Drug Side Effects Data Representation and Full Spectrum Inferencing using Knowledge Graphs in Intelligent Telehealth

By
Saravanan Jayaraman B.Tech, M.B.A

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Professional Studies
in Computing

at

School of Computer Science and Information Systems

Pace University

October 2016

We hereby certify that this dissertation, submitted by Saravanan Jayaraman, satisfies the dissertation requirements for the degree of *Doctor of Professional Studies in Computing* and has been approved.

_____-_____

Dr. Lixin Tao Date
Chairperson of Dissertation Committee

_____-_____

Dr. Ronald Frank Date
Dissertation Committee Member

_____-_____

Dr. Meikang Qiu Date
Dissertation Committee Member

School of Computer Science and Information Systems
Pace University 2016

Abstract

Drug Side Effects Data Representation and Full Spectrum Inferencing using Knowledge Graphs in Intelligent Telehealth

By
Saravanan Jayaraman

Submitted in partial fulfillment
of the requirements for the degree of
Doctor of Professional Studies
in Computing

October 2016

Drug adverse reaction data contains important constraints about side effects and conflict avoidance of component and compound drugs. We observe that many of these constraints are transitive in nature due to the relationship between drug and drug classes. Current drug side effects representations in XML does not have a proper knowledge representation mechanism to clearly specify all kinds of dependencies among the drug components and drugs. Even the recently introduced OWL based approach for medical drug side effects data representation still suffers from several shortcomings inherent to the OWL restrictions like using "is-a" relationship and usage of object property emulations.

In this research, we propose a model *Drug - Side Effects Representation And Inferencing (D -SERI)* built using Knowledge Graph (KG) and enhanced *PaceJena* to represent multiple custom relationships allowing domain experts to capture the transitive nature of the relations in an inference friendly way. The research also developed a concept demonstrator for checking out prescriptions to avoid complications. The research outcome shows that the proposed model allows the doctors and caregivers to derive dynamic information about side-effects avoiding costly errors caused by human interpretation.

Acknowledgements

I would like to dedicate my dissertation work to my parents whose endless love and affection gave me the power to work harder everyday with determination and perseverance to complete the research.

I would like to thank my beloved wife Viji who stood with me in every step of the way and endured countless difficulties and challenges while supporting my family during my absence. Her encouraging words and motivation throughout was a constant energy booster for me during the DPS years.

I affectionately thank my children Lakshaanya and Siddesh whose curious questions and supportive behavior gave me the zeal to strive for excellence. I wish to see them successful in life in their own way and surpass their parent's accomplishments one day.

THANK YOU to my dissertation advisor Dr. Lixin Tao for believing in me and providing constant encouragement and showing the knack to breakdown the dissertation topic and the technique to solve complex research problems. I greatly benefited from his years of guidance experience and incredible counsel. I am and will always be forever GRATEFUL!

Thanks to Dr. Frank and Dr. Qiu for all the great advice, and for being on my dissertation committee.

Special Thanks to my fellow researcher Ning Jiang who offered expert technical assistance and code review with me.

Finally, I would like to sincerely acknowledge my cohorts from the class of 2016 for supporting each other and staying together during the DPS journey.

"*Dream is not what you see in Sleep, dream is something which does not let you sleep*" – Dr A P J Abdul Kalam. Like many others before me, I definitely had many sleepless nights to accomplish my dream. During this, I also realized that the Seeking higher level of knowledge is a journey not a destination and I intend to continue this journey throughout my life.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Study of Medical Drug side effects on Humans was first documented by the Greek Physician Hippocrates of Kos in 460 BC who first studied varied effects of Aspirin as a migraine to relieve pain and suffering on his patients. His followers from Hippocrates school of Medicine further continued the study of clinical study summing up the medical knowledge and wrote the everlasting drug side effect data on early human races. Along with these early studies by doctors in Graco-latino world, several similar advanced civilizations conducted studies in asymmetric manner setting up strong data spanning over centuries to be used by mankind.

1.1 Drug Side Effects Inference Problem

With the advent of the Web advancement with research and development, numerous domains have been remarkably driven by the emerging techniques, such as semantic web [1] [2] cloud computing [3][4][5], and big data[6]. This is even more relevant in the area of Drug development and monitoring in Telehealth [7] [8] where Domain experts depend upon these emerging technologies to encode knowledge. Recently, there is an explosion in the number of drugs approved for treatment and the effect they caused on human

population. In the US market alone, there are 2,000 medications, 94,450 health products and around 175,950 health packages with different active ingredients and these numbers are growing steadily every year. With these huge volume of drug combinations comes the challenge to have a relevant knowledge mechanism to capture the huge volume of side effects data and enable the patients and doctors a way to derive meanings about the potential outcome quickly. For example, when a doctor treats a patient with diabetic condition with any one of the drugs approved for treatment, he or she wants to make sure the benefit of prescribing the drug exceeds the side effects caused by the drug itself, in that case of side effect showing up, he or she wants to relate that to the prescribed drug and then quickly change the treatment option to suit the patient's responses to drugs. While there has been huge advances in the drug side effects data capturing globally including the nationally mandated options like MedWatch or EU drug watch and the availability of that information to public domain, the Information Technology processes behind have not been coping up due to several factors.

1.1.1 Side Effects Awareness

The first is the explosion of the drug side effects data itself. Look at the growth of the side effects in the last 10 years growing exponentially. This data is extremely dynamic. Most of the Institutes and Research organizations spend enormous amount of time in just getting the data downloaded and synced periodically. As per the recent look on the MedWatch system there are 7 million data entries in the system. Figure 1 shows the continuous growth in the data. The interesting question to ask is what happens when a new

adverse event data is released and how it is synchronized with the previously stored data thereby staying relevant.

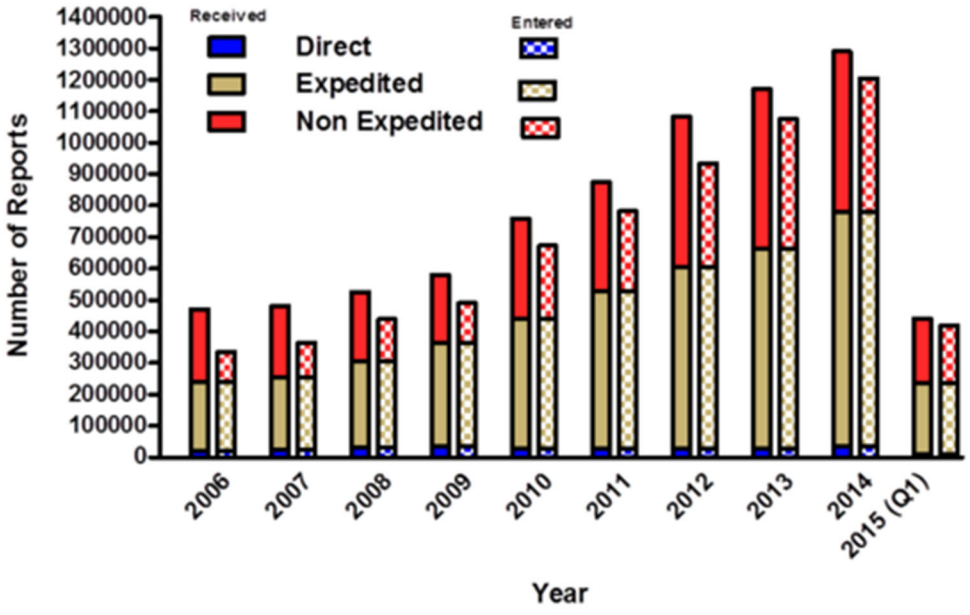


Figure 1 Aggressive Growth of side effects data in the last decade

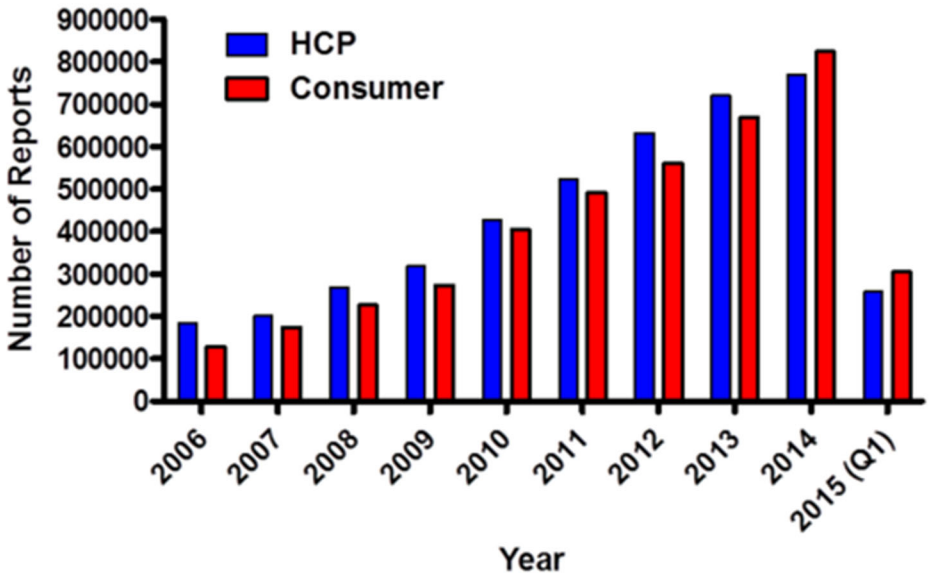


Figure 2 Side effects data reporting by HCP vs Consumers

A recent study by Mayo Clinic Survey [9] points that nearly 7 in 10 Americans take some form of Prescription drugs, and medication errors are surprisingly common and costly to the nation [10]. As per a study conducted by NIH.gov in a North Indian City [27], only 33% of patients knew about the side effects produced by the concerned drug, and only 15.68% knew how to recognize them. This is also confirmed by the CDC report where 70% of adults over 65 years take prescription drugs and 1 in 4 senior Americans take more than 3 prescription drugs.

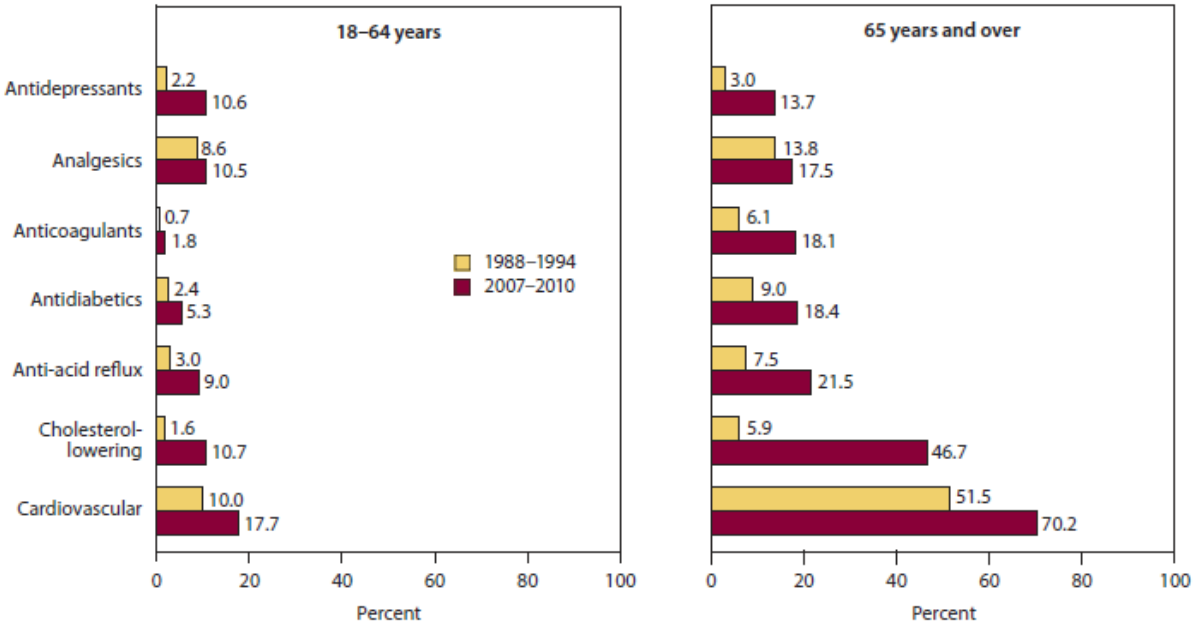
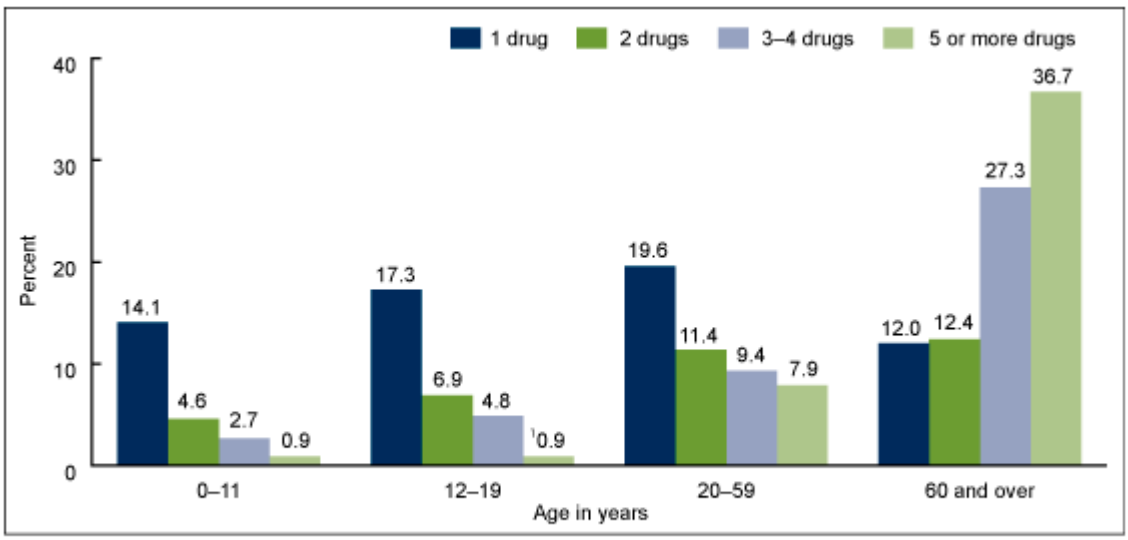


Figure 3 Prescription drug use in the past 30 days among adults aged 18 and over, United States, 1988–1994 and 2007–2010



[1]Estimate is unstable; the relative standard error is greater than 30%.
SOURCE: CDC/NCHS, National Health and Nutrition Examination Survey.

Figure 4 Prescription drug use in combinations among adults aged 18 and over

1.1.2 Growth of Drug Side Effects Data

As per the CDC study the usage of prescription drugs increased by 10% last decade and the use of multiple prescription drugs increased by 20% and the use of 5 or more drugs increased by 70%. As pointed by the same CDC source the usage of prescription drug is directly related to availability of regular healthcare and as more and more people have access to regular healthcare the usage of prescription drugs is expected to grow linear as the necessity to treat multiple diseases at the same time especially on the senior population. This shows the problem with the data representation is not a one-time issues rather a long term issue requiring clarity and newer approaches from knowledge scientists.

1.1.3 Maintaining the Side Effects Data Current

Whenever newer treatment options are available in the market, healthcare providers prefer to prescribe the advanced treatment option to their patients so they get the advanced therapeutic benefits and often doctors are looking for a better way to understand the knowledge about these new drugs to help treat the patients.

Here is one scenario. A patient suffering from uncontrolled diabetes is looking to utilize some of the newer drugs in the category. The doctor, while aware of the potential side effects of a new drug is looking forward to understand side effects of the drug better so that it does not conflict with the current treatment option for the patient. Here the doctor relies on several types of information like FDA's MedWatch data [20], or any other data provided by the manufacturer or labelling information. While these information provide some levels

of details about the drug side effects, the studies have pointed that the study of the drug side effects is one of the complex task to perform by a common man or physicians requiring enormous levels of investigation which is simply impossible due to lack of time or resources. According to the recent study by Harvard medical analytic group the knowledge mechanism of the drug data domain is not catching up with the growth pace of the data itself due to the restrictions on the current format and the lack of validation mechanisms.

1.1.4 Motivating Example

This section shows a real time example where the proposed methodology can make a difference in the drug side effects data representation.

Table 1 exhibits different scenarios where the research can help doctors to avoid human interpretation errors.

Table 1 Doctor's action during a patient visit

Case	Patient observation for a drug	Doctor's action
1	Patient does not report side effect with drug	Continue to prescribe the drug
2	Patient reports side effects but it's not known with drug	Continue to prescribe the drug while reporting new side effects.
3	Patient reports side effects and its known with the drug	STOP the drug and look for alternatives.

While Doctors prescribe drugs during routine visits, they always strive to make sure the benefits of prescribing the drug outweighs the risk caused by side effects. Here they always seek in-depth knowledge about the side effects keeping in mind the ultimate safety of the patient's life at the forefront of the drug prescription strategy. We take three typical scenarios often faced by doctors for this dissertation. There are several other variations of the scenarios are possible but for the case study perspective we restrict it to the three choices.

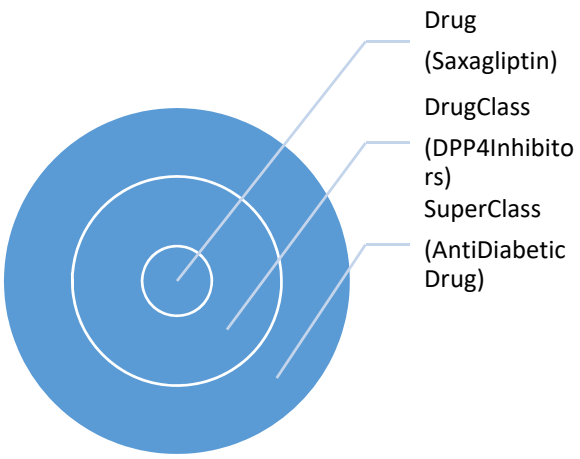


Figure 5 Saxagliptin drug class hierarchy [26] in knowledge graph

Saxagliptin drug is used to treat *diabetes* and they are newest treatment options for patients who are not responding well for other diabetic treatment options. While Saxagliptin----causes--- set of side effects (|A| - *Abdominal Pain, Motor Dysfunction, Hyperhidrosis, malaise, Nasal Congestion, Increased Blood Sugar, Arrhythmia, rash, Cerebro Vascular Accident*) which is well known, doctors often find it that it does not represent the full list possible side effects. In this case, Saxagliptin's parent class DPP4Inhibitors (gliptin) causes its own set of common side effects – (|B| - *nausea,diarrhea,stomach pain,headache,runny nose,sore throat,pancreatitis,severe join pain*).

When Doctors look for side effects caused by a drug against reported by patents, they often rely on the direct side effect list |A| as primary source as the full spectrum of the side effects is not available to them due to several reasons. This could cause them to overlook side effects like *pancreatitis* which is in |B|. Studies show that while emulations with object properties are used to capture these additional relations they often cause other problems like high cost of maintenance on data modelers.

1.1.5 Benefits of dynamic Side Effect Inference

Linking the component and compound drugs using the proposed knowledge graph based approach allows the domain experts to capture the full spectrum side effects of the drug (|A| + |B|) by including all possible side effects while reducing syntax burden to knowledge modelers compared with any other workarounds like object properties.

Such a dynamic data representation model will also provide a full spectrum side effects to the doctors and patient helping them immensely to benefit to either adapt newer treatment models without fear or just to choose a suitable treatment model beneficial to the patient. Even when the knowledge about the drug side effects is available the current format makes it harder medical drug side effects is neither flexible nor suitable in the full spectrum nuisances of the drug side effects data drug being a chemical component.

1.2 Representation of Drug Side Effects Data in Knowledge Graph

A Knowledge graph describes the concepts in the domain and also the relationships that hold between those concepts. Different knowledge graph languages provide different facilities. It makes it possible for concepts to be defined as well as described. Complex concepts can therefore be built up in definitions out of simpler concepts.

1.2.1 Knowledge Graph Usage in Healthcare

The usage of Knowledge Graph is well accepted especially in the semantic web [16][17] area as a primary way of disseminating the information to users or machines even though it's still evolving in the medical domain. For example, Google's knowledge vault [50] is enriched with information about 570 million objects of data and 18 billion facts making the world's largest public knowledge graph vault.

In Knowledge graph, classes are interpreted as sets that contain individuals. They are described using formal (mathematical) descriptions that state precisely the requirements for membership of the class. Unlike traditional approaches where the focus is storage of

the data with less consideration of the timely interpretation or reasoning, the primary goal of the knowledge graph is to enable timely retrieval of the knowledge in this case the use by Doctors or Patients to retrieve time sensitive data. The key once again is the modular ability of the knowledge graph to extend and grow making it an ideal option to store drug adverse data and making it highly suitable for capturing drug side effects data due to the dynamic nature of the domain.

1.3 Problem statement

Drug adverse reaction data contains important constraints about side-effects and conflict avoidance of component and compound drug. These are critically important in checking out prescriptions to avoid complications. Although MedWatch FAERS drug data are in XML, it doesn't have a proper knowledge representation mechanism to clearly specify all kinds of dependencies among the drug components and drugs. Therefore one has to depend on human interpretation to check prescriptions which can be error-prone. The newly introduced OWL based approach for medical drug data representation still suffers from several shortcomings inherent to the OWL restrictions like using “is-a’ relationship and usage of object property based workarounds losing the clarity and dynamic relationship building expected by domain experts to represent knowledge. It’s often difficult for the domain experts to process and derive meaningful information quickly for patients or doctors due to the following reasons.

Drug side effect relationships are highly transitive in nature and often inherit, important adverse reaction constraints from their compound drugs, which is not represented in the current OWL or FAERS format. Since OWL only supports is-a relationship the domain experts are left with the option to use workarounds using object properties which don't behave the same way as true relationships during interpretations.

Drug side effect relationships should be kept updated current always, as new relationships emerge on component or compound drugs. Though this is possible with the current approach using workarounds like object properties, it requires a very high effort for both updates and maintenance. In essence it also forces static nature on the relationship while the need is to keep them dynamic -

Drug side effect relationships should be interpreted easily by users (patients or doctors) allowing them to access the full spectrum of side effects. The current approach of using object property emulation is error prone in its mechanism for patients to interpret the data due to the artificial annotations and intermediate concept usage.

1.4 Proposed Solution Methodology

This dissertation proposes a new approach *Drug-SideEffectsRepresentation AndInferencing (D-SERI)* to encode drug side effects data using knowledge graph with custom OWL relationships.

Using this approach, the dissertation further developed a methodology *Drug-GetParentRelations (D-GPR)* to derive drug adverse meanings quickly for doctors and patients avoiding costly errors caused by human interpretation.

The major motivation for this research is generating an Extendable Knowledge-Graph based approach for drug domain data that can be easily understood by machines allowing them to process and derive meanings for doctors. The Proposed dynamic inferencing model to extract the full spectrum side effects for the drug relies on PaceJena which is a Pace University extended version of open source java framework for building knowledge graphs. PaceJena is specifically built to support the extended version of owl using custom relationships. This proposed approach will greatly extend the capabilities of OWL to be used for drug side effects domain.

This model can play a significant role in the convergence of several healthcare services using digital health in the big data cloud environment and mHealth services empowering doctors and caregivers to better track and manage their patient's health.

1.5 Expected Key Contributions

The feasibility study strives to prove the benefit of developing a Patient focused Knowledge Graph using custom relationships Takes, Cause and part of. This approach addresses the shortcoming identified in capturing the constraints of compound and component drug thereby providing a strong framework for drug adverse reactions knowledge capture**.** As shown in the benefits of "Extending OWL relationships" [18] these

Knowledge Graph (KG) representation allows relationships to stay dynamic in nature thereby reducing the burden on maintaining syntax while processing. This dissertation also helps to show how the framework can be extended to other drug domain information using the modular nature of the knowledge created.

This concept demonstrator strives to shows how these Knowledge Graphs can be parsed iteratively to bring out full spectrum of drug side effect information for patients and doctors who are the central focus of this study reducing errors caused by the human interpretation.

In Summary, the proposed research work strives to prove the benefit of developing a patient focused Knowledge Graph development for DARs data using custom OWL relationships instead of object properties and proves the concept using a web portal deriving meaningful drug adverse data quickly for patients or doctors. The proposed model also allows DARs to be extended for other drug / pharmaceutical domain areas like protein to drug relationships or drug to treatment relationships.

1.6 Dissertation Road Map.

The dissertation is outlined with the following way.

Chap 2 provides existing approaches and concepts published so far for drug side effects data representation.

Chap 3 describes the research methodology in detail with flow charts, logical explanations, algorithm and concept

Chap 4 describes the implementation of the concept demonstrator and outcomes for chosen test cases.

Chap 5 talks about validation of the proposed methodology with results to prove the feasibility of the proposed model.

Chap 6 is the conclusion and outlines the roadmap for using the proposed methodology including the future work.

Chapter 2

Current State of Drug Adverse Relationships Data Representation

Discovering relevant drug adverse relations is always considered an important requirement for a successful drug adverse relationship repository. The following chapter analyses the current state of the data representation and the inferencing techniques widely used in the Industry.

2.1 Industry wide Initiatives

2.1.1 MedWatch Initiative

This Initiative was created by FDA to increase the discovery of adverse events in the general population using a standardized safety information and adverse event reporting service (AERS) called MedWatch. The outcome was a free and easy to use service which can be utilized by general public and physicians to report the occurrence of any adverse event. This Initiative provides multiple purpose with key focus on allowing medical professionals and the public to report side effects and other medical product injuries. This Initiative covers not only prescription drugs, but all other medical products including over-the-counter drugs, devices, nutritional products, dietary supplements, and infant formulas covering all facets of the medical products. This easy to report and no-frill service has enabled public and physicians to report millions of events to the repository. Last year alone

FDA received more than 1 million Adverse Event cases reported through the MedWatch initiative and the number of reports received every year is increasing annually.

The following depicts the data flow of the side effects data from Patients / manufacturers into FAERS database.

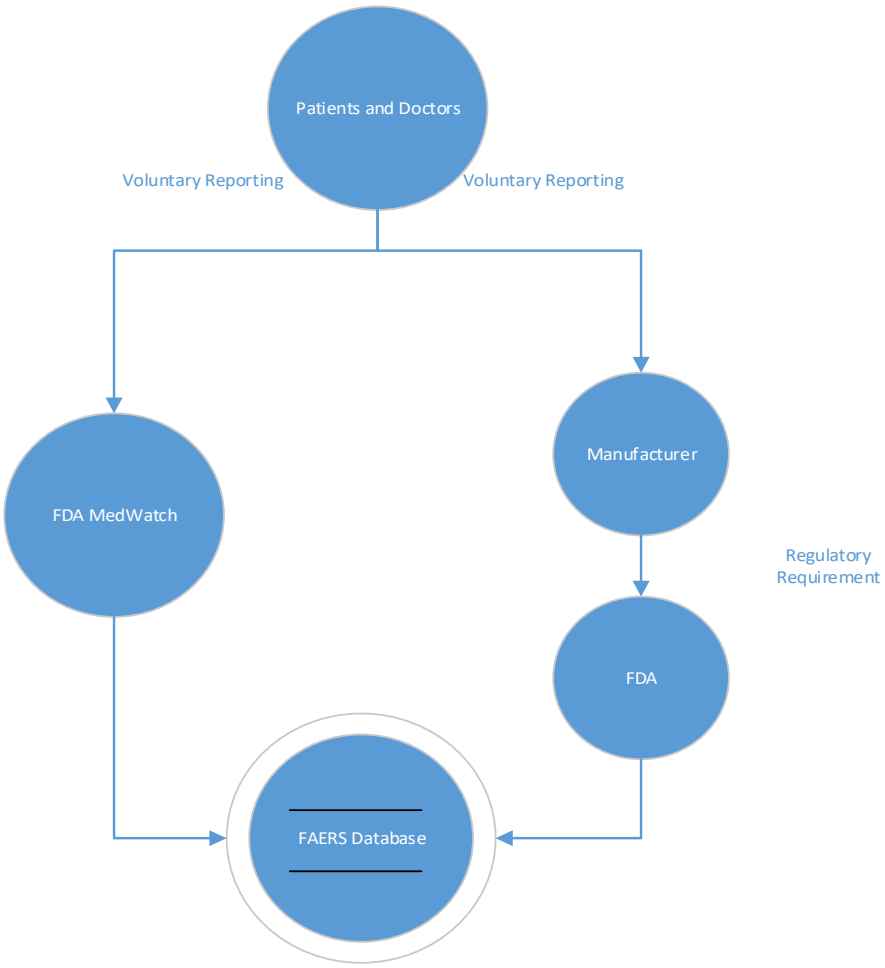


Figure 6 FDA Adverse Data reporting – data flow

The Side effects data collected is then released in Quarterly data format which contains information about the side effects in ASCII and XML format even though FAERS system is itself in relational database. The Quarterly data files also provides summary data and the FAERS data dictionary for the consuming systems to understand the format.

Table 2 Pros and Cons FAERS Data Model

Strength	Weakness
Simple XML / ASCII	Events with high background rates
Simple, Relatively Inexpensive	Not able to clearly inference, when Side effect is not directly caused by Dug
Easy to Report	Identifying Drug Interactions with others
Includes all data reported direct or indirect	Not good for Comparing drugs in same class
Acts to capture side effects missed in early clinical trials (First level of detection)	Identification of trends and other significant safety concerns

OMAP (Observational Medical Outcomes Partnership) Common Data Model's purpose is to standardize the format and content of the observational data so that applications can use them in standardized way. FAERS is a good example of the CDM.

2.1.2 EU-ADRS Data Initiative

Adverse Drug Reaction Reports System (ADRRS) as part of EudraVigilance is a European Economic Area specific database created by European Medicines Agency to capture and dissipate suspected drug side effects across EU zone. It is the central drug side effects repository for EU zone even though the occurrence of the events might have happened anywhere in the world. This system has been in use since 2001 serving the primary purpose of collecting reports of suspected side effects. The intent is to further use this information to monitor the safety of the drugs approved in the EEU zone. This single window system drastically reduces the complexities in reporting the side effects in individual EU countries while bringing in the standards in reporting across the zone.

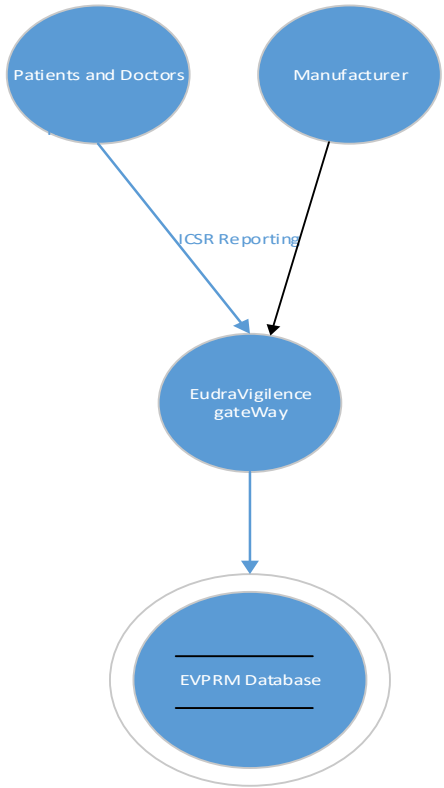


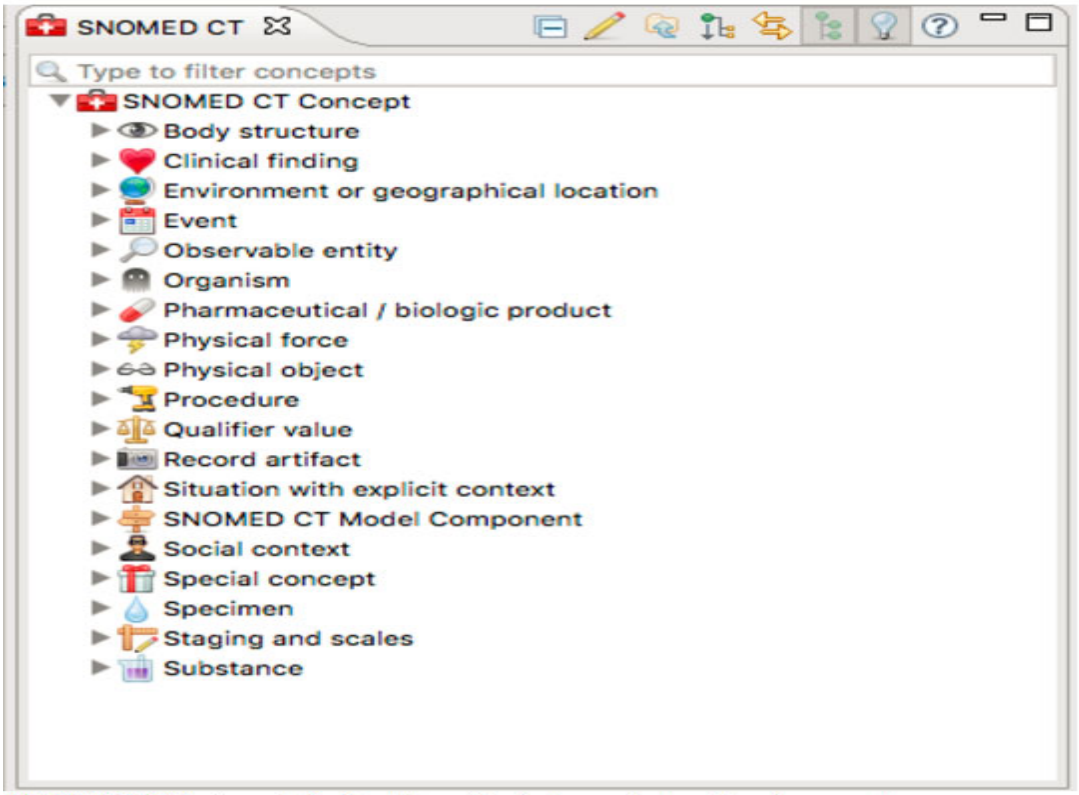
Figure 7 EudraVigilance ICSR Data reporting

The Side effects data collected using EudraVigilance is available for download at the website which contains information about the side effects in ASCII and XML format even though EVPRM system is itself in relational database. The Quarterly data files also provides summary data and the EVPRM data dictionary for the consuming systems to understand the format. In General, EudraVigilance (EU) system has several similarities with the FDA system in terms of providing an easy to report framework and storing in a centralized database for further analysis and consumption. In essence inheriting same strength and weaknesses as listed earlier.

2.1.3 Other Industry Wide Alternatives, SnowMed

SNOWMED CT (Systematized Nomenclature of Medicine -- Clinical Terms)[49] is a standardized vocabulary of clinical terminology that is used by physicians and other health care providers for the electronic exchange of clinical health information[25] .

SNOWMED CT organizes concepts into a tree by their IS A relationships.



SNOMED CT view, default setting displaying only top-level concepts

Figure 8 SNOWMED - standardized clinical terminology

This body of knowledge provides a simple and expandable structure of standardized clinical terminology for the medical domain experts allowing exchange of healthcare information real time. The System while providing standards for naming conventions by using a unique unambiguous fully specified name (FSN), is just a part of the overall landscape to provide the solutions for electronic data exchange for clinical terminologies. Its primary use remains clinical documentation and any further extension of SNOWMED CT for meaning based retrieval requires careful consideration of the data structure and not yet attempted for side effects inferencing.

2.2 DARs representation in XML

XML is the current data format of choice when it comes to the representation of the drug side effects data due to the factors like simplicity, openness, W3c standard, Extensibility, Self-descriptive nature of the tags and supports multilinguals documents and Unicode.

Both FAERS and EVPRM format offer DAR's in XML format due to factors like rapid adoption by the Industry and machine-readable context. In general, MedWatch website's FAERS data has several limitations when it comes side effects inferencing and is not suitable to calculate the incidence of an adverse event. As found with sources including Int J Med [4] The FAERS data when used for pharmacovigilance the ASCII or SMGL files not amendable to query and analysis. Even if any inferencing needs to be done it's nearly impossible to collect all these transaction data in quarterly format and coming up with a relevant feedback to doctors/ patients. For example, when an update is made to the side

effect to show the following new relations the knowledge cannot be updated to the historic data, rather released as new data. Even when this knowledge is released, quicker inferencing of the side effects is nearly impossible without an extensive set of processes and complex background operations

Let's look at the knowledge representation limitations with the FAERS data in detail. As displayed in Figure **9** the data is entered either by physician or patient most of the time and it reflects the transactional nature of the data reporting. Here every quarterly data is available in ASCII or SGML format representing the following types of information about the adverse event

- *demographic and administrative information and the initial report image ID number (if available);*
- *drug information from the case reports;*
- *reaction information from the reports;*
- *patient outcome information from the reports;*
- *information on the source of the reports;*
- *A "README" file containing a description of the files.*

As seen with the exponential growth of drug side effects in the last decade, lot of the side effects were originally stored in FAERS database which still acts as the foundational source for storing drug side effects data reported to Food and Drug Administration agency. This is a cumulative set of data collected by Food and Drug Agency and released quarterly basis to the general public. This data is often the primary source of Information for data analytic purposes. Due to the historic and transactional nature of the data, often it reflects the state of the side effects at that point of time thereby providing a strong foundation to understand the trend of the side effects. Here is one sample format (snippet)

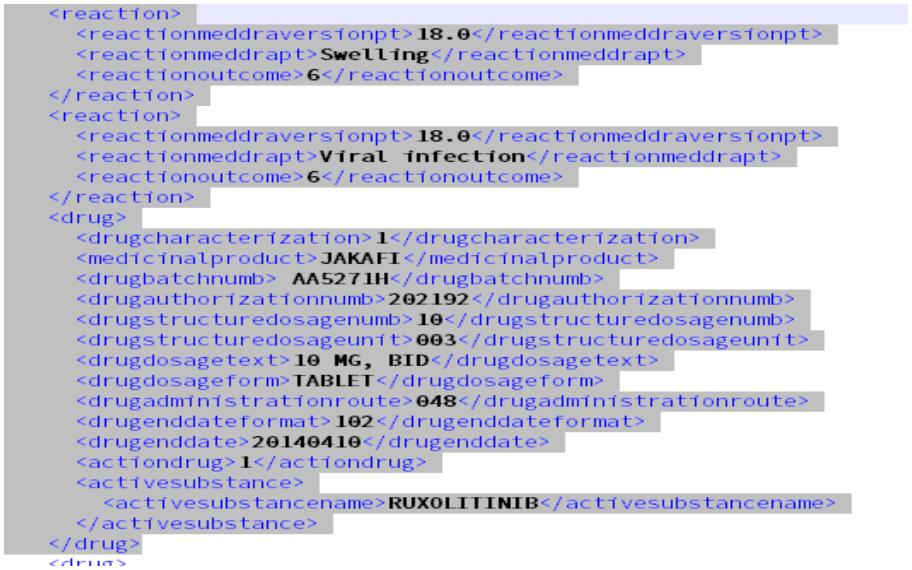


Figure 9 FAERS data in xml representation.

Drug RUXOLITINIB is known to cause "**swelling**" and "**Viral Infection**".

For example, when an update is made to the side effect to show the following

Drug RUXOLITINIB is known to cause "**swelling"** and "**Viral Infection**" AND '"**Bone pain"**

When this new knowledge is released, quicker inferencing of the side effects is nearly impossible without an extensive set of processes and complex background operations

2.3 DARs Representation in OWL

The introduction of Web Otology Language (OWL) to encapsulate drug side effects knowledge has trigged enormous interest in solving the challenges associated with efficient capturing the intricate details especially in capturing the relationships of "drug – drug" domain and "drug-side effects" domain enabling the discovery of relevant information.

This representation of drug medical knowledge using standardized ontologies using OWL has developed several medical ontologies such as SnowMed CT, Galen Ontology, DrOnt, DrugOnt and GeneOntology. Each of these ontologies have focused on certain subdomain of the medical field or a combination of the sub-domains within medical field allowing logical reasoners to tap into the existing ontologies as shown in the picture below.

For example, Earlier approaches were taken to solve the problem using standard drug ontologies relationships as proposed in "Drug ontologies by Samson and others [4]" .These studies focused on using the workaround '**has adverse effect**' object property, to link adverse effects with drugs , but the relationship between component and compound drug is still not represented. One usage example is,

Drug – <**has adverse effect**> – cough

As shown in "Extending OWL to support custom relationships" [18] these workarounds using object properties, forces a static nature on the relations as well as adding complexity in interpreting the data ,due to artificial intermediate notations. In addition with the syntax burden to knowledge modelers caused by the above representation, there are also other impediments like, mappings between patient record to drug data and the drug class information. All these constraints make current OWL only mode of drug adverse data knowledge representation lack the ability to maintain dynamic drug adverse relations and fail to provide an error free interpretation for doctors and patients.

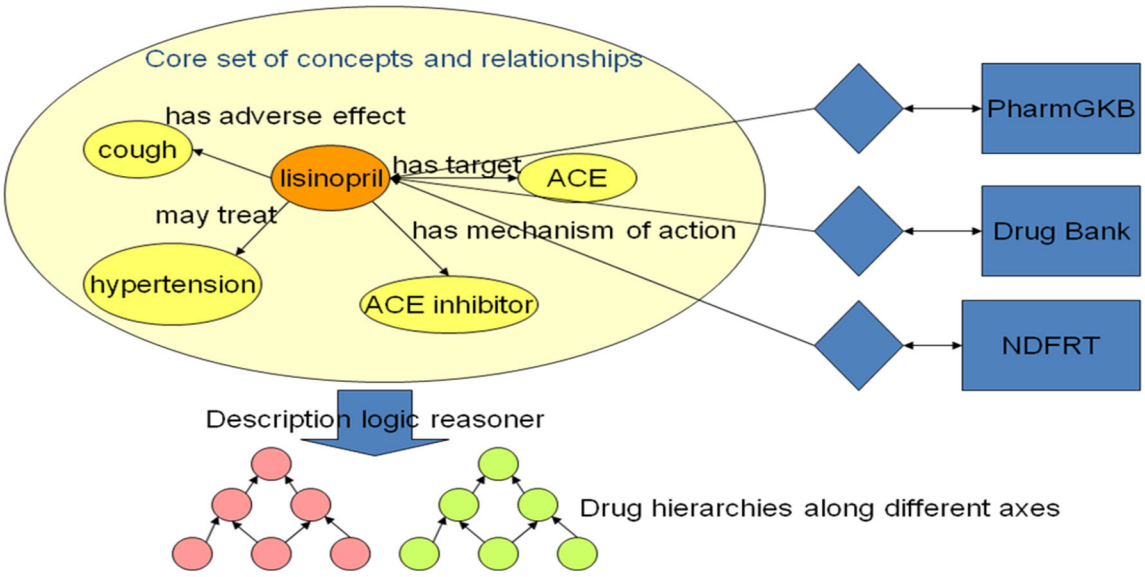


Figure 10 Literature survey 1: Drug Ontology by Samson et al – quoted from drug ontology documentation

While the approach to capture medical drug information using OWL is path breaking, the approach itself is restricted due to the known limitations of OWL in supporting only "is-a" relationship which causes medical knowledge experts to find workaround solutions using object property based approaches as shown in extending owl to support custom relations. As seen earlier, A drug may be classified by the "chemical type of the active ingredient" or by the way it is used to treat a particular condition meaning the current is-a or subclass approach to capture drug data is insufficient. Many approaches try to find workaround for this using object properties like "has mechanism of action" or "parent of". Then these workarounds force static nature around these relationships while in fact they are expected to be dynamic in nature. Each drug can be part of one or more drug classes adding more complexity as Drug adverse events marked with a compound (parent drug

class) applies to all its component drugs but the reverse is not true. All these relations are dynamic in nature requiring flexible approach to capture the knowledge. For example, using object properties we cannot express the relation that drug Lisinopril can only be part of at most one drug class or part of two drug classes. These kind of domain level restrictions won't work with the object property. In addition the custom relationship emulation using workarounds like object properties adds undue burden on domain experts to create the knowledge expression and validation.

For example, Drug dapagliflozin [29]from drug class SGLT-2 inhibitors can cause the following adverse effect: diarrhea, increased weight, increased blood sugar, urinary tract infection and fungal infection. While Drug class SGLT-2 inhibitor itself can cause the following side effects, breathing, nausea, abdominal pain, confusion, vomiting, sleepiness, unusual fatigue. Due to the OWL's restriction to only allow "is-a" relationship, the representation of such as information requires workaround object properties like "part-of" or "has mechanism of action" to represent the knowledge.

2.4 Restrictions with Current Approaches

Let's look deeper into the current representation in FAERS and OWL format

Figure 11 FAERS sample

"51 year old Female in USA who is taking Humira for Arthritis has reported "Injection Site Pain" as an adverse reaction since Apr 11, 2011. She is also taking other medication SIMVASTATIN and PREVASTATIN."

Here we can see that some of the key knowledge representation details are lacking. For example

1) Side Effects Data is transactional and does not represent the relationship between component and compound drugs. Since component drugs can derive some additional side effects from parent class, there is an important knowledge gap to be filled

2) Hard to interpret by knowledge experts or knowledge processors. The transactional format as it appears is more suitable for easier data collection than smarter knowledge inference. This is understandable considering the primary goal for these systems is to provide a welcoming data collection platform for capturing side effects.

This is highly error prone during interpretation due to the missing knowledge mechanism framework. For example, the relationship between side-effect to drug is captured but the relationship between drug to drug classes or its super class is not fully captured. This will lead to end users like doctors coming to inconsistent inference causing errors during interpretation

2.5 Comparative Study

Most of the drug adverse data is released by FDA as FAERS database in the form of relational database and transactional in nature. The systems are built to facilitate a simple way for patients / doctors to report any drug side effects by proving an easy to enter online forms or other communication channels to capture as much data as possible. While this approach is welcomed and brings in more data, the quality of this data often is found to be lacking detailed information about the event making it harder for data scientist to derive at. With the principle of X in X out, we can’t expect to get quality data as outcome of the

MedWatch putting enormous burden in maintain syntax and relationships. For example, the following four side effects could be reported and even though they are all same, the system treats all as different, “I have got headache”, “Pain in the head” and “head hurts” or “migraine”. Another simple example is that the following three conditions “My nose bleeds”, “I see blood in nose”, and “bloody nose” may mean the same thing. This simple examples proves that while the current approach works as a bulk data capture mechanism it’s extremely complex for even data scientist to derive meanings lest the actual users like doctors or patients.

The current approach also brings issues with knowledge relevance to handle any changes to the drug data to keep them relevant. Drug information data is highly susceptible to change due to various industry changes or regulator advice. Assuming few metadata about the drug is changed by regulatory authorities and now all the historic information with the metadata has to be corrected, the challenges in going back and updating the data is highly costly. There is no flexibility to handle any changes due to the fact that all data is tied up with older metadata. This could develop into a situation where the cost of updating historic data is more than cost of importing current date making it less attractive for agencies and organizations to do so.

Table 3 Summary of Known Ontology-Based Drug Knowledge Representation

Author	Year	Representation Approach	Ideal Use-case	Comments	Language
Charalampos Doulaverakis et al	2012	GalenOWL [31] – A Semantic enabled online service which extends the current GALINOS drug database.	This service can be used to query drug database and get information from GALINOS.	OWL is used for expression and representation of the ontologies with focus on drug- drug interactions. Some challenges due to limitations of standard OWL expressions.	OWL, ICD-10
Josh Hanna et al	2013	DrOn [32]– DrugOntology-Simpler extension and Development of the ontology to allow reasoning and construction to scale. Built using RxNorm as primary source.	Helps to reuse drug information from existing sources using standardized ontology artifacts built from RDBMS sources.	The focus seems to be on drug and drug dose forms and little work on drug side effects. Suffers from the same limitations on the standard owl approaches.	OWL2.0
Garry Merrill et al	2008	SafetyWorks – based on UMLS/SNOWMED based Drug Ontology[48]	Development of an Integrated set of methodologies enabling the use of large observational data sources in data exploration and analysis applications.	Intended goal is to works on the principle of Extract, Annotate, Normalize, and Evaluate Drug data. Some problems identified in correctly classifying Drug forms with Drugs.	Hybrid Approach using OWL, RDBMS
Yongqun He et al.	2013	OAE[46]: The Ontology of Adverse Events	Community-driven ontology developed to	Goal is to Improve representation and	OWL

			standardize and integrate data. OAE has over 3,000 terms with unique identifiers, including terms imported from existing ontologies and more than 1,800 OAE-specific terms	organization of adverse event information using different vocabulary resources like Medical Dictionary for Regulatory Activities (MedDRA)[4], the Common Terminology Criteria for Adverse Events (CTCAE)[5], and the World Health Organization (WHO)'s Adverse Reaction Terminology (WHO-ART)	
Joanne Luciano et Al.	2011	Translational Medicine Ontology (TMO) [47]	Integration of knowledge using heterogeneous data from health care to the life sciences	Ontology to integrate chemical, genomic and proteomic data with disease, treatment, and electronic health records	OWL, SPARQL
National Library of Medicine	2005	RxNorm[32]	Provides normalized names for clinical drugs and links its names to many of the drug vocabularies commonly used in pharmacy management and drug interaction software	RxNorm is the source for many different applications in Healthcare area.	Unified Medical Language System

2.6 Summary of findings

The above analysis proves that Expression, Representation and Interpretation of the Drug side effects data continues to remain a challenge to Domain experts and there is an unmet need for a knowledge representation which can address some of these shortcomings allowing flexibility to update the side effects data relationships without incurring enormous cost on system resources. For example, often the transitive nature of the drug-drugclass-drug side effects is not captured in the current format making it unusable to derive full set of meanings. With all these data management problems with the current format, Doctors and Patient suffer the most in finding out what drug side effects are actually caused by which drug groups.

This dissertation intends to address this issue by introducing knowledge graph with custom OWL relationship model to drug adverse data domain and validate its usability with a web based application to be used by doctors during prescription check visits.

2.7 Key Tools and Methodology

In the following section we will look at the some of the key Tools and Methodologies to be used in this research.

2.7.1 Knowledge Graph

A Knowledge Graphs describes the concepts in the domain and also the relationships that hold between those concepts. Knowledge Graphs are often used to capture knowledge about some domain of interest. It makes it possible for concepts to be defined as well as

described. Using Knowledge graphs, Complex concepts can therefore be built up in definitions out of simpler concepts in a modular fashion.

2.7.2 Usage of Knowledge graphs in Healthcare

The usage of Knowledge graph is will accepted especially in the semantic web area as a primary way of disseminating the information to users or machines even though it's still evolving in the medical domain. For example, Google's knowledge vault is enriched with information about 570 million objects of data and 18 billion of facts making the world's largest public knowledge graph vault.

2.7.3 RDF and RDFS

RDF (Resource Description Framework) is a standard for metadata which offers a standard way of specifying metadata about any *resource*. So a *resource* in turn is anything which is described by RDF expressions. A resource can be any real world item like a medical product or anything or something like a webpage, part of webpage. Each resource in semantic web is identified by a Uniform resource identifier (URI), and this URI is used as the global name or the resources to uniquely identify them.

Here is an example often used in semantic web. The following URI uniquely identified a resource.

http://www.yuchen.net/photography/SLR#Nikon-D70

An RDF statement as described is used to describe a property of resources in the triple format.

Resource (subject) +property (predicate) + property value (object)

A resource drug, drug name, product in RDFS is a data model for representing information about resources in the web.

RDF is intended for semantic web where information about web resources needs to be processed by applications rather than being displayed to people.

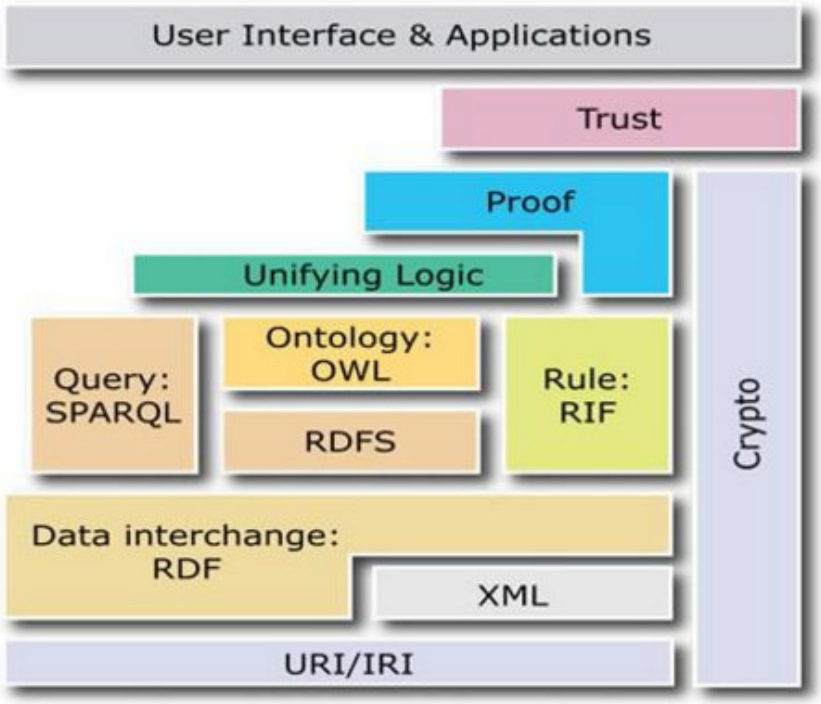


Figure 12 Semantic Web Layer

RDF statements can be written down using triple notations [19]

Subject	Predicate	Object
mySLR:Nikon-D70	mySLR:weight	1.4 lb

2.7.4 OWL and Custom Relationships

OWL classes are interpreted as sets that contain individuals. They are described using formal (mathematical) descriptions that state precisely the requirements for membership of the class. For example, the class Cat would contain all the individuals that are cats in our domain of interest. Classes may be organized into a superclass-subclass hierarchy, which is also known as a taxonomy.

As shown in "Extending OWL to support custom relationships [18]" these emulation of class relationships using object properties, forces a static nature on the relations as well as adding complexity in interpreting the data ,due to artificial intermediate notations. In addition with the syntax burden to knowledge modelers caused by the above representation, there are also other impediments like, mappings between patient record to drug data and the drug class information. All these constraints prove that the current knowledge representation of drug adverse data lack the ability to maintain dynamic DAR (drug adverse relation) and provide an error free interpretation for doctors and patients.

2.7.5 Apache Jena

Apache Jena is an open source semantic web framework for Java providing an API to extract data and write to RDF graphs.

2.7.6 Pace Jena

Pace Jena is an extension of Apache Jena open source semantic framework specifically written to cover the complexities in supporting custom OWL relationships with Protégé. Pace Jena is specifically useful for parsing the knowledge graph with custom OWL relationships which is demonstrated in the inferencing the knowledge graph section of this research.

2.7.7 OWLVIZ

OWLVIZ is an extension framework proving the tree visualization of the sematic relationships for the domain expert users helping them to visualize the semantic relationships between classes. OWMVIZ visualization is typically useful to get a VIEW of the Knowledge graph of reasonable size. Once the Knowledge graph is fully built spanning several pages then a tree visualization tool is often required to see the knowledge graph. The screenshots used in the research is often taken from the OWLVIZ window.

2.7.8 Protégé Tool with Pace Jena

Protégé [28] is an IDE provided from Stanford University which provides suites of tools to construct domain models and knowledge based applications with ontologies. This is a Project compiled using Maven script and once compiled it launches the IDE for the OWL editors to start building the Classes and Relationships in a WYSIWYG editing mode.

The Extensions added by Pace University adds “Relations” tab to Entities window thereby enabling Domain experts to relate the Classes using new custom relationships.

This dissertation uses Pace enhanced Protégé to build the Domain relationship using selected custom relationships.

2.7.9 Knowledge Association

Association is a (*a*) relationship between two classes, which allows one object instance to cause another to perform an action on its behalf. It defines a *has-a* relationship between two classes where there is no particular ownership in place. Association is the more general term that defines the relationship between two classes, where as aggregation and composition are relatively special. Aggregation is a weak type of Association with partial ownership.

2.7.10 OWL is a relationship

One of the most important relationships among objects in the real world is specialization. Specialization can be described as the “*is-a*” relationship. The statement, “A dog is a mammal”, means that the dog is a specialized kind of mammal. Having all the characteristics of any mammal, (the fact that it bears live young, nurses with milk, has hair etc.), it specializes these characteristics to the familiar characteristics of canis domesticus.

2.7.11 OWL part of relationship

The *PartOf* relationship is more formally known as composition. Composition is a strong type of Association with full ownership. The term used for a Composition relationship, is **owns** or **part_of* to imply a strong **has-a** relationship. For example, a department **owns** courses, which means that any course's life-cycle depends on the department's life-cycle. Hence, if a department ceases to exist, the underlying courses will cease to exist as well." Relationships with no ownership in place are regarded as just an Association and the term used is **has-a**, or sometimes the verb describing the relationship. "For example, a teacher **has-a** or **teaches* a student. There is no ownership between the teacher and the student, and each has their own life-cycle". Protégé has emerged as one of the most popular interfaces for creating knowledge. The standard open source protégé primarily uses the inheritance relationship, *is_a* to represent knowledge. Even though current OWL files can emulate the custom relations by using this approach, relying solely on the is_a inheritance relationship construct to describe knowledge is rather restrictive and some domain experts find it awkward to understand and to use it to validate their knowledge representations.

In the quest for a more expressive way of representing knowledge, this research dispenses with these complexities and adds to *is_a* constructs for custom relations using a Knowledge Graph, which is a more natural and flexible way to represent knowledge, and can be used to overcome the limitations and restrictiveness of the sole *is_a* relation.

2.7.12 Conclusion

In Summary, This chapter covered key concepts, terminologies and tools like Semantic Web, Knowledge Graph, RDF, RDFS, Protégé, Pace Jena and OWL relationships ("is a " and "part of") to support this research. These tools and technologies will be later used to demonstrate the Knowledge graph and full spectrum inferencing using a sample of data spanning several classes.

Chapter 3

Solution Methodology

3.1 Knowledge Graph - Drug Side Effects Data

A Knowledge graph describes the concepts in the domain and also the relationships that hold between those concepts. Different knowledge graph languages provide different facilities. It makes it possible for concepts to be defined as well as described. Complex concepts can therefore be built up in definitions out of simpler concepts.

3.1.1 Knowledge Graph Usage in Healthcare

The usage of Knowledge graph is will accepted especially in the semantic web area as a primary way of disseminating the information to users or machines even though it's still evolving in the medical domain. For example, Google's knowledge vault [50] is enriched with information about 570 million objects of data and 18 billion of facts making the world's largest public knowledge graph vault.

In Knowledge graph, classes are interpreted as sets that contain individuals. They are described using formal (mathematical) descriptions that state precisely the requirements for membership of the class. For example, the class SGLT2 inhibitors would contain all the prescription drugs that are SGLT2 type of drugs in our domain of interest. Classes may be organized into a superclass-subclass [26] hierarchy, which is also known as a taxonomy. The modularity of the knowledge graph makes it a perfect fit for knowledge representation

of medical domain especially drug side effects data. Unlike traditional approaches where the focus is storage of the data with less consideration of the timely interpretation or reasoning, the primary goal of the knowledge graph is to enable timely retrieval of the knowledge in this case the use by Doctors or Patients to retrieve time sensitive data. The key once again is the modular ability of the knowledge graph to extend and grow making it an ideal option to store drug adverse data making it highly suitable for capturing drug side effects data due to the dynamic nature of the domain.

3.1.2 Knowledge Graph and Pace Jena Extension

The introduction of Web Otology Language (OWL) to encapsulate drug side effects knowledge has trigged enormous interest in solving the challenges associated with efficient capturing the intricate details especially in capturing the relationships of "drug – drug" domain and "drug-side effects" domain enabling the discovery of relevant information. This representation of medical using standardized ontologies using OWL has developed several medical ontologies such as Snow med CT, Galen Ontology, DrOnt, DrugOnt and GeneOntology. Each of these ontologies have focused on certain subdomain of the medical field or a combination of the sub-domains within medical field allowing logic reasoners to tap into the existing ontologies.

While the approach to capture medical drug information using OWL is path breaking, the approach itself is restricted due to the known limitations of OWL in supporting only "is-a" relationship which triggers to find workaround solutions using object property based approaches as shown in extending owl to support custom relations. For example, A drug

may be classified by the “chemical type of the active ingredient” or by the way it is used to treat a particular condition meaning the current is-a or subclass approach to capture drug data is insufficient. Many approaches try to find workaround for this using object properties like “has mechanism of action” or “parent of”. Then these workarounds force static nature around these relationships while in fact they are expected to be dynamic in nature. Each drug can be part of one or more drug classes adding more complexity as Drug adverse events marked with a compound (parent drug class) applies to all its component drugs but the reverse is not true. All these relations are dynamic in nature requiring flexible approach to capture the knowledge. For example, using object properties we cannot express the relation that drug Lisinopril can only be part of at most one drug class or part of two drug classes. These kind of domain level restrictions won’t work with the object property. In addition the custom relationship emulation using workarounds like object properties adds undue burden on domain experts to create the knowledge expression and validation.

For example, Drug dapagliflozin [29],[30]from drug class SGLT-2 inhibitors can cause the following adverse effect: Diarrhea, Increased weight, increased blood sugar, Urinary tract infection and fungal infection. While Drug class SGLT-2 inhibitor itself can cause the following side effects, breathing, nausea, abdominal pain, confusion, vomiting, sleepiness, unusual fatigue. Due to the OWL’s restriction to only allow “is-a” relationship, the representation of such as information requires workaround object properties like “part-of” or “has mechanism of action” to represent the knowledge.

As shown in “Extending Owl to support custom relationships [18]” these emulation of class relationships using object properties, forces a static nature on the relations as well as adding complexity in interpreting the data ,due to artificial intermediate notations. In addition with the syntax burden to knowledge modelers caused by the above representation, there are also other impediments like, mappings between patient record to drug data and the drug class information. All these constraints prove that the current knowledge representation of drug adverse data lack the ability to maintain dynamic DAR (drug adverse relation) and provide an error free interpretation for doctors and patients.

3.2 Proposed Framework

3.2.1 DARs (Drug adverse relationships) using custom relationships

The Research proposes a Knowledge graph based drug side effects representation model to drastically improve the knowledge representation on Drug Adverse Relations (DARs) using the custom relations as in “Extending OWL to support custom relations” as well as providing a concept demonstrator mechanism to extract the DARs dynamically. This methodology relies heavily on linking the drug (component) vs drug class (compound) and drug vs side effects using custom OWL relationship based approach as well as providing a proof of concept application. The research also strives to prove how the DARs data can be represented in knowledge graph and demonstrate that it brings out meaningful interpretations to doctors and care givers.

This proposed knowledge graph (KG) based representation model (D-SERI) uses custom OWL relationships *Takes, Cause and PartOf* to capture constraints of compound and

component drug thereby providing a strong framework for Drug Adverse Reactions (DARs) knowledge capture**.** As shown in the benefits of "Extending OWL relationships" these KG representation also allows relationships to stay dynamic in nature thereby reducing the burden on maintaining syntax while processing. In essence it uses three custom relationships as shown below.

Patient *takes* Drug, Drug *cause* side effects, Drug *partOf ParentClass*

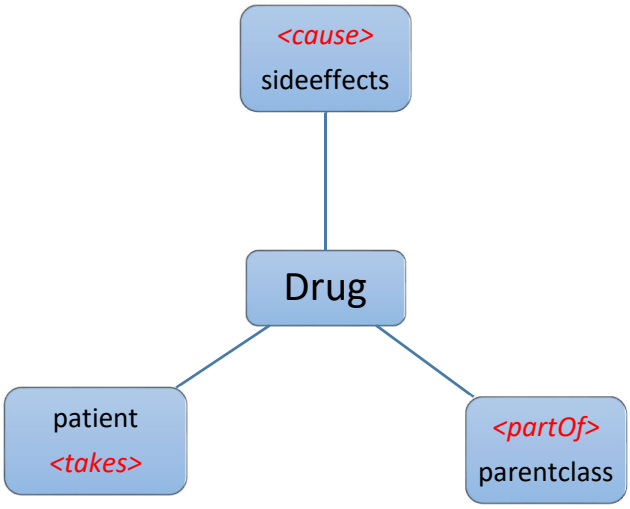


Figure 13 Foundational knowledge representation - using custom OWL relations *takes, cause and partOf* in drug domain data.

The key entities for this model are *patient, drug, side effects and parent class*. These four OWL entities are linked using custom OWL relations *takes, cause and partOf* allowing a linking of knowledge for a drug which is modular in nature. This modular data is easily expandable and as shown in Figure **13**, it can be expanded to cover the drug data for any specific drug. The approach brings issues with Syntax relevance to handle any changes to

the drug data to keep them relevant. Drug information data is highly susceptible to change due to various industry changes or regulator advice. Assuming few metadata about the drug is changed by regulatory authorities and now all the historic information with the metadata has to be corrected, the challenges in going back and updating the data is highly costly. There is no flexibility to handle any changes due to the fact that all data is tied up with older metadata. By using the custom relationship instead of workarounds like object properties, the model allows the flexibility to update the data without incurring the cost. For example when the Syntactical relevance of the knowledge is not captured in the current format the problem gets complicated in its usability to derive meanings.

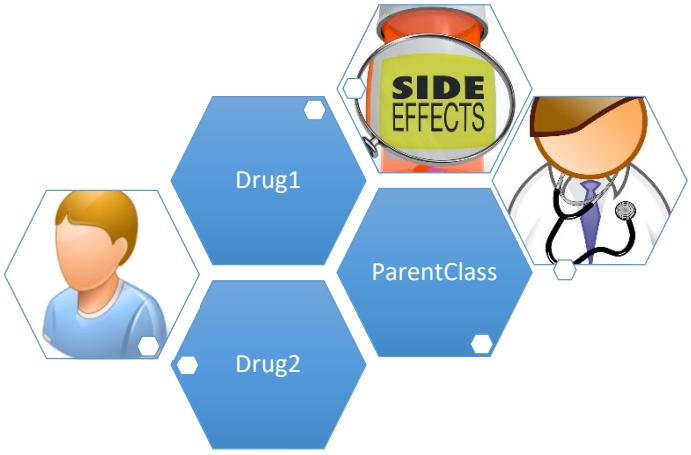


Figure 14 Knowledge Graph linked data model - seamless integration of Patient, Drugs, Side effects, Drug Class and Doctors

3.2.2 Syntax Definition

The Research uses three custom relations for this feasibility study which were defined in Protégé with Pace extension as follows.

Syntax for "cause" custom relationship in Pace Protégé

<rdf:RDF xmlns:rel=http://www.pace.edu/rel-syntax-ns# >
<rel:NewRelation
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#cause"/>

Syntax for "partOf" custom relationship in Pace Protégé

<rdf:RDF xmlns:rel=http://www.pace.edu/rel-syntax-ns# >
<rel:NewRelation
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#partOf"/>

Syntax for "takes" custom relationship in Pace Protégé

<rdf:RDF xmlns:rel=http://www.pace.edu/rel-syntax-ns# >
<rel:NewRelation
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#takes"/>

3.2.3 Knowledge Graph Development

For the purpose of research, this study considers to analyze the drug side effects caused by saxagliptin along with its parent class DPP4Inhibitors and Antidiabetic Drugs. Saxagliptin being the class of gliptin automatically derives the side effects associated with its class. The data used for the KG development is a snapshot of the publically available drug side effects data sources. In order to show the reusability of the concept we will apply this on two different classes/parent classes to repeat the results.

First we take a drug class “DPP-4 inhibitors” which are primarily used to lower blood sugar in adults with type 2 diabetes [45]. Medicines in the DPP-4 inhibitor class include sitagliptin, saxagliptin, linagliptin, and alogliptin. They are available in the market as single-ingredient products and in combination with other diabetes medicines. When untreated, type 2 diabetes can lead to serious problems, including blindness, nerve and kidney damage, and heart disease. DPP-4 inhibitors lower blood sugar by helping the body increase the level of the hormone insulin after meals. Insulin helps move sugar from the blood into the tissues so the body can use the sugar to produce energy and keep blood sugar levels stable. In addition to severe joint pain, other possible side effects of DPP-4 inhibitors include inflammation of the pancreas, low blood sugar when this class of medicines is combined with other prescription medicines used to treat diabetes, and allergic reactions.

The research defines DPP4Inhibitors under the Class DrugClass→AntiDiabeticDrug and lists Drugs→Saxagliptin,Sitagliptin and Aloglipton.

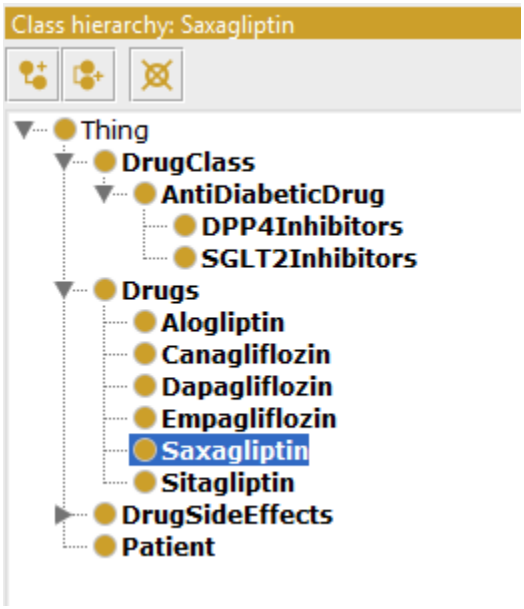


Figure 15 Protégé - Class Hierarchy definition

Once the Parent Drug Class and Drug are defined for DPP4Inhibitors in Protégé, the research further captures the side effects to the component and compound drug using the part of custom relations as shown below in xml.

Syntax for "AntiDiabeticDrug" Drug Class and its side effects in Pace Protégé

```
<!--  http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#AntiDiabeticDrug -->
  <owl:Class rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#AntiDiabeticDrug">
    <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#DrugClass"/>
<rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Hypoglycemia"/>

  </owl:Class>
```

Here are the OWLViz visualizations for the above the syntax.

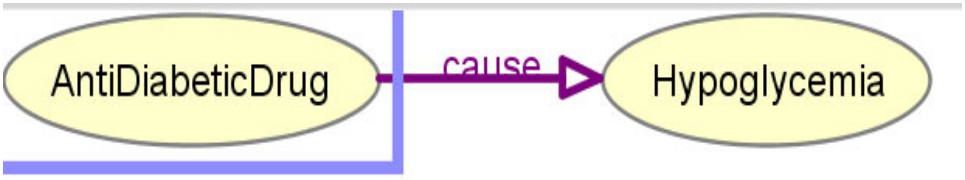


Figure 16 OWLViz - AntiDiabeticDrug

Syntax for ”DPP4Inhibitors” Drug Class and its side effects in Pace Protégé

```
   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#DPP4Inhibitors -->

   <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DPP4Inhibitors">
      <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AntiDiabeticDrug"/>
     <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#StomachPain"/>
      <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Pancreattis"/>
      <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Nausea"/>
      <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Diarrhoea"/>
      <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Headache"/>
      <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#SevereJointPain"/>
      <rel:partOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AntiDiabeticDrug"/>
```

Here are the OWLViz visualizations for the above the syntax.

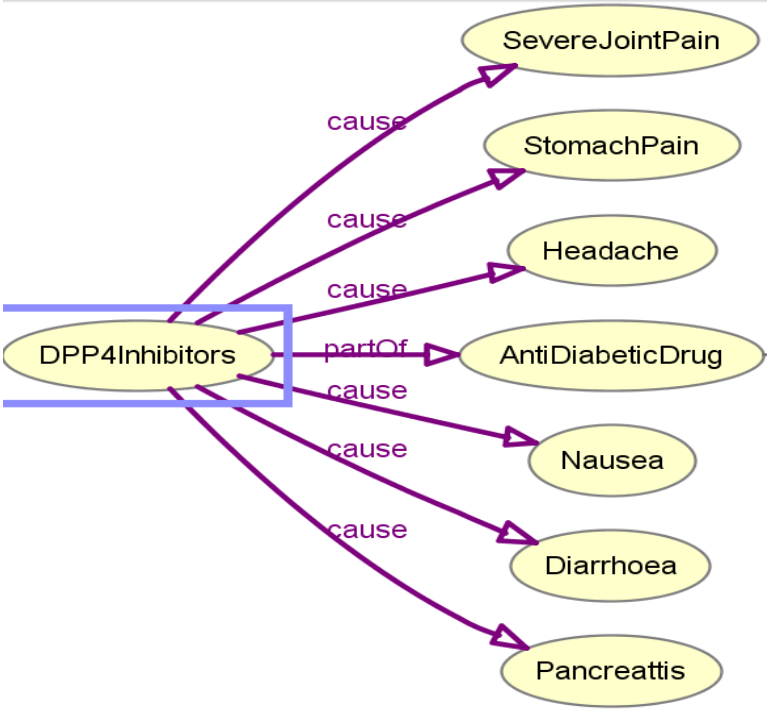


Figure 17 OWLViz – DrugClass

Syntax for ”Saxagliptin” Drug linked with its Parent Class “DPP4Inhibitors” along with its side effects in Pace Protégé

```
<!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Saxagliptin -->

<owl:Class rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Saxagliptin">
    <rdfs:subClassOf rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Drugs"/>
        <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Rash"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Arrhythmia"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Malaise"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#NasalCongestion"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#Hyperhidrosis"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#IncreasedBloodsugar"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#AbdominalPain"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#MotorDysfunction"/>
    <rel:cause rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#CerebroVascularAccident"/>
    <rel:partOf rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#DPP4Inhibitors"/>
</owl:Class>
```

Here are the OWLViz visualizations for the above the syntax.

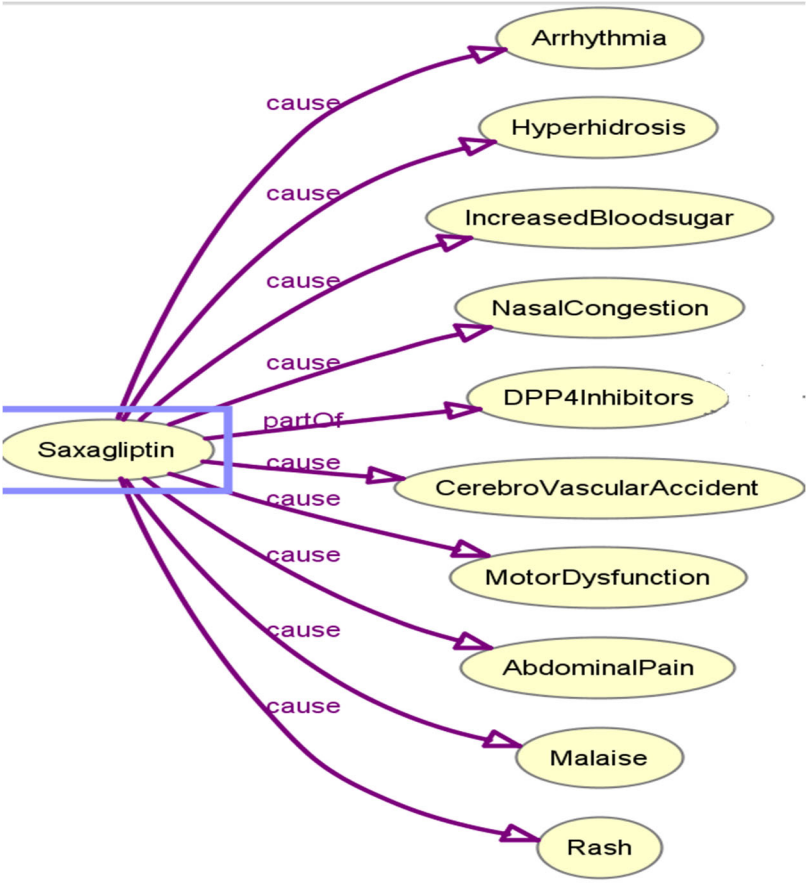


Figure 18 OWLViz – Drug

3.2.4 Inferred Results

Now the research arrive at the critical step of Inferring the side effects data for Saxaglipton where the side effects caused directly by the drug is listed as well as other possible derived outcomes for the drug is captured using the partof custom relationships in the Protégé OWLViz. In the following we expect the side effects caused directly by saxagliptin drug and the side effects derived by its component and parent drug classes all to be directly inferred using the linked data model used in this research. This OWLViz was able to infer the "combined" adverse reactions of component and compound drug class using the newly introduced part of relationship into the drug side effects domain.

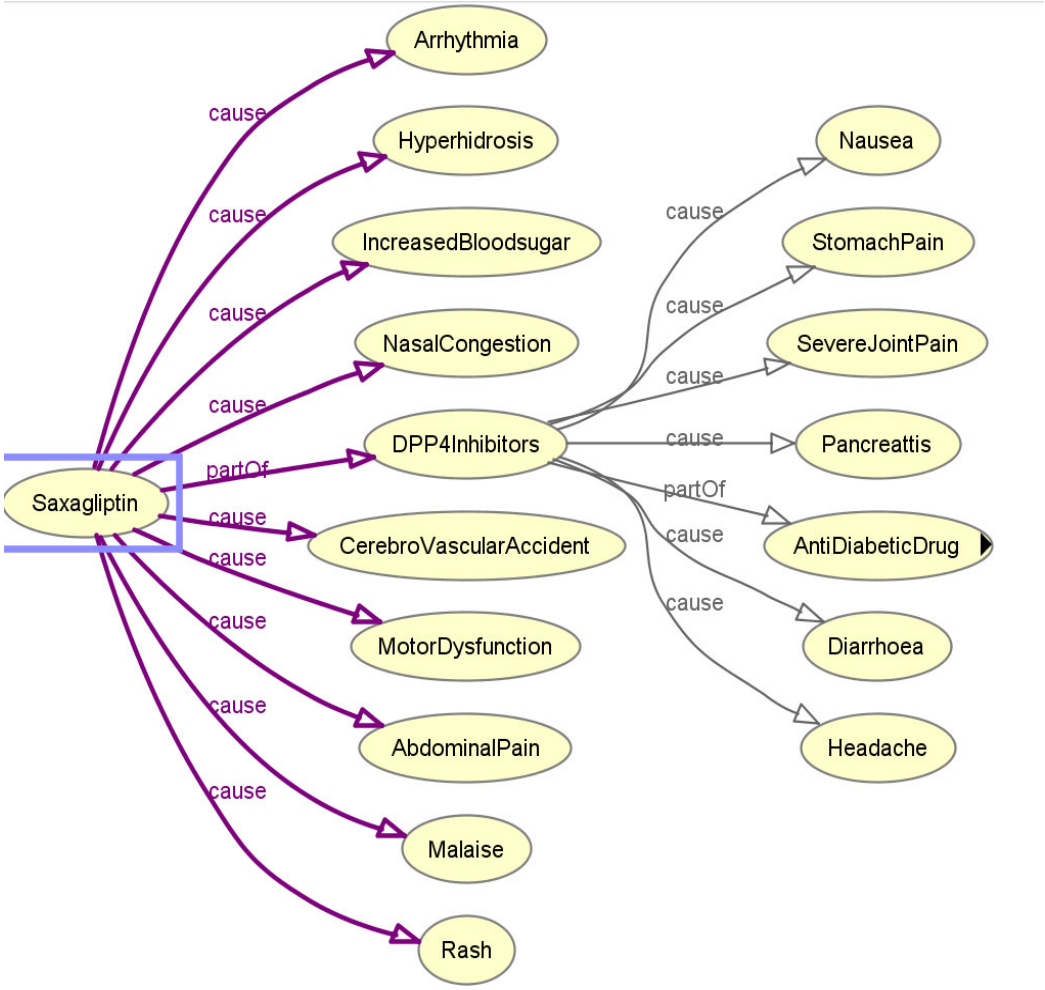


Figure 19 Adverse reactions caused by Saxagliptin drug with class DPP4Inhibitors and AntiDiabeticDrug

This is a significant improvement to the current state in building the knowledge graph for the drug side effects domain. Due to the current limitations of owl to extend beyond “is-a” relationship this relationship is only possible using object property emulation which is error prone and causes syntax relevance issues to domain experts. This new approach seamlessly enables the domain experts to represent complex medical side effects domain knowledge

like part of transitive relations using the Pace enhanced Protégé while reducing the complexities in maintaining syntax and processing the outcomes for patients and doctors.

3.3 Solution Methodology - Outcome

As shown in Figure **14**, the key entities for this model are *patient, drug, sideeffects and parentclass*. These four owl entities are linked using custom owl relations *takes, cause and part of* allowing a linking of knowledge for a drug which is modular in nature. This modular data is easily expandable and shown to work to cover the drug data for any specific drug. Drug information data is highly susceptible to change due to various industry changes or regulator advice. The approach allows the knowledge representation model to handle any changes to the drug data to keep all relationships relevant.

Assuming key metadata about the drug side effects is released by regulatory authorities frequently and due to this, all the historic information with the metadata has to be updated to reflect the new relationship, the challenges in going back and updating the data is highly costly as there is no flexibility to handle any changes due to the fact that all data is tied up with older metadata. By using the custom relationship instead of workarounds like object properties, the model allows the flexibility to update the data without incurring the cost. For example when the Syntactical relevance of the knowledge is not captured in the current format the problem gets complicated in its usability to derive meanings.

The Knowledge Graph developed using the proposed D-SERI model clearly addresses the shortcoming identified allowing the capture of constraints of compound and component drug thereby providing a strong expandable semantic framework for side effects knowledge capture. This solution derives new adverse reactions about component drugs using compound-component knowledge representation which was not possible before.

In Summary, by linking all the key entities like patients, drugs, side effects, doctors and drug classes using custom owl relationships, the proposed ***D-SERI*** model allows domain experts a model to represent the complex and dynamic knowledge associated with the drug side effects data in an inference friendly way. These Newly derived knowledge information can be used to develop automated prescription check processes using Pace Jena which can be used by patients and doctors. The Concept demonstrator in Section 4 will demonstrate this usability of the knowledge information.

3.4 Solution Methodology Files

3.4.1 Snippet of DrugSummary.owl

This section shows a section of the knowledge graph developed for sample purposes.

```
<!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#DrugClass -->

<owl:Class rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-18#DrugClass">
```

```
        <owl:disjointWith
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DrugSideEffects"/>
        <owl:disjointWith
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Drugs"/>
        <owl:disjointWith
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Patient"/>
    </owl:Class>


   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#DrugSideEffects -->

    <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DrugSideEffects">
    </owl:Class>



    <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#Drugs -->

    <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Drugs">
    </owl:Class>




   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#DrugSideEffects -->

    <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DrugSideEffects">
    </owl:Class>
```

```
   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#Drugs -->

   <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Drugs">
   </owl:Class>

   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#AntiDiabeticDrug -->

   <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AntiDiabeticDrug">
      <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DrugClass"/>

                  <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Hypoglycemia"/>

   </owl:Class>


   <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#Patient -->

   <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Patient">
      <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Saxagliptin"/>
      <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Sitagliptin"/>
      <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Canagliflozin"/>
      <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Dapagliflozin"/>
```

```
            <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Alogliptin"/>
        <rel:takes
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Empagliflozin"/>
    </owl:Class>



    <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Alogliptin">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Drugs"/>

            <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#CardiacArrest"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AbnormalLiverFunction"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Rash"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Pancreattis"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Constipation"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Fever"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#HepaticPain"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#IncreasedAppetite"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#PancreaticCarcinome"/>
```

```
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Nasopharyngtitis"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#IncreasedCholestrol"/>
        <rel:partOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DPP4Inhibitors"/>

</owl:Class>

    <!-- http://www.semanticweb.org/sar/ontologies/2015/10/untitled-ontology-
18#DPP4Inhibitors -->

    <owl:Class
rdf:about="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#DPP4Inhibitors">
        <rdfs:subClassOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AntiDiabeticDrug"/>
                <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#StomachPain"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Pancreattis"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Nausea"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Diarrhoea"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#Headache"/>
        <rel:cause
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#SevereJointPain"/>
        <rel:partOf
rdf:resource="http://www.semanticweb.org/sar/ontologies/2015/10/untitled-
ontology-18#AntiDiabeticDrug"/>
    </owl:Class>
```

3.5 Conclusion

This Chapter showcased the benefit of developing a patient focused KG development for DARs using custom OWL relationships instead of object properties allowing drug adverse relationships to stay dynamic in nature thereby reducing the burden on maintaining syntax while processing. Introduction of custom OWL relationships like PartOf allows to discover adverse reactions about component drugs in a much simpler way. These derived Semantic triples can now be used automating the prescription check process through checking constraints using Pace Jena by patients and doctors in the next chapter.

This research solution to the DAR's data representation is unique in introducing custom OWL relationships to DARs, to enable seamless linked data integration, automatically identifying full set of potential side effects for any given drug.

Chapter 4

Solution Implementation.

4.1 Introduction

Chapter 3 described the solution methodology in detail and showed how the key concepts of custom OWL relationships, PaceJena and Pace Protégé help build the knowledge graph to capture the drug domain information. This chapter will further demonstrate how the knowledge graph serves the key purpose of providing full spectrum inferencing to doctors and patients using a dynamic on demand web interface for inferencing the data run time. This section will also describe the development of the PaceJena extension code which helps to provide the tool for inferencing the knowledge graph dynamically using a simple standardized Java interface.

As we saw earlier in Chapter 3, this research introduced the custom OWL relationships to capture drug domain data thereby overcoming OWL's traditional lack of support for custom relationships and limited inferencing capabilities in the traditional OWL. This section specifically addresses the problem identified earlier where its error prone in its mechanism for patients to interpret the data due to the artificial annotations and intermediate concept usage.

4.2 Implementation Details

This section describes the proposed dynamic inferencing model to extract the full spectrum side effects for the drug using PaceJena which is a Pace university extended version of

open source java framework for building knowledge graphs. PaceJena is specifically built to support the extended version of owl using custom relationships. This research greatly extends the capabilities of owl to be used extend the support of custom relationships for drug side effects domain. Since the drug side effects are linked with its compound and other classes, we propose to create an iterative approach to derive the full set of side effects for any drug.

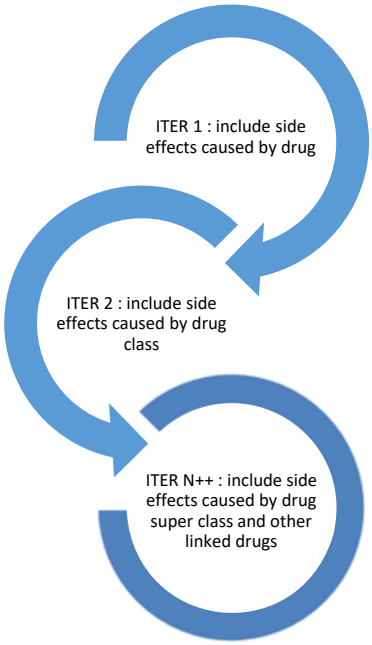


Figure 20 D-GPR – An Iterative approach to get full spectrum side effects using knowledge graph

As shown in Figure 20 it's critically important to link the drug to all possible side effects using an iterative model built on the framework of PaceJena allowing domain experts like Doctors to understand all possible outcomes about the drug side effects. We restrict the iterations to 2 levels for this research even though this could be expanded as needed based

on the complexity of the situation. The aim of the study is to prove the concepts using the model using a sample set of data created using knowledge graph.

In high level, ***D-GPR*** algorithm is designed in the following sequence

1) Input Drug Name D1
2) For the Drug D1, Enter a loop and Look for the Side effects S1-Sn
3) Then for Drug D1; Look for the Parent P1
4) For the Parent P1, Enter a loop and Look for the Side effects S11-Snn
5) Combine the Side effects S1-Sn and S11-Snn into a single set
6) Continue the Iteration until there are no more parents or siblings found
7) Output a combined Side effects list to display

This D-GPR algorithm is designed to work with D-SERI knowledge graph representation.

The process starts when a drug name is keyed in or selected from a dropdown list. Our Algorithm works in an iterative way by starting with the drug and finding its side effects and then the side effects from its superclass until no more higher level classes exist in the knowledge graph. The Algorithm is also designed to handle dynamic nature of the drug-side effects relationship. As the Iteration is performed, the side effects are consolidated into array list with the source/origin details. Once the process is completed, the side effects are displayed as an easy to view table in a web page for doctors/care givers to access. The side effects inquiry process can also be automated by invoking a java web service based approach where the output will be an xml file to be processed by a machine.

4.3 Research Configuration and Data Capture

Considering the primary end users of the model could be doctors or caregivers accessing health information, we chose to test the model against upcoming scenarios linked with antidiabetic drug *classes DPP4Inhibitors and SGLT2Inhibitors.* In each category we chose 3 drugs randomly from the Med Watch's publically listed drug side effects data feed released quarterly. Then we also included side effects knowledge from other sources including other MedWatch news feeds to augment the sample data set.

4.4 Research Equipment

This model was developed keeping in mind the scope to function in big data clouds as the primary source of location storing and retrieving this enormous amount of data. Since MedWatch and other sources remove any privacy or protected data and release only anonymous data, this study chose to ignore any privacy related configurations.

The server used for this configuration is HP server with Ubuntu 15.04, Xeon E5 2.4 GHz 6-Core CPU, and 16 GB memory.

4.5 Concept Demonstrator Web

The concept demonstration aims to focus on finding the full spectrum side effects based on the data model created by ***D-SERI*** knowledge graph using a web based sample tool which can be used by a doctor or patient to retrieve full spectrum side effects analysis.

4.5.1 Concept demonstrator design

The concept demonstrator works as standard home web page where the Upload Handler service will help to upload Knowledge graph xml into the application. The users will have the option to choose the file they want to choose. Once the knowledge graph is loaded, the user will have the ability to make specific request based on the drug name, parent class or simply view all possible knowledge representations.

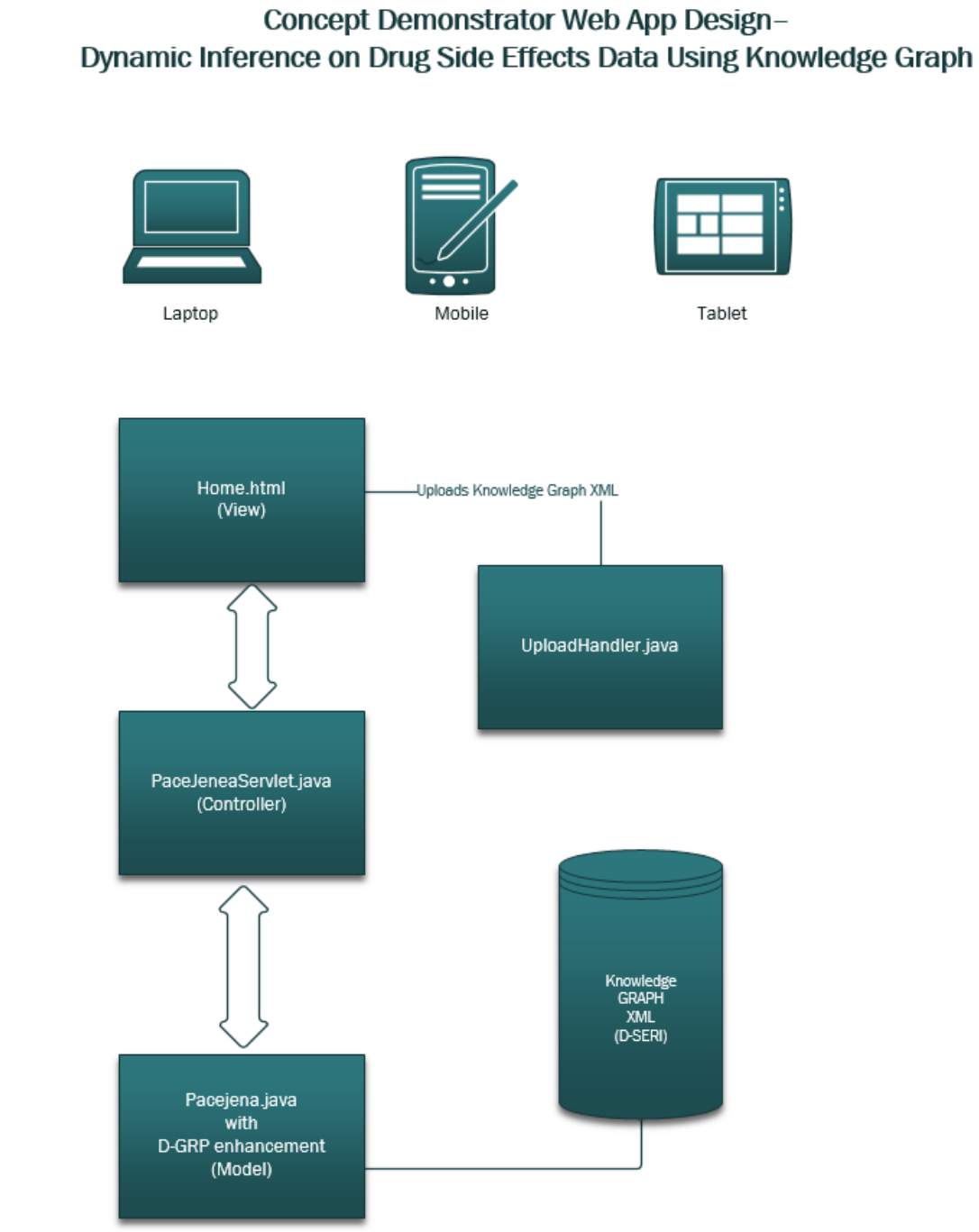


Figure 21 Concept demonstrator web app

In order to accomplish the development, the research uses the following Technologies and Tool

Drug adverse data download from public sources – FAERS and MedWatch

Knowledge graph development using public data - Pace Protégé

Parsing Knowledge graph with PaceJena Enhancement method - PaceJena

Concept demonstration web application development - Eclipse Java EE IDE

Full Spectrum Side Effect Inferencing – Trial Run in browser

4.5.2 Data download from public sources

The research will be using 100% publically available data downloaded from sources like MedWatch which is the most reliable set of real time information available to researchers. This data is fully stripped of any patient specific details by MedWatch before it's made available to researchers and agencies. It's important to point out this research used a subset of the millions of records available considering the limited resource and time available for the research.

4.5.3 Drug adverse data knowledge graph development

For the concept demonstrator, the research will use the predeveloped knowledge graphs from chapter 3 with focus on DPP4Inhibitor class as shown below.

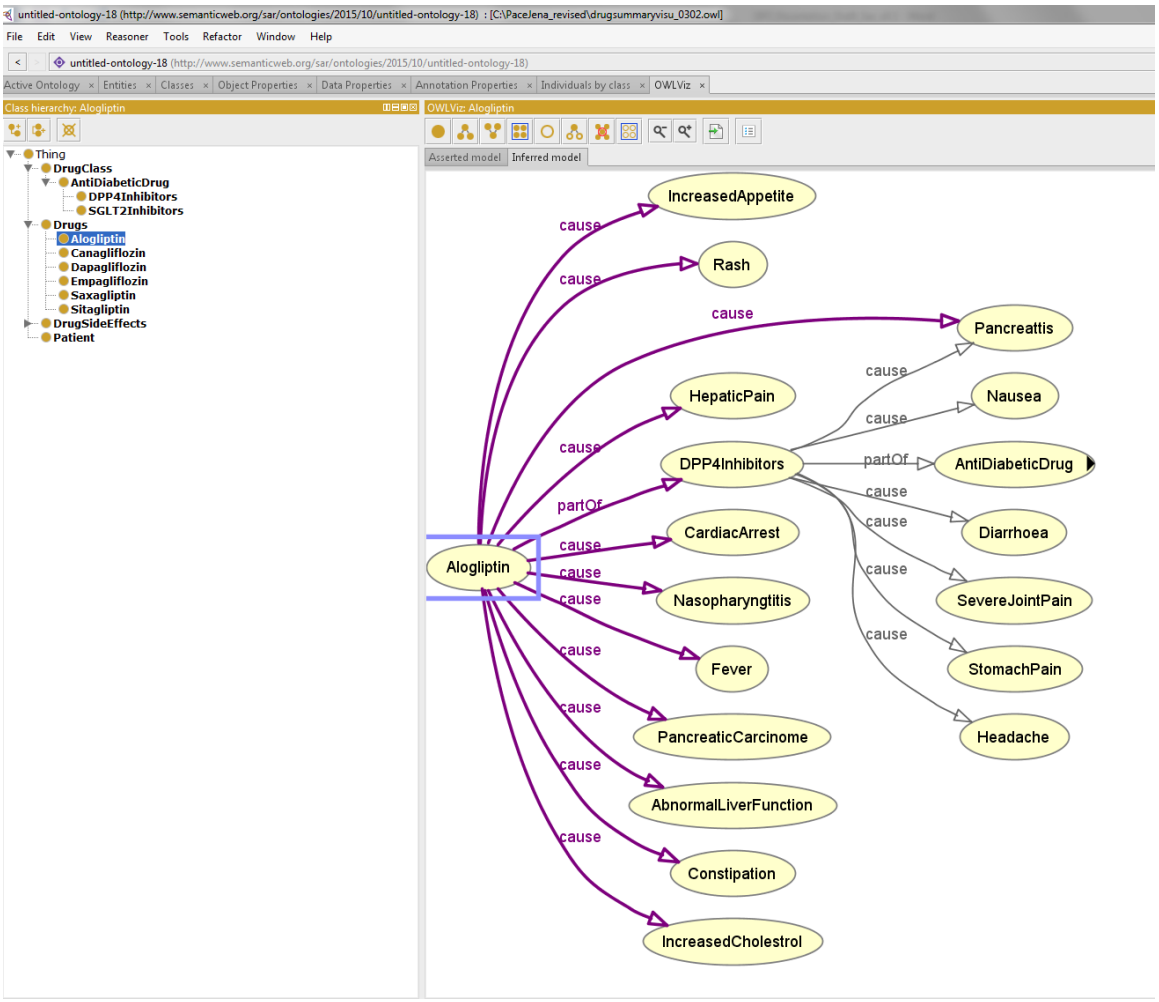


Figure 22 Drug ontology – UseCase Validation

4.5.4 Parsing ontologies with an enhancement method - PaceJena Code

Parsing the knowledge graph for full set of side effects requires the development of the D-GPR (Drug-GetParentRelations) capabilities to fetch the ontologies in an iterative way. For that purpose, the research has developed a new method getParentRelations into PaceJena as described in Chapter 4.2 which is specifically designed to bring out all side

effect combinations for the drug from the knowledge graph. The Algorithm for this enhanced functionality is well explained under section 4.2.

Figure 23 getParentRelations – Extension to PaceJena

4.5.5 Protégé Web Application - Eclipse Java EE IDE

In order for the knowledge graph to be used by Doctors during prescription check visits, The knowledge graph developed using this research should be easily queried using a simple web interface. To prove the workability of this concept, The reseach has developed a simple web front end using Eclipse Java EE IDE to parse the information and visually display to the Doctors and Patients.

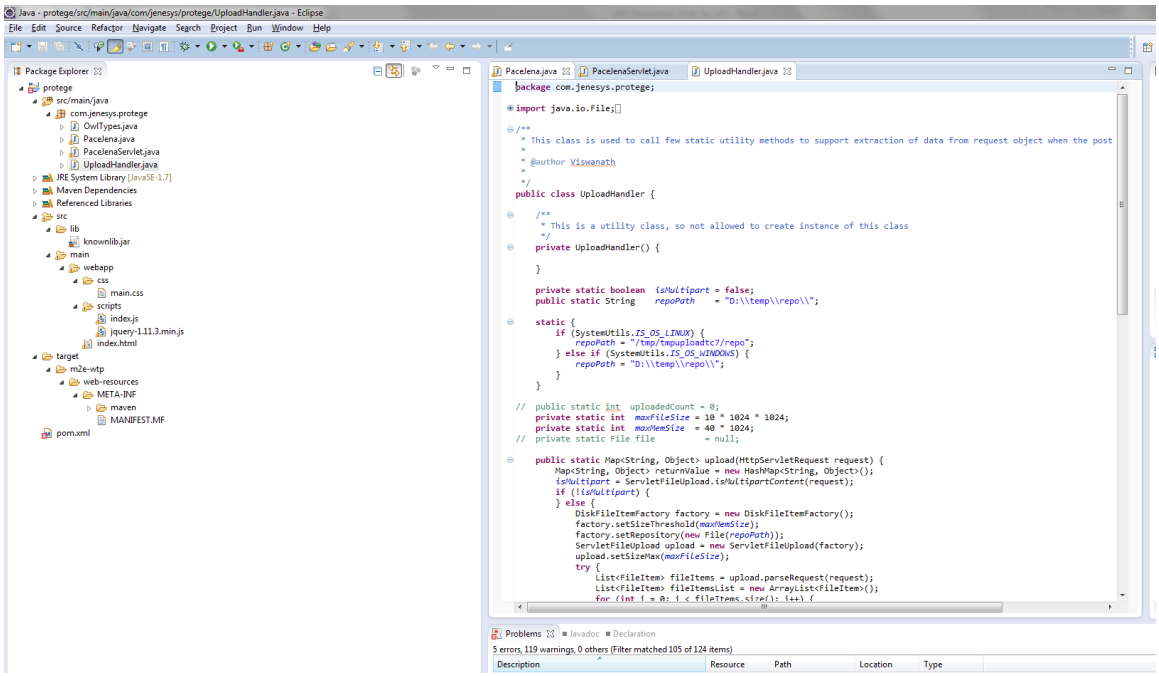


Figure 24 Eclipse IDE Project

The concept and design of this web frontend are explained using Figure 22. This web front serves the primary purpose of visually displaying the side effects to the doctors and patients based on their query. The purpose of this tool is to simply demonstrate the feasibility of the proposed graph graph appraoch.

4.5.6 Trial Run

In this section we simply checked the functioning of the web interface by accessing the home page of http://localhost:8080/protege . The actual Test run and side effect analysis will be demonstrated in the next chapter.

4.6 Key Benefits of the Concept Demonstrator

In the problem statement, the research has previously identified the need for patients and doctors to access full spectrum side effects by easily interpreting the vast drug domain knowledge with the focus on patients. This concept demonstrator simply addresses this problem and further validates the D-SERI knowledge graph model and its suitability to be the primary source of knowledge for Dynamic Side effects inferencing engines like D-GPR.

4.7 Summary

This chapter described the proposed concept demonstrator design, background, D-GPR algorithm details, and Technical details of the Implementation, successful Trail run and key benefits of the Solution in detail. In the chapter 5, the research will further show that using the D-SERI model, new adverse reactions about component drugs could be derived from compound drug's data giving the tool needed by domain experts to represent the drug side effects knowledge in a way which was not possible before. These newly derived knowledge could further fuel advanced applications like automating the prescription check process through checking constraints using Pace Jena by patients and doctors.

Chapter 5

Experimental Validation

5.1 Full Spectrum Drug Side Effect Inferencing – Trial Run 1 (Saxagliptin)

In this section, we will demonsrrate the usability of the drug side effects knowledge graph from the perspective of the doctors and patients in their day to day life to better make sense of the side effects they might be noticing. Here we provide a web interface where the preset knowlegde graph is already loaded for them or they get to choose it. In real world we expect the knowledge graph to be already loaded for the tool to start parsing the information right away.

Step 1 : Doctors access the webfront using URL http://localhost:8080/protege

Result : Home page is displayed with the option to either see ALL drug side effects from the list or use the Show dropdown to filter specific drug.

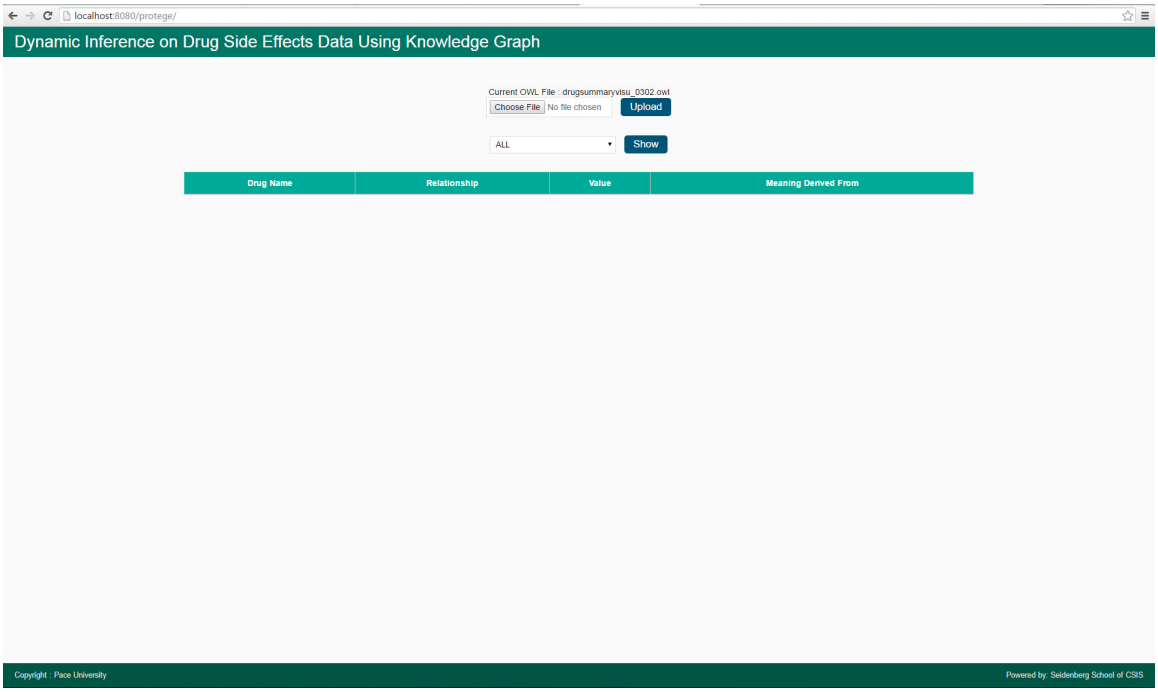


Figure 25 Test Run – Home Page

Step 2 : Doctors use the “Show” button to retreive full side effects for a drug of their intereast. Since knowledge graph is hierarchial and modular, our design allows the flexibility to grow the knowledge over aperiod of time. For this test run, we will display side effects for Drug “Saxaglipin”

Result : This page displays the full spectrum side effects for the chosen drug with details of Drug name, Relationship, Value and MeaningDerivedFrom column as shown below.

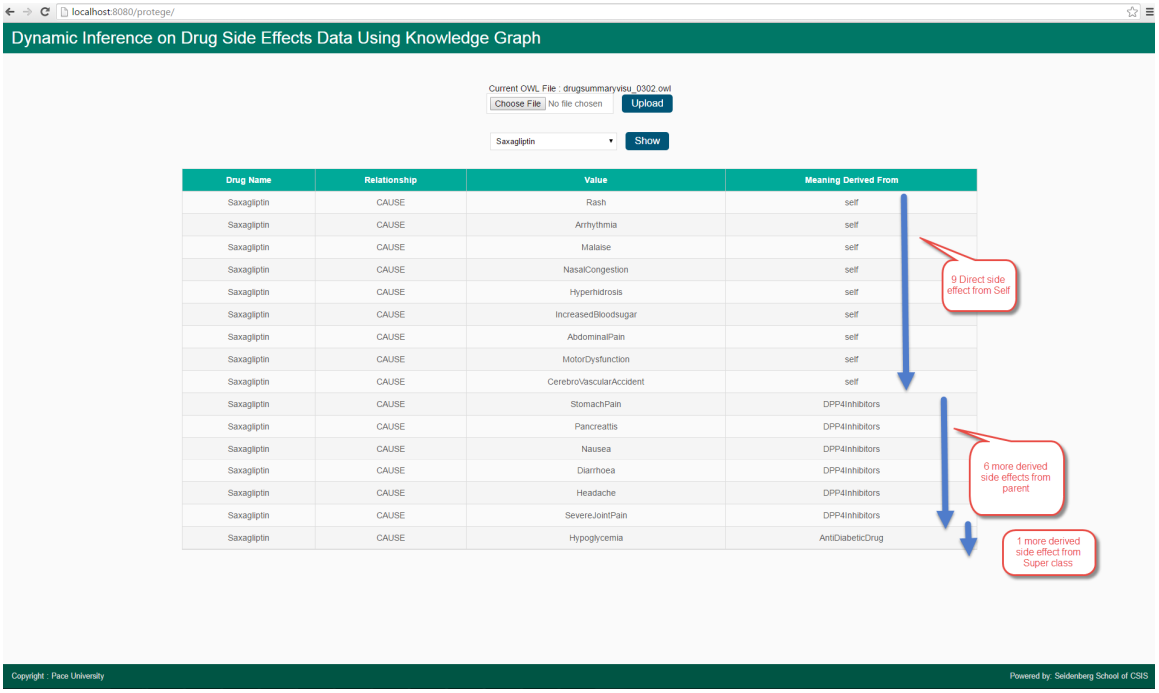


Figure 26 Test Run – Outcome Page

This snapshot view provides the summarised view of all side effects from multiple sources which is made possible by the D-SERI knowlegde model with PaceJena enhancement.

5.2 Full Spectrum Drug Side Effect Inferencing – Trial Run 2 (Alogliptin)

In this section, we will demonsrrate the usability of the drug side effects knowledge graph from the perspective of the doctors and patients in their day to day life to better make sense of the side effects they might be noticing. Here we provide a web interface where the preset knowlegde graph is already loaded for them or they get to choose it. In real world we expect the knowledge graph to be already loaded for the tool to start parsing the information right away.

Step 1 : Doctors access the webfront using URL http://localhost:8080/protege

Result : Home page is displayed with the option to either see ALL drug side effects from the list or use the Show dropdown to filter specific drug.

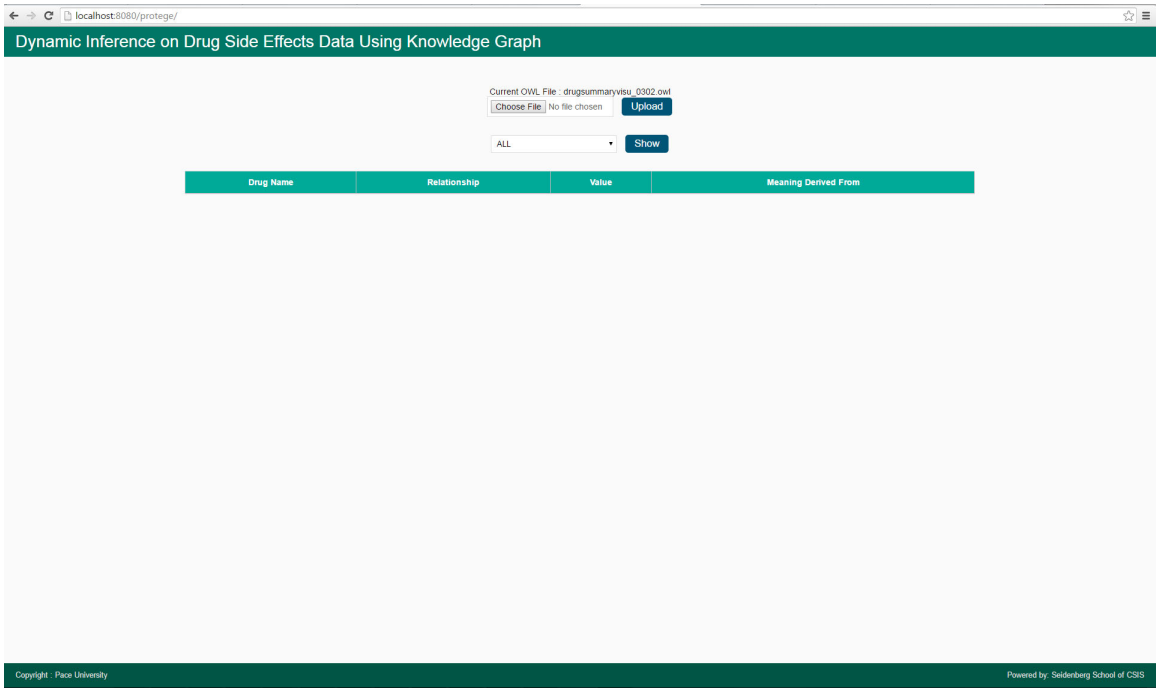


Figure 27 Test Run – Home Page

Step 2 : Doctors use the “Show” button to retreive full side effects for a drug of their intereast. Since knowledge graph is hierarchial and modular, our design allows the flexibility to grow the knowledge over aperiod of time. For this test run, we will display side effects for Drug “Saxaglipin”

Result : This page displays the full spectrum side effects for the chosen drug with details of Drug name, Relationship, Value and MeaningDerivedFrom column as shown below.

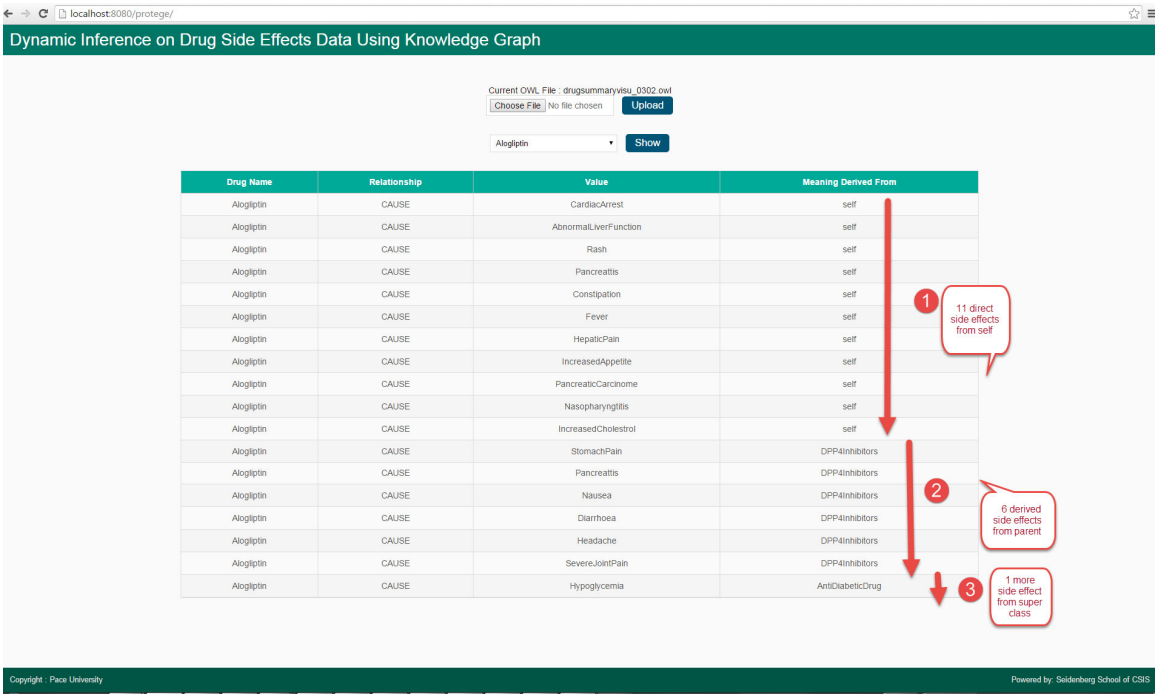


Figure 28 Test Run – Outcome Page

This snapshot view provides the summarised view of all side effects from multiple sources which is made possible by the D-SERI knowlegde model with PaceJena enhancement.

5.3 Experimental Results Analysis

We represent our experimental results in this section. The results have to be looked at from two diverse perspectives of ability to represent the side effects knowledge in a syntax simple way as well as ability to derive the meanings to the doctors and caregivers to avoid human interpretation errors

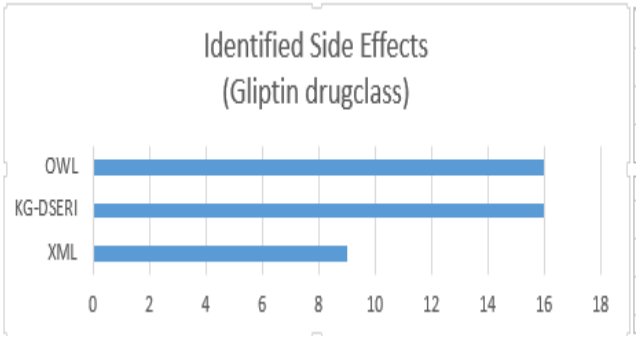


Figure 29 Drug side effects identified by D-SERI (Gliptin drug class)

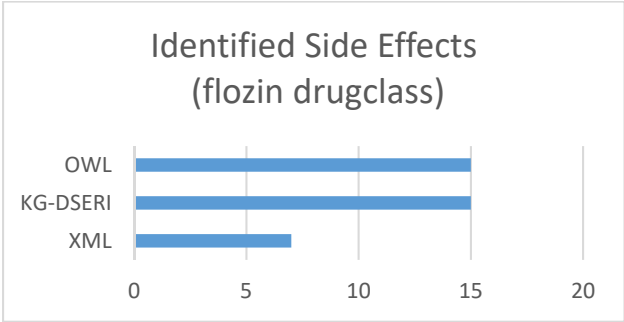


Figure 30 Drug side effects identified by D-SERI (Flozin drug class)

The proposed D-SERI model of representing drug side effects using knowledge graph found out far more number of side effects compared with the original XML approach. It also matched the side effects found using the OWL only way of representation using object property workarounds. Where the D-SERI model beats the OWL only way of representation is by simplifying the knowledge representation for storing the drug –side effects and drug-drug class using custom OWL relationships.

The dissertation validated the model outcomes Figure 22. We used a Java servlet based web application which uses the algorithms depicted in D-GPR to derive full spectrum side effects to doctors and caregivers avoiding the human interpretation errors. This proves that the data integrity and usability for doctors and caregivers avoiding costly interpretation errors using a dynamic and flexible model.

5.4 Validate D-SERI model to larger set of drug domain data

So far the research has demonstrated the solution methodology in two trial runs using different set of drugs and side effects. Having proved that, the next step is to prove the

usability of the research into the Intelligent Telehealth domain using a larger data set of drugs, side effects and drug classes. This is significant as the drug side data domain in ever expanding and so far 6 million knowledge relations are captured overall.

To validate the usability of the knowledge graph for larger data sets, the research built the knowledge graph using multiple drugs, classes and side effects and strives to prove the usability from the concept demonstrator app.

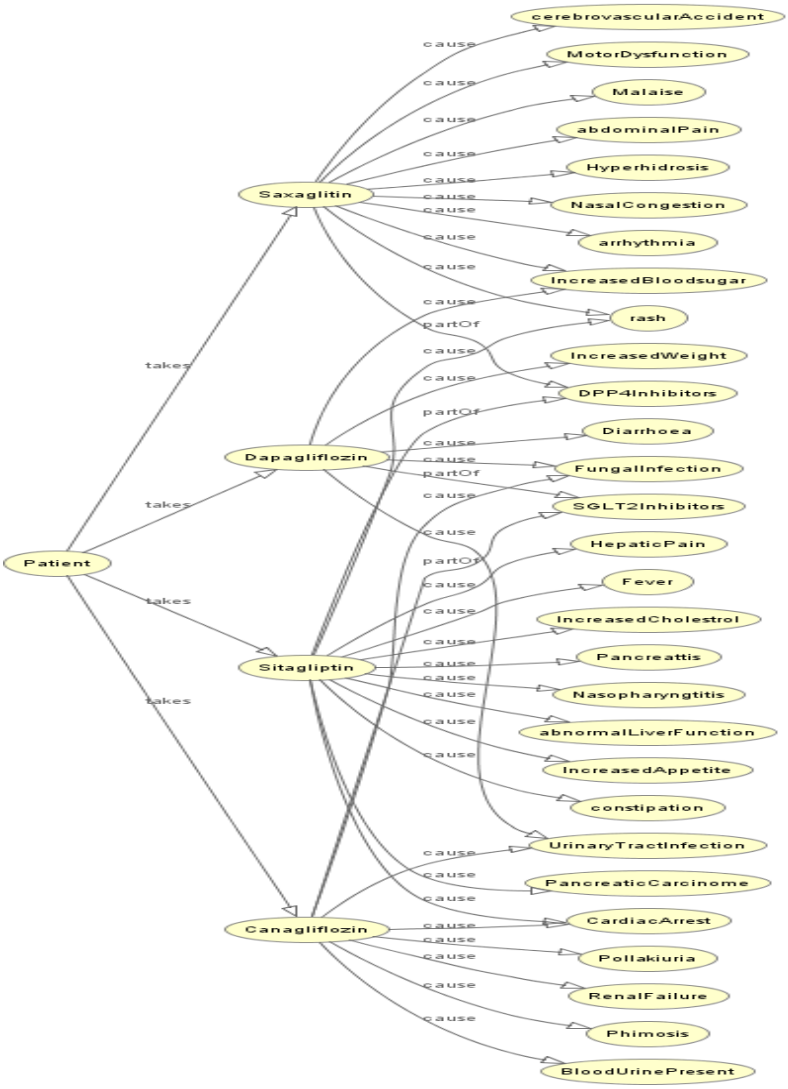


Figure 31 Knowledge graph with larger data set 1

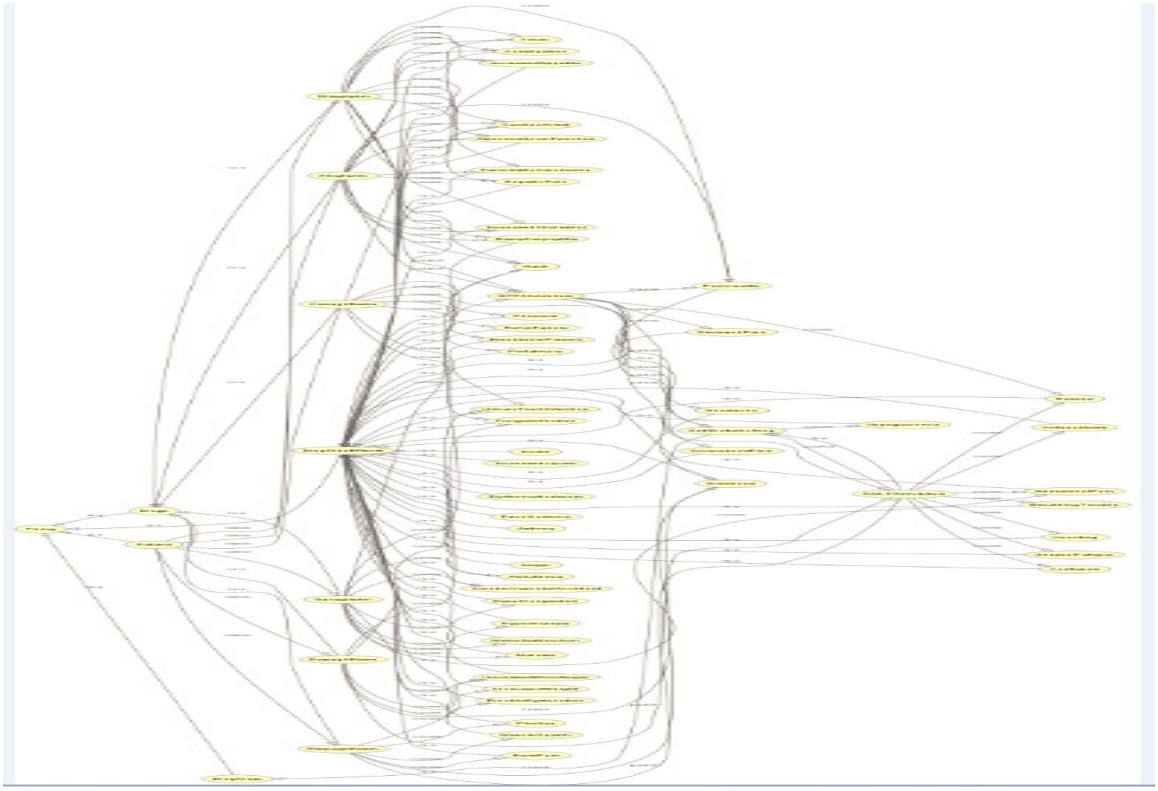
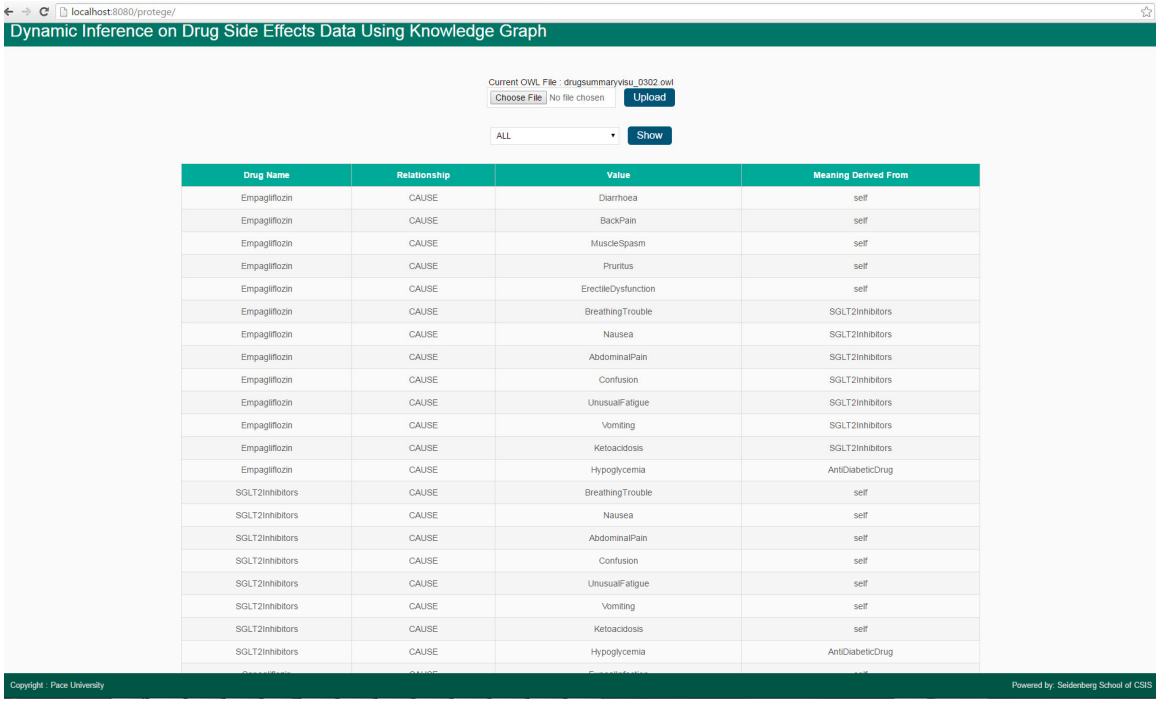


Figure 32 Knowledge graph with larger data set 2

The following page shows the validation of the outcome with the large data sets thus proving the suitability of the knowledge graph for prescription checks.


localhost:8080/protege/
Dynamic Inference on Drug Side Effects Data Using Knowledge Graph
Current OWL File : drugsummaryvisu_0302.owl
Choose File No file chosen
Upload
ALL
Show
Drug Name | Relationship | Value | Meaning Derived From
Empagliflozin | CAUSE | Diarrhoea | self
Empagliflozin | CAUSE | BackPain | self
Empagliflozin | CAUSE | MuscleSpasm | self
Empagliflozin | CAUSE | Pruritus | self
Empagliflozin | CAUSE | ErectileDysfunction | self
Empagliflozin | CAUSE | BreathingTrouble | SGLT2Inhibitors
Empagliflozin | CAUSE | Nausea | SGLT2Inhibitors
Empagliflozin | CAUSE | AbdominalPain | SGLT2Inhibitors
Empagliflozin | CAUSE | Confusion | SGLT2Inhibitors
Empagliflozin | CAUSE | UnusualFatigue | SGLT2Inhibitors
Empagliflozin | CAUSE | Vomiting | SGLT2Inhibitors
Empagliflozin | CAUSE | Ketoacidosis | SGLT2Inhibitors
Empagliflozin | CAUSE | Hypoglycemia | AntiDiabeticDrug
SGLT2Inhibitors | CAUSE | BreathingTrouble | self
SGLT2Inhibitors | CAUSE | Nausea | self
SGLT2Inhibitors | CAUSE | AbdominalPain | self
SGLT2Inhibitors | CAUSE | Confusion | self
SGLT2Inhibitors | CAUSE | UnusualFatigue | self
SGLT2Inhibitors | CAUSE | Vomiting | self
SGLT2Inhibitors | CAUSE | Ketoacidosis | self
SGLT2Inhibitors | CAUSE | Hypoglycemia | AntiDiabeticDrug
Copyright : Pace University
Powered by: Seidenberg School of CSIS


localhost:8080/protege/
Canagliflozin | CAUSE | FungalInfection | self
Canagliflozin | CAUSE | UrinaryTractInfection | self
Canagliflozin | CAUSE | Phimosis | self
Canagliflozin | CAUSE | BloodUrinePresent | self
Canagliflozin | CAUSE | CardiacArrest | self
Canagliflozin | CAUSE | RenalFailure | self
Canagliflozin | CAUSE | Pollakiuria | self
Canagliflozin | CAUSE | BreathingTrouble | SGLT2Inhibitors
Canagliflozin | CAUSE | Nausea | SGLT2Inhibitors
Canagliflozin | CAUSE | AbdominalPain | SGLT2Inhibitors
Canagliflozin | CAUSE | Confusion | SGLT2Inhibitors
Canagliflozin | CAUSE | UnusualFatigue | SGLT2Inhibitors
Canagliflozin | CAUSE | Vomiting | SGLT2Inhibitors
Canagliflozin | CAUSE | Ketoacidosis | SGLT2Inhibitors
Canagliflozin | CAUSE | Hypoglycemia | AntiDiabeticDrug
Dapagliflozin | CAUSE | Diarrhoea | self
Dapagliflozin | CAUSE | IncreasedBloodsugar | self
Dapagliflozin | CAUSE | UrinaryTractInfection | self
Dapagliflozin | CAUSE | FungalInfection | self
Dapagliflozin | CAUSE | IncreasedWeight | self
Dapagliflozin | CAUSE | BreathingTrouble | SGLT2Inhibitors
Dapagliflozin | CAUSE | Nausea | SGLT2Inhibitors
Dapagliflozin | CAUSE | AbdominalPain | SGLT2Inhibitors
Dapagliflozin | CAUSE | Confusion | SGLT2Inhibitors
Dapagliflozin | CAUSE | UnusualFatigue | SGLT2Inhibitors
Dapagliflozin | CAUSE | Vomiting | SGLT2Inhibitors
Dapagliflozin | CAUSE | Ketoacidosis | SGLT2Inhibitors
Dapagliflozin | CAUSE | Hypoglycemia | AntiDiabeticDrug
DPP4Inhibitors | CAUSE | StomachPain | self
Copyright : Pace University
Powered by: Seidenberg School of CSIS

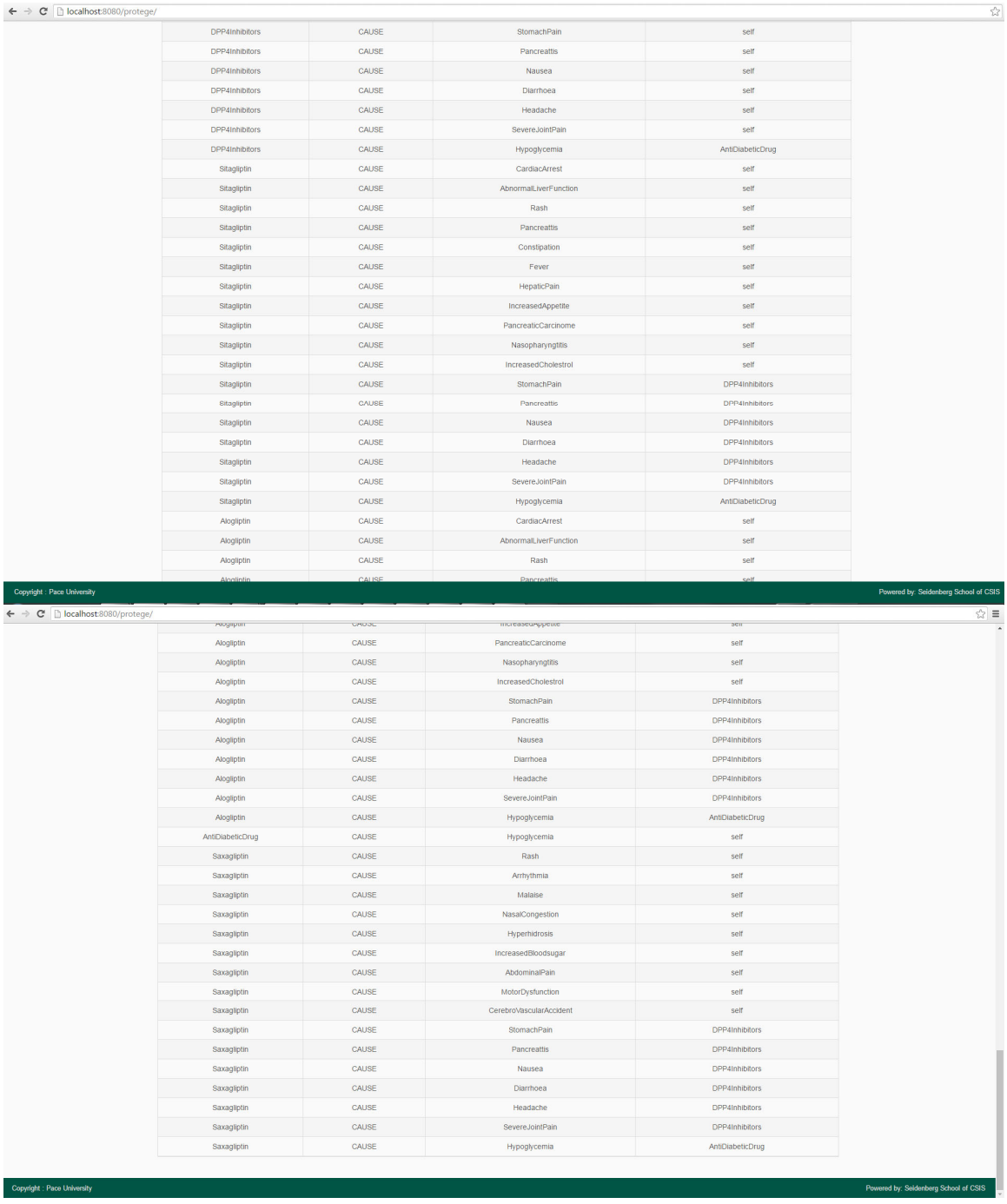


Figure 33 Test Run – Show ALL Page

5.5 Summary Findings and Concept Validation

The research validated the proposed D-SERI model outcomes by using a Java servlet based web application which uses the algorithms depicted in D-GPR to derive full spectrum side effects to doctors and caregivers avoiding the human interpretation errors. This proves that the data integrity and usability for doctors and caregivers avoiding costly interpretation errors using a dynamic and flexible model.

5.6 Conclusion

In this chapter, we set up a platform for experimental demonstration to validate this research methodology. Chapter 6 will provide a summary of the major contributions, followed by suggested future work.

Chapter 6

Conclusion and Future work

6.1 Conclusion of the dissertation

In Summary, This research focused on the issue of finding full spectrum drug side effects by introducing knowledge graphs to describe drug side effects domain concepts in big data clouds. The proposed mechanism was D-SERI that was created to allow the dynamic knowledge representation model necessary to display drug side effects relations. The main algorithm was D-GPR which extended Pace-Jena to retrieve full spectrum side effects to patients and doctors reducing the errors in understanding the side effects. Our experimental evaluations had proved the efficiency of the proposed model.

6.2 Future Work

Future work could address the limitations of the research in this dissertation, such as expanding the scope to cover sub domains within drug data like protein to drug relationship or drug to treatment relationship and rationalizing the data.

Appendix A

Abbreviations

AEO:

Adverse event ontology

AERO:

Adverse event reporting ontology

BFO:

Basic formal ontology

BSPO:

Spatial ontology

CDISC:

Clinical data interchange consortium

CODAE:

Ontology-based detection of adverse events

CTCAE:

Common terminology criteria for adverse events

DOID:

Disease ontology

FAERS:

FDA adverse events reporting system

FDA:

Food and Drug Administration (FDA)

FMA:

Foundational Model of Anatomy

IAO:

Information artifact ontology

IDO:

Infectious disease ontology

MedDRA:

Medical dictionary for regulatory activities

OAE:

Ontology of adverse events

OBI:

Ontology for biomedical investigations

OBO:

Open biomedical/biological ontologies

OGMS:

Ontology for general medical science

PATO:

Phenotypic quality ontology

PRR:

Proportional reporting ratio

RO:

Relation ontology

TIV:

Trivalent inactivated influenza vaccine

LAIV:

Live attenuated influenza vaccine

UBERON:

Uber anatomy ontology

VAERS:

Vaccine adverse event

VO:

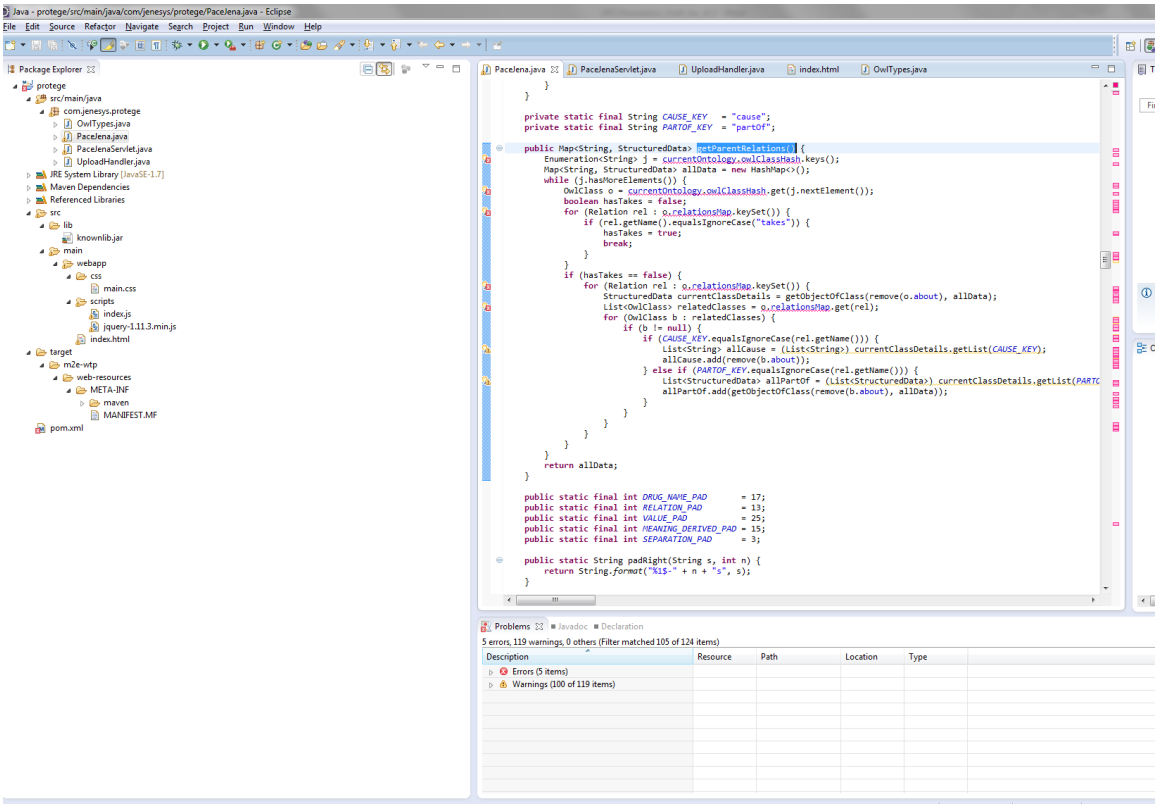
Vaccine ontology

WHO-ART:

WHO’s adverse reaction terminology database

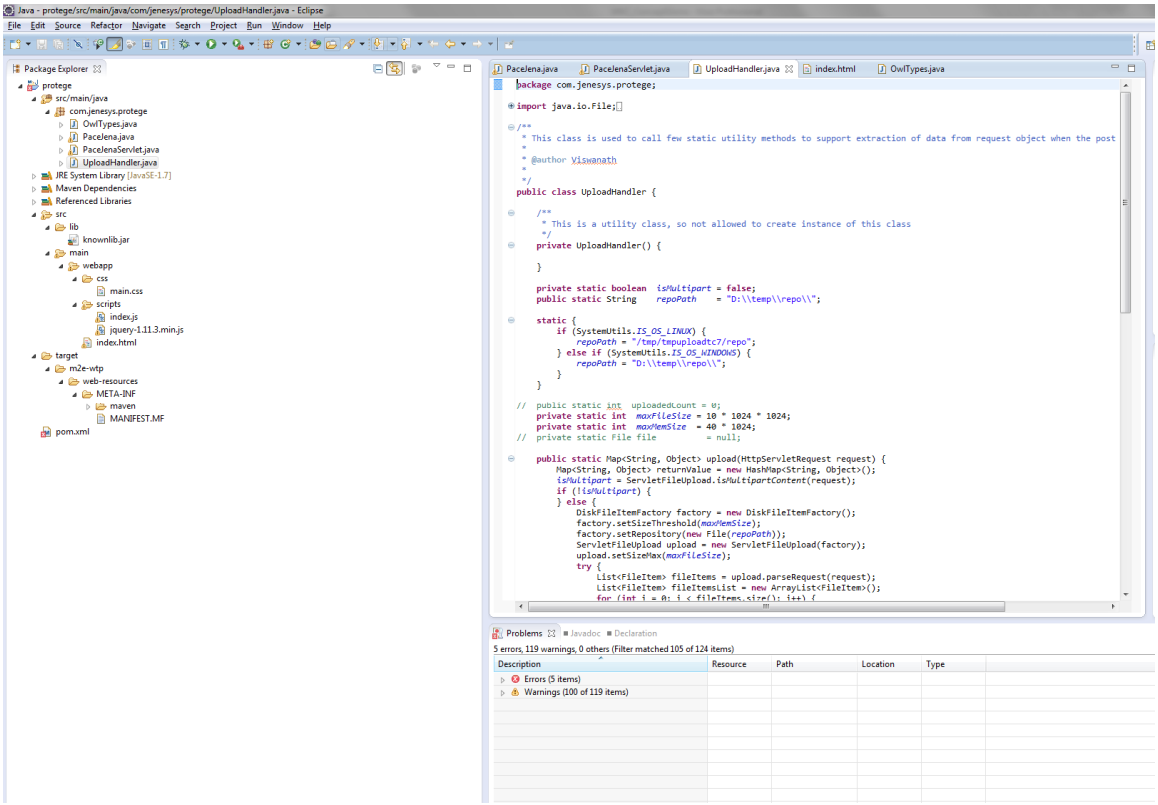
Appendix B

PaceJena.java getParentRelations()



Appendix C

UploadHandler.java code



References

[1]Y. Li, K. Gai, Z. Ming, H. Zhao, and M. Qiu. Intercrossed access control for secure financial services on multimedia big data in cloud systems. ACM Transactions on Multimedia Computing Communications and Applications, PP(99):1, 2016.

[2]M. Qiu, K. Gai, B. Thuraisingham, L. Tao, and H. Zhao. Proactive user-centric secure data scheme using attribute-based semantic access controls for mobile clouds in financial industry. Future Generation Computer Systems, PP:1, 2016.

[3]M. Qiu, M. Zhong, J. Li, K. Gai, and Z. Zong. Phase-change memory optimization for green cloud with genetic algorithm. IEEE Transactions on Computers, 64(12):3528 – 3540, 2015.

[4]K. Gai and S. Li. Towards cloud computing: a literature review on cloud computing and its development trends. In IEEE Fourth Int'l Conf. on Multimedia Information Networking and Security, pages 142–146, Nanjing, China, 2012. IEEE.

[5]K. Gai, M. Qiu, H. Zhao, L. Tao, and Z. Zong. Dynamic energy-aware cloudlet-based mobile cloud computing model for green computing. Journal of Network and Computer Applications, 59:46–54, 2016.

[6]K. Gai, M. Qiu, L. Chen, and M. Liu. Electronic health record error prevention approach using ontology in big data. In 17th IEEE International Conference on High Performance Computing and Communications, pages 752–757, New York, USA, 2015.

[7]K. Gai, M. Qiu, S. Jayaraman, and L. Tao. Ontology-based knowledge representation for secure self-diagnosis in patient-centered telehealth with cloud systems. In The 2nd IEEE Int'l Conf. on Cyber Security and Cloud Computing, pages 98–103, New York, USA, 2015. IEEE.

[8]L. Tao, S. Golikov, K. Gai, and M. Qiu. A reusable software component for integrated syntax and semantic validation for services computing. In 9th Int'l IEEE Symposium on Service-Oriented System Engineering, pages 127–132, San Francisco Bay, USA, 2015.

[9]Prescription Drug Usage among Americans

http://newsnetwork.mayoclinic.org/discussion/nearly-7-in-10-americans-take-prescription-drugs-mayo-clinic-olmsted-medical-center-find

[Accessed on Aug 24, 2016]

[10]Preventing medical errors: quality chasm series. The National Academies Press, Washington DC, 2006.

http://www.nationalacademies.org/hmd/~/media/Files/Report%20Files/2006/Preventing-Medication-Errors-Quality-Chasm-Series/medicationerrorsnew.pdf

[Accessed on Apr 20, 2016]

[11]Mannette-Wright, Anne I, A Feasibility Study of Ontology-Based Automatic Document Transformation, 2009, Seidenberg School of Computer Science, Pleasantville, NY

[12]T Berners-Lee, J Hendler and O Lassila Semantic Web, Scientific American.com online [Accessed on July 18,2015]

[13]K. Gai, M. Qiu, B. Thuraisingham, and L. Tao. Proactive attribute based secure data schema for mobile cloud in financial industry.In The IEEE International Symposium on Big Data Security on Cloud; 17th IEEE International Conference on High Performance Computing and Communications, pages 1332–1337, New York, USA, 2015.

[14]H. Zhao, K. Gai, J. Li, and X. He. Novel differential schema for high performance big data telehealth systems using pre-cache. In The IEEE International Symposium on Big Data Security on Cloud; 17th IEEE International Conference on High Performance Computing and Communications, pages 1412–1417, New York, USA, 2015. IEEE.

[15]Berges, J. Berm´udez, and A. Illarramendi. Toward semantic interoperability of electronic health records. IEEE Trans. on Info.Tech. in Biomedicine, 16(3):424–431, 2012.

[16]H. Fujita, I. Rudas, J. Fodor, M. Kurematsu, and J. Hakura. Mental ontology model for medical diagnosis based on type of intuitionistic fuzzy functions. Mental, 57(71), 2012.

[17]J. Kim, J. Kim, D. Lee, and K. Chung. Ontology driven interactive healthcare with wearable sensors. Multimedia Tools and Applications,71(2):827–841, 2014.

[18]K. Patel, I. Dube, L. Tao and N. Jiang. Extending OWL to Support Custom Relations In the 2nd IEEE Int'l Conf. on Cyber Security and Cloud Computing, pages 98–103, New York, USA, 2015. IEEE.

[19]L Yu, Introduction to the Semantic Web and Semantic Web Services, Chapman and Hall/CRC 2007, ISBN: 978-1-58488-933-5

[20]FDA Adverse Event Reporting System (FAERS)

http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/default.htm

[Accessed on July 18, 2015]

[21]FDA Drug Approval Process

http://www.fda.gov/Drugs/ResourcesForYou/SpecialFeatures/ucm279676.htm

[Accessed on July 18, 2015]

[22]G. Henriques, L. Lamanna, D. Kotowski, H. Hlomani, D. Stacey,P. Baker, and S. Harper. An ontology-driven approach to mobile data collection applications for the healthcare industry. Network Modeling Analysis in Health Informatics and Bioinformatics, 2(4):213–223, 2013.

[23]P. Shvaiko and J. Euzenat. Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering, 25(1):158–176, 2013.

[24]What Attorneys Should Know About FDA's MedWatch Data, 2014

http://www.analysisgroup.com/uploadedfiles/content/insights/publishing/what_attorneys_should_know_about_fdas_medwatch_data.pdf

[Accessed on July 18, 2015]

[25] Axway Interchange - A solution for exchanging documents between trading partners.

https://esg.fda.gov/help/wwhelp/wwhimpl/common/html/wwhelp.htm#context=help&file=ebxml.57.34.html

[Accessed on July 18, 2015]

[26]Drug Classes definition Tree View (relational)

https://www.drugs.com/drug-classes.html?tree=1

[Accessed Jan 28, 2016]

[27]Awareness about prescribed drugs among patients attending Out-patient departments

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3678681/

[Accessed on Feb 22, 2016]

[28]R. Goncalves, M. Horridge, M. Musen, et al, "Protégé," Stanford.edu, May 2009.

http://Protege.stanford.edu

[Accessed: Nov 25, 2015]

[29]H. Lambers, D. Zeeuw, L. Wie, B. Leslie, and J. List. Dapagliflozin a glucose-regulating drug with diuretic properties in subjects with type 2 diabetes. Diabetes, Obesity and Metabolism, 15(9):853–862, 2013.

[30]Pattanawongsa, N. Chau, A. Rowland, and J. Miners. Inhibition of human udp-glucuronosyltransferase enzymes by canagliflozin and dapagliflozin: Implications for drug-drug interactions. Drug Metabolism and Disposition, 43(10):1468–1476, 2015.

[31]Doulaverakis, G. Nikolaidis, A. Kleontas, and I. Kompatsiaris.GalenOWL: Ontology based drug recommendations discovery. J.Biomedical Semantics, 3:14, 2012.

[32]J Josh, E Joseph, M Brochhausen and W Hogan. "DrOn - Building drug ontology based on RxNorm and other sources, Journal of BioMedical Semantics" 2013

[33]C Doulaver, G Nikolaidais , A Kleontas and I Kompatsiaris, "GalenOWL : Ontology-based drug recommendation discovery" , Journal of BioMedical Semantics – 2012

http://www.jbiomedsem.com/content/3/1/14

[Accessed on July 18, 2015]

[34]Z Pan , A Qasem and J Heflin, An Investigation into Feasibility of the Semantic Web,LehighUniversity2012

[35]T Sakaeda, A Tamon, K Kaori and Y Okuno, Data Mining of the Public version of the FDA Adverse Event Reporting System, International Journal of Medical Sciences,2013

[36]A Ali and L Tao , Syntax Agnostic Semantic Validation of XML Documents using Schematron, Seidenberg School of Computer Science Research Day, 2014

[37]Truth about Triple Stores, Onto Text

http://ontotext.com/documents/white_papers/The-Truth-About-Triplestores.pdf

[Accessed on July 18, 2015]

[38]K Alexander, The difference between a Triple Store and a relational database –

http://krisalexander.com/innovation/2013/07/16/the-difference-between-a-triplestore-and-a-relational-database/

[Accessed on July 18, 2015]

[39]Assertion, Condition and providence, "A beginner's guide to Nano publication"

http://nanopub.org/original/index.html

[Accessed on July 18, 2015]

[40]The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3102889/

[Accessed on July 18, 2015]

[41]Semantic Web in Systems Chemical Biology

http://chem2bio2rdf.wikispaces.com/

[Accessed on Sep 25, 2015]

[42]Mappings from Clinical Trial Eligibility Drug Information to Prescription in Patient Data Using Drug Ontology

http://www.w3.org/wiki/HCLS/ClinicalObservationsInteroperability/DrugMapping.html

[Accessed on Aug 26, 2016]

[43]FDA Guide - DRUG INTERACTIONS: WHAT YOU SHOULD KNOW

http://www.fda.gov/downloads/Drugs/ResourcesForYou/UCM163355.pdf

[Accessed on Aug 28, 2016]

[44]Type 2 Oral Diabetes Medications (Oral) overview

http://www.medicinenet.com/antidiabetics-oral/article.htm

[Accessed on June 25, 2016]

[45]Type 2 Diabetes Diagnosis, Treatment, and Medication

http://www.medicinenet.com/type_2_diabetes_pictures_slideshow/article.htm

[Accessed on June 25, 2016]

[46]H Yongqun, S Sarntivijai, Y Lin, Z Xiang, A Guo, S Zhang, D Jagannathan, L Toldo, C Tao. OAE the Ontology of Adverse Events. J. Biomedical Semantics, 5:29, 2014.

https://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-5-29

[Accessed on June 25, 2016]

[47]J S Luciano et al. The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside J Biomed Semantics. 2011; 2

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3102889/

[Accessed on June 25, 2016]

[48]Safety Works - Construction and Annotation of a UMLS/SNOMED-based Drug Ontology for Observational Pharmacovigilance.

https://www.researchgate.net/publication/242259534_Construction_and_Annotation_of_a_UMLSSNOMED-based_Drug_Ontology_for_Observational_Pharmacovigilance

[Accessed July 26, 2016]

[49]SNOWMED - Systematically organized computer process able collection of medical terms

http://www.opencimi.org/tag/SNOWMED%20CT

[Accessed Jan 08, 2016]

[50]Google “Knowledge Vault” To Power Future of Search

http://searchengineland.com/google-builds-next-gen-knowledge-graph-future-201640

[Accessed July 28, 2016]