign off the report. Information such as this is important in assessing the turn around time of a report. It can also be refered to whenever there are any medical legal issues.

When submitting the form, it will validate whether the basic information has been provided. If not, the form will not be allowed to submit. However, if the clinical history field has been left blank for example, the system will remind the radiologists to complete it in once. If the field has been left blank for the second time, the system will allow the form to be submitted since sometimes there is no clinical history available. It is important that the form be flexible because there are times where not all information is available to the radiologists. If the form does not allow them to submit until all fields are completed, they might not be able to submit the report at all.

The ultrasound report should include all relevant areas including both examined and not examined areas [22, 66]. If the relevant areas have not been examined, the reason for it should be stated [66]. To ensure that flexibility is given to the radiologists in completing the report, they have the option to add or remove fields when reporting the relevant areas (see Figure 3.6). All findings and observations should be reported

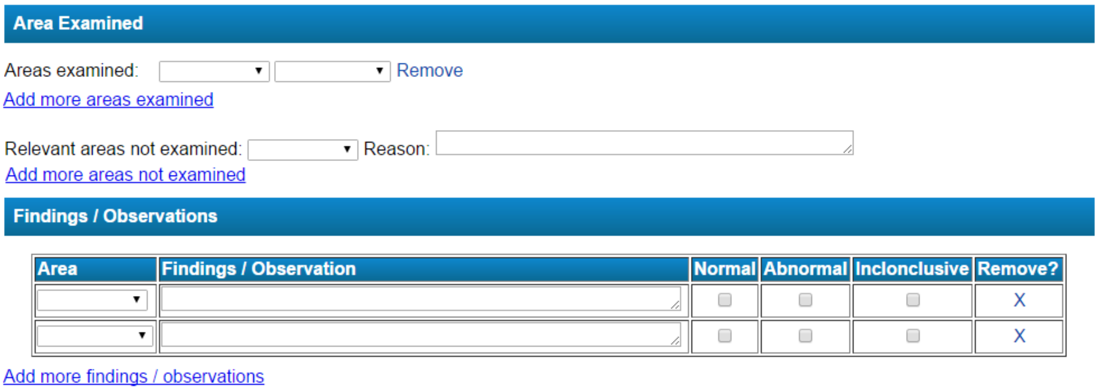


Figure 3.6: Fields for relevant areas and observations / findings

and classified as either normal or abnormal whether it is expected or not [66]. The discussion and the survey conducted during UKRC 2015 has prompted us to design a table as displayed in Figure 3.6 where findings are separated by areas and can be classified as “normal”, “abnormal” or “inconclusive”. When selecting the “normal” checkbox, the radiologists have a further option to indicate whether it is a normal variant. If the radiologists select “abnormal”, they can then indicate if it is an incidental finding. The discussion and reasoning that has led to this design will be elaborated further in Section 3.4.2.

The reason that the radiologists are given the option to classify their findings is to make their report more meaningful. Edwards, Smith and Weston [22] have suggested that radiologists should avoid measurements in their reports unless it is stated whether the measurement is normal or abnormal. The rows in the “findings / observations” table can also be added or removed based on the number of areas that the radiologists have examined. The structured form also includes a field for the radiologists to give a conclusion to the report that answers the clinical question. It also includes a field for the radiologists to suggest any further examinations. The complete structured form interface is displayed in Figure C.6 in Appendix C.

Though this structured report form still requires a lot of mouse control and clicking, the radiologists will always have the option of using the free-form report if they find it hard to use the structured form. However, if they choose to use the structured report form, the benefits are that it would help guide the radiologists in giving enough information about the ultrasound examination and ensure that the report they produced will be of good quality. This reduces the probability of them needing to edit the report to conform to the quality guidelines.

3.3.5 Upload Report Page

Another option that the radiologists have in creating a standardised ultrasound report is by uploading a report that they have written elsewhere. The report could be in formats such as .doc and .pdf and could be written in any way that they prefer. This option is not only for reports that have been recently written. The radiologists could also use this option to upload a free-form report that has been written some time ago in order to convert it into a standard structured form so that it can be used for research and education purposes. If the report does not contain enough information, the generator will notify the radiologists to add the necessary information before allowing it to be signed off.

3.3.6 Structured Report Generator

All reports whether created using our ultrasound reporting system or by uploading existing reports will go through a structured report generator which will transform them into standardised structured reports. The structured report generator uses AUO as its knowledge base where it consists of hundreds of relevant medical terms commonly used in abdominal ultrasound reports. The ontology helps the system to understand what is written in the report and will use that information to create a standardised report. Chapter 4 of this thesis will explain further on the development of AUO. The transformation of the free-form report to structured form is executed using RST whereby the free-form report will be annotated with classes from AUO. RST will then use this annotated report together with a set of rules and discourse markers to segment the report and identify rhetorical relations within it. Chapter 5 of this thesis will give a full description on the working system of this mechanism.

Before the standardised report is displayed to the radiologists, it will go through a quality checker in order to ensure that the report meets the standard quality measure as provided in Appendix A. If it does meet the measure, the standardised report will be displayed for the radiologists to check and sign off. However, if the report does not meet the minimum requirement, a standardised report will still be displayed but with notifications for the radiologists to add more information. These notifications serve as a reminder on what is expected from a report that satisfies the quality requirements. However, if the information is not available, the system will still allow for the report to be displayed and signed off since the information needed is not always available. Considering that this is a medical related system, the electronic signature of the radiologist writing the report is very important since it acknowledges that all information on the report is true.

3.4 Interface Preferences Survey

During the 2015 UK Radiological Congress held in ACC Liverpool, thirty sets of questionnaires were prepared and given out to the attendees. Before answering the questionnaire, each respondent was given a brief explanation on the proposed ultrasound reporting system. The questionnaire consisted of fifteen questions which can be grouped into three different sections (see Appendix B). The first section asked about the demographics of the respondent; their industry or field of work, their age as well as the average time per week they spent in using healthcare or medical systems in their work.

The second section aims at understanding the respondents' view on the variations in reporting styles. A 5-point Likert scale was used to find out whether the respondents agree that variations in reporting style impacts report interpretation and/or patient diagnosis. The third section was the main part of the questionnaire where it focuses on getting the respondents' feedback and opinion on the user interface design of the structured report page in the proposed ultrasound reporting system. In answering this section, respondents were asked to refer to a screenshot of the structured report (see Figure B.1 in Appendix B). Questions in this section are in the form of a 5-point Likert scale and open-ended questions.

Question six, seven and eight of the questionnaire asked about the general look of the structured form; whether the form allows for information to be read easily, contains enough information and whether it is easy to fill. Question nine was an open-ended question that requires the respondents to classify an example of an ultrasound finding as normal or abnormal. Question ten and thirteen seek to get the respondents opinion on whether flexibility is an important feature for an electronic structured report. Other questions in this section provided free text area for respondents to give further comments on the interface or to give justification on their answers. The respondents were also asked if they would like to be contacted for further feedback and updates on the research.

Since the questionnaire used a 5-point Likert scale, the frequency, mean and standard deviation of each question was calculated. In discussing the result of the questions that used the Likert scale, we combined the responses for "strongly agree" and "agree" as a positive result while “strongly disagree” and “disagree” as a negative result. This is to simplify the understanding of the result. The complete result for the questions that uses a 5-point Likert scale is summarised in Table 3.1.

Question No.	Strongly Agree (%)	Agree (%)	Neutral (%)	Disagree (%)	Strongly Disagree (%)	Mean	Standard Deviation
4	37.5	56.3	6.3	0	0	4.31	0.60
5	18.8	56.3	25	0	0	3.94	0.68
6	37.5	62.5	0	0	0	4.38	0.50
7	37.5	50	12.5	0	0	4.25	0.68
8	18.8	68.8	12.5	0	0	4.06	0.57
10	56.3	43.8	0	0	0	4.56	0.51
13	50	50	0	0	0	4.50	0.52

Table 3.1: Summarised result of the questions that uses a 5-point Likert scale

3.4.1 Summary of the Results

Demographic

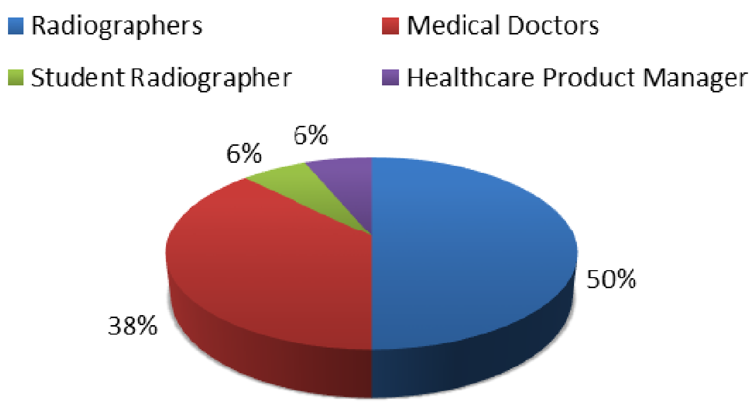


Figure 3.7: Respondents’ profession

Out of the thirty questionnaires circulated during the congress, there were sixteen responses from four types of professionals related to radiology (see Figure 3.7). From the sixteen respondents, eight were radiographers (50%), six were medical doctors (37.5%), one was a student radiographer (6.25%) and another one was a healthcare product manager (6.25%). The average age of the respondents was between 25 to 44 years old with 56.25% of all respondents spent more than 20 hours per week using healthcare or medical systems in their work.

Variations in Reporting Style

Generally, the respondents agreed that the variations in reporting styles do have its implications. For question number four, fifteen out of sixteen respondents (93.8%) agreed that variations in reporting styles impact report interpretation. Only one respondent was neutral regarding this. When asked if the variations in reporting styles affect patient diagnosis (question five), twelve respondents (75%) agreed while the other four (25%) were neutral about it. Following this, we can conclude that the respondents believed that the variations does impact report interpretation but not necessarily the patient diagnosis.

Structured Report Form Interface

The user interface of the structured report form consists of information vital to an ultrasound report such as the patient information, administrative information, clinical history, areas examined, findings and observations as well as conclusion and further management. Each of this information was grouped under one section to facilitate report reading. When asked if this allows for information to be read easily (question six), all sixteen of the respondents agreed with six strongly agreed (37.5%) and the other ten (62.5%) only agreed. The respondents were also asked whether they agree that the form contains enough information for an ultrasound report (question seven). Fourteen respondents (87.5%) gave a positive reponse while the other two (12.5%) have a neutral opinion. Fourteen out of sixteen respondents (87.5%) also agreed that the form was easy to fill (question eight). This means that generally, most respondents agreed that the interface of the structured report form was easy to read and fill in and contains enough information.

The questionnaire also includes an example of an ultrasound finding; “The liver exhibits appearances in keeping with patch fatty infiltration”. The respondents were asked to classify whether this finding falls under “abnormal” or “normal”. Only one respondent classified it as “normal” but under “normal variant” while seven other respondents classified it as “abnormal”. The other half of the respondents found it difficult to classify the finding as neither “normal” nor “abnormal”. One respondent commented that this finding should be worded better in order to be able to classify it. This showed that it is important for ultrasound reports to be standardised so that better interpretation could be made. One respondent suggested that another term other than “normal” or “abnormal” should be included in order to solve this problem of classifying the findings. This is a major issue that was brought up by most of the respondents which we will discuss further in Section 3.4.2.

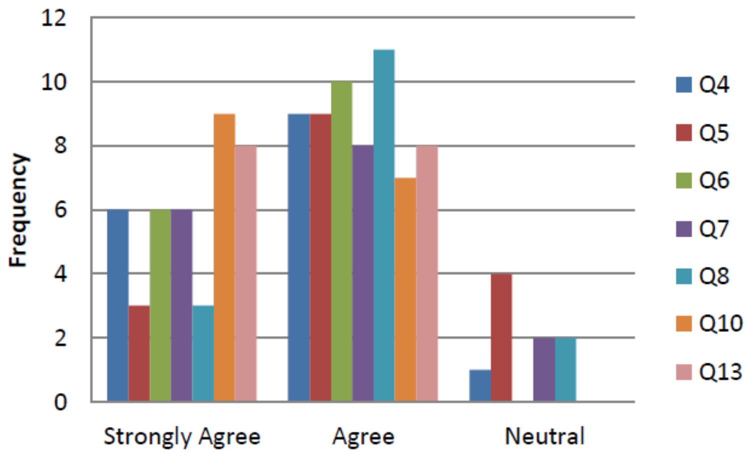


Figure 3.8: Frequency of the “strongly agree”, “agree” and “neutral” answers for each question

Question number ten asked the opinion of the respondents on how important is the flexibility of the system while question number thirteen asked whether the flexibility to write ultrasound report in both free form and structured form would attract practitioners to use the system. For both questions, all sixteen respondents agreed. The respondents felt that the flexibility to add or delete fields in the form according to the needs of the ultrasound report is very important. They also believed that the flexibility to write ultrasound reports in both free form and structured form would better attract practitioners to use the system. The respondents were also asked if they were likely to use the reporting system. However, this question was not applicable to the respondents because not all of them were practising abdominal ultrasound. Hence, we have decided not to discuss the result of this question. Figure 3.8 shows the frequency of the “strongly agree”, “agree” and “neutral” answers for each question.

Further Comments

Most of the respondents gave positive feedback on both the interface of the structured report form as well as the whole idea of the proposed ultrasound reporting system.

"User friendly". "Good and easy to follow". "Clearly sign-posted". "Very clear to edit or read". "Make reporting images more user friendly and simplistic". These comments showed that the respondents were happy with the look of the interface and found that it was easy to use. However, they also had some suggestions that could be looked into. "Good, maybe start with normal appearances first". This respondent suggested that in the findings and observations section, it would be better to present the normal findings first instead of abnormal. This is because the respondent believes that if the report presents the abnormal findings first, the person reading the report would be too focused with the abnormalities and would abandon the normal findings.

Another respondent suggested that there is a need for regional variances in the findings and observation section on top of the normal and abnormal findings. This is an issue that needs some attention which will be discussed in the next section. "System needs to be able to learn patterns and grow as more knowledge is added". This respondent suggested that some parts of the system could be filled automatically after some time of using it and be personalised according to the radiologist who is using it. This is an added value feature that could be included in the system later in the future. Other than these suggestions, the respondents were positive about the system with one respondent commented "As a breast radiologist I would be very keen to see this initiative applied to breast and auxiliary ultrasound".

3.4.2 Discussions

One main concern of the respondents was regarding the findings and observations section of the structured report as seen in Figure 3.9. A few respondents believed that it is better to display the normal findings first before abnormal findings because they believed that if practitioners read the report they would be very concerned about the abnormal findings and not pay much attention to the normal findings. A few other respondents felt that it is good that the structured report distinguishes normal and abnormal findings because it forces the radiologists to decide on their findings. However, they believed that there should be more options in reporting the findings and observations in the structured report form especially for special cases like normal variants. This is because, it would be inaccurate to classify normal variant as abnormal since it is actually normal for some people but classifying it as normal without indicating that it is a normal variant would confuse the referring clinician reading the report.

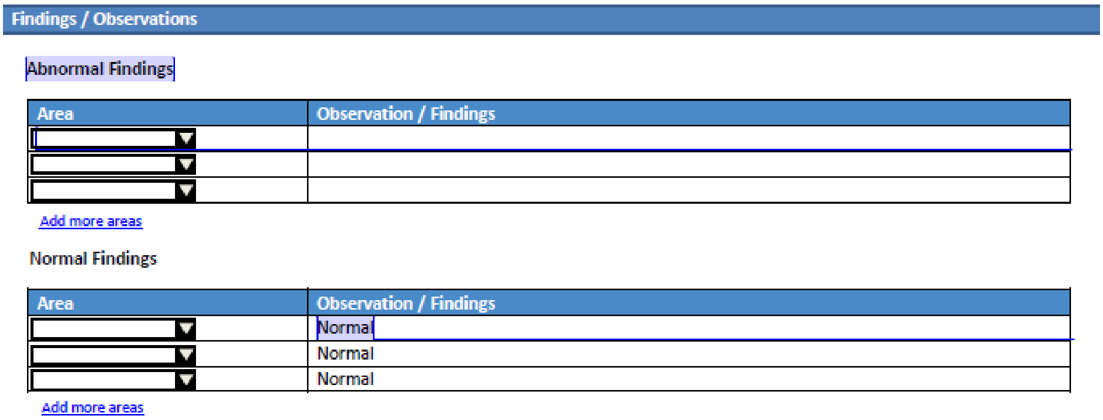


Figure 3.9: The original version of the findings / observations section

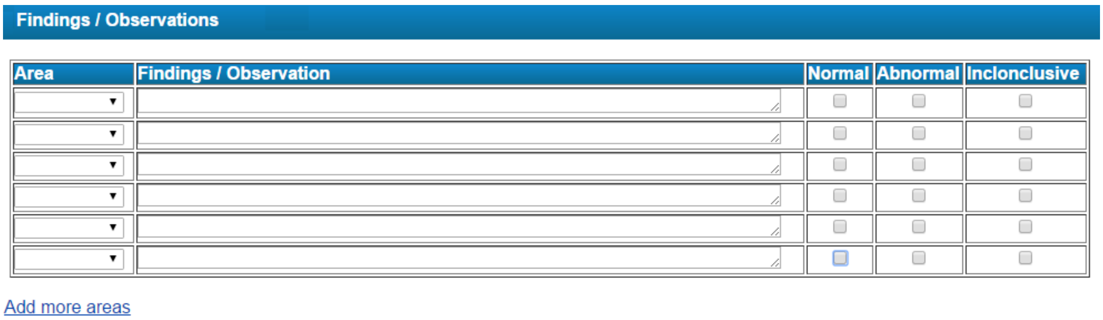


Figure 3.10: The revised version of the findings / observations section

Taking into account all of these responses, we have revised the findings / observations section of the structured report page to accommodate their suggestions (see Figure 3.10). In the revised version, the position of abnormal or normal findings will not be an issue anymore because it is now arranged according to the areas that have been observed. Radiologists will have an option to classify the findings as "normal", "abnormal" or "inconclusive". If they choose "normal", they can also indicate if it is a normal variation and if they choose "abnormal", they can indicate whether it is an incidental finding. The radiologists will also have the option to classify their findings as "inconclusive" for cases where the image was not clear or there were obstructions in examining the relevant areas. However, the radiologists will be required to state their reason for classifying their finding as inconclusive. This forces the radiologists to make decisions on their findings and at the same time giving them flexibility and more options in justifying their classification. It also allows for all findings or observations to be recorded in the report even if the areas are not clearly seen.

The high percentage of participants unable to classify the finding in question nine of this questionnaire as “normal” or “abnormal” proves that it is hard for referring clinicians to make a decision based on just a sentence or paragraph. Even though the sentence or paragraph has enough information such as the measurement of a finding, referring clinician would not be able to make any decision since measurements of organs or biospecimens can give different meanings for different patients. Thus, forcing the radiologists to state such findings as normal or abnormal will give meaning to their report and help the referring clinicians in making a decision in diagnosing the patients.

3.5 Chapter Summary

This chapter presents the system architecture model for the medical ultrasound reporting system. In this system architecture model, a solution was proposed where radiologists are allowed to choose any reporting style that they are most comfortable with and the system will automatically generate a standardised structured version of the report they wrote. The chapter explains each components of the system architecture model together with their purposes. The architecture model uses AUO as a knowledge based together with RST and discourse markers to transform free-form reports to structured form. This chapter also presented the result of the ultrasound reporting system’s interface preference survey. This survey was carried out to understand what would attract radiologists to consistently use an ultrasound reporting system. Overall, the radiologists were happy with the proposed system and the interface design of the structured report page. Several issues were raised during the survey especially on the appearance of the findings / observations section. The revised version of the section as a result of the discussion was also presented.

Chapter 4

The Knowledge Base for the Standardised Reporting System

4.1 Introduction

Biomedical systems are an integral part of today's medical world. Systems such as electronic patient records and clinical decision support systems (CDSS) have played an important role in assisting the works of medical personnel. One area that could benefit from the development of biomedical systems is ultrasound reporting. Variations in ultrasound reporting impacts the way a report is interpreted as well as in decision making. Therefore, it is important that these reports be standardised. In order to achieve this goal, ontologies are used to understand the reports and structure them according to a certain format [94]. They are also used to recognise the relationships between the parts of the text composing the report and to standardise the terminologies used.

Although there exist several biomedical ontologies that cover most of the general concepts in the domain, there are two disadvantages in adopting the whole ontology for reuse in a specific system. First, the ontology will have a lot more classes than required which means that the size will be large. Second, the coverage of the ontology might not be the most optimal since large ontologies have a general domain. In this research for example, using NCIT as its knowledge base would require a large amount of storage because of its size. NCIT contains as many as 118,941 classes and requires more time and effort to process. This would not be efficient since only a small portion of the ontology will be used. On the other hand, to build a new domain specific ontology from scratch will only cause redundancy and requires extra time and manpower. Hence, it is better to select and reuse relevant classes from several ontologies.

This chapter first reviews three established biomedical ontologies that are recognised to be potentially suitable for reuse which are FMA, RadLex and SNOMED CT. This is to assess the efficiency of reusing one of these ontologies in the medical ultrasound reporting system. After arguing against adopting only one whole ontology, we propose a methodology to reuse several biomedical ontologies together with the existing tools that can be used to ease the reuse process. In this chapter, we describe the development of the Abdominal Ultrasound Ontology (AUO) which will serve two purposes in this research: (i) it will be used to standardise the development of ultrasound reports and enforce the use of a standard terminology and (ii) to analyse the reports written in Natural Language (English free-text) with the aim of automatically transforming them into a structured format. Lastly, the result of using the ontology reuse methodology will be discussed. A large portion of this chapter has been published in the proceedings of the 21st International Conference on Application of Natural Language to Information Systems [95].

4.2 Review of Existing Biomedical Ontologies

The biomedical field is among the fields that have a number of existing ontologies because of the amount of words and terminologies being used in the field. To ensure whether it is necessary for us to reuse several ontologies instead of simply adopting an existing one, we need to review all the potential existing ontologies that can be applied in the medical ultrasound reporting system that we would develop. To be clear, when we mention adopting an ontology, we mean that the ontology was used in the system without making any changes to it. This is not to be confused with reusing an ontology where classes in the ontology were added to enhance its coverage or pruned to reduce its size.

After careful consideration, three ontologies have been selected which are FMA, SNOMED CT and RadLex. This selection was purely based on their coverage, language and popularity in the biomedical community. In selecting one ontology to be adopted we first look at the domain of each ontology. FMA covers the concepts and relationship in the structural organisation of the human body from the macrocellular to microscopic levels [73] while SNOMED CT covers more than that and also includes clinical findings, chemical substance scales and other miscellaneous health information [6]. RadLex on the other hand contains many complex radiology related domains from basic science to imaging technology [74]. This shows that both FMA

	FMA	SNOMED CT	RadLex
Number of classes	104,258	324,129	46,140
Number of individuals	0	0	46,635
Number of properties	172	152	96
Format	OWL / OBO	OWL	OWL

Table 4.1: Comparison between FMA, SNOMED CT and RadLex

and SNOMED CT is a broad and more general ontology compared to RadLex which is more domain specific in relation to developing an ontology that covers the abdominal ultrasound process.

To try and select only one ontology to adopt in the system, we then looked at the size of all three ontologies. Table 4.1 shows the comparisons of all three ontologies. This comparison showed that all three ontologies were too large to be implemented in one specific system which could then result in slow computing time and taking up a lot of space. Thus, we have decided to reuse only the relevant concepts from all three ontologies and merge them into one new ontology with the aim of achieving bigger coverage and less space consumption. To ensure that FMA, SNOMED CT and RadLex are the most suitable ontologies to be reused, further assessment needs to be completed and several more criteria need to be set up.

4.3 The Proposed Methodology

In developing a new ontology by reusing existing biomedical ones, proper planning and execution are important in order to ensure the modularity of the concepts reused. One important part in reusing ontologies is to select the suitable ones to be reused. In order to make sure that FMA, SNOMED CT and RadLex are the most suitable, we need to evaluate these ontologies by comparing them with our corpus. This section will present an ontology reuse methodology that will first extract terms from our corpus in order to find a list of recommended ontologies to be reused. This methodology adopts the general four steps mentioned in Section 2.3.3 which are summarised in Figure 4.1. The methodology presented in this research will allow for the development of a new ontology by reusing multiple existing ontologies and suggest tools that would help in each step of the methodology.

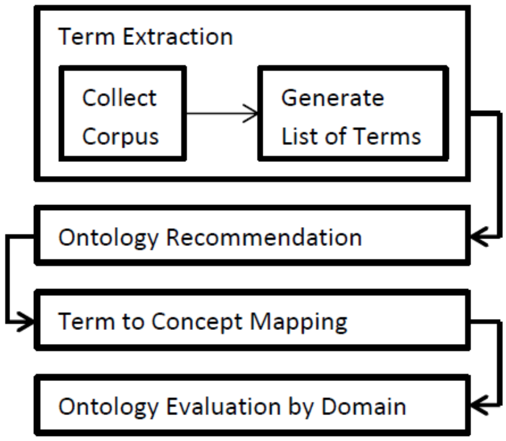


Figure 4.1: Ontology reuse methodology

4.3.1 Term Extraction

The first step in ontology reuse or even in developing an ontology from scratch is to decide on its scope and domain. In this case of developing a medical ultrasound reporting system, the scope and domain of the ontology that we are concerned with is abdominal ultrasound where it includes the taxonomy, pathology, equipment and other terms related to the domain. 100 sample abdominal ultrasound reports have been collected and used as the basis of our corpus. The scope and domain of AUO developed in this research were defined by the terms used in the 100 sample reports.

The first step that needs to be taken once we have our corpus is to extract all the relevant terms from it to generate a list of biomedical and technical terms that need to be in the ontology. During the extraction process, two biomedical extraction applications have been evaluated, TerMine[2] and BioTex[3] to select the most suitable application for this research. 49 out of the 100 sample reports were submitted to both applications and the results are shown in Table 4.2.

From this evaluation, BioTex was chosen as the biomedical term extraction application to be used because of its ability to extract more terms (761 terms) compared to TerMine (241 terms). BioTex is an automatic term recognition and extraction application that allows for both multi-word and single-word extraction [50]. TerMine on the other hand is only able to extract multi-word terms and not single-word terms. In this research it is important that the term extractor is able to extract not only multi-word but also single-word terms in order to obtain the best coverage.

[2] http://www.nactem.ac.uk/software/termine
[3] http://tubo.lirmm.fr/biotex/

	TerMine	BioTex
Language	English	English
License	Open	Open
POS Tagger	GENIA Tagger / Tree Tagger	Tree Tagger
Terms Found	241 (GENIA Tagger) 232 (Tree Tagger)	761
Extraction Type	Multi-word extraction	Multi-word and Single-word extraction

Table 4.2: Comparison of biomedical term extraction using TerMine and BioTex

For example, if the sentence "Unremarkable appearances of the liver with no intrahepatic lesions" was submitted to both applications, TerMine will only extract two multi-word terms "Unremarkable appearance" and "intrahepatic lesion" while BioTex will extract not only the two multi-word terms but also "liver" which is a single word term. If single-word terms such as "liver", "kidney" and "spleen" are not extracted, the ontology developed would be incomplete. Terms which were extracted from BioTex were also validated using the Unified Medical Language System (UMLS) [50] which is a set of documents containing health and biomedical vocabularies and standards. Using BioTex we managed to extract 1119 terms from the 100 sample ultrasound reports.

4.3.2 Ontology Recommendation

The next step after obtaining a list of terms for ontology reuse would be to select the suitable ontology to be reused. Three important criteria were used for selecting the ontology in this research:

1. Ontology coverage - To which extend does the ontology covers the terms extracted from the corpus?

2. Ontology acceptance - Is the ontology being accepted in the biomedical field and how often is it used?

3. Ontology language - Is the ontology written in OWL, OBO or other semantic languages?

Ontology coverage has the highest weightage when determining whether an ontology is suitable to be reused. This is so that we can identify one ontology that contains most of the concepts needed and preserve the model of the ontology so that it can be followed by concepts taken from other ontologies.

The next important criteria in determining whether an ontology is suitable to be reused is the acceptance of the ontology within the biomedical community. This criteria is also evaluated when choosing an ontology because the level of acceptance indirectly shows the quality of the ontology [56]. The higher the acceptance score, the more the ontology is being used thus promoting interoperability between different systems. Finally, the semantic language used to develop the ontology is also an important factor so that less time is needed because there is no need to translate the ontology to another semantic language.

Initial review resulted in choosing FMA, SNOMED CT and RadLex as suitable candidates because of their domain coverage, acceptance in the biomedical community and language which is OWL. In order to verify this, we have decided to use the ontology recommender provided by BioPortal which is an open ontology library that contains ontologies with domains that range from anatomy, phenotype and chemistry to experimental conditions [61] since all three ontologies to be investigated are available in BioPortal. The ontology recommender available on its portal gives suggestions on suitable ontologies to be reused based on the terms needed. It makes a decision based on the following three criteria [37]:

1. Coverage - Which ontology provides most coverage to the input text?
2. Connectivity - How often the ontology is mapped by other ontologies?
3. Size - How many concepts are there in the ontology?

Users are able to submit a paragraph or a list of terms into the recommender and it will give a recommendation of 25 ontologies which are ranked from the highest to the lowest scores (see Figure 4.2). The ranking of all the ontologies available in the BioPortal repository that meets all the three criteria is computed by giving scores to four metrics namely coverage, acceptance, knowledge detail and specialisation. The final score is calculated based on the following formula:

$$\begin{aligned} FinalScore = {} & (CoverageScore * 0.55) + (AcceptanceScore * 0.15) \\ & + (KnowledgeDetailScore * 0.15) + (SpecialisationScore * 0.15) \end{aligned} \tag{4.1}$$

The coverage score is given based on the number of terms in the input that are covered by the ontology. It is given the highest weightage which is 0.55 compared

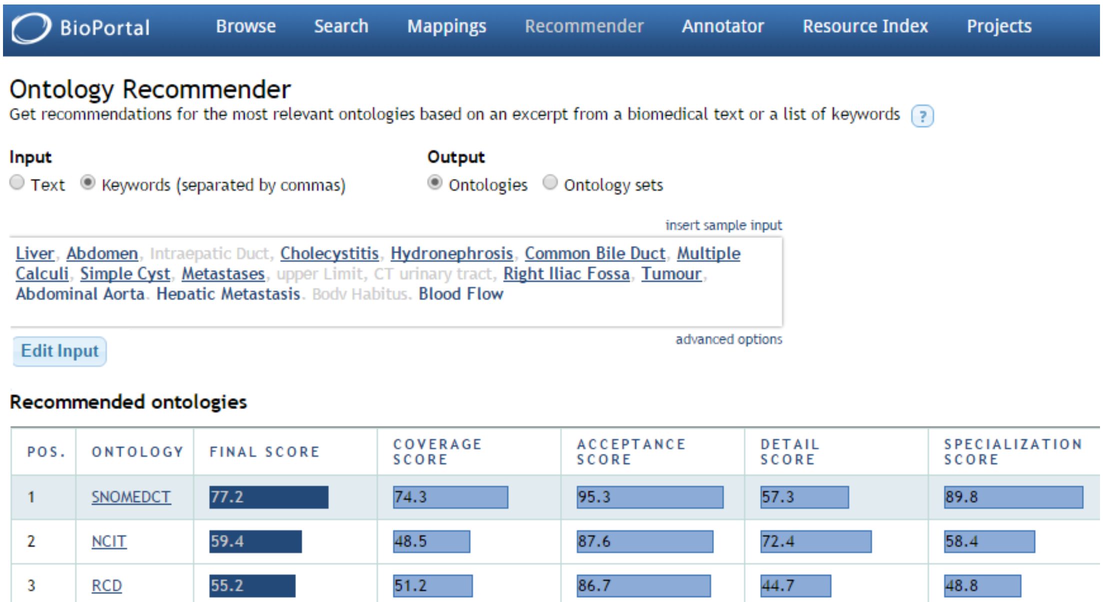


Figure 4.2: BioPortal's ontology recommender

to the other three metrics which are all given a weightage of 0.15 because ontology coverage is seen as the most important factor in determining the suitability of reusing an ontology for a certain corpus. The acceptance score indicates how well-known and trusted the ontology is in the biomedical field. It is given based on the total of website visits to the ontology as well as whether or not the ontology exist in UMLS.

Knowledge detail score on the other hand indicates the level of details in the ontology; i.e. does the ontology have definitions, synonym or other details. This is also an important metric in determining the ontology quality because an ontology that only has a hierarchy of terms and contains no other details such as definitions would merely be seen more as a taxonomy rather than an ontology. Thus, it would have less purpose as a knowledge base in a system or application. Lastly, the specialisation score is given based on how well the ontology covers the domain of the input. This is different from the coverage score because the size of the ontology is taken into consideration since the specialisation score ranks domain specific ontologies higher compared to general ontologies. In this research, the specialisation score is not as important as the coverage score since the aim is to have the highest coverage even though we would need to reuse more than one ontology. An example of this scoring system in action is shown in Figure 4.2 where 18 terms where submitted.

There is however a limitation in using it on the portal whereby only 500 words can be submitted. This limitation has prompted us to develop our own recommender system by manipulating the data from BioPortal's ontology recommender API [7] to accommodate the 1119 terms that have been extracted from our corpus. We first developed the recommender system that would submit all 1119 terms to BioPortal's API and give a recommendation of 25 ontologies ranked according to their final score exactly like the BioPortal's recommender. However, it seems that the 1119 terms were too big for the recommender's server to handle and causes it to crash midway without giving any result. Thus, we decided that the best way to deal with this is by getting the recommender system to submit all terms one by one to the API and get recommendations for each term instead. Algorithm 1 summarises the method used in doing this.

Algorithm 1: Getting ontology recommendations for term list

```
Input: A list of terms, T[n]
Output: 1. Ontology recommendation, O[n] and score, S[n] for each term, T[n]
        2. Ranking of ontology recommendation, O[n] with its frequency, F
foreach T[n] do
    submit URL = "http://data.bioontology.org/recommender?apikey=
      &input=" & T[n];
    if Result != Empty then
        return O[n] = 0;
    else
        return O[n] = acronym [0] and S[n] = evaluationScore[0];
    end
end
foreach O[n] do
    count F of each O[n];
    sort O[n] from highest F to lowest;
end
```

Figure 4.3 shows an excerpt of the result from processing 1119 terms using the recommender system we have developed. The recommender system allows us to submit quite a huge number of terms to be processed and the result will then be stored for analysis. This is seen as a better alternative because it is impossible to straightaway submit all 1119 terms and get a result since the API server would not be able to handle such a huge number. Figure 4.3(a) shows an excerpt of all the list of terms submitted and the ontology recommended for each term based on the final score they obtained. After all terms have been submitted to the recommender, the frequency of each ontology recommended will be counted and sorted from highest to lowest. The recommender system has ranked NCIT as the ontology with the highest frequency (476) followed by

(a)

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

(b)

1	Ontology Name	Frequency
2	NCIT	476
3	SNOMEDCT	207
4	RADLEX	48
5	LOINC	44
6	MESH	42
7	RCD	25
8	MEDDRA	16
9	OMIM	12
10	NIFSTD	10
11	CRISP	9
12	SWEET	9
13	CCO	7
14	DCM	7
15	SOPHARM	6

Figure 4.3: (a) Ontology recommendation for each term (b) Ranking of ontology recommended

SNOMED CT (207) and RadLex (48) as shown in Figure 4.3(b). This has proven that our initial selection of reusing FMA, SNOMED CT and RadLex was not very accurate since FMA does not even appear in the top 15 of the ontologies recommended for this corpus. The reason for this could be because FMA is a very large and broad ontology so it loses a lot of points in the specialisation score. Its broad domain also means that it covers only the general terms available in the corpus causing it to also loose points in the coverage score hence resulting in it not being recommended for reuse.

4.3.3 Term to Concept Mapping

Once the ontologies for reuse have been selected, the next step in building AUO is to map the terms extracted from the corpus to concepts available in the ontology. This was achieved by referring to the results acquired from BioPortal's Search API [7]. The API allows us to insert several parameters to perform concept search. In this research, the parameters used were "q" to specify the term that we would like to search for, and "ontologies" where it specifies the ontology in which we would like to look for the term. Once these parameters have been submitted, the API will return concepts that match the term. The concepts will be returned with several other properties such as the preferred label, definition, synonym, match type and the terms relationship with its children, descendants, parents and ancestors. If several concepts were returned, the concept that has the closest semantic meaning to the term submitted will be chosen.

In previous works by Mejino, Rubin and Brinkley [58] as well as Shah et al. [80], term to concept mapping was performed by going through all the concepts in an existing ontology and deleting irrelevant concepts to retain only the relevant ones. Missing concepts are then added to the ontology to make it complete. This rigorous work consumes a lot of time especially if the existing ontology is huge and it is also hard to reuse more than one ontology this way. BioPortal's Search API requires less time and works as the terms are queried according to the provided parameters. This means that we only need to query terms that are of interest instead of going through the whole ontology. This also ensures the accuracy of the relationship between the concepts and their children, descendants, parents and ancestors since there are links that can be clearly viewed in the API result.

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

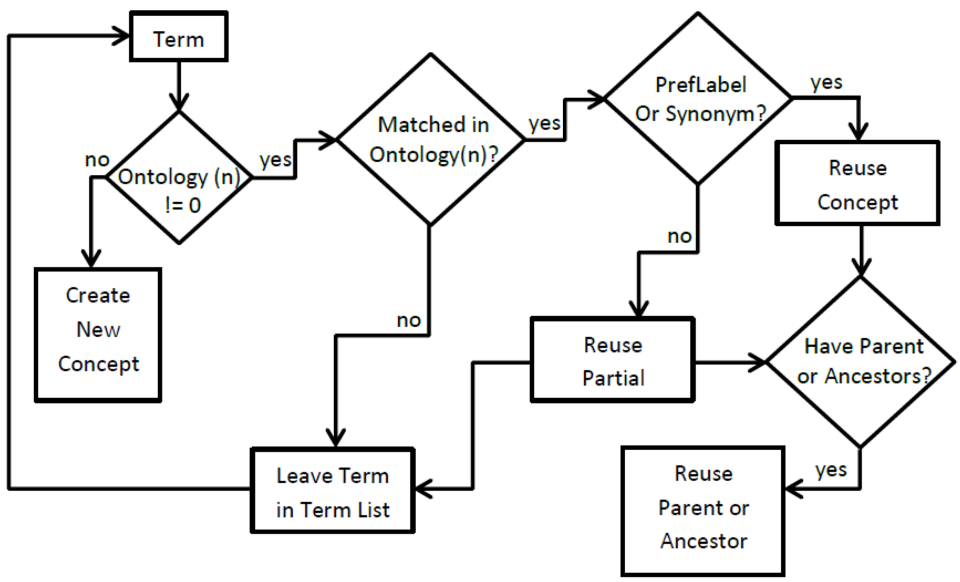


Figure 4.4: Term to concept mapping guide

In deciding whether a term should be reused or not, the term to concept mapping guide (see Figure 4.4) was referred to. First, a term from the list of terms extracted from the corpus will be queried using the Search API. The query will be performed on the ontology with the highest frequency first which in this case is NCIT. If there is a match, we will check whether the match is a preferred label, synonym or partial match. A preferred label (PrefLabel) match means that the API has found a concept that has an exact match to the term. Synonym match on the other hand indicates that the term is found as a synonym to the concept while partial match means that there is no exact match for the term but there are at least two concepts that match the term. If the match is a PrefLabel or synonym match, the concept will be reused and the term will be removed from the list. If the match is partial, the concepts that make up the term will also be reused. However, the term will still be remained in the list so that it can be compared to concepts in other ontologies.

After the concept has been reused, we will find out if the term has a parent or ancestors. Doran, Tamma and Iannone [20] in their work suggested that in order to maintain modularity of an ontology, a concept should be extracted together with its subclasses or children instead of parents and ancestors. They argued that immediate parents and ancestors are unimportant and extracting them would increase the risk of creating an ontology that is equal to the ontology being reused. However, we believe that parents and ancestors are important in connecting concepts so that they would not be floating. If we were to take "spleen" and its subclass as one module, "kidney" and its subclass as another module, as well as "millimeter" and its subclass as another module, it would be hard to group these modules under the same category. Furthermore, the ontology being developed is very specific to the abdominal ultrasound domain. Thus, reusing parents and ancestors of a concept reduce the risk of the ontology being as large as the original one. Once all terms have been searched, this process will then be repeated for the remaining recommended ontologies which are SNOMED CT and RadLex.

A walkthrough example: Consider a list of terms that contains three words which are "gallbladder", "duct dilation" and "gallstone". We first take the word "gallbladder" and query if the concept exist in NCIT. This returns an exact match where there exist a concept in NCIT with the preferred label "gallbladder". Hence, this concept together with all its knowledge details will be reused. We then check to see if the concept has any parents or ancestors. "Gallbladder" has a parent "organ" and an ancestor "anatomic structure, system, or substance" which will both be reused. Since "gallbladder" has an exact match in NCIT, it is removed from the term list.

We then query the second word "duct dilation" in NCIT which returns a partial match with the word "duct" and another word "dilation". Both concepts will be reused together with their knowledge details as well as their parents and ancestors. However, different to "gallbladder", "duct dilation" would not be removed from the term list since it is only a partial match. The final word in the list, "gallstone" is then queried which gives a synonym match to the concept "gallbladder stone" in NCIT. "Gallbladder stone" together with all its knowledge detail including synonyms will be reused. All its parents and ancestors will also be reused and the term will be removed from the term list. Once all terms have been queried in NCIT, we then check to see if there are any terms left in the term list. This give us the term "duct dilation" which is then queried to see whether there is any match with the second ontology recommended which is SNOMED CT. This returns a synonym match with the concept "dilation of duct". Thus, the concept "dilation of duct" will be reused and the term to concept mapping process is finally complete.

Katsumi and Gruninger [42] in their work discussed ontology modelling in ontology reuse where they stressed that for an ontology to be considered as reused, it should have at least a small fragment of the existing ontology. In determining which model an ontology should follow, Katsumi and Gruninger state that it depends on the strength of the existing ontology. If the existing ontology is weaker than the new ontology that is being developed, then the new ontology may only map some parts of the existing ontology's model. However, if the existing ontology is stronger than the new ontology being developed, then the new ontology may follow the model of the existing ontology. In the case of several ontologies being reused, Katsumi and Gruninger suggested that some parts of the new ontology should follow parts of the existing ontology being reused. They however, failed to define what are the criteria that classifies an ontology as stronger or weaker compared to another ontology.

In this research, three ontologies were reused namely NCIT, SNOMED CT and RadLex. Seeing that NCIT has the highest frequency by far (478), compared to the other two (207 and 48 respectively), we assumed than NCIT is stronger than SNOMED CT and RadLex. This has prompted us to follow the modelling of NCIT in the development of the Abdominal Ultrasound Ontology (AUO). When merging concepts that were reused from SNOMED CT and RadLex with the ontology developed by reusing NCIT, we would first find suitable parents for the concepts in the ontology. If no such parent exists, the parent and ancestors of the concepts will then be reused according to

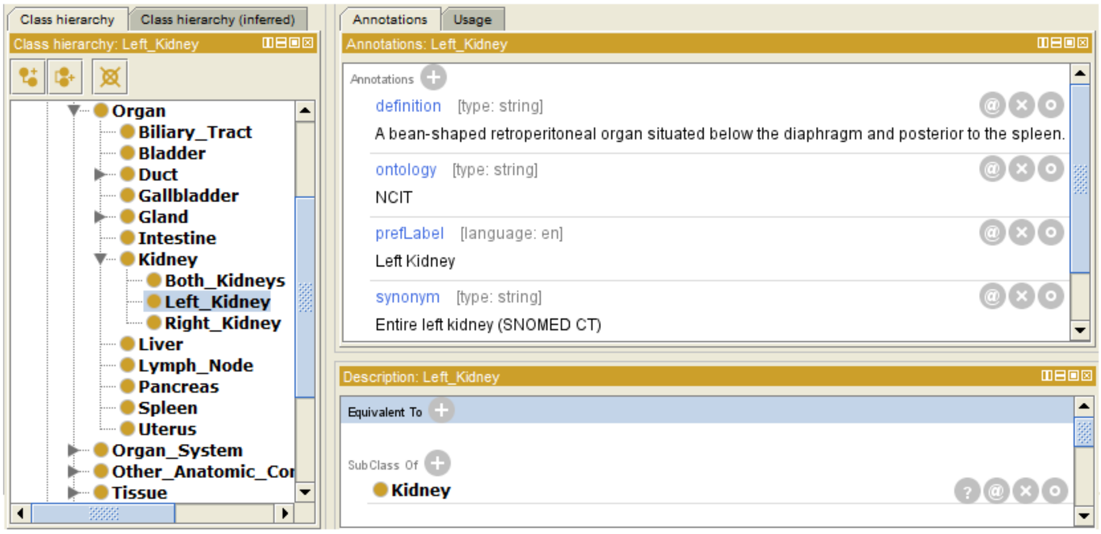


Figure 4.5: Snapshot of the Abdominal Ultrasound Ontology

the modelling of NCIT. If no match is found in any of these three ontologies, a new concept will then be created with the help of domain experts. The new concepts created will be carefully integrated into AUO and contain the same level of knowledge details as other concepts in AUO. The competence from the domain experts are needed at this level to give the correct definition and add relevant synonyms to these new concepts. All concepts in AUO were annotated with their original ontology for easy reference in the future if necessary. Figure 4.5 shows a snapshot of the complete Abdominal Ultrasound Ontology.

4.3.4 Ontology Evaluation by Domain Expert

Once a complete AUO has been developed using the ontology reuse methodology, it is important that the ontology be evaluated by a domain expert in order to verify that the relationship between the terms and their definitions are correct. In evaluating this ontology, we have sat down together with a domain expert and went through the whole ontology. There are some corrections that need to be done but overall, the domain expert believes that the 92.7% ontology coverage is enough to cover all the important concepts that an abdominal ultrasound report would need. For the other 7.3% terms that have no match in the ontology, some of it were caused by human error whereby spelling mistakes were made by the reporter. As for the rest of it, the domain expert will help in giving definitions and suggestions on where it would fit in the ontology.

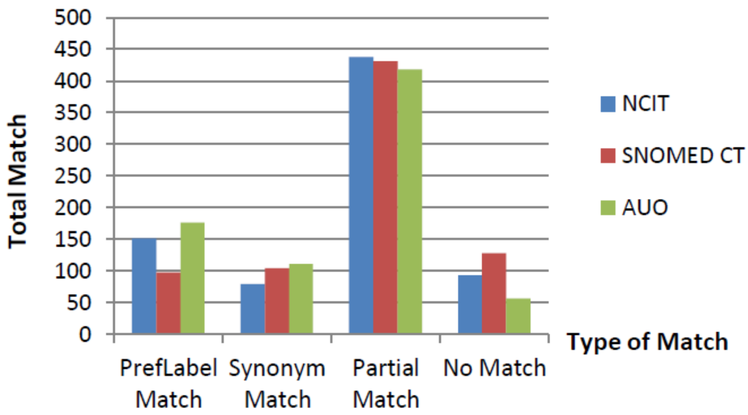
Out of the 7.3% terms that have no match in the ontology, there were also several terms that the domain expert believed we can omit because these words should not be in an ultrasound report for good practice. Examples of such words are “comet tail”, “NAD”, and “hepato petal”. These words might be understood by the radiologists but make no sense to others [22]. The main objective of using this ontology reuse methodology is to achieve as much coverage as possible and reduce the need for domain experts in developing the ontology. If an ontology were to be build from scratch, domain experts will be needed from the very first step in designing the ontology. However, with ontology reuse, domain experts are only needed at the end of the process to verify that everything is correct and to assist in adding new concepts into the ontology.

4.4 Results and Discussions

Phase 1: 49 Sample Ultrasound Reports

The coverage of AUO was first tested on a set of 49 sample ultrasound reports with a total of 761 terms in comparison with two existing ontologies which are NCIT and SNOMED CT. A term to concept matching was completed using the 761 terms extracted from the sample ultrasound report corpus and the result can be seen in Figure 4.6. AUO which was developed using the ontology reuse methodology that combines the three ontologies gave the highest number of concept matches compared to reusing only one ontology. Between NCIT and SNOMED CT, NCIT has the highest concept match total with 151 PrefLabel matches, 79 synonyms matches and 438 partial matches. SNOMED CT on the other hand has only 98 PrefLabel matches, 104 synonyms matches and 431 partial matches. The reason SNOMED CT has lower PrefLabel matches compared to synonyms is because of its naming convention. For example, the preferred label for “kidney” is “kidney structure” and “entire gallbladder” for “gallbladder”. When writing reports, radiologist often use simpler words like “kidney” and “gallbladder” instead of “kidney structure” and “entire gallbladder”. As a result of this, when the term to concept matching was performed, SNOMED CT returned more synonym matches compared to PrefLabel.

Compared to NCIT and SNOMED CT, AUO returns the highest total match where it has 176 PrefLabel matches, 111 synonym matches and 418 partial matches. The reason AUO returns the most number of matches is because the ontology reuse methodology selects the best match from different ontologies and merge it into the AUO. Its



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

Figure 4.6: Breakdown of total match according to type against NCIT, SNOMED CT and AUO

exhaustive mapping in several ontologies based on the ontology rank has ensured that almost all terms in the corpus are covered by AUO. Whenever possible, a PrefLabel match will be inserted in the ontology. If not, either a synonym match or partial match will be included to ensure the ontology has a wide coverage of the corpus.

From the analysis, it can be concluded that it is better to reuse from several ontologies compared to just one. This is because reusing several ontologies offers better term coverage compared to reusing just one. Figure 4.7 shows the percentage of total match and no match in all three ontologies. If ontology reuse was performed by mapping the 761 terms against NCIT, there will only be an 87.8% of coverage. If the mapping were performed against SNOMED CT, the percentage of coverage will be only 83.2% which is lower than NCIT. However, the percentage of coverage increases to 92.6% when several ontologies were reused; in this case are NCIT, SNOMED CT, and RadLex. The percentage of no match is also very small (7.4%) indicating that AUO covers almost all the terms in the corpus. The reason there is still 7.4% of no match is because there are several term in the corpus that the domain experts believed are poor usage of terms to describe findings in an ultrasound report. The domain expert believed that this is a bad practice and the medical ultrasound experts are now slowly cutting

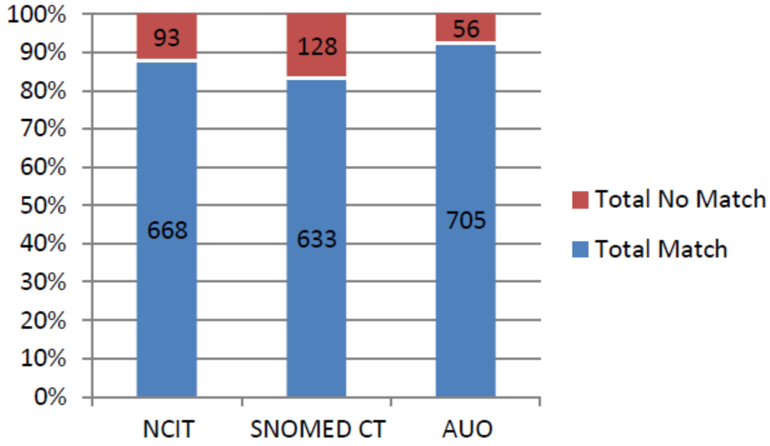


Figure 4.7: Percentage of total match and no match in NCIT, SNOMED CT and AUO

down the usage of such words thus making it irrelevant to be in the AUO. Another reason for the 7.4% of no match is spelling errors made by ultrasound reporters. This is not a concern for now but for future work, we could consider using the ontology to also correct and understand these errors.

NCIT has a total of 118,941 classes while SNOMED CT has 324,129 classes. However, there are only 668 and 633 matches respectively for each NCIT and SNOMED CT regarding abdominal ultrasound terminology. On the other hand, AUO has only 509 classes which is less than 0.5% of either NCIT or SNOMED CT but still managed to have 705 matches which is more than the matches NCIT and SNOMED CT each gets. This is because of the specialisation of the ontology. Since the ontology has an intended purpose in an application, it is much better and more efficient to build a domain specific ontology through reuse. It definitely would not be efficient to store a large ontology such as NCIT and SNOMED CT and use only less than 0.6% of it. This is because it would take a lot of storage space and also slows down the application since it will need to go through the whole ontology to find a match. Therefore, the better way to develop an ontology-based application is to build a new domain specific ontology through ontology reuse methodology.

Phase 2: 100 Sample Ultrasound Reports

Phase 1 of the testing and evaluation process has proven that ontology reuse provides a wider coverage compared to using only one existing ontology. However, this testing was done in comparison to the 49 sample ultrasound reports corpus which was also used as the training data in developing AUO. Phase 2 looks at how well AUO performs in relation to the term to concept mapping when being evaluated using 100 sample reports; where 49 of the sample reports were used for training purpose and the other 51 were a new sample of ultrasound reports which have not been used before in developing AUO. Phase 2 also looks at how much work is needed to update the ontology to include additional concepts based on the requirements of the new corpus.

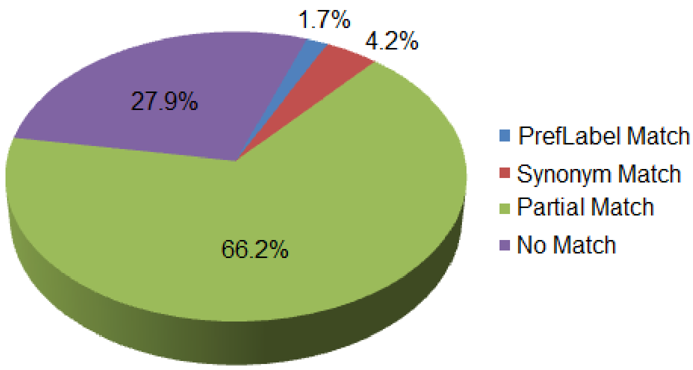


Figure 4.8: Percentage of total prefLabel, synonym, partial and no match for the 358 new terms when compared to AUO

Terms extraction which was executed on the 100 sample reports returned 1119 terms which is 358 more than the terms extracted in Phase 1. The number of matches for the 761 terms has already been found out in Phase 1. To investigate the number of matches for the rest of the terms, we checked to see whether the remaining 358 terms exist in AUO. Figure 4.8 shows the percentage of each type of matches as well as no match for all the 358 terms when compared with AUO.

Overall, there is a quite high percentage of total match which is 72.1% where 237 out of the 358 terms were found as partial match (66.2%), 15 as synonym match (4.2%) and six as prefLabel match (1.7%). This shows that the AUO is adequate to cover most terms in the biomedical domain since the total of no match is quite small which is 27.9%. In abdominal ultrasound reporting, the content of the reports usually contains words which are more or less similar in semantic even though different terms were used especially in the case of normal ultrasound reports. Only in abnormal ultrasound reports we could find a higher variety of words being used. With the existence of AUO as a knowledge base, the usage of different words in writing ultrasound reports would not have mattered because the semantic of the words are more important than the words themselves. Because of the similarity of the contents, a small number of the corpus used as training data still gives quite a good number of total match percentage.

After comparing the 358 terms with AUO, we found that there are 100 terms without a match in AUO. A good methodology should be able to allow for it to adapt to any changes when required. Thus, in order to reduce the percentage of no match, the ontology reuse process was repeated on all 358 terms to include new concepts into AUO. First, we need to check whether the addition of 51 new sample reports changes the ontology recommended for reuse. Fortunately, NCIT still comes up as the best ontology for reuse followed by SNOMED CT and RadLex. Next, all 358 terms were then compared with NCIT so that the term to concept mapping could be carried out. Once the process was finished, it was then repeated with SNOMED CT and RadLex. After the addition of the new concepts and synonyms such as "adnexal masses", "incidental finding" and "morphology" to AUO, the total number of terms without a match was reduced from 100 to 26, which is a 74% reduction. Figure 4.9 shows the percentage of total match of all three types of matches as well as the percentage of no match for the 1119 terms after new concepts have been added to AUO.

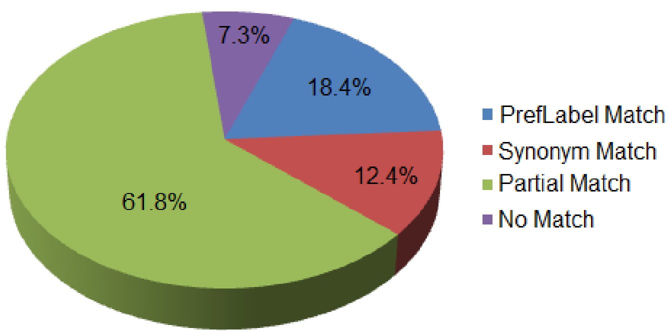


Figure 4.9: Percentage of total prefLabel, synonym, partial and no match for all 1119 terms when compared to AUO

The total match percentage after the addition of new concepts has increased from 72.1% to 92.7% where 692 out of the 1119 terms were found as partial match (61.8%), 139 as synonym match (12.4%) and 206 as prefLabel match (18.4%). Before the addition of the new concepts, the percentage of partial match was higher which is 66.2% compared to the current percentage of 61.8%. This number was reduced after the term to concept mapping was repeated because some of the partial matches were found as either a synonym or prefLabel match in NCIT, SNOMED CT or RadLex. For example, the word “right ovary” was first labelled as partial match since AUO does not have a concept called “right ovary” but it has two concepts which are “right” and “ovary”. However, after performing term to concept mapping again, “right ovary” was found as a prefLabel match in NCIT thus the concept was added to AUO. This causes the percentage of synonym and prefLabel match to also increase from 4.2% and 1.7% to 12.4% and 18.4% respectively.

Phase 2 of the testing and evaluation of AUO showed that the number of sample ultrasound reports used as testing data was adequate in ensuring that AUO is able to provide coverage for most abdominal ultrasound sample reports. This also proved that the domain expert’s assumption that the 92.7% total matches of the first version of AUO is indeed enough to cover most abdominal ultrasound sample reports. It also showed that the ontology reuse methodology proposed in this research is adaptable since it is fairly easy for new concepts to be added according to new requirements from additional sample reports.

4.5 Chapter Summary

This chapter first reviewed three existing biomedical ontologies that have the potential to be used in the reporting system which are FMA, SNOMED CT and RadLex. Following this, it is found that adopting one whole ontology in a specific system is inefficient because of its size and coverage. Therefore, this chapter proposed a methodology to reuse ontology together with supporting tools that would make the process much easier. Furthermore, the development of AUO using this methodology was explained and evaluated. It is proven that ontology reuse is favourable in developing a small domain specific ontology that has a wide coverage of the domain compared to using a large and general domain ontology. Finally, the evaluation of the proposed methodology in this chapter proved that it is adaptable because it allows for changes to be made easily whenever there are any new requirements from the corpus. AUO developed in this research will serve as a knowledge base for the medical ultrasound reporting system where it will standardised the terminologies used in the report and support the transformation of free-form report to structured form.

Chapter 5

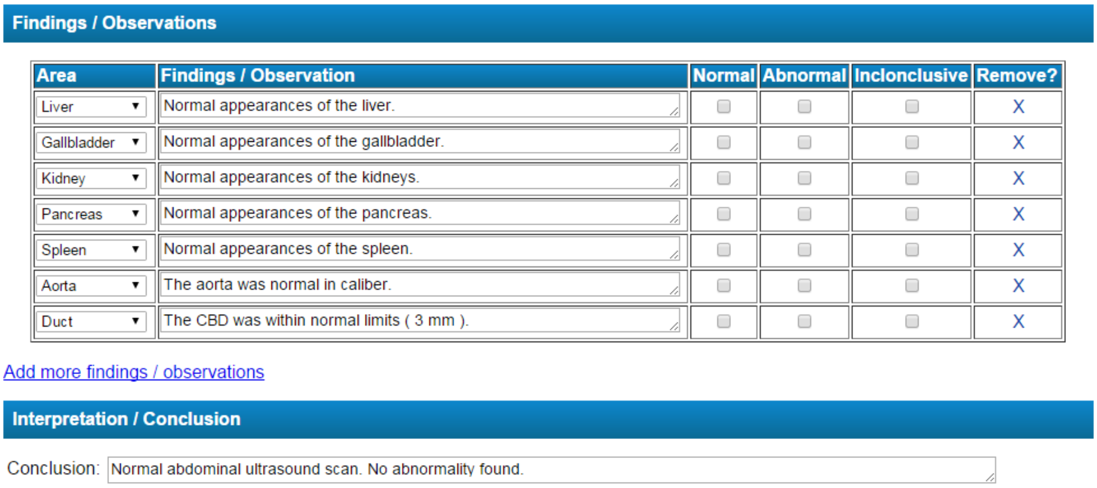
Structured Report Generation

5.1 Introduction

Although it was acknowledged that standardisation in the form of structured reporting will improve the quality of ultrasound reports, many radiologists still opt to write their reports in free-form because of its familiarity and simplicity. To give a better chance for the adoption of structured reporting, we argue that those who prefer to use free-form reports should be allowed to do so. However, there should be a mechanism in the ultrasound reporting system that will transform these reports into a structured form. Following this, we investigate the possibility of employing RST in performing the transformation. RST has the ability to segment texts into smaller text spans and recognise the rhetorical relations that exists between them. This allows for the content of the reports to be dissociated and regrouped under suitable headings.

A lot of researches have been using RST in parsing texts such as newspaper articles, personal letters, political newsletters and journal abstracts. However, as far as we are concerned, there have not been any researches that have used RST to parse ultrasound reports as yet. Therefore, we would like to explore this possibility. RST uses discourse markers and sentence structure as indicators to segment texts and identify rhetorical relations. Since ultrasound reports are often not grammatically correct and written in abbreviations, we proposed a novel rule-based method to segment texts and identify rhetorical relations using AUO where the reports will be annotated with relevant classes in the ontology. This method will then be applied in the ultrasound reporting system to transform free-form reports to structured form. Figure 5.1 shows an output example of the ultrasound reporting system where it transforms an excerpt of a free-form report to a structured form as follows:

> "US Abdomen: Normal appearances of the liver, gallbladder, kidneys, pancreas and spleen. The aorta was normal in caliber. The CBD was within normal limits (3mm). Main or principal diagnosis: Normal abdominal ultrasound scan. Conclusion: No abnormality found."



Findings / Observations

Area	Findings / Observation	Normal	Abnormal	Inclonclusive	Remove?
Liver	Normal appearances of the liver.				X
Gallbladder	Normal appearances of the gallbladder.				X
Kidney	Normal appearances of the kidneys.				X
Pancreas	Normal appearances of the pancreas.				X
Spleen	Normal appearances of the spleen.				X
Aorta	The aorta was normal in caliber.				X
Duct	The CBD was within normal limits (3 mm).				X

Add more findings / observations

Interpretation / Conclusion

Conclusion: Normal abdominal ultrasound scan. No abnormality found.

Figure 5.1: Example of a structured report generated from a free-form report

This chapter will first give a brief introduction on classic RST and how discourse parsing was performed using it. Then, it will explain on how this theory could be adapted for use in medical reports. The chapter then presents the rules of seven rhetorical relations that have been recognised in the 100 sample ultrasound reports that were reviewed. It will also explain the development of a medical discourse parser that implements these rules together with RST and AUO. Finally, it will discuss the result of implementing RST on sample ultrasound reports by comparing the automatic parsing of the system to the manual parsing completed by natural language processing experts. The development of a medical discourse parser presented in this chapter will serve as the basis of transforming free-form reports to structured form in the medical ultrasound reporting system.

5.2 Classic RST Discourse Parsing

Among all the existing linguistic theories, RST has been shown to be effective in many computational linguistic applications [87]. RST allows for large texts to be broken into smaller text spans where each individual segments has their own role to play in ensuring the coherency of a text. This coherency is ensured by the rhetorical relations that exist between each of these non-overlapping text identified during the text analysis process. There are two subtasks to be performed when analysing texts using RST. These subtasks are the text segmentation process and rhetorical relations identification process.

The first task requires texts to be segmented into text spans or discourse units where each unit has an independent functional integrity [52]. Classic RST segments texts according to its sentence structure as well as using cue words and punctuation. Marcu [54] for example segments texts into EDUs by using discourse markers as an indicator to recognise the possible rhetorical relations and set boundaries to perform segmentation. Mann and Taboada [93] also used the same method in their work where they segment a sentence using the discourse marker "and" to signal a JOINT relation as seen in Figure 5.2. A similar approach was also taken by Carlson, Marcu and Okurowski [17] where texts are segmented into EDUs using lexical and syntactic clues as well as parts of speech to help determine the boundaries.

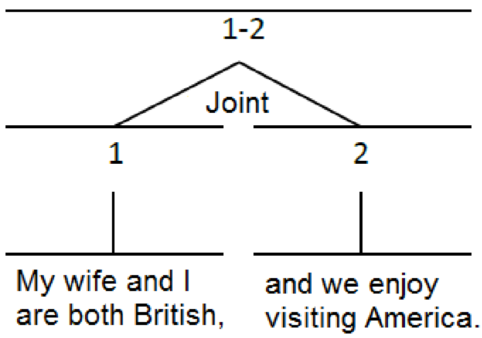


Figure 5.2: JOINT relation signalled by "and" [93].

Using discourse markers, lexical clues, syntactic clues and parts of speech clues to segment text in an ultrasound report is a challenge since most ultrasound reports were not written in complete sentences. Sentences written in the reports are often short and straight forward, thus making it hard for parts of speech rules to be applied in segmenting them. In order to apply RST in ultrasound reports, some modification needs to be applied in the approach. Instead of using lexical and syntactic clues as well as part of speech clues to segment text in ultrasound report, this chapter discusses the possibility of using an ontology, which in this case is AUO, in addition to discourse markers such as "but", "otherwise" and "or" to segment a text. Correct segmentation of texts is important in developing a quality RST tree as mentioned by Soricut and Marcu [86] whereby correct segmentation will ease the process of identifying relevant relations between the segmented texts.

This comes to the second subtask of analysing texts using RST which is to identify rhetorical relations between segmented text. In classic RST, 30 rhetorical relations were defined by Mann and Taboada [93] such as EVIDENCE, ELABORATION, JOINT and CONCESSION relations. In previous works [17, 52, 54], these relations were recognised using discourse markers as well as POS tags where each word in the text is assigned with its sentence parts such as nouns, verbs and adjectives. Even though discourse markers helped in giving indication on the type of relations that exists between two text spans, not all relations are signalled with discourse markers [93]. This is more pertinent in ultrasound reports where words used in the reports are limited and often contained medical terms. Since ultrasound reports are often not written in proper syntactic structure, rhetorical relations identification is much more difficult using clues from POS tags. Thus, this chapter explores the possibility of identifying these relations using a rule-based approach that applies discourse markers together with classes from AUO.

5.3 Text Analysis of the Sample Ultrasound Reports

A text analysis was conducted on all 100 sample ultrasound reports to understand their characteristics by looking at the average word count, average sentence count, average maximum sentence length, average minimum sentence length and average token size. All 100 reports have an average word count of 70.82, average sentence length of 11.15 and average token size of 6.45. The average maximum sentence length was 14.63 which indicates that in each report, the longest sentence has at most 14 words. The average minimum sentence length on the other hand was 1.34 which means that the shortest sentence in each report has at least one word. An example of such short sentence is the sentence “Conclusion:”.

Out of the 100 reports, 60 reports were randomly selected as training data to recognise text span boundaries as well as rhetorical relation that exists between these text spans. These training data have an average word count of 62.88, average sentence length of 9.15 and average token size of 6.48. Their average maximum sentence length was 12.48 while their average minimum sentence length was 1.53. On the contrary, the remaining 40 reports were used as testing data to evaluate the performance of both

	Average Word Count	Average Sentence Count	Average Max Sentence Lenght	Average Min Sentence Lenght	Average Token Size
Training Data	62.88	9.15	14.28	1.53	6.48
Testing Data	82.73	14.15	15.12	1.07	6.41
Both Data	70.82	11.15	14.63	1.34	6.45

Table 5.1: Summary of the text analysis of all 100 sample ultrasound reports

the segmentation and the rhetorical relation process. These testing data have an average word count of 82.73, average sentence length of 14.15 and average token size of 6.41. They have an average maximum sentence length of 15.12 and average minimum sentence length of 1.07. The summary of the statistics is presented in Table 5.1.

As a comparison, the training data mostly have fewer words per report as compared to the testing data. This means that the training data is generally shorter compared to the testing data because it contains less sentences per report as well as less words per sentence. Even so, the selection of the reports in both types of data was purely random.

5.4 Rules in Identifying Rhetorical Relations in Ultrasound Reports

The 60 training data reports were parsed to recognise their structure and how the texts in these reports can be segmented into text spans. Discourse parsing, which is a process of identifying discourse relations between discourse units [27] was performed on these reports where they were first annotated with relevant classes from AUO before they were segmented and drawn into RST trees and rhetorical relations were recognised between their text spans. From the discourse parsing, seven rhetorical relations were recognised which are PREPARATION, RESTATEMENT, JUSTIFY, ELABORATION, LIST, JOINT and CONTRAST relations based on the definitions by Mann [52].

PREPARATION, RESTATEMENT, JUSTIFY and ELABORATION relations are all mononuclear while LIST, JOINT and CONTRAST relations are all multinuclear. Table 5.2 and 5.3 show the summary of the definitions of all seven relations as defined by Mann [52] where S is the satellite, N is the nucleus, R is the reader and W is the writer. These relations have been validated by an ultrasound expert that we were working with to ensure their relevance in ultrasound reports. These reports were then analysed again together with the relations that have been identified to design a set of rules that are able to automatically identify all seven relations using discourse markers and relevant classes from AUO. In the following subsections, we will describe how these relations were identified and what they represent and in the later sections, we will assess the accuracy of these rules when tested using the remaining 40 sample ultrasound reports.

Relation Name	Constraints on either S or N individually	Constraints on N + S	Intention of W
Preparation	None	S precedes N in the text; S tends to make R more ready, interested or oriented for reading N	R is more ready, interested or oriented for reading N
Restatement	None	on N + S: S restates N, where S and N are of comparable bulk; N is more central to W's purposes than S is	R recognises S as a restatement of N
Justify	None	R's comprehending S increases R's readiness to accept W's right to present N	R's readiness to accept W's right to present N is increased
Elaboration	None	S presents additional detail about the situation or some element of subject matter which is presented in N or inferentially accessible in N	None

Table 5.2: Definitions of mononuclear relations - PREPARATION, JUSTIFY, and ELABORATION relation [52]

Relation Name	Constraints on each pair of N	Intention of W
List	An item comparable to others linked to it by the List relation	R recognises the comparability of linked items
Joint	None	None
Contrast	No more than two nuclei; the situations in these two nuclei are (a) comprehended as the same in many respects (b) comprehended as differing in a few respects and (c) compared with respect to one or more of these differences	R recognises the comparability and the difference(s) yielded by the comparison is being made

Table 5.3: Definitions of multinuclear relations LIST, JOINT and CONTRAST relation [52]

5.4.1 Preparation Relation

PREPARATION relation is one of the seven relations identified in the sample ultrasound reports. It is a mononuclear relation where PREPARATION relation serves as a precedence in order to prepare the readers on what they are about the read. In a normal text or paragraph, one example of a PREPARATION relation is between the title and the rest of the text where the rest of the text is the nucleus and the title is the satellite. This is the same in the case of ultrasound reports where 89 PREPARATION relations have been recognised where each title prepares the audience for the content. From the analysis of all 60 sample ultrasound reports, 87 out of 89 PREPARATION relations that have been recognised in the reports started with a title and were followed by colons (:). Consider the example below:

S1. US Abdomen :

S2. Normal liver echo pattern with no focal lesion demonstrated.

S3. No evidence of gall stones or dilatation of the bile ducts.

S4. Both kidneys are normal in size and echo pattern with no mass lesion or evidence of obstruction.

S5. Conclusion:

S6. Normal examination.

In this example, it is clear that S1 is the title of the report which enables the readers to know that they would be reading a report on an ultrasound examination that was conducted on the abdominal area. S1 has a PREPARATION relation with a list of findings which are S2 until S4. S5 on the other hand has a PREPARATION relation with S6 where it prepares the readers for the conclusion. Therefore, a rule can be stated that if a certain text is followed by a semicolon then it has a PREPARATION relation with all the texts following it until there is another PREPARATION relation or it is the end of the paragraph.

5.4.2 Restatement Relation

Another type of relation that exists in the 100 sample ultrasound reports is the RESTATEMENT relation. This relation is a mononuclear relation where the writer re-expresses a sentence using another sentence. The RESTATEMENT relation was found in only 11 out of the 100 acquired sample ultrasound reports. This is often signalled by the appearance of a "main or principal diagnosis" title together with a "conclusion" title in an ultrasound report. Consider the example below:

S1. US Abdomen :

S2. Normal appearances of the liver, gallbladder, kidneys, pancreas and spleen.

S3. The aorta was normal in caliber.

S4. The CBD was within normal limits (3mm).

S5. Main or principal diagnosis:

S6. Normal abdominal ultrasound scan.

S7. Conclusion:

S8. No abnormality found.

The ultrasound report above details the list of findings of an abdominal ultrasound with S2, S3 and S4. It then reports the main finding in S6 using S5 as the title that prepares the readers for it. The report then gives a conclusion of the report in S8. It can be perceived that in this report, the "main or principal diagnosis" is actually already the conclusion of all the findings listed above it which is a "normal abdominal ultrasound scan". This statement is then repeated with a "conclusion" of "no abnormality found". Thus, it can be inferred that the "conclusion" restates the "main or principal diagnosis" in the report. Most reports that have been acquired have the same pattern when both the “main or principal diagnosis’ and “conclusion” were recorded in the report. If one of these titles is absent, then the RESTATEMENT relation will not hold.

5.4.3 Justify Relation

The JUSTIFY relation allows a reader to accept the nucleus based on the justification given by the satellite. In the context of an ultrasound report, the JUSTIFY relation gives reason as to why there is such finding. In identifying the JUSTIFY relation, there are two rules that can be applied which will denote whether the finding is normal or abnormal. The rules are stated as below:

(i) Normal Findings
negative + biospecimen / disease /disorder / finding $\xrightarrow{justify}$
organ / body part + positive

(i) Abnormal Findings
biospecimen / finding + organ / body part $\xrightarrow{justify}$ disease / disorder

In the first rule, if a text span contains a negative word (no, without etc.) and a word annotated with the class “biospecimen”, “disorder”, “disease” or “finding” followed by or preceded by another text span that contains a word annotated with “organ” or “body part” and a positive word (normal, unremarkable etc.) then it denotes a JUSTIFY relation in a normal finding. Consider the example of a finding below:

TS1. Liver is normal in echo pattern

TS2. with no focal lesion.

In this example, the word “no” shows the absence of “lesion”; which is annotated with the class “finding” in TS2. This justifies the “normal” (positive word) condition of the “liver”; which is annotated with the class “organ” in TS1. Figure 5.3 illustrates the RST tree for this relation.

The second rule for a JUSTIFY relation denotes an abnormal finding. This rule states that if a text span contains a word annotated with the class “biospecimen” or “finding” and a word annotated with the class “organ” or “body part” followed by or preceded by a text span with a word annotated with “disorder” or “disease”, then there is a JUSTIFY relation. Consider the example given below:

TS1. Subtle dilatation of the collecting system of the left kidney,

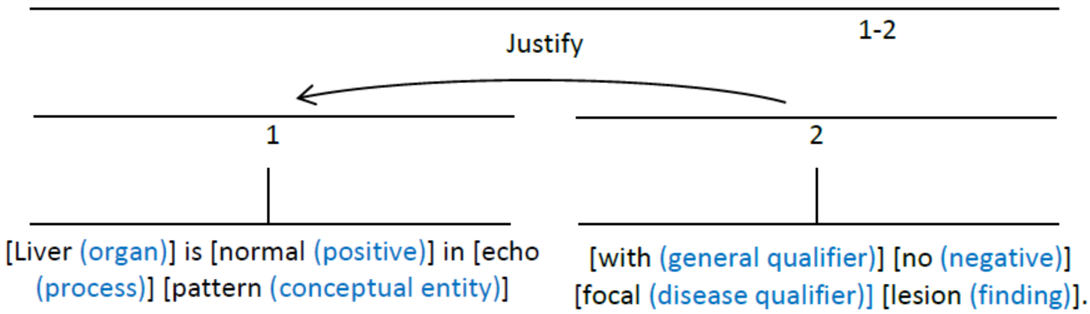


Figure 5.3: Example of a JUSTIFY relation in a normal finding

TS2. hydronephrosis grade 1.

In this example, the “dilatation” of the “left kidney” in TS1 which was annotated with “finding” and “organ” respectively justifies the “disorder” found in TS2 which is “hydronephrosis” as seen in Figure 5.4. Out of all 60 sample ultrasound reports, there were a total of 72 JUSTIFY relations found where 48 of them follow these two rules. This number includes both direct JUSTIFY relations as well as JUSTIFY relations that were nested into either themselves or other relations. Another 24 JUSTIFY relations found in the reports were the relation between a list of findings and the conclusion of the report where the list of findings justifies the conclusion. This is for instance in Subsection 5.4.1 where the findings in S2 to S4 justified the conclusion in S6.

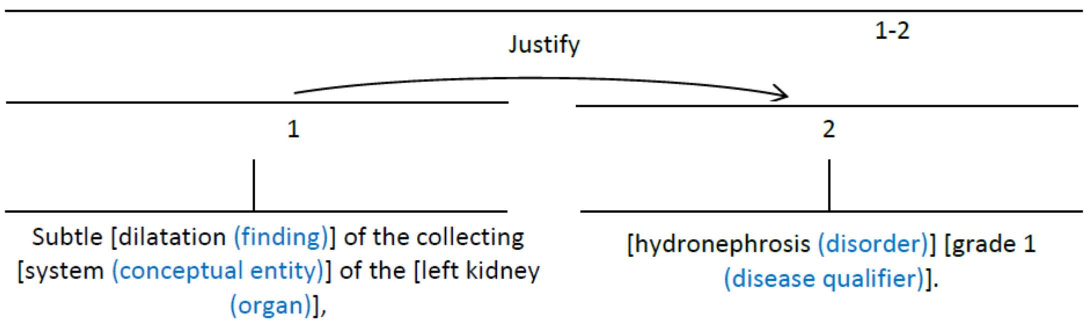


Figure 5.4: Example of a JUSTIFY relation in an abormal finding

5.4.4 Elaboration Relation

The ELABORATION relation gives further information on the text span. Unlike other RST relations, ELABORATION relation has many subtypes as it can be found in many text spans without being signalled by any discourse markers [2]. There are several subtypes of ELABORATION relation that could exist in an ultrasound report where each has different signals or cue words and gives elaboration on different aspects of the text span.

However, there are three prominent subtypes that have been identified in the 60 sample ultrasound reports used as training data. The first subtype is an ELABORATION relation that gives further information on the location of a finding. This is often signalled by a spatial qualifier such as the word "within". This type of ELABORATION relation is represented by the relation rule below:

(i) Spatial Qualifier
spatial qualifier + organ / body part $\xrightarrow{elaboration}$ biospecimen / finding

In this rule it is stated that a combination of a spatial qualifier and an organ or body part elaborates a biospecimen or a finding by letting us know the location of the biospecimen or finding. To understand this better, consider the following example:

TS1. No gall stones are seen

TS2. within the gallbladder.

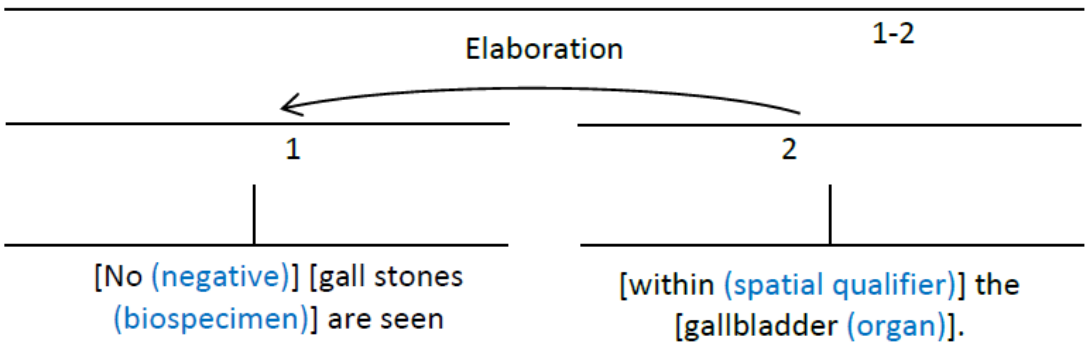


Figure 5.5: Example of an ELABORATION relation with spatial qualifier

In this example, we can see from the RST tree in Figure 5.5 that TS1 can be understood by the reader without TS2. However, TS2 elaborates further on the finding in TS1 by letting the reader know the exact location of where the "gall stones" were found. The second subtype of ELABORATION relation can be defined by the rule as follows:

(ii) Unit of Measure

unit of measure $\xrightarrow{elaboration}$ biospecimen / finding / organ / body part

This subtype of ELABORATION relation is the easiest one to recognise and gives further information on the measurement of the “finding” or the “organ”. It is often signalled by words such as “measures” and “measuring” or parenthesis that contains a measurement or simply a unit of measure such as “mm” and “cm”. Consider the following example:

TS1. The gallbladder contains a single stone

TS2. measuring 7mm.

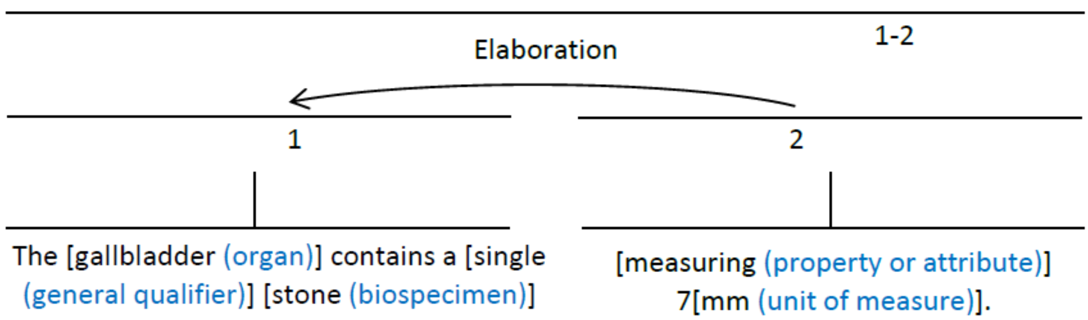


Figure 5.6: Example of an ELABORATION relation with unit of measure

Just like the example before, TS1 can be understood without TS2 where readers would know that there is a “single stone” in the “gallbladder” (see Figure 5.6). TS2 however extends the information on TS1 by giving the measurement of the stone. The last subtype of ELABORATION relation is when a “finding” is elaborated by another “finding” as stated by the rule below:

(iii) Finding within a finding

biospecimen / finding $\xrightarrow{elaboration}$ biospecimen / finding + organ / body part

Consider the following example as illustrated in the RST tree in Figure 5.7:

TS1. Multiple small gall stones at GB neck

TS2. with no signs of inflammation.

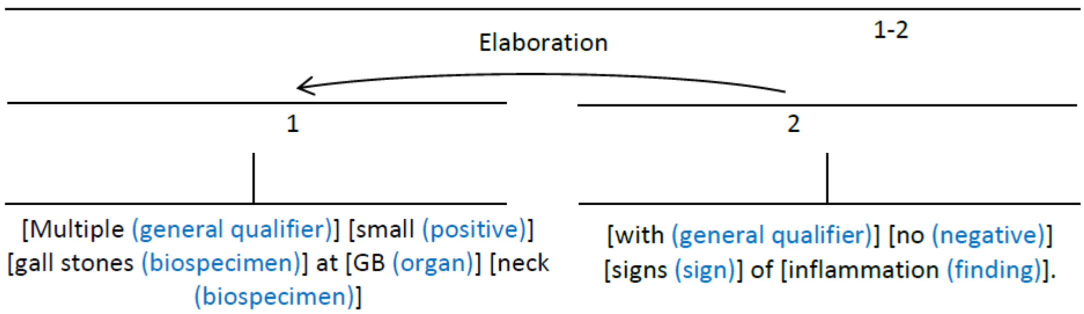


Figure 5.7: Example of an ELABORATION relation where a finding elaborates another finding

In this example, TS1 reports a “finding” of “multiple small gall stones at GB neck”. Instead of elaborating on where the “gall stones” are or the size of it, this subtype of ELABORATION relation describes another finding in relation to the finding in TS1. Instead of giving the reason for the finding in TS1, TS2 further elaborates the finding with another finding which makes it different from the JUSTIFY relation explained in the previous section.

5.4.5 List Relation

The LIST relation is a multinuclear relation which gives a list of items to the readers. In a multinuclear relation, each text span is a nucleus and plays an important role. There are no specific rules in identifying this relation. However, in ultrasound reports, LIST relation usually starts after the title. Consider a sample ultrasound report below:

S1. US Abdomen :

S2. There were multiple tiny calculi in the neck of the gallbladder.

S3. The CBD appeared normal.

S4. The pancreas was obscured by gas.

S5. No abnormality was seen in relation to the spleen or kidneys.

S6. There was no ascites.

S7. Conclusion:

S8. Tiny calculi within the neck of the gallbladder.

In this example, S1 provides the title of the report which is “US Abdomen”. The LIST relation usually comes after this title which in this example is S2 until S6. This relation gives the list of findings in the abdominal ultrasound report. S8 however does not have a LIST relation with S7 since it is single, the only relation that exists between S7 and S8 is that S7 prepares the reader for the conclusion which is given in S8.

5.4.6 Joint Relation

The JOINT relation is another multinuclear relation that can be found in an ultrasound report. This relation is important in ensuring that when a text is segmented, it does not lose its meaning. The JOINT relation was found 157 times which makes it the most used relation in this research. JOINT relation is often signalled by “AND” or “OR”. However, not all text span segmented with “AND” or “OR” denotes a JOINT relation. Some examples of the rules for identifying the JOINT relation are:

$$\text{organ} \xleftrightarrow{\textit{joint}} \text{organ + finding}$$
$$\text{biospecimen} \xleftrightarrow{\textit{joint}} \text{biospecimen + organ}$$
$$\text{anatomy qualifier} \xleftrightarrow{\textit{joint}} \text{anatomy qualifier}$$

Consider the following example:

TS1. There were multiple calculi

TS2. and sludge within the gallbladder.

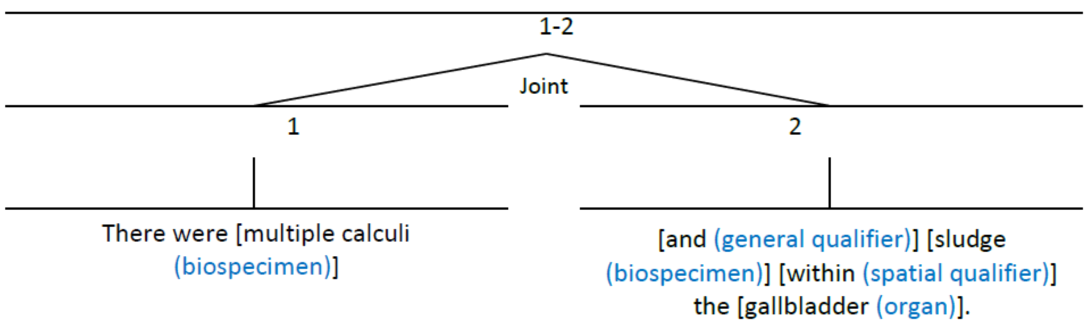


Figure 5.8: Example of a JOINT Relation

In the example above (see Figure 5.8), TS1 contains a “biospecimen” but without an “organ” whereas TS2 contains a “biospecimen” together with an “organ”. This denotes a JOINT relation where the “organ” in TS2 is jointly referred to by the “biospecimen” in both TS1 and TS2. With the JOINT relation, it is possible to assert that there are two findings in the sentence which are “There were multiple calculi within the gallbladder” and “There were sludge within the gallbladder” without losing any important information. Most sentences that contain the discourse marker “AND” and “OR” signals JOINT relation unless it follows the rule given below:

TS1: organ + finding
TS2: organ + finding

If each text span has a pair of “organ” and “finding” or “organ” and “biospecimen”, then it does not have a JOINT relation since both text span have enough information without needing to share any other information from the other text span. This is instead a LIST relation. Consider the following example:

TS1. The bile ducts were not dilated

TS2. and the liver texture appears satisfactory.

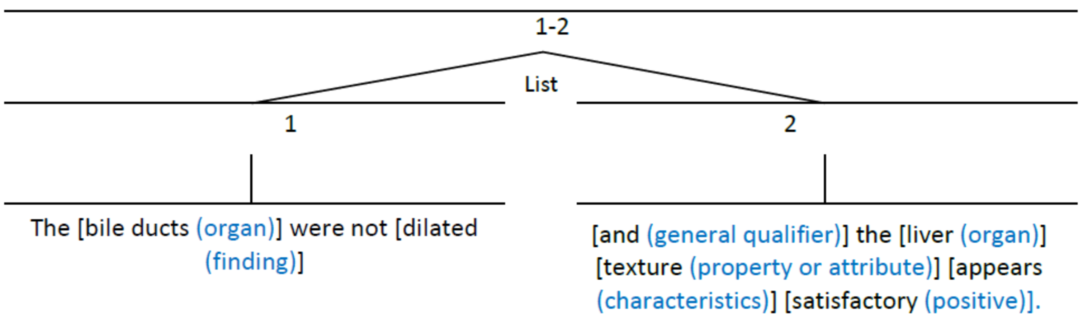


Figure 5.9: Example of the occurence of AND/OR which does not signals a JOINT Relation

In TS1, the “bile duct” is an “organ” and “dilated” is a “finding” while in TS2, the “liver” is an “organ” and “satisfactory” is a “positive finding”. Since each text span has its own pairs of organ and finding, thus this is not a JOINT relation but only a list of findings (see Figure 5.9).

5.4.7 Contrast Relation

The CONTRAST relation is another rhetorical relation that can be found in ultrasound reports. However, the frequency of use of this relation is small as only 7 relations were found. The CONTRAST relation is often signalled by the word “but” and “otherwise” and it gives contradicting findings. Consider the following example:

TS1. The liver shows fatty change

TS2. but no focal lesion.

TS1 and TS2 shows a CONTRAST relation with the existence of the word “but”. This relation is different from the JOINT relation because in JOINT relation the findings are not contradicting.

All the examples provided in this section demonstrated that it is possible to identify rhetorical relations using a rule-based approach that applies ontology. These rules were generated from the set of 60 sample ultrasound reports used as training data. The next section will explain further on how these rules can be used together with the ontology to perform discourse parsing on medical reports.

5.5 Applying RST and Ontology in Medical Discourse Parsing

Discourse parsing is a task whereby texts are segmented based on certain word boundaries in order to find rhetorical relations between them which signals the coherency of the text. Bärenfänger et al. considered discourse parsing as an iterative task whereby annotated texts are submitted as an input that produces an output text with another annotation [2]. Traditionally, discourse parsing was carried out using texts that have been annotated with POS tags whereby these tags helps in identifying the relations between text spans based on syntactic structures. This method is common in discourse parsing using RST [35, 38, 86] on corpus related to newspaper and magazine articles as well as academic journals. Bärenfänger et al. [2] however, tried implementing discourse parsing in a different approach where in their work, two OWL ontologies were used to perform certain tasks of discourse parsing. In this research, we aim to use a similar approach whereby an ontology was used instead of POS tags or treebanks in annotating texts as well as in implementing the rhetorical rules defined in previous sections.

5.5.1 Annotating Relevant Classes

The first step in executing discourse parsing using RST and ontology is to annotate the reports with relevant classes in AUO. Not all words will be annotated since we are only interested with classes that are involved in the rhetorical relation rules. Punctuations such as full stops (.), colons (:), and commas (,) were first separated from the text before splitting the paragraph into single texts. Algorithm 2 details the method of annotating an ultrasound report with ontology classes using XPath in PHP.

Algorithm 2: Annotating an ultrasound report with classes in AUO

```
Input: Paragraphs of ultrasound report, P
Output: Annotated version of P, A
Seperate P into single words, T[n];
Load results array, R[n] with ontology classes and synonym using XPath;
foreach n = 0; n <count(T[n])-2; n++ do
    Combine three words and store as array T3[n];
    foreach R[n] do
        if R[n] == T3[n] then
            Annotate T[n], T[n+1] and T[n+2] with relevant parent of matched class;
        end
    end
end
foreach n = 0; n <count(T[n])-1; n++ do
    Combine two words and store as array T2[n];
    foreach R[n] do
        if R[n] == T2[n] then
            Annotate T[n] and T[n+1] with relevant parent of matched class;
        end
    end
end
foreach n = 0; n <count(T[n]); n++ do
    foreach R[n] do
        if R[n] == T[n] then
            Annotate T[n] with relevant parent of matched class;
        end
    end
end
implode T[n] as A;
return A;
```

The classes in the ontology have at most three words combination. Therefore, once the paragraph has been split into single words, the next step is to combine the single words into three words combination and it will then be compared with the classes in AUO to see if there is a match. Once a match is found, the text will then be annotated with the class or its parent or ancestors that are relevant to the rules. The following sample ultrasound report will be used throughout this section as an example to explain the implementation of RST and ontology in medical discourse parsing:

> "US Abdomen : There were multiple tiny calculi in the neck of the gallbladder. The CBD appeared normal. The pancreas was obscured by gas. No abnormality was seen in relation to the spleen or kidneys. There was no ascites. Conclusion: Tiny calculi within the neck of the gallbladder."

"[US Abdomen | #Imaging_Technique|]: There were multiple tiny [calculi | #Biospecimen|] in the [neck | #Body_Region|] of the [gallbladder | #Organ|]. The [CBD | #Organ|] [appeared | #Characteristic|] [normal | #Clinical_or_Research_Assessment_Answer|]. The [pancreas | #Organ|] was [obscured | #Visibility_Descriptor|] by gas. No [abnormality | #Finding|] was seen in relation to the [spleen | #Organ|] or [kidneys | #Organ|]. There was no [ascites | #Disease_or_Disorder|]. Conclusion: Tiny [calculi | #Biospecimen|] [within | #Spatial_Qualifier|] the [neck | #Body_Region|] of the [gallbladder | #Organ|]."

Figure 5.10: Sample ultrasound report that has been annotated with relevant classes in AUO

The report will first be annotated by the parser with relevant classes from AUO as shown in Figure 5.10. This can be exemplified using the word "CBD" where the direct parent of "CBD" is "Extrahepatic Bile Duct". However, because we are more interested in knowing the general class of "CBD", it is annotated with the class "Organ" instead, which is the ancestor of both "CBD" and "Extrahepatic Bile Duct'. This annotation makes it possible to compare the text spans in the report with the rhetorical relations rules to identify the relations that exist. The same process will then be repeated for two words and one word combination to find the exact match of the words or their synonyms. Once the matching process has been completed, the annotated single words will then be combined to produce a complete annotated paragraph which will then be returned to the system for further processing.

5.5.2 Segmenting Ultrasound Reports

The next step in parsing the ultrasound reports would be to segment the annotated paragraph syntactically. Soricut and Marcu [86] defined discourse segmentation as a process where texts are split into non-overlapping text spans where each have their own roles in rhetorical relations. The segmentation task presented in this research loosely follows the work of Marcu [53] where he developed a shallow analyser that uses discourse markers instead of traditional POS tagging technique to segment texts into smaller text spans. He argued that knowing the usage of discourse markers, whether it is sentential or not, is enough to determine the boundaries of texts and detect its rhetorical relations with the assumption that the texts are well-formed [53].

Punctuation / Signal Words	Corresponding Rhetorical Relation
Fullstop (.) / Question Mark (?)	End of Sentence
Colon (:)	PREPARATION, LIST
Comma (,)	JOINT, JUSTIFY, ELABORATION
Parenthesis	ELABORATION, JUSTIFY
with / compatible with / in keeping with / associated with	JUSTIFY
suggest / suggestive of	JUSTIFY
could be	JUSTIFY
measuring / measuring with / measures	ELABORATION
however / but / otherwise	CONTRAST
and / or	JOINT
which / within	ELABORATION

Table 5.4: Punctuation and signal words together with the rhetorical relations they often signal

In our work, the same technique was used where segmentation was conducted based on punctuations and signal words which act as the text span boundaries. However, in segmenting ultrasound reports, we assumed that most texts in an ultrasound report are not well-formed. This is based on the review performed on the 100 sample reports that have been collected. Further details on the review will be discussed in Section 5.6.1. Most reports were written in short but straight to the point sentences at the cost of sentence structures being sometimes disregarded. Even so, this technique can still be applied in our work because the rhetorical relation rules presented in previous sections serve as a constraint in determining the relations between the segmented text spans. Table 5.4 states the list of punctuations and signal words that have been

Segmented Text

TS1. [US Abdomen | #Imaging_Technique|] :

TS2. There were multiple tiny [calculi | #Biospecimen|] in the [neck | #Body_Region|] of the [gallbladder | #Organ|] .

TS3. The [CBD | #Organ|] [appeared | #Characteristic|] [normal | #Clinical_or_Research_Assessment_Answer|] .

TS4. The [pancreas | #Organ|] was [obscured | #Visibility_Descriptor|] by gas .

TS5. No [abnormality | #Finding|] was seen in relation to the [spleen | #Organ|]

TS6. or [kidneys | #Organ|] .

TS7. There was no [ascites | #Disease_or_Disorder|] .

TS8. Conclusion :

TS9. Tiny [calculi | #Biospecimen|]

TS10. [within | #Spatial_Qualifier|] the [neck | #Body_Region|] of the [gallbladder | #Organ|] .

Figure 5.11: Sample ultrasound report that has been annotated and segmented

recognised from the 60 sample ultrasound reports used for training. It also gives the corresponding relation that these punctuations and words often signal.

The process of segmenting paragraphs using this technique is quite straightforward. The paragraph will first be parsed to detect the occurrence of punctuation that follows another punctuation (e.g. ?.), a signal word that follows another signal word (e.g. compatible with, measure within), a punctuation that follows a signal word (e.g. ,and) or vice versa. If there are any, a dash (-) symbol will be inserted between the words as a flag so that over segmentation does not happen. The parser will then try to locate a punctuation and set a boundary after the punctuation so that it can be split. Next, it will try to locate a signal word and a boundary will be placed before the signal word for it to be split. Figure 5.11 shows an example of a segmented and annotated ultrasound report. In the example, it is demostrated that in TS1, the text span boundaries have been set after the colon (:) symbol whereas in TS6, the boundaries of the text span was before the signal word “or”.

5.5.3 Identifying Rhetorical Relations

After the texts have been segmented, the next task in discourse parsing is to identify relations that exist between these text spans. Marcu [54] performed this by hypothesising the relations based on the appearances of discourse markers. If a discourse marker is absent, the co-occurrence of similarity will then be measured [54]. In this research, discourse markers were also used in identifying possible rhetorical relations that exist between text spans. The difference is that this research applied a rule-based approach whereby the text spans need to follow a set of rules as explained in Section 5.4. Discourse markers are sometimes ambiguous to which rhetorical relation it is signalling [54], hence the rule-based approach was proposed to reduce this ambiguity. Even though discourse markers were highly used in this work, it does not depend on them entirely. Discourse markers signals rhetorical relation, however, the ontology classes and relation rules confirm it.

Discourse Relations

TS1 PREPARE LIST (TS2-TS7)
TS5 JOINT TS6
(TS1-TS7) JUSTIFY (TS8 PREPARE TS9-TS10)
TS10 ELABORATE TS9

Figure 5.12: List of relations identified using ontology and the rhetorical relation rules

In identifying rhetorical relations between text spans, our discourse parser takes the annotated and segmented text spans as an input and outputs a list of all possible relations between these text spans. Figure 5.12 shows the output of the discourse parser where discourse relations have been identified for the ultrasound report that were used in this section as an example. The text spans will be parsed one by one and the parser will first look for the existence of any PREPARATION relation which is often signalled by colons (:). Almost all sample ultrasound reports that were used as training and testing have a title as its first text span. Most reports have between one to two titles that prepare the readers to understand its content and whenever there is a conclusion, it is often the last title in a report. The discourse parser takes particular attention on conclusions because they do not only have a PREPARATION relation with the text spans that comes after it but it also has a JUSTIFY relation with the text spans before it. This is because all the findings stated before the conclusion justifies it. This is clearly demonstrated in the complete RST tree for the example ultrasound report built using O'Donnell's RST Tool [63] (see Figure 5.13) where TS1 - TS7 JUSTIFY TS8 - TS10.

Once the PREPARATION relation has been recognised, the parser will then identify LIST relation which is all the text spans that comes after the title up until before the next title or until the last text span, whichever comes first. Then, the parser will look for the JOINT relation which is often signalled by "and", "or" and comma (,). Marcu [53] in his work has considered "and" as a highly ambiguous discourse marker. Even in our

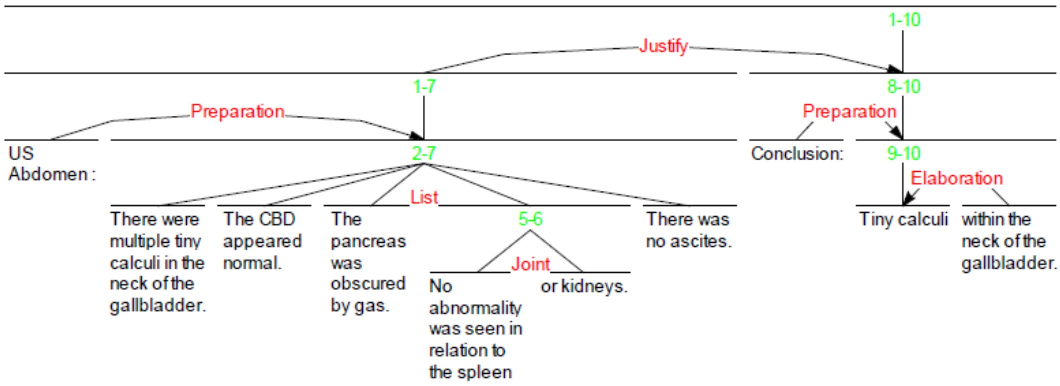


Figure 5.13: RST tree of the sample ultrasound report

work, all three “and”, “or” and comma (,) discourse markers were the most common signals whereby there were at least 165 occurrences. However, not all occurrences signal a JOINT relation as mention in Section 5.4. Hence, before a JOINT relation can be identified, the parser will first need to recognise the occurrences of “and” and “or” which are not a JOINT relation before attempting to identify the JOINT relations. Finally, the discourse parser will search for the remaining three relations which are JUSTIFY, ELABORATION and CONTRAST.

In most cases, a text span has a rhetorical relation with a text span immediately before or after it. A text span with a mononuclear relation often only have a relation with either a text span before or after it. This is unless it has a nested relationship whereby a text span has a relation with an immediate text before it and both of them have a relationship with another text span. The following sentence that has been segmented into three text spans can be used as an example to illustrates this:

> “[The gallbladder is well distended]$_{\mathrm{TS1}}$ [and contains at least 4 stones]$_{\mathrm{TS2}}$ [measuring just over 1 cm in diameter.]$_{\mathrm{TS3}}$”

In this case, TS2 has an ELABORATION relation with a text span right after it, TS3 where TS3 elaborates TS2. Both TS2 and TS3 then have a nested relation with TS1 where TS1 JOINT (TS3 ELABORATE TS2). As for a text span with a multinuclear relation, it is common for the text span to have the same relation with both text spans before and after it. From the training and testing work completed on all the sample ultrasound reports, we did not come across a text span, say TS1 skipped another text span, say TS2 and has a relation with TS3. The evaluation of both the segmentation and rhetorical relation processes will be presented in the next section.

5.6 Results and Discussions

This section presents the results of applying RST and AUO in the discourse parsing of 100 sample ultrasound reports. Out of the 100 reports, 60 were used for training data while the remaining 40 were used for testing and evaluation. It will first introduce the pre-processing phase and explain what was accomplished during this phase and how it impacts on the overall result of the parsing. It then presents the result of the report segmentation and identification of rhetorical relations where the system parsing was compared to the manual parsing performed by natural language processing experts. For the training data, the manual parsing was completed by a single expert. However, for the testing data, three experts were asked to manually parse all 40 sample ultrasound reports. The parsing which has a two third majority were then selected to be compared with the system parsing.

5.6.1 Pre-processing Phase

In the initial stage, RST and AUO were applied to the ultrasound reports without any cleaning or pre-processing phase. After the annotation and segmentation stages were performed, the accuracy rate was found to be quite low. The reason was found to be caused by human errors such as spelling mistakes, missing punctuations, missing spaces between words as well as abbreviations and symbols that the ontology does not recognise. The number of occurences of each error found in both training and testing data is summarised in Table 5.5.

Types of Human Errors	No of Occurrences in Training Data	No of Occurrences in Testing Data
Spelling mistakes	3	4
Missing or wrong punctuations	7	4
Missing spaces	1	5
Unrecognised abbreviations and symbols	1	5
Incorrect sentences	1	5
Total	13	23

Table 5.5: Summary of the human errors found in all 100 sample ultrasound reports

Even though the total number of occurrences of these errors were not alarmingly high, it still had a huge impact on the evaluation result. Therefore, we have decided to include a pre-processing phase before submitting the ultrasound reports to the system for it to be parsed. During the pre-processing phase, we have manually corrected spelling mistakes as well as adding spaces between words and punctuations such as full stops and commas in places we believed appropriate. For example, the sentence "The spleen the kidney and pancreas is normal" should be segmented into three text spans which are "the spleen", "the kidney" and "and pancreas is normal". However, because there is no comma in between "the spleen" and "the kidney", the system failed to segment it.

There were also several instances of incorrect sentences where they were incomplete or two sentences were combined without using any conjunctions. An example of an incomplete sentence is "The spleen is of normal size and echotexture, measuring ...". The radiologist writing the report has an intention to give the measurement of the organ but somehow did not. In this case, a full stop was included after the word "echotexture" to make the sentence more meaningful. Other than that, symbols and abbreviations were also automatically converted into words using regular expressions in the system. For example, symbols such as "/" and "&" were automatically changed to "or" and "and" respectively. The pre-processing phase has proven to improve the accuracy rate of the parser significantly. The result of this will be presented further in the next section.

5.6.2 Report Segmentation Result

The accuracy of the system's report segmentation based on RST was evaluated in two stages. The first stage was without the introduction of the pre-processing phase while the second stage was with the pre-processing phase performed on the reports. The evaluation was undertaken separately on the 60 training data reports and the 40 testing data reports that were both compared to a gold standard which is the expert's manual parsing. Both the training and testing data were evaluated by comparing the total number of text spans that were produced by the system for each report against the total number of text spans that were produced by the expert's manual parsing. If the total number of text spans segmented by the system matches the total number of text spans segmented by the expert, then the report is viewed as accurately segmented. Table 5.6 displays the result of the evaluation.

Without pre-processing, the training data achieved a 78.33% accuracy rate where 47 out of the 60 reports were segmented correctly. However, as for the testing data, the accuracy rate was very low where only half of the reports were segmented correctly. This result was worrying thus prompting us to adopt a pre-processing phase. When

	Without Pre-processing Phase		With Pre-processing Phase	
	Accurately Segmented Reports	Percentage (%)	Accurately Segmented Reports	Percentage (%)
Training Data	47	78.33	53	88.33
Testing Data	20	50.00	33	82.50
Both Data	67	67.00	86	86.00

Table 5.6: Comparison of the accuracy percentage for report segmentation performed with and without pre-processing phase

the pre-processing phase was introduced, significant improvement of the accuracy rate was observed for both sets of data especially for the testing data. The accuracy rate increased from 78.33% to 88.33% for the training data and from 50% to 82.5% for the testing data.

The reason for the increment was because there were a lot of reports which were under segmented in the initial stage where pre-processing was not conducted. This signifies that there were certain boundaries that the experts believe should be segmented but because there are no signal words or punctuation, the system fails to recognise them. This is demonstrated in the following example "Normal calibre aorta measuringe 1.5cm in diameter." This sentence should be segmented into two text spans which are "'Normal calibre aorta" and "measuring 1.5cm in diameter" where there is an ELABORATION relation between them. However, because the word "measuring" was misspelled, the system failed to recognise it and did not execute the segmentation. Accordingly, the percentage increased significantly when the pre-processing phase was introduced.

Nevertheless, the accuracy rate was still less than 90%. This was substantially contributed by the over segmentation of texts caused by the appearance of discourse markers that signals text boundaries. Over segmentation of texts can be demonstrated in an example of the sentence "The CBD was within normal limits". In this sentence, the word "within" is a signal word for the ELABORATION relation. Therefore, it serves as a text boundary whereby the sentence will be split before the word "within". This produces two text spans which are "The CBD was" and "within normal limits". The segmentation of this sentence should not happen as both text spans do not have any meaning on its own and therefore, could not act as the nucleus in the relation.

The same is exhibited in the sentence "Spleen with normal size." where it was split into two text spans, "Spleen" and "with normal size". This is because of the cue word "with" which usually signals the JOINT relation. This implies that, for the problem of over segmentation to be reduced, further rules should be defined to avoid sentences to be segmented in certain cases although they contain a signal word.

5.6.3 Rhetorical Relation Identification Result

The identification of the rhetorical relations between text spans in the sample ultrasound reports was also evaluated separately between the 60 training data and the 40 testing data. The evaluation was performed using the common information retrieval measurement methods which are precision, recall and F-score. In this research, precision will be the ratio of the relevant relations identified by the system to the total number of relevant and irrelevant relations identified by the system as illustrated in Formula 5.1. Recall on the other hand will be the ratio of the relevant relations identified by the system to the total number of relevant relations identified by the experts as stated in Formula 5.2. Finally, the F-score will be measured as the harmonic mean of both the precision and recall based on Formula 5.3. Table 5.7 presents the precision, recall and F-score for both the training and testing data.

$$Precision = \frac{relevant\ relations\ identified\ by\ the\ system}{relevant + irrelevant\ relations\ identified\ by\ the\ system} \tag{5.1}$$

$$Recall = \frac{relevant\ relations\ identified\ by\ the\ system}{relevant\ relations\ identified\ by\ the\ experts} \tag{5.2}$$

$$F\text{-}score = 2\frac{(precision)\ (recall)}{precision + recall} \tag{5.3}$$

In the 60 training data, a total of 325 relations have been identified by the experts. When these reports were submitted to the system, it managed to identify 310 relations between the text spans in the reports. Out of the 310 relations identified, 302 relations were similar to the ones identified by the experts in their manual parsing. This has resulted in a very high precision and recall of 97.42% and 92.92% respectively. As a result, the training data achieved an F-score of 95.12%.

The system has failed to correctly recognise 23 relations that were identified by the experts. This is particularly because there were certain words which were not annotated

	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Training Data	325	310	302	97.42	92.92	95.12
Testing Data	274	263	240	91.25	87.59	89.38
Both Data	599	573	542	94.59	90.48	92.49

Table 5.7: Evaluation of the rhetorical relation identification process

with the relevant classes from AUO. Two text spans which are "The liver is slightly hyperechoic" and "in keeping with fatty liver disease." can be taken as an example to illustrate this problem. There exist a JUSTIFY relation between the two text spans where the first text span justifies the second. For a JUSTIFY relation in an abnormal finding to be recognised, one of the words in the text span needs to be annotated with the class "finding" or a "biospecimen". However, because the word "hyperechoic" was not associated with any of the two classes, the system failed to recognise that it justified the existence of a "liver disease".

The precision and recall for the testing data were slightly lower compared to the training data which were 91.25% and 87.59% respectively. It also has a lower F-score of 89.38%. In the testing data, the experts have managed to identify 274 relations between the text spans in the 40 reports while the system managed to identify only 263 relations. Out of these relations identified by the system, 240 relations were similar to the ones identified by the experts. This means that the system failed to correctly recognise 34 relations between the text spans.

A reason for this was the same as in the training data whereby there were some words which were not annotated with the relevant classes in AUO. In addition to this, another reason that the system failed to correctly recognise all 34 relations was because most of these relations were a first occurrence. Therefore, the system was not trained to recognise such patterns. Accordingly, in order to improve the score of the system, the ontology will need to be enhanced to accommodate the words which were not annotated. In addition to that, more sample ultrasound reports should also be added to the training data so that the system is able to recognise more patterns of the relations.

In both the training and testing data, there were several relations that have been identified by the system but were not recognised by the experts. This happened mostly because of over segmentation of the texts. One example of this is the sentence "Spleen with normal size". As mentioned in the previous section, this sentence was wrongly segmented into two text spans which were "Spleen" and "with normal size". This has resulted in the system identifying an ELABORATION relation between the two text spans because there is a measurement and no mention of an "organ" in the second text spans and there is an "organ" in the first text span without any "findings", "disease" or measurement. To avoid such problems in the future, the rules used to perform the segmentation need to be improved so that over segmentation of texts can be reduced.

Breakdown of the Result Based on Each Relation

The rhetorical relation identification result has also been analysed based on each relation to recognise which of them have been identified better by the system. This is also to recognise which relations that have been poorly identified so that it can be improved in the future. The breakdown of each relation has been divided into two types which are single and nested. Examples of a single type relation are "TS5 JOINT TS6" and "TS3 JUSTIFY TS2" where the relation was direct between the two text spans. Nested type relations on the other hand combine two or more relations from the same or different relation. Examples of this are "TS1 PREPARE LIST (TS2-TS9)" where it combined the PREPARE and LIST relations and "TS5 JOINT TS6 JOINT TS7" where it combined two JOINT relations. Table 5.8 presents the precision, recall and F-score for the single type of each relation in the training data.

In the 60 training data, the single type of the JUSTIFY relation has the lowest recall and F-score which are 63.64% and 77.78% even though it has a 100% precision. This was caused by the lack of annotated classes that is required to identify the JUSTIFY relation. For example, there is a JUSTIFY relation between the text span "Within the gallbladder wall there is at least one comet tail artefact," and "suggestive of a tiny crystal deposit". However, because there is no "disease" class annotated, the system failed to recognised the relation. Although the system managed to identify only some of the JUSTIFY relations identified by the experts, all the relations that it identified were correct, hence the 100% precision rate. As for the single type of the ELABORATION relation, the system achieved a high precision, recall and F-score of 97.5%, 95.12% and 96.3% respectively. This is the same for the single type of the JOINT relation where the system achieved a precision, recall and F-score of 95.45% each.

In the training data, all three CONTRAST relations have been identified correctly with a precision, recall and F-score of 100% each. The high accuracy is due to the low

	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	0	0	0	-	-	-
Restatement	0	0	0	-	-	-
Justify	33	21	21	100	63.64	77.78
Elaboration	41	40	39	97.50	95.12	96.30
List	0	0	0	-	-	-
Joint	88	88	84	95.45	95.45	95.45
Contrast	3	3	3	100	100	100

Table 5.8: Evaluation of the single type of each relation in the training data

	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	89	89	89	100	100	100
Restatement	2	2	2	100	100	100
Justify	45	43	42	97.67	93.33	95.45
Elaboration	21	20	19	95.00	90.48	92.68
List	60	60	60	100	100	100
Joint	66	62	62	100	93.94	96.88
Contrast	5	5	5	100	100	100

Table 5.9: Evaluation of the nested type of each relation in the training data

volume of this relation identified by the experts in the training data. As for the PREPARATION, RESTATEMENT and LIST relations, none of them has been identified as a single type in the training data. Therefore, the precision, recall and F-score for these relations could not be calculated.

All 89 PREPARATION relations that have been identified in the training data by the experts were of the nested type. The system managed to correctly identify all 89 relations, hence, achieving 100% precision, recall and F-score. The system also does not have any problem identifying the RESTATEMENT, LIST and CONTRAST relations which were also of the nested type, where all three achieved 100% precision, recall and F-score. As for the JUSTIFY relation, the experts have identified 45 nested relations. Out of the 45 relations, the system has managed to identify 43 relations where 42 of them were similar to the ones identified by the experts. This has resulted to a precision of 97.67%, recall of 93.33% and F-score of 95.45%.

The system also performed well in identifying the ELABORATION relation where it achieved a precision, recall and F-score of 95%, 90.48% and 92.68% respectively. This is because the system managed to identify accurately 19 out of the 21 nested relations identified by the experts. Finally, the system achieved a 100% precision, 93.94% recall and 96.88% F-score when identifying the nested JOINT relation. Out of the 66 nested relations identified by the experts, 62 were identified correctly by the system. The summary of the precision, recall and F-score of all the nested type of the seven rhetorical relations is summarised in Table 5.9.

Generally, the system performed well in identifying both the single and nested type of all seven relations in the training data. The only low score was recognised in the single type of the JOINT relation. However, this low score can be increased if the ontology used in the discourse parser is improved. On the other hand, the precision, recall and F-score for both the single and nested type of all seven relations in the testing data was slightly lower. This is presented in Table 5.10 and Table 5.11.

In the 40 testing data, all PREPARATION, RESTATEMENT, LIST and CONTRAST relations achieved a 100% precision, recall and F-score where all relations identified by the experts were also correctly identified by the system. The PREPARATION, RESTATEMENT and LIST relations were only found as nested types while the CONTRAST relation was found as both single and nested type. In the testing data, the experts have also identified a total of 17 single JUSTIFY relations. On the other hand, the system has managed to identify 16 relations where 15 of them were identical to the ones identified by the experts. This gives a precision of 93.75%, recall of 88.24% and F-score of 90.91%. As for the nested JUSTIFY relations, the system achieved a slightly lower precision, recall and F-score of 89.8%, 93.62% and 91.67% respectively.

The ELABORATION relation was the most poorly identified rhetorical relation in the testing data. In the single type relation, the experts have identified a total of 24 relations while the system has identified 25 relations. However, out of the 25 relations identified by the system, only 18 were accurate. This resulted in a precision of 72%, recall of 75% and F-score of 73.47%. The result is much poorer for the nested type ELABORATION relation where the system managed to identify only 8 out of the 22 relations identified by the experts. The precision, recall and F-score of the nested relation are 87.5%, 31.82% and 46.67% respectively. The main reason that the system performed very poorly in identifying nested ELABORATION relations is because these different types of relations were never trained before. One example is the relation

	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	0	0	0	-	-	-
Restatement	0	0	0	-	-	-
Justify	17	16	15	93.75	88.24	90.91
Elaboration	24	25	18	72.00	75.00	73.47
List	0	0	0	-	-	-
Joint	72	67	66	98.51	91.67	94.96
Contrast	4	4	4	100	100	100

Table 5.10: Evaluation of the single type of each relation in the testing data

	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	94	94	94	100	100	100
Restatement	9	9	9	100	100	100
Justify	47	49	44	89.80	93.62	91.67
Elaboration	22	8	7	87.50	31.82	46.67
List	49	49	49	100	100	100
Joint	81	75	71	94.67	87.65	91.03
Contrast	8	8	8	100	100	100

Table 5.11: Evaluation of the nested type of each relation in the testing data

between these three text spans “The right kidney has a bipolar length of 117mms”$_{\text{TS1}}$, “there are two simple cortical cysts seen largest at the midpole”$_{\text{TS2}}$ and “measuring 47ms”$_{\text{TS3}}$ where (TS3 ELABORATE TS2) ELABORATE TS1. However, because this pattern of nested relation was not in the training data, the system failed to identify it correctly.

Finally, the JOINT relation achieved a better score as compared to the ELABORATION relation where the system managed to identify 67 out of the 72 single relations identified by the experts. This caused it to achieve a precision of 98.51%, recall of 91.67% and F-score of 94.96% for the single type. The system also performed well in identifying the nested JOINT relation where it correctly identified 71 out of the 81 relations identified by the experts. As a result, the precision, recall and F-score for the nested JOINT relation are 94.67%, 87.65% and 91.03% respectively. This breakdown of the result based on each relation enables us to better identify where the problem lies in ensuring accuracy. As a result of this breakdown, it is realised that the precision and recall can be improved by enhancing the ontology as well as acquiring more sample reports in order to train the system with more relation patterns.

5.6.4 Ontology-Agnostic versus Ontology-Informed

Although the RST medical discourse parser developed in this research uses different methods compared to the existing automatic discourse parsers, a comparison between the two works was attempted. An automatic discourse parser called CODRA: A Document-level Discourse Parser for English [39] was developed by Joty et al. as an improvement to the work undertaken by Feng & Hirst [27]. CODRA was made available by them in their research website[1] where users are able to submit texts to be parsed. Before attempting to use CODRA on the sample ultrasound reports that have been acquired, we first tested it by submitting a sentence "My wife and I are both British, and we enjoy visiting America." which was one of the examples given by Mann and Taboada in explaining the JOINT relation [93]. This returns a correct result where CODRA managed to segment the text into two text spans and denotes a JOINT relation between them. This is depicted in Figure 5.14.

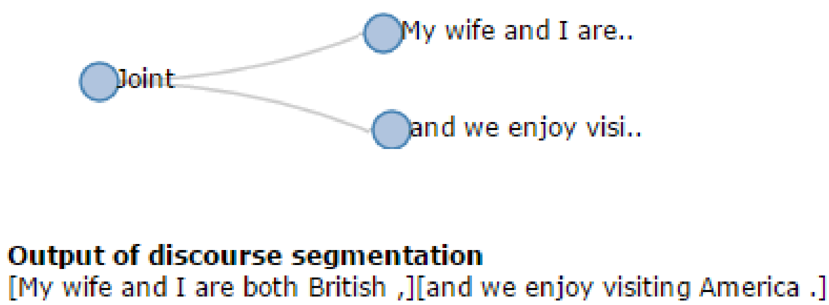


Figure 5.14: Discourse parsing of the sentence "My wife and I are both British, and we enjoy visiting America." as performed by CODRA

Following this, we have submitted all 100 sample ultrasound reports to see how CODRA performs on our data. However, the result for both the segmentation and the rhetorical relation identification was bad and could not be compared with the system

[1] http://alt.qcri.org/demos/Discourse˙Parser˙Demo/

we have developed. An example of how CODRA performs with regards to our data can be seen using the following ultrasound report as an example:

> "US Abdomen: Normal sonographic appearance of the liver. No focal lesion. Normal thin-walled gallbladder. No gallstones. No biliary dilatation. Normal direction of portal venous flow. Both kidneys are normal in size and cortical depth. RK = 96mm LK = 100mm. Normal appearance of the pancreas, aorta and spleen. Main or principal diagnosis: Normal study. No cause for abnormal LFTs identical."

The output from the discourse segmentation was 13 text spans as seen in Figure 5.15. The segmentation, however, was not accurate as it is only performed on sentence boundaries signalled by full stops without taking into consideration other cues such as commas and colons as well as signal words like "and". This method of segmentation will not be beneficial if it were implemented in the transformation of free-form reports to structured form because it will not be able to separate the findings for two different areas examined which were recorded in the same sentence.

Output of discourse segmentation

```
[US Abdomen :][Normal sonographic appearance of the liver .]
[No focal lesion .]
[Normal thin-walled gallbladder .]
[No gallstones .]
[No biliary dilatation .]
[Normal direction of portal venous flow .]
[Both kidneys are normal in size and cortical depth .]
[RK = 96mm LK = 100mm .]
[Normal appearance of the pancreas , aorta and spleen .]
[Main or principal diagnosis :][Normal study .]
[No cause for abnormal LFTs identical .]
```

Figure 5.15: Output of discourse segmentation performed by CODRA

Other than segmenting the reports, CODRA is also able to produce a complete RST tree that shows the rhetorical relations identified between the text spans. In Figure 5.14, it can be deduced that CODRA correctly identifies a JOINT relation between the two text spans. However, when the free-form ultrasound report was submitted, CODRA produced an RST tree that identified only two rhetorical relations between the text spans which were TOPIC-COMMENT and ELABORATION relation. The RST tree for this parse is depicted in Figure 5.16. This demonstrates that CODRA failed

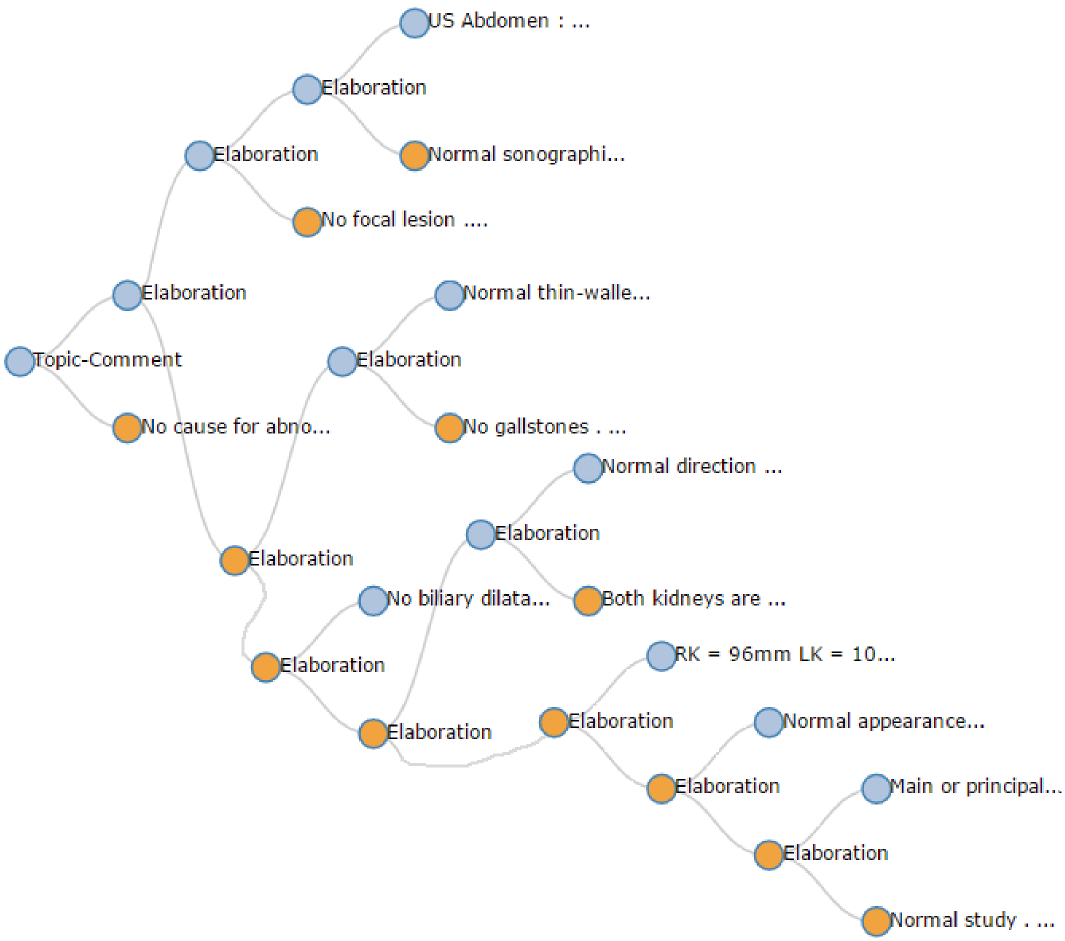


Figure 5.16: RST tree produced by CODRA

to correctly segment and identify rhetorical relations in a medical ultrasound report. Therefore, we believed that it would not be beneficial to do a comparison between our system and CODRA.

As a result of this, we have decided to do an ontology-agnostic and ontology-informed comparison using the same code used for the system but with minor changes. The ontology-informed system was the system that we have developed whereby the decision taken was based on the information provided by AUO. As for the ontology-agnostic system, the same code as our system was used but the annotation part was left out. This means that the code uses the same rules as our system but without the information provided by AUO as the knowledge base. This is to determine the importance of including ontology in the discourse parsing of ultrasound reports. For the segmentation of the ultrasound reports, there was no difference at all in the comparison between the ontology-agnostic and ontology-informed system. This is because ontology was not used in the segmentation process of the reports. However, there was a clear difference in the result of the rhetorical relations identification for the ontology-agnostic and ontology-informed system. Table 5.12 compares the precision, recall and F-score for the ontology-agnostic system against our system.

	Ontology-Agnostic			Ontology-Informed		
	Training Data	Testing Data	Both Data	Training Data	Testing Data	Both Data
No of Relations Identified Manually	325	274	599	325	274	599
No of Relations Identified by the System	250	227	477	310	263	263
Similarities	224	211	435	302	240	542
Precision (%)	89.60	92.95	91.19	97.42	91.25	94.59
Recall (%)	68.92	77.01	72.62	92.92	87.59	90.48
F-score (%)	77.91	84.23	80.86	95.12	89.38	92.49

Table 5.12: Comparison of the precision, recall and F-score for the rhetorical relations identification of the ontology-agnostic and ontology-informed parser

In the table, it can be seen that the ontology-agnostic system achieved a precision of 89.60% for the training data, 92.95% for the testing data and 91.19% for overall data. This is lower than the precision achieved by our system which is 97.42% for the training data, 91.25% for the testing data and 94.59% for both data. The same result was achieved when the recall for each data was evaluated where the ontology-agnostic system gives a recall of 68.92% for the training data, 77.01% for the testing data and 72.62% for both data. This is a lot lower than our system which has a recall of 92.92% for the training data, 87.59% for the testing data and 90.48% for the overall data. The F-score for the ontology-agnostic system was also a lot lower than our system where it only achieved 77.91% for training data, 84.23% for testing data and 80.86% for both data, as compared to our system which achieved 95.12% for training data, 89.38% for testing data and 92.49% for both data. This result was calculated by comparing the number of relations identified by both the ontology-agnostic and ontology-informed system with the number of relations identified by the experts. Based on this result it can be seen that ontology-informed discourse parsing resulted in a much better precision, recall and F-score for ultrasound discourse parsing using RST as compared to

		Ontology -Agnostic					Ontology-Informed				
	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	89	89	89	100	100	100	89	89	100	100	100
Restatement	2	2	2	100	100	100	2	2	100	100	100
Justify	78	28	28	100	35.9	52.83	64	63	98.44	80.77	88.73
Elaboration	62	6	3	50	4.84	8.82	60	58	96.67	93.55	95.08
List	60	60	60	100	100	100	60	60	100	100	100
Joint	154	155	131	84.52	85.06	84.79	150	146	97.33	94.81	96.05
Contrast	8	5	5	100	62.50	76.92	8	8	100	100	100

Table 5.13: Breakdown of the precision, recall and F-score of ontology-agnostic and ontology-informed parser based on rheorical relations for the training data

		Ontology -Agnostic					Ontology-Informed				
	No of Relations Identified Manually	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)	No of Relations Identified by the System	No of Similarities	Precision (%)	Recall (%)	F-score (%)
Preparation	94	94	94	100	100	100	94	94	100	100	100
Restatement	9	9	9	100	100	100	9	9	100	100	100
Justify	64	49	49	100	76.56	86.73	49	44	89.80	93.62	91.67
Elaboration	46	1	1	100	2.17	4.26	8	7	87.50	31.82	46.67
List	49	49	49	100	100	100	49	49	100	100	100
Joint	153	141	127	90.07	83.01	86.39	75	71	94.67	87.65	91.03
Contrast	12	11	11	100	91.67	95.65	8	8	100	100	100

Table 5.14: Breakdown of the precision, recall and F-score of ontology-agnostic and ontology-informed parser based on rheorical relations for the training data

ontology-agnostic.

In order to investigate the reason behind this, the result was broken down based on each of the seven rhetorical relations. The result for the training data is summarised in Table 5.13 while the result for the testing data is summarised in Table 5.14. In this table, the result for the identification of the single and nested type for each relation was combined and the precision, recall and F-score were calculated. From the result of both the training and testing data, it can be seen that there were no difference in the identification of the PREPARATION, RESTATEMENT and LIST relation for both ontology-agnostic and ontology-informed where both managed to achieved 100% precision, recall and F-score. The reason for this is because all three relations do not need the annotation of the classes from AUO in order for the system to recognise them.

The significant difference between using ontology as the knowledge based for the rhetorical identification process and without using it can be seen in the identification of the ELABORATION relation. With ontology, the training data achieved a precision, recall and F-score of 96.67%, 93.55% and 95.08% while the testing data achieved a precision, recall and F-score of 87.5%, 31.82% and 46.67%. Without ontology, the result is significantly lower whereby the training data achieved a precision of 50%, recall of 4.84% and F-score of 8.82% while the testing data achieved a precision of 100%, recall of 2.17% and F-score of 4.26%. The reason that the precision of the ontology-agnostic system is a lot higher compared to the recall and F-score is because most of the relations that the system managed to identify is correct. However, the recall and F-score were low because the ontology-agnostic system failed to identify a lot of the relations identified by the experts.

The reason that the ELABORATION relation performed poorly in the ontology-agnostic system is because almost all the rules used in identifying it need the annotation from AUO. For example, in identifying an ELABORATION relation that elaborates the measurement of the finding, the "unit of measure" is needed. In the sentence "The gallbladder contains a single stone measuring 7mm.", without the annotation of the "mm" as the "unit of measure" and "stone" as the "biospecimen" the system will fail to recognise that the "7mm" elaborates the size of the "stone". This is the same for relations such as JUSTIFY and CONTRAST where both depend on the annotation of the report to understand the semantics of the sentence. This is different in the case of the JOINT relation. According to the rule, as long as there is an "and" or "or" word in the sentence, it signals a JOINT relation. However, there are circumstances where this is not true. One of it is when there exist a pair of organ and each has their own descriptions but were combined using the word "and". Without the information provided by AUO, the system will fail to recognise this situation and resulting in the identification of a non-JOINT relation as a JOINT relation. As a conclusion to this evaluation, it is clear that there was an improvement in the identification of rhetorical relations in ultrasound reports when ontology was used to inform the parsing as compared to when ontology was absent.

5.6.5 Discussions

The implementation of RST with the combination of an ontology such as AUO in the medical domain is a new method in computer science that has not been explored before. For this reason, there are a lot of issues that can be discussed in order to improve its accuracy. An example of an issue is regarding the term "intra- extra hepatic ducts" and "intra or extra hepatic ducts". Currently, both terms are not separated into "intra hepatic ducts" and "extra hepatic ducts". However, there was a suggestion from the ultrasound specialists who were evaluating the system to separate this term into two so that it can be associated with two different areas of examination which are "Liver" and "Duct".

Another issue that can be discussed is regarding the usage of words in the reports. This has been brought up by the ultrasound specialist where words such as "Fossa" can be understood depending on the context of the report. However, if it was separated in one sentence, the word will be meaningless. Consequently, the implementation of RST using an ontology in the medical domain can be massively improved with further involvement of experts from both the natural language processing and the medical fields. Limitations of the current implementation can also be further reduced if more sample ultrasound reports can be gathered. This is because they will allow for the system to be trained using more data and at the same time learn new patterns of rhetorical relations.

5.7 Chapter Summary

RST is a well-established theory in computational linguistics used to recognise relations between text spans in a coherent text where it uses discourse markers and sentence structure to recognise relations. This chapter presented our approach in using an ontology and discourse markers to identify RST relations in ultrasound reports. Several rules have been designed as a guide to segment and recognise rhetorical relations in ultrasound reports. These rules have been designed based on an analysis performed on 60 out of 100 sample reports collected. From the analysis, seven rhetorical relations have been identified which are PREPARATION, RESTATEMENT, JUSTIFY, ELABORATION, LIST, JOINT and CONTRAST.

These seven rhetorical relations were then applied in the discourse parsing of the sample ultrasound reports. From the evaluation, the system achieved an accuracy of 88.33% and 82.50% for both the training and testing data for the segmentation process. As for the identification of rhetorical relations, the system achieved a precision of 97.42%, a recall of 92.92% and an F-score of 95.12% for the training data and a precision of 91.25%, a recall of 87.59% and an F-score of 89.38% for the testing data. This proves that implementing ontology in the discourse parsing process of RST made it possible for RST to also be applied on medical reports. A comparison between the system and another system that does not use ontology to inform the parsing was also presented. From the comparison, it was clear that there were improvements in the identification of rhetorical relations when using an ontology-informed system as compared to an ontology-agnostic system.

Chapter 6

The Development of the Standardised Reporting System

6.1 Introduction

Thus far this thesis has presented the development and evaluation of a knowledge base called AUO to standardise the ultrasound reports. It has also explored the implementation of RST on medical data using AUO and discourse markers which gave promising results. Therefore, the next step is to apply both of these correctly in the medical ultrasound reporting system to automatically generate a standardised structured report from the submitted free-form report.

This chapter presents the development of an ultrasound reporting system, focusing on the structured report generator component. It will explain the approach in implementing RST together with AUO in transforming free-form reports to structured form. It first reviews the 60 training data reports to identify their structure and the types of information they contain. This review is vital in deciding which RST relations and rules that should be applied in reconstructing the reports according to the requirements of a structured report. Next, the chapter will evaluate the implementation.

There was an idea to automatically classify findings as “normal”, “abnormal” and “inconclusive”. However, after evaluations and discussions with a pair of specialists, it was concluded that it is best to abandon this idea for the moment because of the lack of information. This chapter will present the rule-based approach in executing the classification and discuss the specialists’ evaluation and opinions regarding this matter. Finally, this chapter will give an overall discussion regarding the standardisation of ultrasound reports based on the results and feedbacks that have been obtained.

6.2 Development Tools

The medical ultrasound reporting system was built using several different tools. The ontology component of the system, AUO was developed in OWL language using an ontology editor called Protègè[1]. This open-source editor provides a friendly user interface for the development and maintenance of ontologies. It also allows for the ontology to be stored in various formats other than OWL such as RDF and XML.

The overall system including the RST component of the system was developed in PHP while the login information as well as the patients' information were stored in a MySQL database. Since the system is a web application, a web server is also required. In this research, XAMPP[2] was chosen as the development environment because it supports PHP and MySQL as well as providing a local server to host the syste. XAMPP is also open-source which makes it cost effective. The graphical interface of the system was designed using the Cascading Style Sheets (CSS) language and the Adobe Dreamweaver web development tool.

6.3 Implementation of the Structured Report Generator

Almost all of the components in the system architecture model presented in Chapter 3 have been implemented in the overall system. The development of the ontology component has been explained in Chapter 4 while the RST component has been presented in Chapter 5. The rest of the components have also been described in Chapter 3. Therefore, this chapter will focus on the implementation of the structured report generator component that transforms free-form reports to structured form by implementing the RST component with the support of AUO as its knowledge base.

In the medical discourse parser presented in Chapter 5, there were seven rhetorical relations that have been identified in the 60 reports training data. However, in implementing RST in the transformation process, not all seven relations will be needed. Therefore, it is important to first identify which relations that are relevant. This was performed by reviewing the structure and content of the training data to recognise the types of information available and how it can be reconstructed to fit the requirements of a structured report.

[1] http://protege.stanford.edu/

[2] https://www.apachefriends.org/index.html

6.3.1 Structural and Content Review of the Training Data

In reviewing the structure of the training data, only two out of the 60 reports were itemised while the rest were written in free-form with the usage of titles to separate the different types of information they included. Almost all reports were written with a main title in the beginning of the reports to indicate the type of ultrasound examination that was conducted. Other than the findings and observations of the examination, three other distinct types of information that can be viewed in the reports were clinical history, conclusion and further management. The occurrences of these types of information in the reports are summarised in Table 6.1.

Types of Information	Total Occurrences	No of Explicit Occurences	No of Implicit Occurences
Clinical history	4	0	4
Conclusion	28	28	0
Further management	9	1	8

Table 6.1: The occurences of the three types of information found in the training data

The number of clinical history found were very minimal as there were only four occurrences. All four of these were implicit, which means that they were written together with the findings, making it hard for the referring clinicians to locate the information. Almost half of the reports include a conclusion at the end of them. All of these conclusions were explicit whereby they were separated from the findings using a title such as "Conclusion:". There were two reports which contains a "Main or principal diagnosis:" which gives an overall summary of the report and then restates it with a conclusion. The review also found nine occurrences of further management where eight of them were implicit and only one was separated with a title "Further investigations or management:".

For an ultrasound report to be considered as written in a structured format, the different types of information in the report needs to be separated with suitable headings [18]. Therefore, in reconstructing the free-form reports to structured form, these three types of information will need to be separated from the findings. Since some of this information was already separated using a title, the PREPARATION and RESTATEMENT relation can be used to identify them other than using signal words. The result of the survey conducted during the 2015 UK Radiological Congress has also prompted the structured report to be designed in such a way whereby the findings are separated according to the area examined. Therefore, the JOINT, LIST and ELABORATION relations can be applied so that the sentences in the reports can be separated by areas of findings without losing any information. The next section will elaborate further on this implementation.

6.3.2 The Role of Ontology and RST in the Structured Report Generator

Now that the structure and content of the training data have been understood, the ontology and the RST components can finally be implemented in the system. This section will explain the steps that need to be taken in transforming the free-form reports into a structured form. Figure 6.1 summarises these steps.

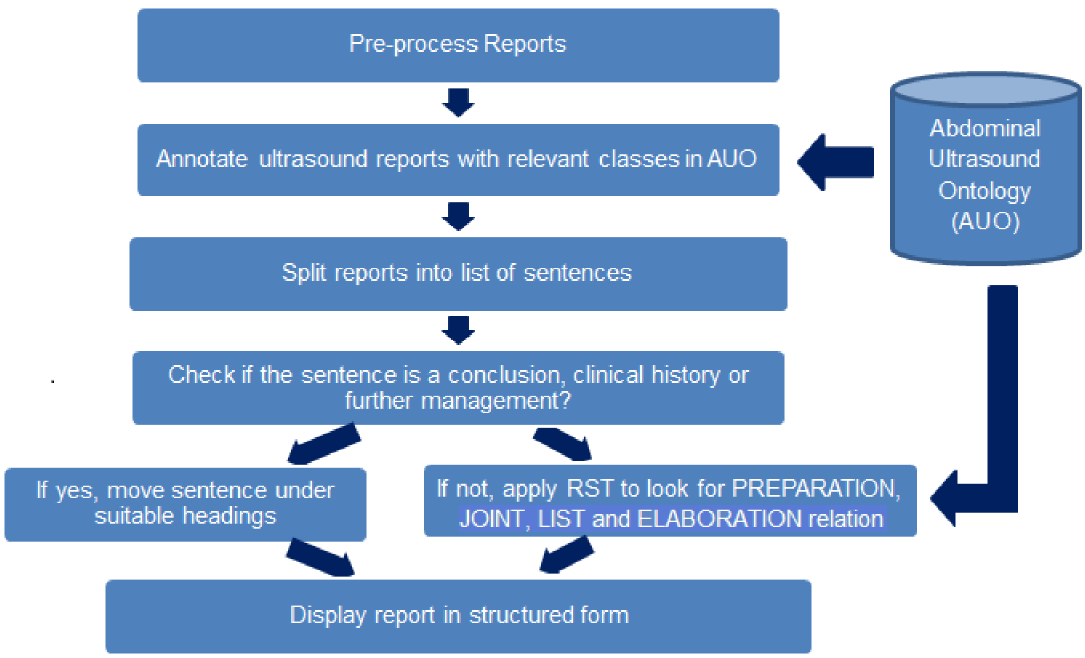


Figure 6.1: The steps in transforming free-form reports to structured form

Pre-processing the Reports

The first step that needs to be taken in transforming the free-form reports to structured form is to pre-process the reports to remove the many obvious errors caused by the radiologists as mentioned in Chapter 5. This was performed by manually going through the reports and correcting the errors. Once this has been completed, the reports will then be submitted to the system to be transformed into a structured form. After submission, another pre-processing phase will be performed automatically by the system whereby all the main titles of the reports such as “US Abdomen” and “Ultrasound: Abdomen” will be removed because the produced reports will be assigned with a standardised title.

Annotating and Splitting the Reports

Subsequently, relevant words from the reports will be annotated with classes from AUO. The annotation process will be performed using the same method as presented in Subsection 5.5.1 where the paragraphs in the report were split into single words and later combined to two and three word combinations before being compared with relevant classes from AUO. The relevant classes are those that have been defined in the rules to identify rhetorical relations in ultrasound reports as described in Section 5.4. When all the words have been annotated, they will be merged to form a paragraph before being split again after each sentence boundaries such as a full stop or a colon to create a list of annotated sentences. The reason these reports were split into single sentences was because it is important to recognise which type of information each sentence contains so that the system knows how to process the sentence.

Identifying the Sentence Type

As mentioned in Subsection 6.3.1, there were four different types of information that have been recognised in the sample ultrasound reports. Thus, the next step would be to identify which of these types of information does each sentences in the reports belongs to. These sentences can be classified as either a clinical history, finding / observation, conclusion or further management. The classification was performed by using signal words, AUO or both as well as by identifying RST relations. Since the information on clinical history were all implicit, it can only be identified using a list of signal words such as “previous”, “history” as well as the AUO class,“Month of the Year”.

Conclusions on the other hand were all explicit, therefore, they can be identified using the PREPARATION relation where they were separated under the title “Conclusion” or “Comment:”. A RESTATEMENT relation can also be used to identify conclusions that were written as the “Main or Principal Diagnosis:”. On the contrary, sentences which are of the type further management can be identified using cue words such as “advise”, “suggest”, “consider” and “recommend”. Since it is also sometimes

Type of Information	Signal Word / Title
Clinical history	previous, history, “Month of the Year” class
Conclusion	“Conclusion:”, “Comment:”, “Main or Principal Diagnosis:”
Further management	advise, suggest, consider, recommend

Table 6.2: The different types of information and its signal words

explicitly written, the further management information can also be identified using the PREPARATION relation where the information is written under the title “Further Management:”. Table 6.2 mentions several examples of the signal words and titles in the PREPARATION relation and the equivalent information they are signalling. All the sentences which were of these three types of information will be extracted from the free-form report and moved under suitable headings in the structured report. However, for all the other remaining sentences that does not fall under these three types will then be regarded as finding / observation and will be further processed.

Applying RST on the Findings and Observations

The aim of applying RST in the transformation process is to group the findings in the reports according to the area examined under the “Findings / Observations” heading as shown in Figure 6.2. For example, if the sonographer has examined the liver, pancreas and spleen of the patient, the findings and observations should be recorded according to the area examined instead of writing all of it in one paragraph. This could be performed by identifying three out of the seven rhetorical relations which are the JOINT, LIST and ELABORATION relations in the findings and observations.

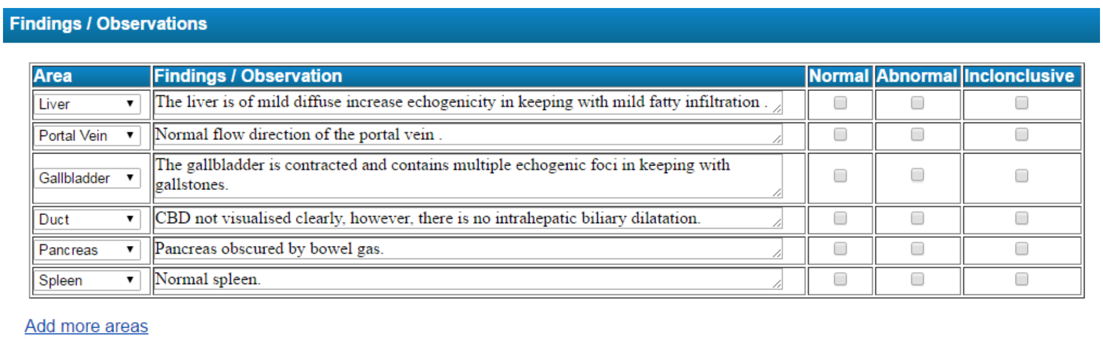


Figure 6.2: An example of a list of findings and observations

The JOINT relation was used to separate several areas examined which were initially reported in one sentence, into several separate sentences. The JOINT relation was the most used relation in transforming the free-form reports into structured reports. For example, one of the sample reports stated that “The kidney, spleen and pancreas is normal”. In this example, it was clear that there were three areas being examined which are “kidney”, “spleen” and “pancreas”. When RST is applied to this sentence, it will recognise that there exist a JOINT relation between the three organs because of the cue word “and” as well as the commas between the organs, although grammatically it looks wrong as the radiographer wrote “is normal” when it should be “are normal”. This allows the system to separate these three organs into separate sentences without losing its observation.

The existence of the JOINT relation informs the system that it should report three different sentences that starts with the three organs and shares the same observation which is “is normal”. This produced three sentences which were “The kidney is normal”, “Spleen is normal” and “Pancreas is normal”. The JOINT relation will only separate a sentence into several other sentences when there is more than one area being reported. If there is only one, it will not separate the sentence. For example, it was reported that “The liver has smooth contour and normal echogenicity”. Even though there was a word “and” that signals a JOINT relation, but since the sentence was reporting about only one area which is the liver, it will not be split into two.

In a sentence where there is the word “and” but there are also two organs and two observations being reported, no JOINT relation will be identified based on the rules defined in Section 5.4. Instead, this sentence will be segmented into two text spans that have a LIST relation between each other. Therefore, the sentence will be reproduced as two sentences that do not share one observation. An example of this is the sentence “The gallbladder wall is very thickened and the liver appear prominent”. In this sentence, there were two areas being reported which are the gallbladder and the liver. However, because there were also two observations (“very thickened” and “appear prominent”), the system will recognise this sentence as having a LIST relation. Therefore, it separates the sentence into two but with both of them having their own observation. The decision whether or not to separate a finding to more than one sentences is summarised in the flowchart in Figure 6.3.

Another relation being used in the transformation process is the ELABORATION relation. This relation is important in ensuring that any other extra information was not lost when it is being separated or joined with another sentence. For example, consider

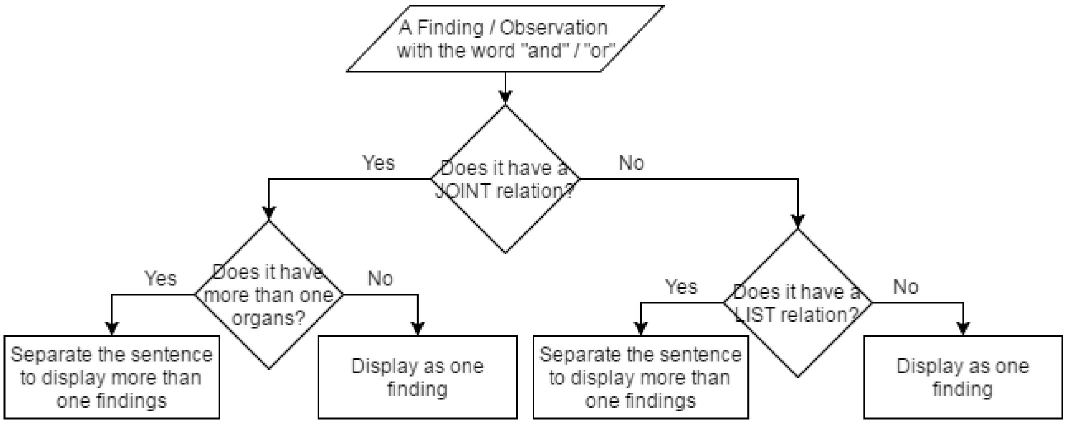


Figure 6.3: Separating a finding based on JOINT and LIST relation

this sentence “Normal appearance of spleen (measuring 12.6cm), head and body of pancreas and aorta (measuring 1.4cm inner to inner)”. This sentence consisted of three areas which were the spleen, head and body of pancreas and aorta. There was only one observation which was “normal appearance”. A JOINT relation that exist between the three areas will separate the sentence into three which are “Normal appearance of spleen”, “Normal appearance of head and body of pancreas” and “Normal appearance of aorta”. Without the ELABORATION relation, it will be hard to retain the information on the measurement of the organs. Therefore, ELABORATION relation is also needed in the transformation process. The ELABORATION relation allows the sentence to be separated without losing any information resulting the two sentences to become “Normal appearance of spleen (measuring 12.6cm)” and “Normal appearance of aorta (measuring 1.4cm inner to inner)”.

Displaying the Structured Report

When all the sentences in the free-form report have been grouped under suitable headings, the complete structured reports will be displayed to the radiologists to verify their accuracy. An example of the generated structured report of the following free-form report is shown in Figure 6.4:

> “US Abdomen : Normal liver echo pattern with no focal lesion demonstrated. No evidence of gall stones or dilatation of the bile ducts. Both kidneys are normal in size and echo pattern with no mass lesion or evidence of obstruction. Normal pancreas, aorta and spleen. Conclusion: Normal examination.”

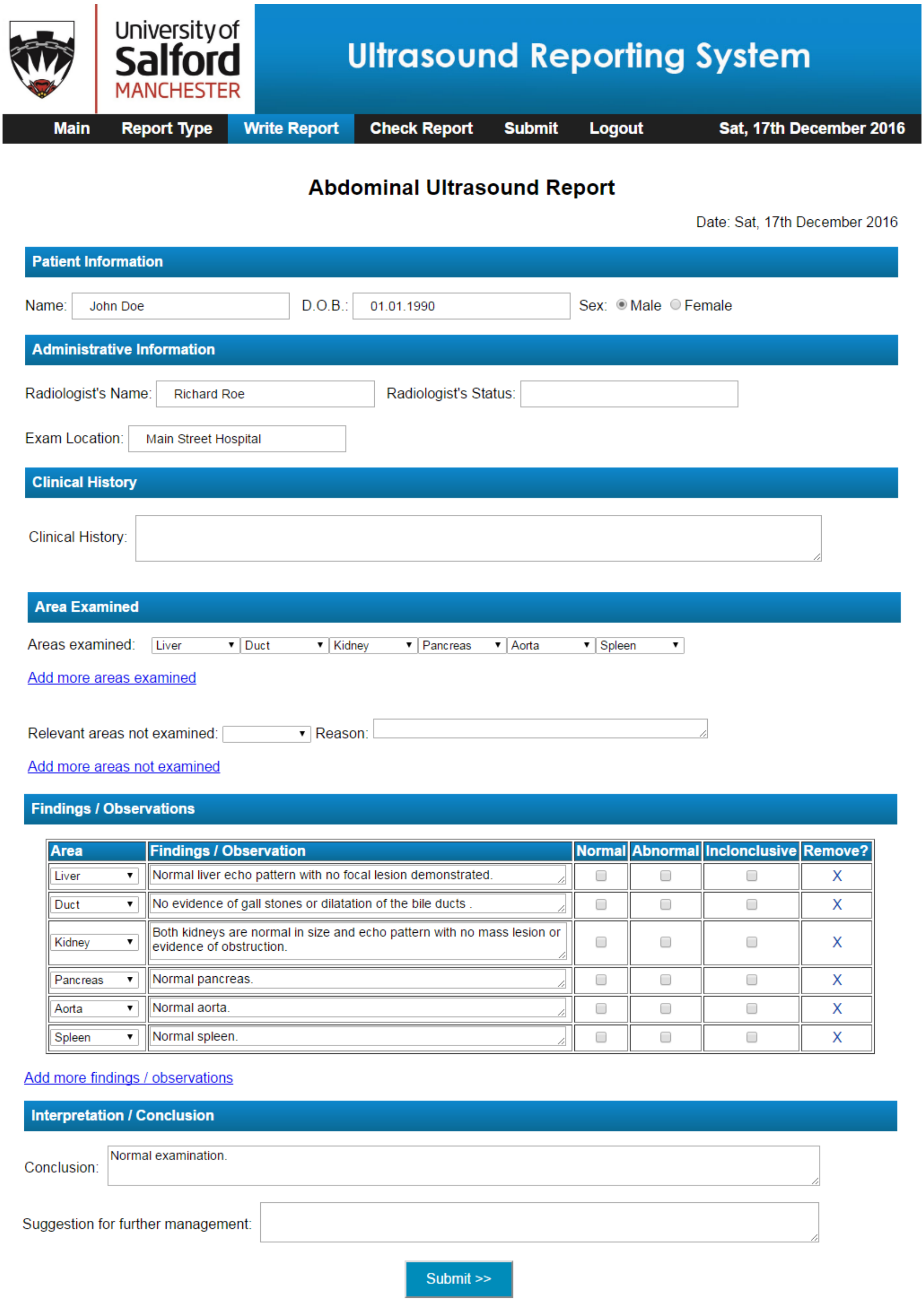


Figure 6.4: Example of a structured report produced from a free-form report

6.4 Results

This section discusses the result of the implementation of the medical ultrasound reporting system. It first presents the software quality of the overall system by going through four of the software quality aspects which are functionality, reliability, usability and security. It then discusses the evaluation of the structured report generator component which was performed by a pair of ultrasound specialists. This section also expresses the issues that were raised by the specialists during the evaluation.

6.4.1 Overall Software Quality

One of the important aspects of software quality is its functionality. The function of this system is to provide a flexible ultrasound reporting system that allows radiologists to prepare their report in any form but still produce a standardised final output. The system performs exactly this which gives the system high functionality. Another important aspect is the system's reliability. Even though the system is able to perform its exact function, however, there were still some errors in the production of the structured report. Thus, the system has low reliability at this development stage. However, since the final report will be checked and signed off by a radiologist before it can be submitted, the low reliability can be accepted. The system is also easy to use with instructions on every page which makes the system to comply with the usability aspect of software quality. Another important aspect of software quality is the security. At the moment, the security of the system is ensured by including a login page that gives access to the system to only authorised users such as the radiologists. However, since the system has not been implemented fully, this aspect cannot be tested. Overall, the ultrasound reporting system is of high quality and was developed by placing importance on the four aspects of software quality which are functionality, reliability, usability and security.

6.4.2 Evaluation of the Structured Report Generator

The evaluation of the structured report generator was conducted by approaching a pair of ultrasound specialists whereby all 100 sample ultrasound reports were provided together with their corresponding structured reports for review. Although all 100 reports were provided, only a sample of the reports was reviewed. However, both specialists assumed that the evaluation may well be applicable for the remainder of the reports.

In general, both specialists agreed that the translation of the free-form reports to structured form was good. Most information in the free-form reports have been transferred correctly into the structured form. However, there were four issues that have been raised by them.

First, the specialists mentioned that the structured reports have faithfully replicated the errors contained in the free-form reports. For example, there was a finding “normal size is normal parenchyma of the liver” where the specialists believed was not making any sense. Another pair of examples were “16 mm solid foci seen inferior to liver” and “The liver is of mild diffuse increase echogenicity” where the specialists argued that they were grammatically wrong and that the better words to be used were “focus” which is a singular of “foci” and “increased echogenicity” instead. This issue was indeed interesting and should not be taken lightly so that the final report produced is of good quality. However, at the moment, we are not concerned with correcting semantics and grammatical mistakes such as this. For this reason, the ultrasound reporting system was developed by ensuring that the final structured report was signed off by a radiologist to acknowledge that the information in the report was correct.

In addition to this, another issue that was raised was in terms of the area allocation of the findings. The specialists argued that the finding “No evidence of gallstones or dilatation of bile ducts” should be grouped under the area “Gallbladder” instead of “Duct”. An example of this finding can be seen in Figure 6.4. Following this, necessary changes have been made to ensure that the findings are appropriately grouped. The third issue that has been brought up by the specialists was regarding the related information that would have been better grouped into the same area. This is for instance in the finding “Liver echogenicity appears normal. No hepatic mass lesion is seen” which should not be separated into two sentences. This is indeed a limitation of our system for the moment because it failed to recognise that the word “hepatic” is related to the word “liver”. The limitation can later be improved by enhancing AUO to include this information. Finally, the specialists mentioned the difficulty of classifying the area examined of the finding “No intra- or extra-hepatic biliary dilatation” where it could fall into both the area “Liver” as well as “Duct”. Further discussions on this matter should be initiated for a decision to be made whether this finding would better be grouped under the area “Liver”, “Duct” or both.

As a conclusion, the specialists believed that the translation of the free-form report to structured form was executed appropriately. Although there were several errors that have been detected, these errors were mostly reproduced from the initial free-form reports which were not within the control of this study. Errors caused by the system however, were able to be reduced by making minor changes to the codes as well as by enhancing the information available in AUO.

6.5 Discussions: Non-Validated Automatic Classification of Ultrasound Findings

The annotation of the ultrasound reports with relevant classes in AUO was not only used to transform free-form reports to structured form. There was also an effort in attempting to utilise these annotation to automatically deduct whether each finding can be classified as "normal", "abnormal" or "inconclusive" depending only on the information available in each sentence. The automatic classification was executed using a rule-based approach with the hope of avoiding ambiguity and enabling the referring clinicians to better understand the reports. The rule-based approach taken in automatically classifying the findings in the reports has been presented as a poster in the British Medical Ultrasound Society (BMUS) Annual Meeting 2016. However, in validating the result of the automatic classification, a pair of ultrasound specialist that have been approached for the validation have raised up several issues which suggested that the automatic classification would not be possible because of the lack of information. The next subsections will explain further on the rule-based approach as well as the discussions with the specialists regarding this approach.

6.5.1 The Rule-Based Approach

In automatically classifying findings as "normal", "abnormal" and "inconclusive", the findings were first annotated with relevant classes from AUO such as "Organ", "Finding" and "Disease or Disorder". Several words were also listed as trigger words such as "normal", "good", "no" and "non". This method was similar to the work undertaken by Morioka et al. [59] where they automatically classify abdominal aorta ultrasound examinations as "pertinent negative", "pertinent positive" and "non-diagnostic" by referring to the list of trigger words. Relevant classes and words in the findings were then assigned a value based on the type of the word. These values were then calculated to determine the classification of each finding based on the rules below:

Normal Finding

All words annotated with the class “Organ” were assigned the value of 1. Positive words such as “normal”, “unremarkable” and “good” were also assigned the value of 1. On the other hand, negation words such as “no”, “not” and “non” were allocated the value of -1. In classifying a finding as normal, the multiplication of the value of each word in the sentence must equal to 1. Consider the examples below:

S1. The [left kidney (Organ) 1] appear [normal (positive) 1].
Total = 1 x 1 = 1

S1 contained the word “normal” which is a positive word with the value of 1 and the multiplication of this value with the word “left kidney” which also has a value of 1, returned a total of 1 which denoted that the sentence has a normal finding.

S2. Liver is [normal (positive) 1] in echo pattern with [no (negative) -1] focal [lesion (Finding) -1].
Total = 1 x (-1) x (-1) = 1

Even though S2 contained a negative word “no” and a word annotated with the negative class “Finding”, the multiplication of its values still resulted in a total of 1 which indicated that the finding was normal.

Abnormal Finding

Negative words such as “abnormality” or words annotated with negative classes, for instance “Finding”, “Disease or Disorder”, and “Biospecimen” were assigned the value of -1. The result of the multiplication returned a total of -1 which indicated that the sentence has an abnormal finding. Consider the examples below:

S1. There were tiny [calculi (Biospecimen) -1] within the [gallbladder (organ) 1].
Total = (-1) x 1 = -1

S1 contained the word “calculi” which was annotated with the class “Biospecimen” and has the value of -1. The multiplication of its value with the value of the organ which is 1, gives a total of -1 and resulted to S1 as being classified as an abnormal finding.

S2. Incidental 1.5 cm [simple cyst (Finding) -1] noted in the [left lobe of the liver (Body Part) 1].
Total = (-1) x 1 = -1

S2 has a negative value caused by the word “simple cyst”. Thus, this sentence was considered as reporting an abnormal finding. However, the word “Incidental” at the beginning of the sentence caused this sentence to be treated as a special case of abnormal finding which was incidental.

Inconclusive Finding

Words which signalled inconclusive findings such as “limited view” and “not seen” or words annotated with the “Visibility Descriptor” class were assigned the value of 10. In classifying a finding as inconclusive, the multiplication result of the values of each word in the sentence must be more than or equal to 10.

S1. Pancreas (organ) 1] [obscured (visibility descriptor) 10] by bowel gas.
Total = 1 x 10 = 10

S1 presents an inconclusive finding where the “visibility descriptor” indicates that the “organ” was “obscured”. Thus, no conclusion can be made regarding the “finding”. The result of the multiplication of the values of each relevant word in S1 equalled to 10 which denoted that this finding was inconclusive.

6.5.2 Reason for Non-Validation of Data

The rules presented in the previous section enabled the ultrasound reporting system to automatically classify findings as “normal”, “abnormal” or “inconclusive” based only on the words that the sentence contained. When applying these rules on each sentence of the findings in all 100 sample ultrasound reports, the system managed to classify most of the sentences without any problem especially for sentences with normal findings. In order to verify the result of the automatic classification, two medical ultrasound specialists have been approached to seek for their help in manually classifying each finding in the 100 reports so that it can be compared with the automatic classification completed by the system. However, they argued that the classification was not possible thus making it impossible to validate our data.

The specialists required the “normal”, “abnormal” and “inconclusive” classification to be defined and given certain criteria for them to be able to classify the findings because they argued that the classification of the findings was not binary. We defined “normal” findings as findings that do not need further attention while “abnormal” findings was defined as findings that needs further observation. “Inconclusive” findings on the other hand was defined as findings that could not be determined whether it is normal or abnormal because of reasons such as the organ was obstructed or was not seen and the ultrasound examination might need to be carried out again.

These definitions however were not enough to classify the findings in the reports as the specialists argued that the classification of the findings strongly depends on the context of the referral which includes information such as the patient's clinical history and the reason for referral. This information is important so that correct interpretation can be made. They provided two examples; (i) a gall bladder polyp could be the cause of symptoms in somebody with pain in that area, but would be an incidental finding in someone asymptomatic and (ii) an endometrial thickness of 6mm would be normal in a 25-year-old but will likely be abnormal in a 70-year-old. Therefore, the absence of the context of referral and other relevant information makes it difficult to classify the findings of the reports. This information was not made available to us because of the privacy and confidentiality of the patient's data. Therefore, as it stands, the specialists were not able to perform the manual classification because many of the information needed was absent.

The automatic classification was intended to assist the radiologists in completing the structured reports as they will have fewer fields to fill. However, the discussions with the medical ultrasound specialists have made it clear that this automation process was not possible because of the absence of the necessary information. If manual classification of the findings in the reports could not be completed by the ultrasound specialists, hence, it would not be possible for the system to do the same. As a result, this feature has been abandoned from the system for the time being.

6.6 Chapter Summary

Various ultrasound reporting styles and format have prompted the need to standardise them so that it is better read and understood by the referring clinicians. This chapter presented the development of the medical ultrasound reporting system including the development tools used. It explored the implementation of both RST and AUO in the transformation of the free-form reports to structured form. It first reviewed the 60 reports training data which recognised four different types of information that existed in the reports. This has resulted in the identification of the PREPARATION, RESTATEMENT, JOINT, LIST and ELABORATION relation as the RST rhetorical relations which are relevant to be applied in the transformation process.

Following this, the chapter presented all the steps taken in translating the free-form reports to structured form and how these relations were applied. Next, the chapter discussed the software quality aspect of the overall system and the evaluation of the structured report generator from a pair of specialists regarding the implementation. Their evaluation found the transformation as generally correct. Finally, the chapter discusses an attempt to automatically classify ultrasound findings into “normal”, “abnormal” and “inconclusive”. However, the data was not validated since the feedback from the ultrasound specialists suggested that this is not possible. This has resulted in the abandonment of the automatic classification process because there were insufficient amount of information for even the specialists to manually classify them.

Chapter 7

Conclusion and Future Directions

7.1 Introduction

The final chapter of this thesis summarises the research and provides a general description of the future directions that can be pursued after this research. This chapter will evaluate the results of this research against its aim and objectives as well as revisit the main contributions of this research. It will also restate the main findings of this research as well as some major discussions that were made.

7.2 Review of Research Contributions

The aim of this research is to overcome the problem of variations in ultrasound reporting by standardising the reports. This has been achieved by the development of a medical ultrasound reporting system that uses domain ontology as its knowledge base as well as RST in transforming free-form reports to structured form. In order to achieve the research aim, a set of research objectives have been established in the early stage of the research as stated in Chapter 1. This chapter will assess whether these objectives have been met by presenting what have been accomplished as well as highlighting the main contributions of this research.

7.2.1 Objective 1

- To determine the variations in ultrasound medical reporting styles.

In order to achieve this objective, 100 anonymised sample ultrasound reports have been collected through the Directorate of Radiology of the University of Salford. An audit style review was undertaken to understand the variations in ultrasound reporting by focusing on the reports' format, content, terminologies used and reporting styles among others. The review have demonstrated many variations in the reports where it can be identified that the reports were written by at least four different radiologists. Even reports that were acquired from the same department have different reporting styles. Several reports were itemised while most reports were written in free-form. The audit performed proved that there were many variations in the medical ultrasound reports.

7.2.2 Objective 2

- To identify the characteristics of a good quality ultrasound report.

A set of quality criteria was developed based on several literatures that consider the criteria which makes a good quality report as well as guidelines that were established by various radiology and ultrasound related bodies in the United Kingdom. These criteria were then used to classify the sample reports that have been acquired according to its quality. The characteristics were also used to construct a system architecture model of the reporting system as well as to design the structured report page in the system. The system architecture model was presented and published as a short paper in the 8th International Conference on Health Informatics (HEALTHINF) [94]. It was also presented as a poster in the 2015 UK Radiological Congress where a survey was also conducted to gather initial feedbacks from the medical practitioners and to validate the architecture of the system and the design of the structured report page.

7.2.3 Objective 3

- To develop and evaluate an abdominal ultrasound ontology using ontology reuse methodology.

For the system to be able to understand the reports that were submitted, an ontology of the abdominal ultrasound domain was needed as the knowledge base. There were several existing ontologies in the biomedical domain such as FMA, SNOMED CT, NCIT and RadLex. However, because of their size and broadness, it was realised that they were not suitable to be implemented in the medical ultrasound reporting system that we are developing. Therefore, an ontology specific to the abdominal ultrasound domain needs to be developed. Instead of developing this ontology from scratch, the possibility of developing it using the ontology reuse methodology was explored. We have proposed a four step methodology that begins with term extraction followed by ontology recommendation, term to concept mapping and ontology evaluation by domain experts. The methodology has allowed us to develop an AUO by reusing existing classes from NCIT, SNOMED CT and RadLex. The result from developing AUO using the ontology reuse methodology has been presented in the 21^{st} International Conference on Applications of Natural Language to Information Systems (NLDB) [95].

7.2.4 Objective 4

- To find out the effectiveness of using RST on medical ultrasound reports.

RST has been seen as a possible theory to transform free-form reports to structured form because of its ability to recognise rhetorical relations between text spans which ensure that reports are being transformed without any important information being lost. RST has been widely applied in texts such as news articles and personal letters but not in medical reports. To find out the effectiveness of using RST on medical ultrasound reports, we first reviewed the 100 sample reports that were obtained to find out which rhetorical relations that might exist between the text spans in the reports. Classic RST approach uses lexical and syntactic clues as well as POS tags to segment texts and assign rhetorical relations. However, ultrasound reports were usually written in grammatically incorrect form which makes the application of the classic RST approach as less effective. Therefore, we have combined ontology with RST and developed rules that will help segment text and recognise rhetorical relations between the text spans in the ultrasound reports. This approach has proven that it is effective to use RST together with ontology on medical ultrasound reports.

7.2.5 Objective 5

- To develop and evaluate the medical ultrasound reporting system that produces standardised ultrasound reports using an ontology and RST.

The accomplishment of Objective 3 and 4 have allowed for the development of a medical ultrasound reporting system that transforms free-form reports into structured form. The system uses AUO as its knowledge base and RST as the mechanism to do the transformation. Objective 4 has resulted in the identification of seven rhetorical relations that were found in the 60 sample reports used for training. However, in developing the system, only three out of the eight relations were needed. We have submitted all 100 reports to the system and the reports were successfully transformed to structured form. The result has been evaluated by a group of ultrasound specialists to verify whether the transformation was correct and no important information went missing. In general, the specialists agreed that the transformation was completed appropriately and that most information was there.

7.2.6 Objective 6

- To investigate the possibility of automatically classifying the findings in the ultrasound reports as normal, abnormal or inconclusive based only on the information available in the report.

In this research, the possibility of automatically classifying findings in the ultrasound reports as normal, abnormal or inconclusive was also investigated. This automatic classification was executed using a rule-based approach which endeavours to classify the findings based only on the information available in the reports. However, evaluations and discussions with ultrasound specialists have concluded that this was not possible because in order to classify findings as normal, abnormal or inconclusive, vital information such as the patient's personal information and clinical history as well as the reason for referral was necessary. This information was unfortunately not made available to us in this research because of its privacy and confidentiality. Therefore, it was not possible for the automatic classification to be performed at present.

7.3 Future Directions

This section will discuss several suggestions of unexplored areas as well as features that can be applied in the system in the future. The aim of these improvements is to further enhance the user experience so that the ultrasound reporting system will appeal to more radiologists. The improvements will also increase the possibility for new knowledges to be discovered.

One obvious enhancement that can be explored in the future to improve the quality of the medical ultrasound reporting system is the possibility of using ontology to automatically correct spelling errors in the reports before it can be processed by the system. Spelling errors can really impact the quality of the structured reports produced by the system. This was demonstrated in the results presented in Chapter 5 where a pre-processing phase needs to be introduced and therefore manually correct these mistakes in order to increase the accuracy rate of the text segmentation process in RST.

Tolentino et al. [88] have previously developed a UMLS-based spell checker in the domain of vaccine safety where their work can be an initial guide in developing a similar spell checker using ontology. In their work, they first cleaned the free-texts with 200 regular expressions that they have established in order to transform abbreviations and medical shortcuts into standard words. The free-text was then submitted to the system for it to recognise spelling mistakes by comparing it to the UMLS dictionary before providing a list of possible words. The list of possible words was then ranked and the one nearest to the word that was wrongly spelt will be selected.

In our case, the same approach could be taken where instead of comparing the spelling mistakes to a dictionary; it can be compared to the ontology. Another possible way that this could be conducted is by storing all the possible spelling mistakes for each word or terminology in the ontology. This way, the system would be able to refer to the ontology to correct the spelling mistakes without needing to generate a list of possible words. It is possible for the ontology to be trained with reports that contain spelling mistakes and whenever there are any new mistakes for the same word, the ontology will then be updated to include the new mistake.

Another improvement that can be included in the ontology is by adding more information regarding certain words which will allow them to be asserted as having a relation with another word. An example of such word is “hepatic” where it is defined as “pertaining to, affecting, or associated with the liver” in the annotation part of AUO. Currently, “hepatic” is a class under “anatomy qualifier” without direct relation to the class “liver”. If the ontology was improved where definitions were included in the description part of the ontology instead of the annotation, it would be possible to automatically assert that “hepatic” is related to “liver”. This will then improve the rhetorical relation assignment process in RST as well as in the structured report. For example, RST will be able to identify a JUSTIFY relation between “The liver has smooth outline and normal echotexture” and “No focal hepatic lesions” which it is currently not able to.

During the 2015 UK Radiological Congress, there was also a suggestion for the system to be more personalised to the user. This is another enhancement that can be included so that future implementation of this system can appeal to more radiologists and other medical practitioners. The system could be trained to learn the user’s pattern and preferences so that it can be personalised according to the reporting style of the user. This enhancement will require quite a substantial amount of study in order to understand the different reporting styles and preferences of radiologists. However, its implementation will not only appeal to the radiologists and medical practitioners using it but it could also reduce the report turnaround time as personalisation would help ease the radiologists' usage of the system.

Finally, the idea of automatically classifying findings into "normal", "abnormal" and "inconclusive" can still be realised. This however will require further studies and discussions with the radiologists and ultrasound specialist to understand what type of information is needed in order for the classification to be made automatically. More sample ultrasound reports need to be collected over time which will allow them to be used to train the system to perform the automatic classification. The collection of more sample ultrasound reports will not only benefit the automatic classification of the findings but they can also improve the RST segmentation process and the identification of rhetorical relations between text spans. The collection of more sample ultrasound reports will also allow for different patterns of texts to be recognised which could lead to the discovery of other rhetorical relations that exist between the texts in the reports.

Bibliography

[1] Harith Alani. Position paper: Ontology construction from online ontologies. In *Proceedings of the 15th International Conference on World Wide Web, WWW '06, Edinburgh, Scotland*, pages 491–495, New York, NY, USA, 2006. ACM.

[2] Maja Bärenfänger, Mirco Hilbert, Henning Lobin, and Harald Lüngen. OWL ontologies as a resource for discourse parsing. In *LDV Forum*, volume 23, pages 17–26, 2008.

[3] Jane Bates. *Abdominal Ultrasound: How, Why and When*. Churchill Livingstone Elsevier, Edinburgh, United Kingdom, 2nd edition, 2004.

[4] Douglas S Bell and RA Greenes. Evaluation of UltraSTAR: performance of a collaborative structured data entry system. pages 216–22. American Medical Informatics Association, January 1994.

[5] Douglas S Bell, RA Greenes, and P Doubilet. Form-based clinical input from a structured vocabulary: Initial application in ultrasound reporting. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 789. American Medical Informatics Association, 1992.

[6] Tim Benson. *Principles of Health Interoperability HL7 and SNOMED*. Health Informatics. Springer, London, 2010.

[7] BioPortal. API Documentation. http://data.bioontology.org/documentation.

[8] Anuradha Bodile and Manali Kshirsagar. Text mining in radiology reports by statistical machine translation approach. In *Communication Technologies (GCCT), 2015 Global Conference on*, pages 238–241. IEEE, 2015.

[9] G. Boland. Enhancing the radiology product: The value of voice-recognition technology. *Clinical Radiology*, 62(11):1127, 2007.

[10] Elena Paslaru Bontas. *A contextual approach to ontology reuse*. PhD thesis, Freie Universität Berlin, 2007.

[11] Elena Paslaru Bontas, Malgorzata Mochol, and Robert Tolksdorf. Case studies on ontology reuse. In *Proceedings of the 5th International Conference on Knowledge Management, IKNOW05*, 2005.

[12] Kalina Bontcheva and Yorick Wilks. Automatic report generation from ontologies: the MIAKT approach. In *International Conference on Application of Natural Language to Information Systems*, number Section 2. Springer, 2004.

[13] JML Bosmans, Lieve Peremans, Maurizio Menni, AM De Schepper, PO Duyck, and PM Parizel. Structured reporting: if, why, when, how and at what expense? results of a focus group meeting of radiology professionals from eight countries. *Insights into imaging*, 3(3):295–302, 2012.

[14] William E Brant and Clyde A Helms. *Fundamentals of diagnostic radiology*. Lippincott Williams & Wilkins, 2012.

[15] Enrico G. Caldarola, Antonio Picariello, and Antonio M. Rinaldi. An approach to ontology integration for ontology reuse in knowledge based digital ecosystems. In *Proceedings of the 7th International Conference on Management of Computational and Collective intElligence in Digital EcoSystems, MEDES '15, Caraguatatuba, Brazil*, pages 1–8, New York, NY, USA, 2015. ACM.

[16] Maria Angeles Capellades. Assessment of reusability of ontologies: A practical example. In *Proceedings of AAAI1999 Workshop on Ontology Management, AAAI Press*, pages 74–79, 1999.

[17] Lynn Carlson, Daniel Marcu, and Mary Ellen Okurowski. Building a discourse-tagged corpus in the framework of rhetorical structure theory. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue - Volume 16*, pages 1–10, 2001.

[18] JA Cramer, LB Eisenmenger, NS Pierson, HS Dhatt, and ME Heilbrun. Structured and templated reporting: An overview. *Applied Radiology*, 2014.

[19] G. H. Danton. Radiology Reporting, Changes Worth Making Are Never Easy. *Applied Radiology*, 5(May):19–23, 2010.

[20] Paul Doran, Valentina Tamma, and Luigi Iannone. Ontology module extraction for ontology reuse: An ontology engineering perspective. In *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management, CIKM '07, Lisbon, Portugal*, pages 61–70, New York, NY, USA, 2007. ACM.

[21] N Reed Dunnick and Curtis P Langlotz. The radiology report of the future: a summary of the 2007 Intersociety Conference. *Journal of the American College of Radiology*, 5(5):626–629, 2008.

[22] Hazel Edwards, Jane Smith, and Michael Weston. What makes a good ultrasound report? *Ultrasound*, 22(1):57–60, 2014.

[23] D. W. Ellis and J. Srigley. Does standardised structured reporting contribute to quality in diagnostic pathology? The importance of evidence-based datasets. *Virchows Archiv*, 468(1):51–59, 2016.

[24] Joseph P Erinjeri, Daniel Picus, Fred W Prior, David A Rubin, and Paul Koppel. Development of a Google-based search engine for data mining radiology reports. *Journal of Digital Imaging*, 22(4):348–356, 2009.

[25] Lorenzo Faggioni, Francesca Coppola, Riccardo Ferrari, Emanuele Neri, and Daniele Regge. Usage of structured reporting in radiological practice: results from an Italian online survey. *European Radiology*, pages 1–10, 2016.

[26] Vanessa Wei Feng and Graeme Hirst. Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics: Long Papers-Volume 1*, pages 60–68. Association for Computational Linguistics, 2012.

[27] Vanessa Wei Feng and Graeme Hirst. A linear-time bottom-up discourse parser with constraints and post-editing. In *ACL (1)*, pages 511–521, 2014.

[28] Mariano Fernández-López, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. Methodological guidelines for reusing general ontologies. *Data & Knowledge Engineering*, 86:242 – 275, 2013.

[29] Jeffrey Friedlin, Malika Mahoui, Josette Jones, and Patrick Jamieson. Knowledge

discovery and data mining of free text radiology reports. In *2011 First IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology (HISB)*, pages 89–96. IEEE, 2011.

[30] Benjamin David Godwin, Vlad Valentin Simianu, Frederick Thurston Drake, Manjiri Dighe, David Flum, and Puneet Bhargava. Is there a need to standardize reporting terminology in appendicitis? *Ultrasound Quarterly*, 31(2):92, 2015.

[31] Tianxia Gong, Chew Lim Tan, Tze Yun Leong, Cheng Kiang Lee, Boon Chuan Pang, CC Tchoyoson Lim, Qi Tian, Suisheng Tang, and Zhuo Zhang. Text mining in radiology reports. In *ICDM '08. Eighth IEEE International Conference on Data Mining*, pages 815–820. IEEE, 2008.

[32] Thomas R Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5-6):907–928, 1995.

[33] Ferris M Hall. The radiology report of the future. *Radiology*, 251(2):313–316, 2009.

[34] Nicholas J Hangiandreou. AAPM/RSNA Physics Tutorial for Residents: Topics in US: B-mode US: Basic Concepts and New Technology. *Radiographics*, 23(4):1019–1033, 2003.

[35] Hugo Hernault, Helmut Prendinger, David A DuVerle, Mitsuru Ishizuka, and Tim Paek. HILDA: a discourse parser using support vector machine classification. *Dialogue and Discourse*, 1(3):1–33, 2010.

[36] Annette J Johnson, Michael YM Chen, J Shannon Swan, Kimberly E Applegate, and Benjamin Littenberg. Cohort study of structured reporting compared with conventional dictation. *Radiology*, 253(1):74–80, 2009.

[37] Clement Jonquet, Mark A Musen, and Nigam H Shah. Building a biomedical ontology recommender web service. *Journal of Biomedical Semantics*, 1(S1), 2010.

[38] Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. A novel discriminative framework for sentence-level discourse analysis. In *Proceedings of the 2012 Joint*

Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, pages 904–915. Association for Computational Linguistics, 2012.

[39] Shafiq Joty, Giuseppe Carenini, and Raymond T Ng. CODRA: A novel discriminative framework for rhetorical analysis. *Computational Linguistics*, 2015.

[40] Shafiq Joty, Giuseppe Carenini, Raymond T Ng, and Yashar Mehdad. Combining intra-and multi-sentential rhetorical parsing for document-level discourse analysis. In *ACL (1)*, pages 486–496, 2013.

[41] Charles E Kahn, Curtis P Langlotz, Elizabeth S Burnside, John A Carrino, David S Channin, David M Hovsepian, and Daniel L Rubin. Toward best practices in radiology reporting. *Radiology*, 252(3):852–6, September 2009.

[42] Megan Katsumi and Michael Gruninger. What is ontology reuse? In R Ferrario and W Kuhn, editors, *Formal Ontology in Information Systems: Proceedings of the 9th International Conference, FOIS 2016*, volume 283, pages 9–22. IOS Press, 2016.

[43] Klaus Kuhn, Thomas Zemmler, Manfred Reichert, Christian Heinlein, and Dietmar Roesner. Structured data collection and knowledge-based user guidance for abdominal ultrasound reporting. In *Proceedings of the Annual Symposium on Computer Application in Medical Care*, page 311. American Medical Informatics Association, 1993.

[44] Anand Kumar and Barry Smith. Oncology ontology in the NCI thesaurus. In *Conference on Artificial Intelligence in Medicine in Europe*, pages 213–220. Springer, 2005.

[45] Curtis P Langlotz. RadLex: a new method for indexing online educational materials. *Radiographics*, 26(6):1595–1597, 2006.

[46] Curtis P Langlotz. Structured radiology reporting: Are we there yet? *Radiology*, 253(1):23–25, 2009.

[47] David B Larson, Alex J Towbin, Rebecca M Pryor, and Lane F Donnelly. Improving consistency in radiology reporting through the use of department-wide standardized structured reporting. *Radiology*, 267(1):240–250, 2013.

[48] Wendy Levinson. Physician-patient communication: a key to malpractice prevention. *JAMA*, 272(20):1619–1620, 1994.

[49] Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. Recognizing implicit discourse relations in the Penn Discourse Treebank. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 343–351. Association for Computational Linguistics, 2009.

[50] Juan Antonio Lossio-Ventura, Clement Jonquet, Mathieu Roche, and Maguelonne Teisseire. BIOTEX: A system for biomedical terminology extraction, ranking, and validation. In *Proceedings of the 2014 International Conference on Posters & Demonstrations Track - Volume 1272, ISWC-PD '14, Riva del Garda, Italy*, pages 157–160, Aachen, Germany, Germany, 2014. CEUR-WS.org.

[51] A. Maedche, B. Motik, L. Stojanovic, R. Studer, and R. Volz. An infrastructure for searching, reusing and evolving distributed ontologies. In *Proceedings of the 12th International Conference on World Wide Web, WWW '03, Budapest, Hungary*, pages 439–448, New York, NY, USA, 2003. ACM.

[52] William C. Mann and Sandra A. Thompson. Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3):243–281, 1988.

[53] Daniel Marcu. The rhetorical parsing of natural language texts. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, pages 96–103. Association for Computational Linguistics, 1997.

[54] Daniel Marcu. The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational linguistics*, 26(3):395–448, 2000.

[55] Dirk Marwede, Matthew Fielding, and Thomas Kahn. RadiO : A Prototype Application Ontology for Radiology Reporting Tasks Institute for Formal Ontology and Medical Information Science. *AMIA Annual Symposium Proceedings*, pages 513–517, 2007.

[56] Melinda McDaniel, Veda C Storey, and Vijayan Sugumaran. The role of community acceptance in assessing ontology quality. In Elisabeth Métais, Farid Meziane, Mohamad Saraee, Vijayan Sugumaran, and Sunil Vadera, editors, *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016,*

Salford, UK, June 22-24, 2016, Proceedings, pages 24–36. Springer International Publishing, 2016.

[57] Hylton B. Meire and Pat Farrant. *Basic Clinical Ultrasound: BIR Teaching Series No. 4*. The British Institute of Radiology, London, 1982.

[58] Jose LV Mejino Jr, Daniel L Rubin, and James F Brinkley. FMA-RadLex: An application ontology of radiological anatomy derived from the foundational model of anatomy reference ontology. In *AMIA Annual Symposium Proceedings*, volume 2008, page 465. American Medical Informatics Association, 2008.

[59] Craig Morioka, Frank Meng, Ricky Taira, James Sayre, Peter Zimmerman, David Ishimitsu, Jimmy Huang, Luyao Shen, and Suzie El-Saden. Automatic classification of ultrasound screening examinations of the abdominal aorta. *Journal of Digital Imaging*, 29(6):742–748, 2016.

[60] Sandeep S Naik, Anthony Hanbidge, and Stephanie R Wilson. Radiology reports: examining radiologist and clinician preferences regarding style and content. *American Journal of Roentgenology*, 176(3):591–598, 2001.

[61] Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, et al. BioPortal: Ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, pages W170–W173, 2009.

[62] NF Noy and DL McGuinness. Ontology development 101: A guide to creating your first ontology. pages 1–25, 2001.

[63] Michael O'Donnell. RSTTool 2.4: a markup tool for Rhetorical Structure Theory. In *Proceedings of the first international conference on Natural language generation-Volume 14*, pages 253–256. Association for Computational Linguistics, 2000.

[64] Society & College of Radiographers and British Medical Ultrasound Society. Guidelines For Professional Ultrasound Practice. (December):36–41, 2015.

[65] European Society of Radiology (ESR et al. Good practice for radiological reporting. Guidelines from the European Society of Radiology (ESR). *Insights into Imaging*, 2(2):93, 2011.

[66] United Kingdom Association of Sonographers (UKAS). Guidelines For Professional Working Standards Ultrasound Practice. (October):45–51, 2008.

[67] Adela Ortiz. Polionto: Ontology reuse with automatic text extraction from political documents. In *Proceedings of the 6th Doctoral Symposium in Informatics Engineering*, pages 309–320, 2011.

[68] JA Overton, Cesare Romagnoli, and Rethy Chhem. Open Biomedical Ontologies applied to prostate cancer. *Applied Ontology*, 6:35–51, 2011.

[69] a a O Plumb, F M Grieve, and S H Khan. Survey of hospital clinicians' preferences regarding the format of radiology reports. *Clinical radiology*, 64(4):386–94; 395–6, April 2009.

[70] Daniel K Powell and James E Silberzweig. State of structured reporting in radiology, a survey. *Academic radiology*, 22(2):226–233, 2015.

[71] Bruce Reiner and Eliot Siegel. Radiology reporting: returning to our image-centric roots. *American Journal of Roentgenology*, 187(5):1151–1155, 2006.

[72] Bruce I. Reiner. Customization of medical report data. *Journal of Digital Imaging*, 23(4):363–373, 2010.

[73] Cornelius Rosse and José L V Mejino. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics*, 36(6):478–500, December 2003.

[74] Daniel L Rubin. Creating and curating a terminology for radiology: Ontology modeling and analysis. *Journal of Digital Imaging*, 21(4):355–62, December 2008.

[75] Thomas Russ, Andre Valente, Robert MacGregor, and William Swartout. Practical experiences in trading off ontology usability and reusability. In *Proceedings of the 12th Workshop on Knowledge Acquisition, Modeling and Management, KAW '99*, pages 16–21, 1999.

[76] Silvano Samartine, Linda White, Dennis McKeon, and Murray Becker. Enhancing structured reporting: Improving quality by tailoring the report to the clinical scenario. *Journal of the American College of Radiology*, 12(8):845–847, 2015.

[77] Lawrence H Schwartz, David M Panicek, Alexandra R Berk, Yuelin Li, and Hedvig Hricak. Improving communication of diagnostic radiology findings through structured reporting. *Radiology*, 260(1):174–81, July 2011.

[78] Leslie M. Scoutt. Preface. *Ultrasound Clinics*, 2(3):xi–xii, July 2007.

[79] Tejal Shah, Fethi Rabhi, and Priyadip Ray. OSHCO: A cross-domain ontology for semantic interoperability across medical and oral health domains. In *Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications Services, HEALTHCOM '13, Lisbon, Portugal, 9-12th October, 2013*, pages 460–464. IEEE, Oct 2013.

[80] Tejal Shah, Fethi Rabhi, Priyadip Ray, and Kerry Taylor. A guiding framework for ontology reuse in the biomedical domain. In *47th Hawaii International Conference on System Sciences, HICSS 2014, 6-9 Jan, 2014*, pages 2878–2887. IEEE, 2014.

[81] Elena Simperl. Reusing ontologies on the semantic web: A feasibility study. *Data & Knowledge Engineering*, 68(10):905 – 925, 2009.

[82] Chris L Sistrom and Janice Honeyman-Buck. Free text versus structured format: information transfer efficiency of radiology reports. *American Journal of Roentgenology*, 185(3):804–812, 2005.

[83] Chris L Sistrom and Curtis P Langlotz. A framework for improving radiology reporting. *Journal of the American College of Radiology*, 2(2):159–167, 2005.

[84] D Sluis, KP Lee, and N Mankovich. DICOM SR-integrating structured data into clinical information systems. *Medicamundi*, 46(2):31–36, 2002.

[85] Barry Smith. Ontology: philosophical and computational. In Floridi L, editor, *The Blackwell Guide to the Philosophy of Computing and Information*. Blackwell Publishers, Oxford, 2003.

[86] Radu Soricut and Daniel Marcu. Sentence level discourse parsing using syntactic and lexical information. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology-Volume 1*, pages 149–156. Association for Computational Linguistics, 2003.

[87] Maite Taboada and William C Mann. Rhetorical structure theory: Looking back and moving ahead. *Discourse studies*, 8(3):423–459, 2006.

[88] Herman D Tolentino, Michael D Matters, Wikke Walop, Barbara Law, Wesley Tong, Fang Liu, Paul Fontelo, Katrin Kohl, and Daniel C Payne. A UMLS-based spell checker for natural language processing in vaccine safety. *BMC medical informatics and decision making*, 7(1):3, 2007.

[89] Leanne Tran, Anupma Wadhwa, and Erika Mann. Implementation of structured radiology reports. *Journal of the American College of Radiology*, 13(3):296–299, 2016.

[90] Mike Uschold, Peter Clark, Mike Healy, Keith Williamson, and Steven Woods. An experiment in ontology reuse. In *Proceedings of the 11th Knowledge Acquisition Workshop, KAW '98*, 1998.

[91] A Wallis and P McCoubrie. The radiology report–are we getting the message across? *Clinical radiology*, 66(11):1015–22, November 2011.

[92] David L Weiss and Curtis P Langlotz. Structured reporting: Patient care enhancement or productivity nightmare? *Radiology*, 249(3):739–747, 2008.

[93] William C Mann and Maite Taboada. Rhetorical Structure Theory. http://www.sfu.ca/rst/index.html.

[94] Nur Zareen Zulkarnain, Gillian Crofts, and Farid Meziane. An architecture to support ultrasound report generation and standardisation. In *Proceedings of the 8th International Conference on Health Informatics, HEALTHINF 2015, Lisbon, Portugal, 12-15th January, 2015*, pages 508–513. SCITEPRESS, 2015.

[95] Nur Zareen Zulkarnain, Farid Meziane, and Gillian Crofts. A methodology for biomedical ontology reuse. In Elisabeth Métais, Farid Meziane, Mohamad Saraee, Vijayan Sugumaran, and Sunil Vadera, editors, *Natural Language Processing and Information Systems: 21st International Conference on Applications of Natural Language to Information Systems, NLDB 2016, Salford, UK, June 22-24, 2016, Proceedings*, pages 3–14. Springer International Publishing, 2016.

Appendix A

Ultrasound Reporting Quality Criteria Guidelines

General Report Style

1. The report should be written in tabulated or itemised form [19, 22, 69]. Example [69]:

Liver	Normal
Gallbladder	Normal. No gallstones seen.
Biliary tree	Normal. No intrahepatic or CBD dilatation.
Pancreas	Normal
Kidneys	Normal. Right 8.4 cm in bipolar length, left 9 cm.
Comment	No cause for pain identified.

2. The style of the report should contain suitable headings, breaking the content into several sections.

3. The structure of the report should be logical, concise, clear, accurate and easily understood [22, 66].

4. Though written in tabulated or itemised form, the report should be detailed. Avoid leaving out important comments [66].

5. Language used should not be ambiguous [66]. Avoid using terms that are too technical and too lay [22].

Report Content

The content of an ultrasound report should include information such as stated below [64, 22, 41, 91].

Administrative Information

1. The report should include the name and status of the sonographer / clinician issuing the report [66].

2. The date and facility of the ultrasound examination should be recorded [41].

Patient Information

1. The report should include basic information of the patient such as name, sex and date of birth [41].

Clinical History

1. If available, medical history should be included [41].

2. This section should include the reason for referral written in concise form [91].

3. If insufficient clinical information is available, this should be recorded in the report [91].

4. The clinical question that needs to be answered by the examination should be identified and recorded [91].

Area Examined

1. All area examined must be recorded in the report [22].

2. If a relevant organ has not been fully examined, the reason(s) should be stated [66].

3. Any limitations or action taken should be reported [66].

Findings / Observations

1. Observations should be classified as either normal, unequivocal abnormal; whether expected or unexpected, equivocal findings and normal variants [64].

2. Avoid including measurements in the report unless it is defined whether the measurement indicates normal or abnormal findings [22].

3. Avoid using any ultrasound terminology (e.g. transonic, echogenic etc.) which are meaningless to non-ultrasound users [64].

4. Irrelevant information should be avoided [66].

Interpretation / Conclusion

1. A succinct conclusion should be included the end of the report [66].

2. The conclusion should directly answer the clinical question identified [22, 91].

3. Where no conclusion is possible, consider to include alternative explanations for the ultrasound appearances [66].

4. The report should include suggestion for further management [22, 66, 69].

5. Where appropriate, the report should also include differential diagnosis taken [22, 91].

6. If there is any uncertainity or doubt in the diagnosis, it should also be made known [64].

Appendix B

Ultrasound Reporting System Survey

Variations in writing style; content; format; and terminology in ultrasound reporting have an impact on the value of the report. In our research, we are developing an ultrasound reporting system that will generate and standardised these reports. One of the main interfaces in the system is the structured form page (see Figure B.1). Through this brief survey, we would like to get your opinion on the structured form interface. Your response will help us understand your preferences and guide us in improving the interface.

1. How would you best describe your industry or field of work? (Select only one)
   - ○ Radiographer
   - ○ Nurse
   - ○ Medical Doctor
   - ○ Healthcare Product Consultant
   - ○ Other (please specify): __________
2. Which of the following groups contains your age?
   - ○ Under 18
   - ○ 18-24 years old
   - ○ 25-34 years old
   - ○ 35-44 years old
   - ○ 45-54 years old
   - ○ 55 and older

3. On average, how many hours per week do you spend using healthcare / medical system in you work?

    - ○ 0-1 hours
    - ○ 1-5 hours
    - ○ 5-10 hours
    - ○ 10-20 hours
    - ○ More than 20

4. Do you agree that variations in reporting styles impact the interpretation of the report? (Please circle one answer)

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

5. Do you agree that variations in reporting styles impact patient diagnosis? (Please circle one answer)

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

For questions 6-11, please refer to the image in Figure B.1. (Please circle one answer)

6. Do you agree that the form allows for information to be read easily?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

7. Do you agree that the form contains enough information for an ultrasound report?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

Is there any other information that needs to be added / discarded from the form? Please state:

8. Do you agree that the form is easy to fill in?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. Below is an example of an ultrasound finding.

   > "The liver exhibits appearances in keeping with patchy fatty infiltration"

   Would you put this finding under abnormal or normal findings? Or do you require another field for this finding?

10. How important is the flexibility to add or delete fields in the form?

Very Important	Important	Neutral	Slightly Important	Unimportant

11. What are your overall comments on the interface of the structured form?

12. How likely are you going to use this interface?

Very Likely	Somewhat Likely	Neutral	Somewhat Unlikely	Very Unlikely

13. One of the features of our ultrasound reporting system is it allows practitioners to write ultrasound report in both free form and structured form. Do you agree that this feature would attract practitioners to use the system?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

14. Please provide any other additional comments that you would like to add:

15. May our research team be in contact with you for further feedback? If so, please enter your contact information below:
    Name:
    Company:
    Email Address:
    Phone Number:

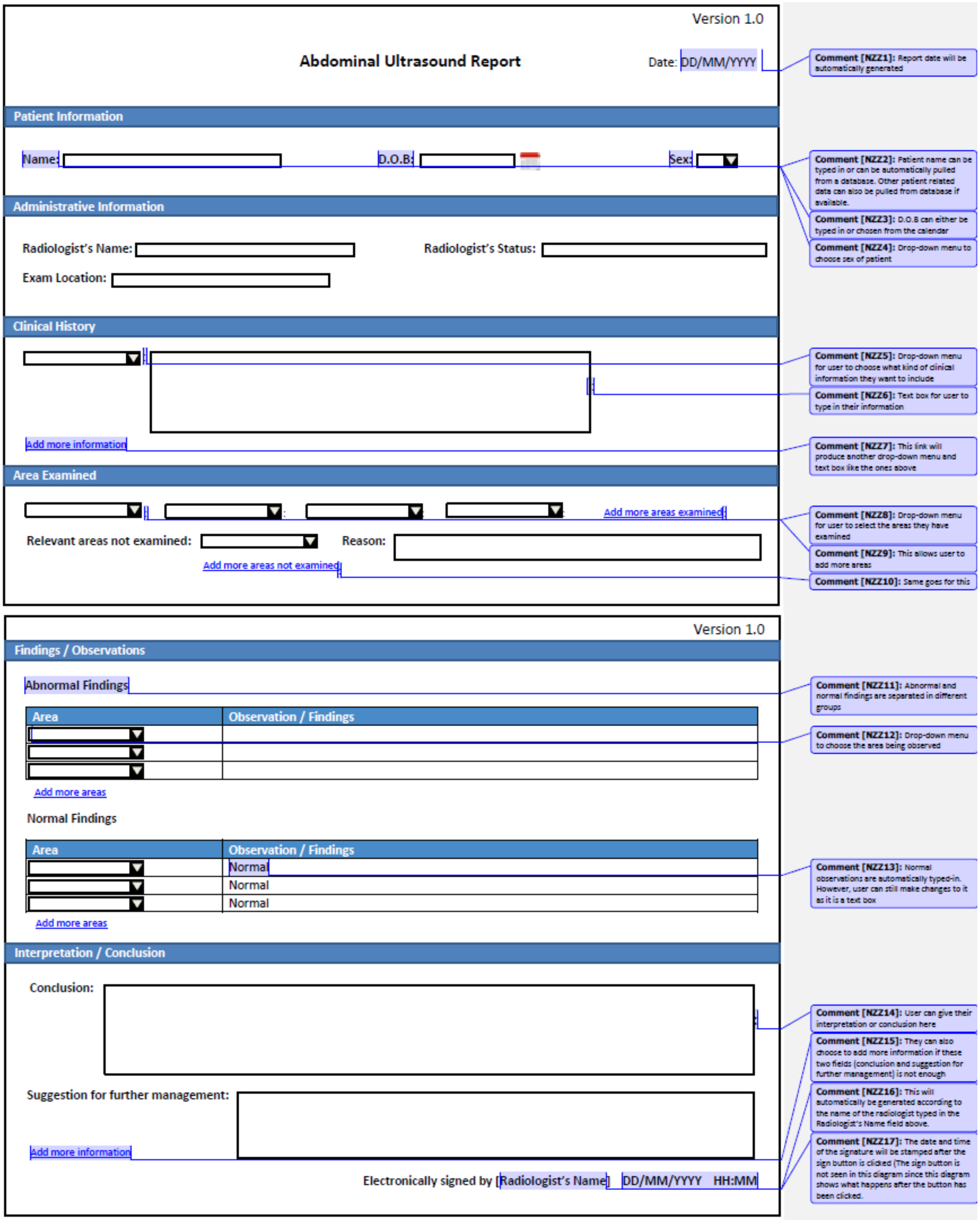


Figure B.1: Screenshot of the structured report for questionnaire

Appendix C

Medical Ultrasound Reporting System User Interface Design

Following are the screenshots of all the pages in the Medical Ultrasound Reporting System. The user interface design follows the quality criteria guidelines presented in Appendix A.

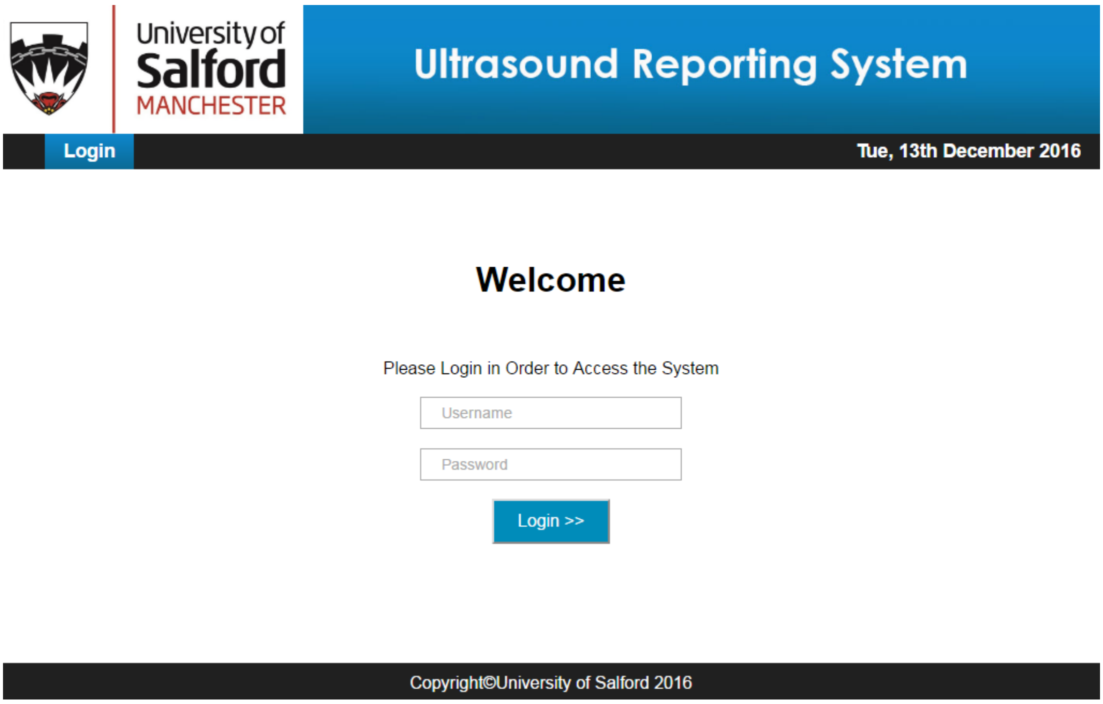


Figure C.1: Login page

Figure C.2: Main page

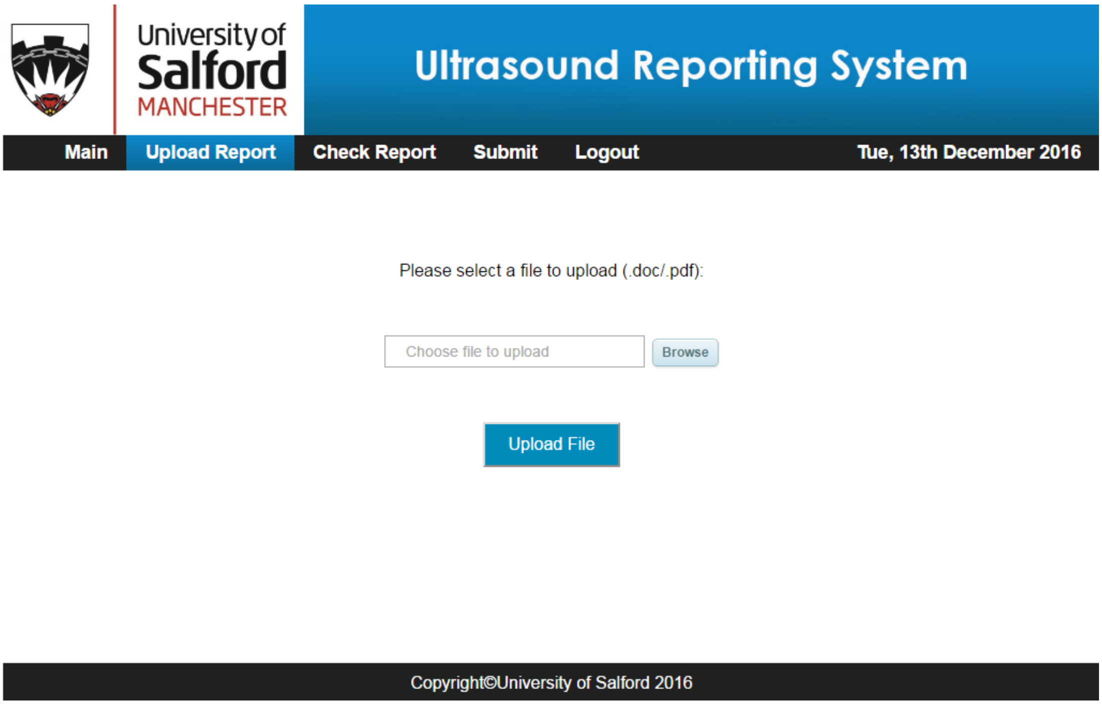


Figure C.3: Upload report page

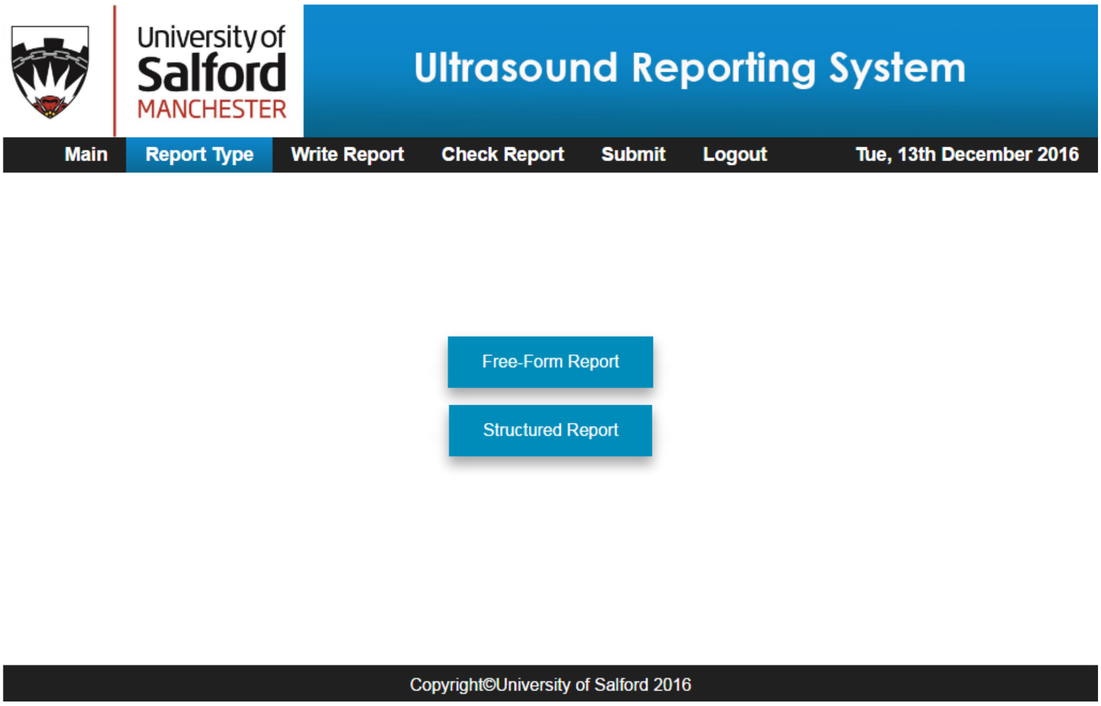


Figure C.4: Create report page

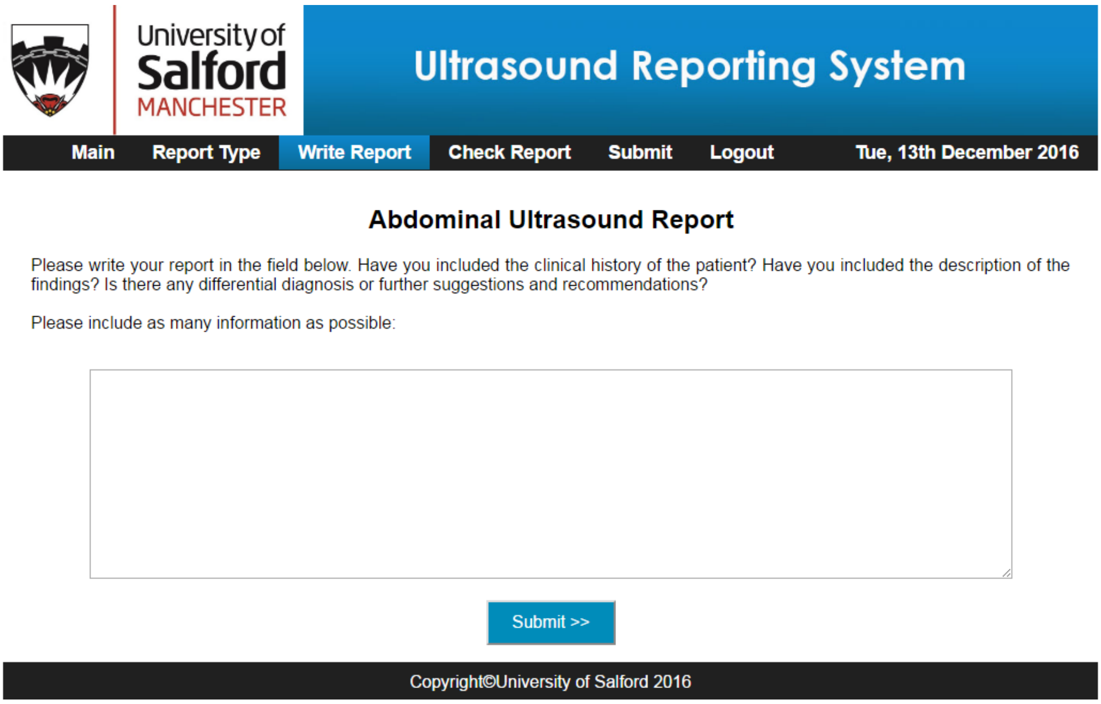


Figure C.5: Free-form report page

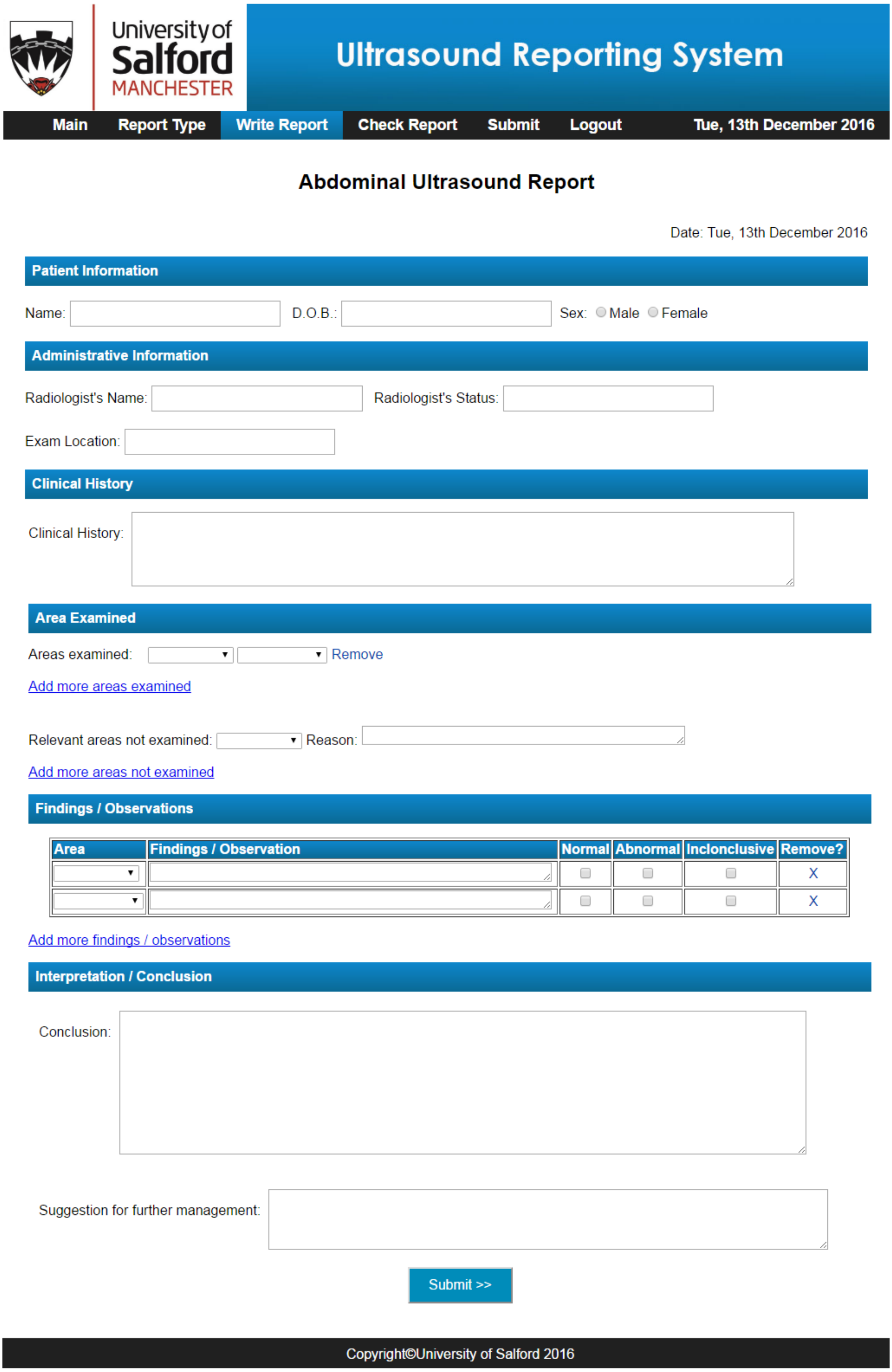


Figure C.6: Structured report page