



**Mapping Adverse Drug Reactions in Canadian Vigilance
Adverse Reaction Database to various standardized
terminologies and determining an ideal terminology to be
used as a supplement to MedDRA**

By:
Soumya Shastri
B00752441
soumyashastri@dal.ca

Performed at:
Faculty of Computer Science, Dalhousie University, 6050 University Avenue,
Halifax, NS B3H 1W5

**In Partial Fulfillment of the Requirements of the Master of Health Informatics Program,
Dalhousie University, Halifax, NS**

***Report of Internship for the Period May 29th, 2017 to August 25th, 2017
Date Submitted: August 18th, 2017***

Acknowledgement and Endorsement

I am using this opportunity to express my deepest gratitude and special thanks to Dr. Samuel Stewart for his supervision and support. His guidance and encouragement were extremely valuable for the project. I would also like to thank my project partner Rahul Thandavan for his support and contribution throughout the project.

I would also like to express my deepest sense of gratitude to Dr. Raza Abidi who took out time to hear and guide me throughout the project and all other faculty members of Health Informatics for preparing me for this internship.

This report has been written by me and has not received any previous academic credit at this or any other institution.

(Electronically signed)

Soumya Shastri

Executive Summary

Canada Vigilance Program is Health Canada's post-market surveillance program that gathers and evaluates reports of suspected adverse drug reactions to various health products. This facilitates Health Canada to monitor safety profiles of these health products once they are introduced in to the market. This helps to keep track of the products that they continue to outweigh the risk associated with them.

This program is supported by seven Canada Vigilance Regional Offices [1] who provide a regional point of contact for health professionals and consumers. All the collected reports and their data are then stored in the Canada Vigilance Adverse Reaction Database. The database has a total of 12 tables consisting of information regarding Drug Products, Drug product ingredients, Indications for which they are used and reactions reported for them and their severity.

Adverse reactions in the database are mostly coded using MedDRA (Medical Dictionary of Regulatory Activities) which has standardized terms for symptoms, signs, diseases, syndromes and diagnoses. For the purpose of the project, only the data related to Crohn's Disease was analyzed. One of the focus of the project was to map these adverse reactions and drug products using BioPortal REST API to various terminologies like MedDRA, SNOMED CT, MeSH, UMLS and RXNorm (mapped to respective codes, their parents and hierarchy levels) and store this data in the database using python programming language. Other focus of the project after mapping was to evaluate the results of the mappings performed and determine which is the ideal terminology which can be used as an adjunct to MedDRA. The reason for finding a terminology as a supplement to MedDRA is that MedDRA has a shallow hierarchical structure with only 4 levels. Due to lack of details in these levels of hierarchy it restricts the scope and lessens the chances of getting more fine-grained information.

The results of these evaluations showed that SNOMED CT gave the best results when used along with MedDRA. SNOMED CT has a more broader scope and very detailed hierarchal structure (14 levels) which classifies the term more finely. When a particular term was searched in MedDRA the results received were based on text matching rather than concept matching. But when the same was searched using SNOMED CT hierarchal structure additional terms were returned based on related terms of the original term. This finding can be very valuable in the process of signal detection of identifying drug-patterns in Pharmacovigilance as it will enhance the matching of the terms and giving more accurate results.

Table of Contents

Acknowledgement and Endorsement

Executive Summary

1. Introduction.....	5
1.1 Background	5
1.2 Project Objectives	7
2. Description of organization.....	8
3. Work Description.....	8
3.1 Major Tasks.....	8
3.1.1 Perform mapping of terms to five lexicons MedDRA, SNOMED CT, MeSH, UMLS and RXNORM.....	8
3.1.2 Determining ideal terminology which can be used as an adjunct to MedDRA to find related terms for improving reporting and signal detection.	13
4. Relation to Health Informatics.....	15
6. Conclusion	17
7. Recommendations.....	17
8. References.....	19
APPENDIX.....	21
APPENDIX 1: Sample of Drugs for Crohn’s Disease	21
APPENDIX 2: Sample of reactions for Crohn’s Disease	22
APPENDIX 3: Python code for retrieving data	23
APPENDIX 4: JSON file used BioPortal.....	24
APPENDIX 5: Examples of additional terms returned.....	25

1. Introduction

An Adverse Drug Reaction (ADR) is described by World Health Organization (WHO) as a 'response to a medicine which is noxious and unintended and occurs at a dose normally used in man' [2]. Previous studies have estimated that 25% of the hospital admissions and Emergency Department visits are drug related [3,4]. These adverse reactions are primarily reported by the health professionals and consumers which are then recorded using controlled dictionaries like MedDRA.

The purpose behind reporting and recorded this data is to help in Signal Detection for pharmacovigilance. Signal detection can be defined as identifying similar reactions to a particular drug as soon as they are reported. This is to promote early detection of any indication caused due to unexpected ADRs. With the help of these results further investigation can be carried out on the reported drug to prevent any other reports of the same reaction.

The project is based on the Canadian Adverse Drug Reaction(ADR) Database which is a repository of the reported ADRs. The database contains detailed information about the reported reactions for different diseases. For the purpose of this project the author focused only on the ADRs and other data related to Crohn's Disease. The focus of the project is to map the existing terms in the database to their respective codes, parents and hierarchy from terminologies like MedDRA, SNOMED CT, MeSH and UMLS. The other focus of the project is to identify an ideal terminology which can be a supplement to MedDRA to give better qualitative results by comparing all the mappings being performed.

1.1 Background

Crohn's Disease

Crohn's Disease is also known as the inflammatory bowel disease. Studies have shown that Crohn's Disease occurs at a rate of 1-2 per 1000 in countries like US, Canada and northern Europe [6]. The onset of the disease varies depending on the age, initial sign and symptoms are seen during teens and twenties and secondary symptoms which are less common seen during fifties or sixties.

In relation to the Canadian Vigilance Adverse Reaction Database, there a total of **3788** adverse reactions reported for Crohn's Disease and almost **166** drugs being used for the treatment. There are **30980** distinct Report IDs for Crohn's and almost double **92017** Reaction IDs. The difference can be due to the fact that each reaction reported in the database has been given a unique number so one Report can have many Reaction IDs.

BioPortal

BioPortal is a tool developed by the National Centre for Biomedical Ontology as a community based biomedical repository [7]. It provides users access to rich collection of over 300 ontologies in the biomedical discipline through REST API, SPARQL and through differentiable URI's [8]. Mgrep serves as the concept mapping engine of NCBO-BioPortal.

This tool is capable of direct concept mapping and can also generate semantically expanded annotations which gives a level of flexibility to this tool. This tool can be configured to produce outputs in the form of text or XML or JSON format [8,9].

Medical Dictionary of regulatory Activities(MedDRA)

MedDRA is a rich and highly specific standardized terminology used for recording and reporting ADRs for both pre-marketing and post-marketing drug surveillance. MedDRA is designed in such a way that it includes terms from most of the dictionaries for ADRs. It allows precise description of the reactions and issues related to them [10].

MedDRA terms are structured into five hierarchal classes:

1. System Organ Class(SOC)
2. High Level Group Terms (HLGT)
3. High Level Terms (HLT)
4. Preferred Terms (PT)
5. Low Level Terms (LLT)



In the above hierarchy, the Lowest Level Term(LLT) is the most specific level or instance/child of the term. PT level represent a single medical concept which can be a symptom, sign, disease, diagnosis, therapeutic indication, surgical or medical procedure or medical family history [10]. The next level HLT is based on anatomy, pathology, physiology, etiology or function and these terms are then linked to the next level which is the HLGT.

This structured hierarchy of MedDRA allows grouping the terms under different classes and facilitate easy navigation. However, this hierarchy structure is also one of the limitation to MedDRA as its restrictive in approach and not very detailed. This point has been further proved and explained in the later sections of the report and also a potential solution of how the efficacy of MedDRA can be improved.

SNOMED CT

SNOMED CT (Systematized Nomenclature of Medicine - Clinical Terms) is the most comprehensive and multilingual terminology [11]. It consists of clinically validated data which can be used for representing clinical data. It covers broad range of clinical data varying from procedures, findings, diseases which are used in Electronic Health Record (EHR). SNOMED CT has a richer hierarchal structure as compared to other terminologies. There are up to 14 levels on the hierarchy which makes it more finer grained than other terminologies.

UMLS

The Unified Medical Language System (UMLS) is a terminology integration system developed at National Library of medicine (NLM). UMLS integrates up to 150 dictionaries including MedDRA, SNOMED CT and MeSH [12]. It also records the relations associated with the terms along hierarchy and mapping relations. The terms in UMLS are given a Concept Unique Identifier (CUI) which is unique to each term.

In this project, the author has used UMLS Metathesaurus software known as MetaMorphosys for retrieving the hierarchies of SNOMED CT, MeSH and MedDRA. The software has all the vocabularies integrated in it as well as the relations and hierarchy present as text files which can be loaded in the any database using MySQL.

RXNorm

RXNorm is the terminology which provides normalized names for clinical drugs both generic and branded. It is also used as tool for supporting semantic interoperation between drugs and pharmacy knowledge based systems [13]. The main purpose of RXNorm is to allow computer systems to efficiently communicate drug-related information between each other.

MeSH

Medical Subject Headings (MeSH) is a vocabulary designed in a way to make the process of indexing and cataloguing biomedical information simpler using a hierarchical structure made from sources like MEDLINE/PUBmed [14].

1.2 Project Objectives

The overall objective of the project was divided into two parts:

- i. Mapping the terms in the Canadian Vigilance Adverse Database to five main lexicons (MedDRA, SNOMED CT, MeSH, UMLS and RXNORM) and compare the outcome of the mapping performed:*
 - Mapping of terms to their respective codes and their parents. Also, mapping the term's hierarchy (in which hierarchy level the term is present).
 - The database terms consist of both drugs and reactions for which different lexicons were taken into consideration. MedDRA is especially designed for the adverse reactions so it was used only for the reactions, RXNorm being designed just for the drugs was used for drugs only. Whereas, SNOMED CT, MeSH and UMLS were used for both drugs and reactions as they include both drugs and reactions.

- ii. *Determining the ideal terminology which can be used in adjunct to MedDRA to improve the grouping of MedDRA terms for signal detection:*
 - Evaluating all the terminologies along with the mappings and comparing which one gives better result when used with MedDRA.

2. Description of organization

Canada Vigilance Program is Health Canada's post-market surveillance program that gathers and evaluates reports of suspected adverse drug reactions to various health products. This facilitates Health Canada to monitor safety profiles of these health products once they are introduced in to the market. This helps to keep track of the products that they continue to outweigh the risk associated with them.

This program is supported by seven Canada Vigilance Regional Offices who provide a regional point of contact for health professionals and consumers. All the collected reports and their data are then stored in the Canada Vigilance Adverse Reaction Database. The database has a total of 12 tables which consist of information regarding Drug Products, Drug product ingredients, Indications for which they are used and reactions reported for them and their severity.

3. Work Description

The author was assigned weekly tasks throughout the internship and was required to present the progress and the work done to the supervisors at least once in 4 weeks.

3.1 Major Tasks

3.1.1 Perform mapping of terms to five lexicons MedDRA, SNOMED CT, MeSH, UMLS and RXNORM.

- I. *Determining a semantic mapping tool for mapping the terms in database*
Performing a literature review for evaluating the current biomedical automatic annotators and then selecting the ideal tool which can easily map the terms to different terminologies.

In the literature review, the author evaluated tools like MetaMap, Aber-OWL, NCBO-BioPortal, Ontology Lookup Service and ConceptMapper. After analyzing the tools mentioned, NCBO-BioPortal was selected as the ideal one for the project. The reason for

this selection was the efficiency of BioPortal in the text mining and concept mapping has already been proved in the previous studies. Also, the ability to be retrieved by using REST API is also a significant feature for constructing an automatic mapping tool. In addition to this, the outputs in the form of JSON files can be easily used in programming languages like Python to retrieve the results and save them to the database.

II. *Identifying and create mapping fields for new data*

Determining the fields (MedDRA code, ID, parentID, CUI and hierarchy) for new tables to be created in the database to which the drugs and will be mapped. Also, finalizing the process by which the new data will be retrieved and added to the database.

METHOD

- *Adding tables to the database with identifies mapping fields*

The database had 12 existing tables to which the author added 20 new tables using MySQL to make the database more efficient. The new 20 tables consist of *mapping tables* (adverseMEDDRAMap, adverseMESHmap, adverseSNOMEDmap, adverseUMLSmap, drugMap) which recorded the term code in different terminologies and linking codes to the existing data and the new data with the REACTION_IDs(already present), *term tables* (MEDDRAterm, advMESHterm, advSNOMEDterm, advUMLSterm) recorded the codes for the terms/reactions, drug term tables (RXNORMterm, MESHterm, UMLSterm, SNOMEDterm) recorded codes for the drugs in different terminologies, tree tables (RXNORMtree, MESHtree, UMLStree, SNOMEDtree, MEDDRAtree) recorded the childID and parent ID of the terms and hierarchy tables (MEDDRA_hier1, SNOMED_hier1, MESH_hier) recorded the hierarchy structure of the reaction/drugs in the terminologies.

III. *Performing mapping for the drugs and reactions for Crohn's Disease*

Mapping the terms to five lexicons i.e. MedDRA, SNOMED CT, MeSH, RXNorm and UMLS.

METHODS

- *Extracting the data using the BioPortal REST API and storing the data*

To extract the data from the BioPortal Annotator, Python programming language was used along with BioPortal REST API to parse the JSON file containing links to all the information about the searched term in a particular terminology. The code extracts the required information (terminology code, Concept Unique Identifier(CUI), parents, mappings) from the JSON file and inserts/store it in to the database under the preferred table. As the project focused only on a small portion of the Canadian Vigilance database, the code was written in a flexible manner that it could be extended to the whole database.

The approach used to retrieve codes from all the terminologies was same. Although, in some cases the mapping had to be performed from going through one terminology to another. The reason for choosing this approach was that few terms were being mapped to terms which did not exactly meant the same as the original term.

- *Conversion of terms to MedDRA*
As most of the reactions in the database were recorded as MedDRA terms, maximum number of terms were mapped directly to MedDRA (almost 99%).
- *Conversion of terms to RXNorm*
Most of the drug names present in the database were brand names rather than generic, therefore RXNorm was able to map maximum number of drugs except few drugs.
- *Conversion of terms to SNOMED CT*
For mapping the terms in SNOMED CT, other terminologies like MedDRA and RXNorm were used. SNOMED CT was also used directly but most of the terms returned were either not accurate or were in split form. This problem has been discussed in detail in the Problems encountered section of the report.
- *Conversion of terms to MeSH*
Similar to SNOMED CT, for MeSH, MedDRA mappings were used to map the terms. Most of the terms were not found in the MeSH terminology when searched directly but when searched through MedDRA more number of terms were mapped. For the drugs, the mappings were done through RXNorm as in majority the drugs had their brand names in the database and only the ones with generic ones were found in MeSH.
- *Conversion of terms to UMLS*
For UMLS a different approach was taken, the Concept Unique Identifier(CUI) code which is unique to each concept was used to map the terms. For example, the term *Abdominal distension* is uniquely identified by the code *C0000731*. This code was then used for mapping this term to different terminologies. The terms were mapped to their Semantic type as the parent of the terms could not be found.

RESULTS

For drugs (**out of 166**), in RXNorm **118** were found and mapped, in SNOMED CT **129** were found, in MeSH only **68** drugs were found and in UMLS **111** drugs were found.

For reactions (**out of 3788**), in MedDRA **3777** reactions were found (highest), in SNOMED CT for mappings done through MedDRA to SNOMED CT **1487** reactions were found whereas for mappings done by using SNOMED CT directly **3028** reactions were found (high number because of terms returned with multiple codes), in Mesh **1535** reactions were found and in UMLS **3749** reactions were found.

IV. Evaluating accuracy of the mappings performed

The accuracy of the mappings performed was evaluated separately for drugs and reactions. The author used MySQL for this task to find the number of exact matches, partial matches and incorrect or no matches.

▪ DRUGS

While evaluating the accuracy of the drugs only 160 drugs were evaluated as 6 drugs excluded were the ones which were recorded in the database incorrectly like '*ACETAMINOPHEN/CAFFEINE/CODEINE PHOSPHATE*'.

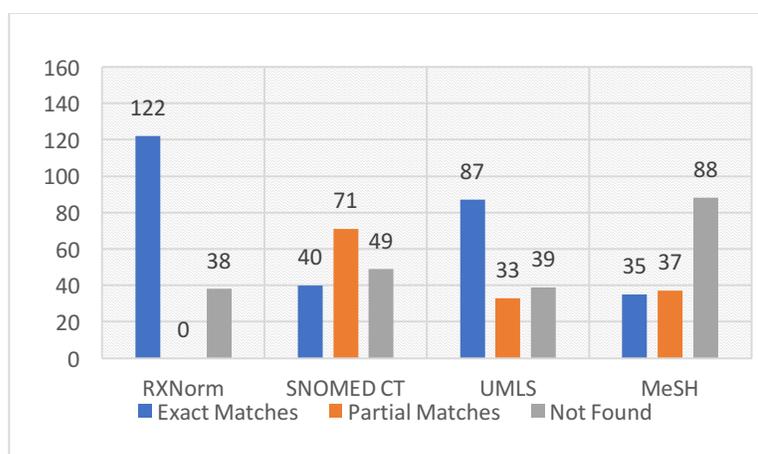


Figure 1: Accuracy of drugs

The author observed that in **RXNorm** almost 77% of the total drugs had exact matches and 23% were not found. RXNorm had the maximum percentage of exact matches as compared to any other terminologies. This can be due to the fact that as most of the drugs had brand names.

On the other hand, in **SNOMED CT** only 25% drugs were mapped accurately whereas majority of the drugs (45%) were partially mapped and the rest were not found. By partial matches, the author means that 2 to 3 different codes were returned for one drug (output received in compositional grammar format/split format). For example, *Pantoprazole Magnesium* was mapped to *Pantoprazole* and *Magnesium* separately.

In **UMLS** 55% of the drugs were accurately mapped, 21% were partially mapped and rest were not found. For example, the drug *Apo-azathioprine* was mapped to the main drug component *Azathioprine*.

MeSH had the least number of percentage for both exact matches and partial matches (only 22%). The reason for the low percentage can be the absence of brand names of the drugs in the terminology.

▪ REACTIONS

As the total number of reactions in the database was 3788, a sample of 200 reactions (random) were taken to test the accuracy.

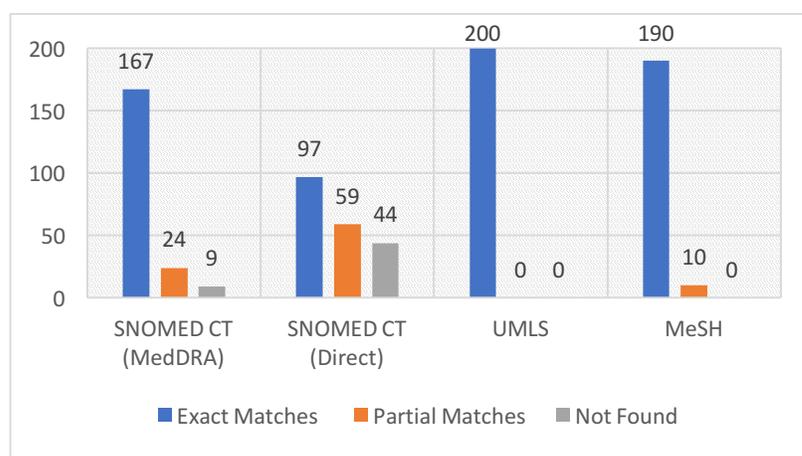


Figure 2: Accuracy of adverse reactions

For **SNOMED CT**, the mappings were done in two ways. First, using the MedDRA mappings and Second, using SNOMED CT directly. For the first one (MedDRA to SNOMED CT), 83.5% were exact matches whereas 12% were partial matches. The partial matches found were either the children or the parent of the original term. For example, *Pyrexia* was mapped to *Body Temperature above reference range*. The mappings done through SNOMED CT directly 48.5% were accurate and 21.5% were partially accurate.

In **UMLS**, all the terms (100%) were accurately mapped because all the terms were found in the terminology and also due to sampling bias as the terms sample was selected randomly.

MeSH was 95% accurate and 5% were partially matched. For example, the term *Pre-syncope* was mapped to *Syncope* which is the child of pre-syncope.

3.1.2 Determining ideal terminology which can be used as an adjunct to MedDRA to find related terms for improving reporting and signal detection.

This task was the second and most important part of the whole project. The purpose and aim of this task was find out which terminology along with MedDRA can be used which will help in returning more related terms using the hierarchal information.

There were three approaches which the author adopted to test and determine the best tool:

- *APPROACH 1*: Utilise the MedDRA to SNOMED CT mappings in the BioPortal.
- *APPROACH 2*: Utilise the SNOMED CT directly through BioPortal.
- *APPROACH 3*: Utilise UMLS Concept Unique Identifier(CUI) in BioPortal along with the SNOMED-CT hierarchy in UMLS database.

METHOD

After the hierarchy was completed for these terminologies, the author decided to take a sample of 800 terms which is approximately 20% of all reactions and divide it further into sets of 3.

- **SET 1**: Top 400 terms having highest reaction counts.
- **SET 2**: Mid 200 terms having reaction counts ranging from 5-7.
- **SET 3**: Bottom 200 terms having instance 1 in the database.

PROCEDURE

The MedDRA terms were used to find related terms in the database using SNOMED CT hierarchy model i.e. similar terms were searched in SNOMED CT hierarchy rather than MedDRA hierarchy.

RESULTS

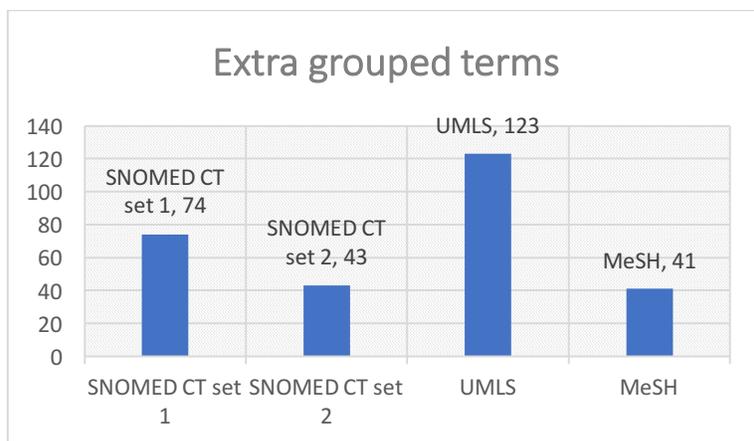


Figure 3: Number of additional terms returned

Out of all the approaches it was observed that the third approach i.e. using UMLS CUI along with SNOMED CT hierarchical information was the best approach as it returned the maximum extra related terms (123 terms) to the searched term. The next approach which returned most terms (74 terms) was the second one which used MedDRA (SNOMED CT SET 1) along with SNOMED CT hierarchy. On the other hand, SNOMED CT direct mappings (SNOMED T SET 2) and MeSH comparatively returned lesser terms (43 and 41 respectively) indicating that when used alone SNOMED CT is not the best tool but when used as a supplement with MedDRA it generates better and more quality results.

To explain this further, when the term “*Intestinal Obstruction*” searched in MedDRA alone returned **5 terms** similar to intestinal obstruction but when the same term was searched in the SNOMED CT hierarchy, **14 extra related terms** were returned which clearly shows the efficacy of using SNOMED CT along with MedDRA.

MedDRA term	Additional terms	MedDRA SOC Class
Intestinal Obstruction	Distal intestinal obstruction syndrome'	Gastrointestinal disorders
	Gastrointestinal obstruction'	Gastrointestinal disorders
	Intestinal obstruction'	Gastrointestinal disorders
	Large intestinal obstruction'	Gastrointestinal disorders
	Small intestinal obstruction'	Gastrointestinal disorders

5 additional terms returned

As it can be seen from the above example, all the additional terms returned for intestinal obstruction are grouped under the SOC Class of Gastrointestinal Disorder. This clearly shows the shallow hierarchical structure of MedDRA as there is no other level in the hierarchy to classify them in a more finer grained way. This also arises a potential possibility of missing the terms while counting the instances as there could be other terms which would fall under the same class.

On the other hand, when the class hierarchy of SNOMED CT is observed 14 additional terms were found i.e. more terms are found under the class of Gastrointestinal disorder which addresses the issue mentioned while explaining the MedDRA hierarchy.

Other than Gastrointestinal disorder two other classes which were found using the SNOMED CT hierarchy ‘*Injury, poisoning procedural complications* and ‘*Congenital, familial and genetic disorders*’.

SNOMED CT term	Additional Terms	Hierarchy
Intestinal Obstruction	Anal stenosis'	Gastrointestinal disorders
	Duodenal obstruction'	Gastrointestinal disorders
	Duodenal stenosis'	Gastrointestinal disorders
	Gastric stenosis'	Gastrointestinal disorders
	Ileus paralytic'	Gastrointestinal disorders
	Intestinal stenosis'	Gastrointestinal disorders
	Intussusception'	Gastrointestinal disorders
	Jejunal stenosis'	Gastrointestinal disorders
	Oesophageal stenosis'	Gastrointestinal disorders
	Postoperative ileus'	Injury, poisoning procedural complications
	Pyloric stenosis'	Congenital, familial and genetic disorders'
	Rectal stenosis'	Gastrointestinal disorders
	Volvulus of small bowel'	Gastrointestinal disorders
	Volvulus'	Gastrointestinal disorders

14 additional terms returned

New classes identified using SNOMED CT hierarchy

Working with this approach when the terms Intestinal Obstruction were searched again the reaction count increased from 3281 to 4144. The difference between the old and new reaction count represents the number of terms that were missed when searched only through MedDRA and also the efficacy of SNOMED CT which returns the terms related to the original terms on conceptual levels.

4. Relation to Health Informatics

Health Informatics(HI) is a multidisciplinary field which uses Information Technology to improve the delivery of care. Out of numerous aspects of Health Informatics, this project relates specifically to the Health Standards portion of HI along with Data Mining and Data Analysis.

Major portion of the project is based on analyzing different clinical terminologies on a clinical database. The techniques used throughout the project were the ones author learned during the previous two terms of HI like MySQL which was used for majority of the data analysis as well as creating and updating data. This helped the author to improve the skills required while handling huge databases. This project also gave the author the opportunity of learning and exploring Python programming and using REST API to retrieve data, both being prominently used for designing and creating online medical tools and applications.

With this project, the author proposes meaningful use of the all the information documented which is one of the major aim in the field of Health informatics. Previously, this database was just used for reporting and recording the adverse reactions but now with the approaches proposed more can be done with the same information. This project promotes converting unstructured data in to a structured one which ultimately can be used for various other researches.

5. Analysis of difficulties/problems encountered

One of the major problem that the author came across during this project was while mapping the terms in the database to SNOMED CT. As mentioned in the earlier sections, SNOMED CT although being a more comprehensive terminology, has its own limitations. In relation to our database it can be explained as follows:

The database consists of many reactions/drugs which have comparatively long names as compared to others. The problem observed while mapping these terms with long names was that, when queried in SNOMED CT these terms were returned in a compositional grammar format or in a split format.

For example, the term "*Human chorionic gonadotropin increased*", when searched in SNOMED CT was returned as '59433001' '*Human chorionic gonadotropin*' and '35105006' '*Increased*' i.e. two different codes received for one term. This split format causes irregularity in the data and in some cases, changes the meaning of term. As there were 3788 reactions in total and for most of them the result received was in the similar format. Similarly, for drugs the results came in same format. For example, the drug *Cortifoam Aerosol* was mapped to two different codes *Cortifoam* and *Aerosol*.

To overcome this MedDRA mappings were used to map these terms. These mappings were found in the JSON file of the searched term and by processing this file codes were retrieved for SNOMED CT. This approach proved to be more accurate in most of the cases and reduced the chances of error and irregularities significantly. The reason why mappings through MedDRA and RXNorm

only were used was that almost all of the reactions could be found in MedDRA and almost all drugs were present in RXNorm. Using these mappings few terms which were not found directly in SNOMED CT were with this approach.

6. Conclusion

With MedDRA being used most commonly while reporting and recording the adverse reactions into the vigilance database, it is very important to analyze whether MedDRA alone is the best tool or there are other tools which can be used along MedDRA or individually to improve the quality of the data recorded.

This report consisted of two main parts, first one focused on how the mappings were performed on the terms present in the Canadian Vigilance Adverse Reaction Database to five lexicons like MedDRA, Mesh, RXNorm, SNOMED CT and UMLS. The second part focused on determining the ideal tool to be used as a supplement to MedDRA.

After the analysis, the author concluded that when the terms in the database were mapped alone to MedDRA there was a potential possibility that not all the terms were included in the grouped term or in other words some terms were missed while grouping. To address this problem, the author evaluated various approaches which could be adopted for improving the efficacy of the MedDRA.

The results clearly showed that using SNOMED CT along with MedDRA and UMLS CUI gave better results both qualitatively and quantitatively supporting the fact that it is the ideal tool to be used along MedDRA. The other terminologies evaluated had other shortcomings which could not be proved helpful in regard to the project.

In relation to Signal detection this approach can be useful as while detecting the possible ADRs more number of reactions would be present which will ultimately improve accurate detections. Also, while reporting ADRs the health professionals as well as the consumers will be able to see more options increasing the accuracy of the reports. SNOMED CT and MedDRA is the choice of terminology for recording data in many clinical settings. If the idea of using them together is adopted it will promote interoperability at a larger scale and it will be easier to integrate this into the current systems.

7. Recommendations

For future work, these methods of mapping and evaluation can be extended to all of the reactions and drugs present for Crohn's Disease as well as to entire database. Every approach including codes and queries are flexible enough to be extended over the entire database.

Another recommendation while mapping or querying could be use of the Active Drug ingredients rather than the drug names and see whether the mapping percentage improves as most of the drugs were brand names which could not be found in many terminologies and using the ingredient may be more useful.

Also, promoting meaningful use of all the new data found during this project for different research purposes or collecting more evidence that this approach could help in determining new patterns of drug-event pairs.

8. References

- [1] Canada, H. (2017). Canada Vigilance Program - Canada.ca. [online] Canada.ca. Available at: <https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada/canada-vigilance-program.html> [Accessed 15 Aug. 2017].
- [2] WHO,1972
- [3] Zed P, et al. Incidence, Severity and Preventability of Medication-Related Visits to the Emergency Department: A Prospective Study. *Canadian Medical Association Journal*. June 3, 2008;178(12): 1563-1569.
- [4] Wu C. Adverse Drug Reactions in the Emergency Department Population in Ontario: Analysis of National Ambulatory Care Reporting System and Discharge Abstract Database 2003-2007. University of Toronto. 2009
- [5] Canada, H. (2017). Instructions for the Canada Vigilance Adverse Reaction Online Database - Canada.ca. [online] Canada.ca. Available at: <https://www.canada.ca/en/health-canada/services/drugs-health-products/medeffect-canada/adverse-reaction-database/instructions-canada-vigilance-adverse-reaction-online-database.html> [Accessed 8 Aug. 2017].
- [6] Randall, C., Vizuete, J., Wendorf, G., Ayyar, B. and Constantine, G. (2012). Current and emerging strategies in the management of Crohn's disease. *Best Practice & Research Clinical Gastroenterology*, 26(5), pp.601-610.
- [7] Bioportal.bioontology.org. (2017). Welcome to the NCBO BioPortal | NCBO BioPortal. [online] Available at: <https://bioportal.bioontology.org/> [Accessed 8 Aug. 2017].
- [8] Salvadores, M., Alexander, P. R., Musen, M. A., & Noy, N. F. (2013). BioPortal as a Dataset of Linked Biomedical Ontologies and Terminologies in RDF. *Semantic Web*, 4(3), 277–284.
- [9] Shah, N., Bhatia, N., Jonquet, C., Rubin, D., Chiang, A., & Musen, M. (2009). Comparison of concept recognizers for building the Open Biomedical Annotator. *BMC Bioinformatics*, 10(Suppl 9), S14. <http://dx.doi.org/10.1186/1471-2105-10-s9-s14>
- [10] Meddra.org. (2017). Vision for MedDRA | MedDRA. [online] Available at: <https://www.meddra.org/about-meddra/vision> [Accessed 8 Aug. 2017].
- [11] snomed.org/snomed-ct (2017).

[12] Nlm.nih.gov. (2017). UMLS Quick Start Guide. [online] Available at: <https://www.nlm.nih.gov/research/umls/quickstart.html> [Accessed 8 Aug. 2017].

[13] Nlm.nih.gov. (2017). RxNorm. [online] Available at: <https://www.nlm.nih.gov/research/umls/rxnorm/> [Accessed 8 Aug. 2017].

[14] Nlm.nih.gov. (2017). Medical Subject Headings - Home Page. [online] Available at: <https://www.nlm.nih.gov/mesh/> [Accessed 8 Aug. 2017].

APPENDIX

APPENDIX 1: Sample of Drugs for Crohn's Disease

'REMICADE' 'ENBREL'
 'SANDOSTATIN LAR 20MG'
 'MORPHINE'
 'METHADONE'
 'TYLENOL'
 'COLESTID'
 'IRON SUCROSE'
 'OXYNEO'
 'HYDROMORPHONE HYDROCHLORIDE'
 'FERRLECIT'
 '5-AMINOSALICYLIC ACID'
 'BUSCOPAN'
 'VITAMIN D'
 '6-MERCAPTOPYRINE'
 'DESOXIMETASONE'
 'OMEPRAZOLE'
 'ACTEMRA'
 'NOVA VITAMIN B 12'
 'SALAZOPYRIN EN-TABS 500 MG'
 'FERROUS SULFATE'
 'SALAZOPYRIN'
 'METHOTREXATE SODIUM'
 'PMS-SULFASALAZINE 500MG/TAB USP'
 'MOXIFLOXACIN HYDROCHLORIDE'
 'MACROBID'
 'HYDROCORTISONE SODIUM SUCCINATE INJ 1GM/VIAL'
 'SOLU-MEDROL'
 'DIPENTUM'
 'ESOMEPRAZOLE MAGNESIUM TRIHYDRATE'
 'ALENDRONATE'
 'FOLIC ACID'
 'METHOTREXATE'
 'CALCIUM'
 'METRONIDAZOLE'
 'TYSABRI'
 'CIPROFLOXACIN'
 'ASACOL ECT 400MG'
 'MELOXICAM'
 'MESALAZINE'
 'SULFASALAZINE'
 'ANUSOL HC /00028604/'
 'APO-NALTREXONE'
 'MYLAN-AZATHIOPRINE'
 'OLESTYR'
 'BUTRANS 5'
 'PANTOPRAZOLE SODIUM'
 'MESALAMINE'
 'TYLENOL 1'

APPENDIX 2: Sample of reactions for Crohn's Disease

Meningitis listeria,
Sepsis"
Nephrostomy"
Pyrexia"
Tachycardia"
Hyperhidrosis"
Confusional state"
CSF test abnormal"
Eye irritation"
Uveitis"
Condition aggravated"
Drug ineffective"
Intraocular pressure increased"
Glaucoma"
Visual acuity reduced"
Vision blurred"
Demyelinating polyneuropathy"
Fatigue"
Muscular weakness"
Hypoesthesia"
Neuropathy peripheral"
Polyneuropathy"
Antinuclear antibody positive"
Lymphocyte count abnormal"
Abdominal pain"
Dysphagia"
Splenomegaly"
Thrombocytopenia"
Hepatic cirrhosis"
Liver disorder"

APPENDIX 3: Python code for retrieving data

```

import urllib.request
import json
import os
import mysql.connector
from pprint import pprint

#connecting to rest api-the url and key are specified below
REST_URL = "http://data.bioontology.org"
API_KEY = "88b92995-13c7-4957-970b-aeecf3fa7b39"

def get_json(url):
    opener = urllib.request.build_opener()
    opener.addheaders = [('Authorization', 'apikey token=' + API_KEY)]
    return json.loads(opener.open(url).read())

# defining an object called print_annotations through which we will get the code and Preferred name of terms by selecti
#file we receive from the rest API.This code is only to get the direct values of SNOMEDCT of terms from the annotator
def print_annotations(annotations, get_class=True):
    values=""
    for result in annotations:
        class_detail1 =get_json(result["annotatedClass"]["links"]["self"]) if get_class else result["annotatedClass"]
        class1=class_detail1
        #taking the fields required from the JSON file and assigning them an object
        strTemp1=class1["@id"].rsplit('/',1)
        strTemp2=class1["@id"].rsplit('/',2)
        strTemp3=class1["cui"]

        # defining values in the format we need the python output by taking the above defined objects
        values= values+ (" "+strTemp1[1] +" "+strTemp3+" "+class1["prefLabel"]+""),"
    return(values)

#loading the JSON file containing drugnames/reactions
with open('/Users/soumyashastry/Downloads/ADR_term.json') as json_data:
    d = json.load(json_data)

#SQL query for inserting data directly in the database
# query = "INSERT INTO MEDDRATERM(meddraCUID,term) VALUES "
# cnx = mysql.connector.connect(user='adr',password='adrPASSWORD',host='129.173.20.222',database='canadianADR')
# cursor = cnx.cursor()
counter = 0
values = ""
for strVal in d:
    text_to_annotate = str(strVal).split(" ")[1]
    annotations = get_json(REST_URL + "/annotator?text=" + urllib.request.quote(text_to_annotate) + "&ontologies=SNO
    values = values + print_annotations(annotations)
    counter += 1
    if counter == 5:break
values= values[:-1].replace(",",""),1][:-1]

```

APPENDIX 4: JSON file used BioPortal

```

- annotatedClass: {
  @id: http://purl.bioontology.org/ontology/MEDDRA/10027248,
  @type: http://www.w3.org/2002/07/owl#Class,
  - links: {
    self: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248,
    ontology: http://data.bioontology.org/ontologies/MEDDRA,
    children: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/children,
    parents: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/parents,
    descendants: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/descendants,
    ancestors: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/ancestors,
    instances: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/instances,
    tree: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/tree,
    notes: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/notes,
    mappings: http://data.bioontology.org/ontologies/MEDDRA/classes/http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248/mappings,
    ui: http://biportal.bioontology.org/ontologies/MEDDRA?p=classes&conceptid=http%3A%2F%2Fpurl.bioontology.org%2Fontology%2FMEDDRA%2F10027248,
  }
  - @context: {
    self: http://www.w3.org/2002/07/owl#Class,
    ontology: http://data.bioontology.org/metadata/Ontology,
    children: http://www.w3.org/2002/07/owl#Class,
    parents: http://www.w3.org/2002/07/owl#Class,
    descendants: http://www.w3.org/2002/07/owl#Class,
    ancestors: http://www.w3.org/2002/07/owl#Class,
    instances: http://data.bioontology.org/metadata/Instance,
    tree: http://www.w3.org/2002/07/owl#Class,
    notes: http://data.bioontology.org/metadata/Note,
    mappings: http://data.bioontology.org/metadata/Mapping,
    ui: http://www.w3.org/2002/07/owl#Class
  }
},
- @context: {
  @vocab: http://data.bioontology.org/metadata/
}
},
hierarchy: [ ],
- annotations: [
  - {

```

APPENDIX 5: Examples of additional terms returned

Surgery	'Aortic surgery'	'Surgical and medical procedures'	1	686	32	718	'Angioplasty'	'Surgical and medical procedures'
	'Arthroscopic surgery'	'Surgical and medical procedures'	18				'Arteriovenous fistula operation'	'Surgical and medical procedures'
	'Breast conserving surgery'	'Surgical and medical procedures'	4				'Carotid endarterectomy'	'Surgical and medical procedures'
	'Cancer surgery'	'Surgical and medical procedures'	1				'Coronary artery bypass'	'Surgical and medical procedures'
	'Elective surgery'	'Surgical and medical procedures'	2				'Haemorrhoid operation'	'Surgical and medical procedures'
	'Eye laser surgery'	'Surgical and medical procedures'	10				'Keratomileusis'	'Surgical and medical procedures'
	'Gastrointestinal surgery'	'Surgical and medical procedures'	20				'Liposuction'	'Surgical and medical procedures'
	'Hip surgery'	'Surgical and medical procedures'	16				'Varicose vein operation'	'Surgical and medical procedures'
	'Laparoscopic surgery'	'Surgical and medical procedures'	3					
	'Metabolic surgery'	'Surgical and medical procedures'	7					
	'Neck surgery'	'Surgical and medical procedures'	1					
	'Neurosurgery'	'Surgical and medical procedures'	2					
	'Oral surgery'	'Surgical and medical procedures'	13					
	'Plastic surgery to the face'	'Surgical and medical procedures'	3					
	'Plastic surgery'	'Surgical and medical procedures'	2					
	'Renal surgery'	'Surgical and medical procedures'	5					
	'Spinal fusion surgery'	'Surgical and medical procedures'	9					
	'Surgery'	'Surgical and medical procedures'	564					
	'Wrist surgery'	'Surgical and medical procedures'	5					
Cellulitis	'Anorectal cellulitis'	'Infections and infestations'	2	561	7	568	'Erysipelas'	'Infections and infestations'
	'Breast cellulitis'	'Infections and infestations'	2				'Ludwig angina'	'Infections and infestations'
	'Cellulitis of male external genitalia'	'Infections and infestations'	3					
	'Cellulitis orbital'	'Infections and infestations'	2					
	'Cellulitis streptococcal'	'Infections and infestations'	7					
	'Cellulitis'	'Infections and infestations'	504					
	'External ear cellulitis'	'Infections and infestations'	13					
	'Incision site cellulitis'	'Infections and infestations'	6					
	'Infusion site cellulitis'	'Infections and infestations'	1					
	'Injection site cellulitis'	'Infections and infestations'	7					
	'Periorbital cellulitis'	'Infections and infestations'	2					
	'Post procedural cellulitis'	'Infections and infestations'	6					
	'Stoma site cellulitis'	'Infections and infestations'	4					
	'Vaginal cellulitis'	'Infections and infestations'	2					
Urticaria	'Angioedema'	'Skin and subcutaneous tissue disorders'	17	1051	51	1102	'Serum sickness'	'Immune system disorders'
	'Circumoral oedema'	'Skin and subcutaneous tissue disorders'	1					
	'Eyelid oedema'	'Eye disorders'	30					
	'Face oedema'	'General disorders and administrative procedures'	38					
	'Injection site urticaria'	'General disorders and administrative procedures'	13					