ISSN: 0976-3104
SPECIAL ISSUE (SCMSA)
Agrawal and Swarnalatha

THE IIOAB JOURNAL

ARTICLE OPEN ACCESS

COMBINING SEMANTIC DATA STORE AND BIG DATA FOR PHARMA USE CASE

Pragya Agrawal* and Swarnalatha P.

School of Computing Sciences and Engineering, VIT University, Vellore, Tamil Nadu, INDIA

ABSTRACT

In the pharmaceutical R&D procedure the increased generation of data has failed to generate the estimated returns in terms of better efficiency and pipelines. The failure of existing integration methodology to systematize and apply the available knowledge to the range of real scientific and business issues which influence on not only efficiency but also transparency of information in crucial safety and regulatory applications. The new range of semantic technologies based on ontologies enables the proper integration of knowledge in a way that is reusable by several applications across businesses, from discovery to corporate affairs. This paper supports the use of Semantic Web technologies across health care, life sciences, clinical research and translational medicine which help to increase the accuracy of information mining, retrieve complex entities, combine structured and unstructured analytical queries and create comprehensive queries.

Received on: 30th-Nov-2015
Revised on: 11th-March-2016
Accepted on: 26– March-2016
Published on: 10th –Aug-2016

KEY WORDS

Semantic web technologies;Real World Evidence; clinical pharmacy database; big data; Allegrograph

***Corresponding author: Email:** ag.pragya796@gmail.com, **Tel:** +91-9944912522

INTRODUCTION

Data analysis acts as significant role in recognizing productive use cases in any field. Analytics in Pharmaceutical industry orbits around determining better drugs, management of supply chain, and other competitive reward. Drug detection has so far been based on the historical records of the company, and approximately market research with help from researchers. Though, there is more information in the real world than what is really available with the company alone, for example, EMRs, Insurance claims, Prescriptions, etc. When these real world data place to utilize for analytics would carry a stronger evidence subsequent in better and more dedicated inventions. In simple words, RWE involves assembling and scrutinizing data in what way a drug is actually utilized in the real world, as contrasting to what occurs in a structured clinical research situation with protocols and highly motivated physicians.

(RWE) Real World Evidence [1] can be stated as "insights from anonymous patient-level data using sound commercial and scientific analytics". (RCTs) Randomized controlled trials is the golden way for demonstrating security and effectiveness before launch, but stakeholders are observing for more. They entail outcomes and information about the holistic patient journey. Growing healthcare data, produced through hospital reports and payer claims EMRs could offer that significant addition to RCTs. Applicable sets of that information, combined with clinical, commercial and scientific expertise could allow organizations to support and prove importance during the product lifecycle.

Factors driving RWE –

- Patient Centricity - A drug must be approved to a patient also based on his medical background instead of the disease for which he is being treated presently.
- Peer Trends - Every corporation wants to outperform their peer, get improved medicine in terms of both efficacy and cost leading to an improved market. Competitors' information requirement to be associated with real world data to fulfill the visions.
- Clinical Trials - are RCTs which doesn't provide the same outcomes of analysis exterior to the tested investigational conditions, for e.g., Medications that were tested on one background may not work on the other.

THE IIOAB JOURNAL
ISSN: 0976-3104

- Observational Data - Each observation of a case (patient) when documented offers more understanding than what is obtainable from the conventional information sources.
- Regulatory Requirements - Pharma corporations compete to enter their drugs into Tier-1 formularies, subsequently drugs in this tier get instant endorsement. To attain this, corporation's intention is on cost and drug effectiveness. Insurance corporations also favor paying for medications in the Tier-1 list which has indication of execution well.

Life sciences corporations are using RWE [1] to provision advances in data technology, external and internal healthcare decision making, engagement and analytics prototypes could further its scientific and commercial influence all functions must study how to join the power of its visions. Fig 1. Represent the multiple data sources that can produce an evidence-based patient journey.

Who Is Using RWE Today?
• Epidemiology, drug safety researchers and HEOR, to increase faster visions from richer patient datasets.
• Pricing and market access colleagues, to inform payers and HTAs with evidence of their products' performance
• Brand and franchise teams, to understand their markets, differentiate their products, and improve stakeholder engagement.
• Clinical development teams, to design trials based on actual treatment practices versus dated and inconsistent guidelines

Fig: 1. Patient Journey: Multiple data sources can create an evidence-based patient journey.

RWE [2] helps us to define subsets of patients who are most benefitted from a medicine, based on their background of genetics, social aspects and variations in disease. RWE [3] may help us decide where the most demanding medical requirements are. It also helps drug companies to terminate the drug candidates that are not going to effectively follow with existing treatments. It also makes us to choose "safety signals" much faster, thereby notifying companies and the public to the hazardous side effects of some medicines.

Real World Data means variety offered in a huge scale streamed at high rates. Data from diverse sources as well as those from patient observations lead to the existence of world data in different forms. Most of these are scarce. The data volume dumped steadily from various sources such as social networks, clinical trials, etc. are so enormous to handle. Semantically identical texts may be characterized differently. So, Natural Language Processing (NLP) needs to implement which is highly complex. NLP will help in bringing diverse data sources together. Semantic web technologies and big data thus comes into the picture as a solution to these.

Semantic Web technologies [4] – is a group of very particular technology standards from the (W3C) World Wide Web Consortium that are reflected to define and relate information inside enterprises and on the web. These standards include:

- a flexible information prototypical (RDF[5-7]),
- ontology languages and schema for defining concepts and relationships (OWL[8] and RDFS),

THE IIOAB JOURNAL
ISSN: 0976-3104

- a query language (SPARQL[9-11]),
- a rules language (RIF),
- A language for marking up data inside Web pages (RDFa) and more.

The term "ontology" can be stated as “an explicit specification of conceptualization”. Ontologies establishes the building of the domain, i.e. conceptualization. This contains the model of the domain with probable limitations. The conceptualization defines knowledge about the domain and not approximately the certain state of relationships in the domain. In other words, the conceptualization is not varying, or is varying very seldom. Ontology is then specification of this conceptualization: the conceptualization is specified by using particular modelling language and particular terms. Formal specification is required in order to be able to process ontologies and operate on ontologies automatically.
Ontology [8][12] defines a domain, while a knowledge base defines certain state of relationships. Each knowledge based system or agent has its own knowledge base, and only what can be communicated using an ontology can be kept and used in the knowledge base. When an agent wants to communicate to another agent, he uses the constructs from some ontology. In order to understand in communication, ontologies must be shared between agents.

Medicine informatics is defined as the “field of information science concerned with the analysis, use and dissemination of medical data and information through the application of computers to various aspect of health care and medicine” [13].

In 2007, the (ASHP) American Society of Health System Pharmacists released a position paper that defined this subspecialty area of pharmacy practice as the use , designates the pharmacist’s role in informatics, integration of knowledge, data, technology, information, and automation in the medication-use process for the determination of refining health consequences [14]. The term “big data” has been coined and is defined “as the emerging use of rapidly collected, complex data “[15]. Big data is a term well-defined in three V’s: volume, velocity, and variety. Other dimensions may also include complexity and variability. The Centre for US Health System Reform released a paper that defines the revolution of big data in health care and cites four major sources of big data that include:

- Pharmaceutical research and development from pharmaceutical companies and academia, clinical trials, and high-throughput screening libraries.
- Clinical data provided by the electronic medical record (EMR) that contain patient-specific data on treatment outcomes.
- Claims and cost data from payers and providers that contain utilization of care and cost estimates
- Patient behaviour and sentiment data that come from consumers and stakeholders outside of health care(for instance, from retail exercise apparel and exercise monitoring equipment) [16]

In this paper the use of Semantic Web technologies and Real World Evidence across health care, life sciences, clinical research and translational medicine will help to increase the precision of information mining, retrieve complex entities, combine structured and unstructured analytical queries and create comprehensive queries.

MATERIALS AND METHODS

Prediction based data aggregation- a survey

The energy management is one of the major issues in wireless sensor networks. A sensor utilizes high energy for communication rather than sensing and processing. The redundant communication in noisy channels causes the depletion of network energy. The prediction based data aggregation approach reduced unnecessary data transmission and so energy expenditure in communication subsystem was minimized. Hyuntea Kim et al., [4] exploited linear data prediction method to improve communication efficiency and to minimize energy consumption with data correlation. As the model is designed considering some factors such as the selective transmission, it reduced data accuracy and adjustments in aggregation period caused the network to meet the additional delay. Guiyi Wei et al., [7] proposed a method that saves network energy and eliminates redundant communication by exploiting prediction based data aggregation protocol. However, in this method synchronization time increased due to synchronization has to be done prior to each transmission. Guorui Li et al., [9] proposed an Auto Regressive Integrated Moving Average Model (ARIMA) that predicts the next time value based on the previous observed values. When the prediction error is less than the preconfigured threshold value the aggregator would not transmit the data sensed by the source node. Otherwise, it transmits the data to sink node. Therefore ARIMA model reduced the amount of data transmitted between the ordinary sensor node and aggregator node. Since this method performed aggregation on the ordinary sensor node and aggregator node it increased the computational complexity and reduced accuracy. Rajesh G et al., [5] proposed the data fusion method using Simpson’s 3/8 rule to forecast next time data based on the early sensed information. When prediction error is

THE IIOAB JOURNAL
ISSN: 0976-3104

greater than the prediction threshold the cluster head transmits the actual sensed value to the base station. Otherwise, it would not transmit data to the base station. This method reduced unnecessary transmission between cluster head and base station. However, this method provides less prediction accuracy since the deviation error is increased between subsequent values. There are several data fusion techniques in Wireless sensor networks. The main features of the proposed work are that it, Improves the performance of the forecast and Performs less computation to obtain the forecasted data.

Chaos theory based data aggregation (CTAg) technique

The typical features of chaos include: 1) Nonlinearity. If it is linear, it cannot be chaotic. 2) Determinism. It has deterministic underlying rules every future state of the system must follow. 3) Sensitivity to initial conditions. Small changes in its initial state can lead to radically different behavior in its final state. Long-term prediction is mostly impossible due to sensitivity to initial conditions. A dynamic system is a simplified model for the time-varying behavior of an actual system [17]. These systems are described using differential equations specifying the rates of change for each variable. A dynamical system of dimension N system first–order differential equations for N variables $x_1(t), x_1(t) \ldots x_N(t)$ evolve with time t according to,

$$\hat{x}_1 = f_1(x_1, x_2, \ldots x_N, t) \quad (3)$$

$$\hat{x}_2 = f_2(x_1, x_2, \ldots x_N, t) \quad (4)$$

$$\hat{x}_N = f_N(x_1, x_2, \ldots x_N, t) \quad (5)$$

Where f_1, f_2 are assigned functions and a dot is a derivative with respect to time.

The system following Characteristics of a Chaotic System:

- Sensitivity to initial conditions
- Non-linear
- Dynamic and mixed topology system and Continuous or periodic time.

So that the Chaos is the aperiodic long–term behavior in a deterministic system that exhibits sensitive dependence on the initial condition. These characteristics enables chaos theory based data aggregation (CTAg) prediction method is suitable for eliminating data redundancy in WSNs.

Considered a hierarchical wireless sensor network G (SN, E) where, SN represents the sensor nodes and E represents links connecting the nodes. These sensor nodes collect weather monitoring data (Temperature, Humidity) periodically. Each node transmits data to sink node through the intermediate node or aggregator node (A). The aggregator (A) will perform data fusion by eliminating redundant data using chaos theory before transmitting the gathered data towards the base station. This will minimize the amount of data transmitted between aggregator node and sink node.

Steps for Ontology Learning or Enrichment-

1. Convert xml document into owl.
2. Perform computational mapping of rawterms to ontology through NLP and get intelligent raw data.
3. Through intelligent raw data get unmapped terms and mapped terms through NLP.
4. Manually add the unmapped terms using Ontology editor. Mapped terms will be added computationally to the ontology.

Fig 2, shows the ontology enrichment where new tuples are inserted or updated as and when required to achieve the desired result.

THE IIOAB JOURNAL
ISSN: 0976-3104

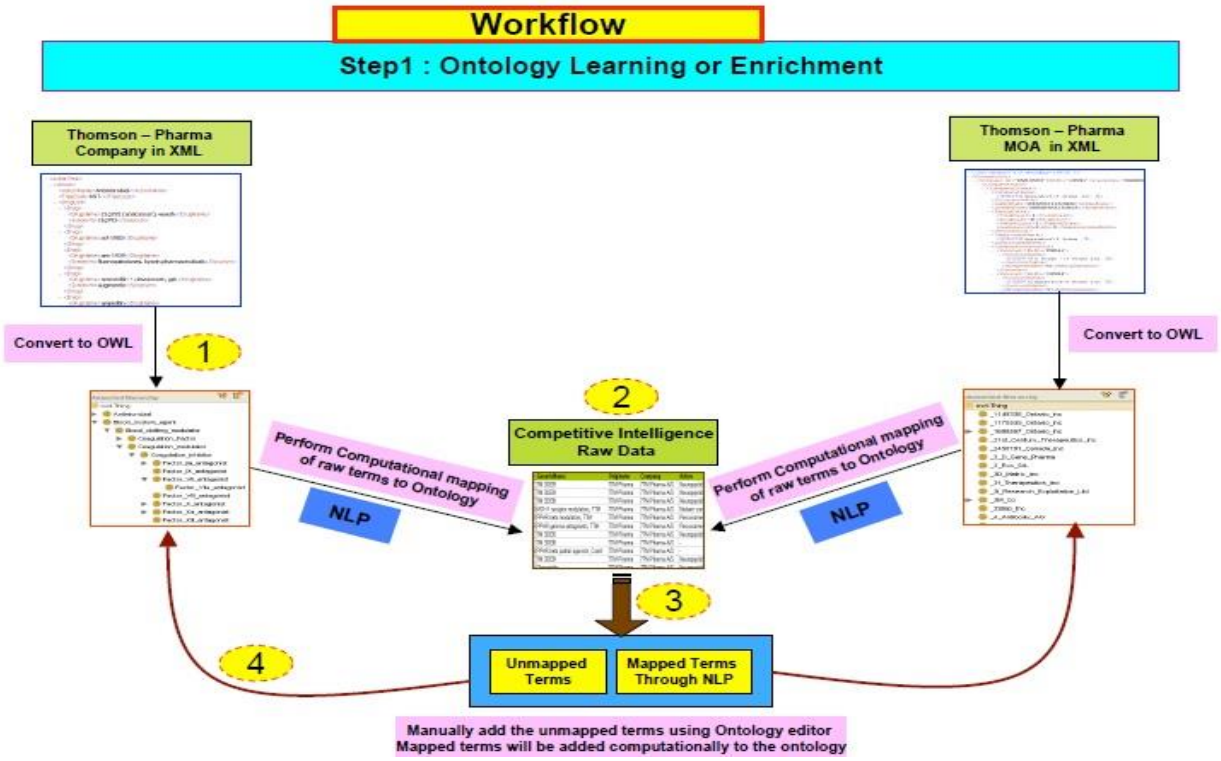

Fig: 2.Ontology Learning or enrichment

……

Steps for RDF conversion, loading and querying-

1. Ontology is coverted to rdf.
2. RDF is coverted using Jena Programming API.
3. Coverted data is uploaded to Allelograph Native RDF triple store.
4. After uploading, data is queried using SPARQL

Figure 3, shows RDF conversion where xml file is converted to rdf file. After conversion, the rdf file is loaded to Allegrograph and sparql query is performed.

THE IIOAB JOURNAL
ISSN: 0976-3104

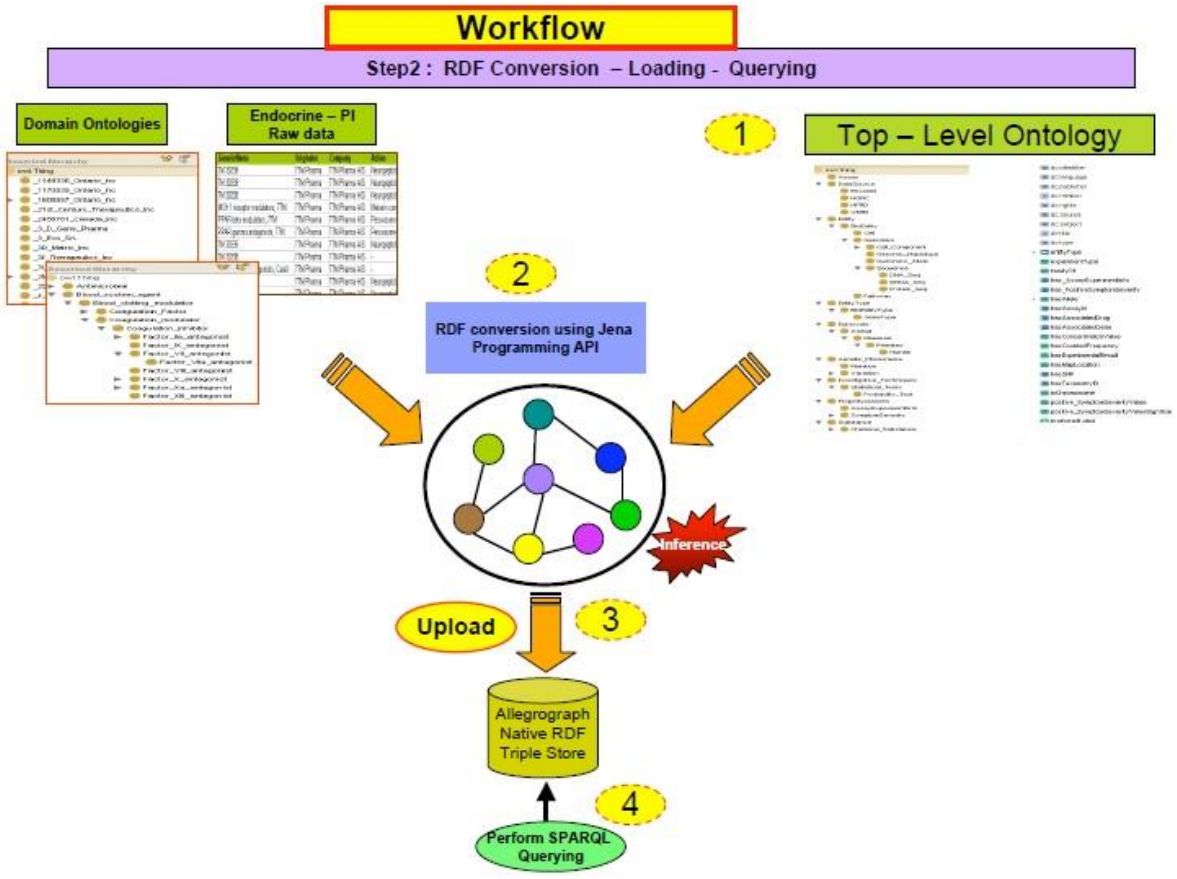

Fig: 3.RDF conversion

……

Finding the market share of drugs.

- Provides details on which brand has a better market share on every state, which could indicate the possibilities of another brand with the same ingredients to improve their market share on that state.
- Data Considered
  - CMS[25] for drug name, total drug cost, provider city, provider state and specialty description.
  - RxNorm[24] for preflabel and tradename.

CMS: As part of the Obama Administration's struggles to create our healthcare system more apparent, reasonable, and responsible. The (CMS) Centers for Medicare & Medicaid Services have organized a public data set, the Part D Prescriber Public Use File ("Part D Prescriber PUF"), with information on prescription drug events (PDEs) incurred by Medicare beneficiaries with a Part D prescription drug plan. The Part D Prescriber PUF is organized by National Provider Identifier (NPI) and drug name and contains information on drug utilization (claim countsand day supply) and total drug costs.

Data Content

- NPI – National Provider Identifier (NPI) for the performing provider on the claim.
- NPPES_ENTITY_CODE – Type of entity reported in NPPES. An entity code of 'I' identifies providers registered as individuals and an entity type code of 'O' identifies providers registered as organizations
- NPPES_PROVIDER_LAST_ORG_NAME – individual (entity type code='I'), this is the provider's last name. Entity type code = 'O', this is the organization name.
- NPPES_PROVIDER_FIRST_NAM
- NPPES_PROVIDER_GENDER
- DESCRIPTION_FLAG – A flag variable that indicates the source of the specialty_description.

THE IIOAB JOURNAL
ISSN: 0976-3104

- DRUG_NAME – The name of the drug filled. This includes both brand names and generic names.
- GENERIC_NAME – A term referring to the chemical ingredient of a drug rather than the advertised brand name under which the drug is sold.
- BENE_COUNT – The total number of unique Medicare Part D beneficiaries with at least one claim for the drug. Beneficiary counts fewer than 11 are not displayed.
- TOTAL_CLAIM_COUNT – The number of Medicare Part D claims. This includes original prescriptions and refills. Claims counts fewer than 11 are not displayed.
- TOTAL_DAY_SUPPLY – The aggregate number of days' supply for which this drug was dispensed.
- TOTAL_DRUG_COST – The aggregate total drug cost paid for all associated claims. This amount includes ingredient cost, dispensing fee, sales tax, and any applicable vaccine administration fees.
- BENE_COUNT_GE65 – The total number of unique Medicare Part D beneficiaries with at least one claim for the drug where the beneficiary is 65 or older. Beneficiary counts fewer than 11 are not displayed.
- TOTAL_CLAIM_COUNT_GE65 – The number of Medicare Part D claims where the beneficiary is 65 or older. This includes original prescriptions and refills. Claims counts fewer than 11 are not displayed.
- DAY_SUPPLY_GE65 – The aggregate number of days' supply for which this drug was dispensed, where the beneficiary is 65 or older.
- TOTAL_DRUG_COST_GE65 – The aggregate total drug cost paid for all associated claims where the beneficiary is 65 or older. This amount includes ingredient cost, dispensing fee, sales tax, and any applicable vaccine administration fees.

npi	nppes_provider_last_org_name	nppes_provider_first_name	nppes_provider_city	nppes_provider_state	specialty_description	description_flag
1003889502	AACHI	VENKAT	SAN JOSE	CA	Physical Medicine and R	S
1336172998	AAMODT	DENISE	ALBUQUERQUE	NM	Family Practice	S
1255569273	AANDERUD	PAUL	CLACKAMAS	OR	Dermatology	S

drug_name	generic_name	bene_count	total_claim_count	total_day_supply	total_drug_cost	bene_count_ge65
ACETAMINOPHEN-CODEINE	ACETAMINOPHEN WITH CODEINE		14	366	$277.42	
ACETAMINOPHEN-CODEINE	ACETAMINOPHEN WITH CODEINE		19	458	$271.90	
ACETAMINOPHEN-CODEINE	ACETAMINOPHEN WITH CODEINE	18	19	39	$70.68	

bene_count_ge65_redact_flag	total_claim_count_ge65	ge65_redact_flag	total_day_supply_ge65	total_drug_cost_ge65
*		*		
*		*		
#		#		
#		#		
#		#		
*	19		307	$358.59

Fig 4: CMS database

……

RxNorm[17] Ontology:

- Created by NIH's National Library of Medicine.
- Combines several different drug vocabularies.
- A standardized nomenclature for drug names.
- Unifies vocabularies around RXCUI, a concept unique identifier

THE IIOAB JOURNAL ISSN: 0976-3104

Data Content-

Properties

- Notation
- PrefLabel
- RXAUI
- RXCUI
- Cui
- Tui
- hassty
- constitutes
- RXN_STRENGTH
- RXN_IN_EXPRESSED_FLAG
- RXN_AVAILABLE_STRENGTH
- altLabel
- RXTERM_FORM
- RXN_HUMAN_DRUG

Properties

- RXN_HUMAN_DRUG
- RXTERM_FORM
- NDC
- RXN_QUANNTITY
- ORIG_SOURCE
- RXN_ACTIVATED
- RXN_OBSOLETE
- contains
- ORIG_CODE
- RXN_BN_CARDINALITY
- ORIG_CODE
- ORIG_TTY
- ORIG_VSAB

Relationship

- Has_ingredient
- Inverse_isa
- isa
- Has_dose_form
- form_of
- precise_ingredient
- has_tradename
- consists_of
- Has_part
- Ingredient_of
- Part_of

Drug-Drug interactions

The aim is to find the drug to drug interactions that lead to adverse reactions among patients and report such cases as a notification to doctors, pharma companies, etc. for alerting them of such interactions which would lead to better prescription knowledge amongthe doctors. To make inflight adjustments, recommend Concomitant drugs, and drug label enhancements by the pharmaceutical companies is achieved. New cases of drug interactions would also be used for enrichment of the Ontologies dealing with drug-drug interactions.
Data Sources required:-
(I) The FAERS data provided by the FDA has information on the adverse events and outcomes, a patient has undergone and the drugs consumed during each of these events by the patient.
There are also information on:-
1. Patient details like age, sex, demography of event, etc.
2. Drug details like name, brand, active ingredient, dose form, etc.
3. Adverse reactions like headache, chest pain, etc.
4. Indications for which the drug is to be taken like headache, nausea, etc.
5. Outcomes like hospitalization, death, etc.
6. Source of Information - Whether it is from the doctor, consumer, distributor, study material, etc.
7. Dates on which the therapy has taken place.
(II) The DrugBank data/ontology which has the information on drug-drug interactions. DrugBank is available in xml. It can be converted to RDF and uploaded into any triple store like Allegrograph.

RESULTS

Market Share of drugs

This evaluation has been done in Allegrograph where a federated session was created between RxNorm and CMS.A drug from CMS was referred in RxNorm to get the GENERIC Name of the drug; other brand drugs that were having the same generic name were then listed; this list was returned to CMS for finding the total drug cost of each of those drugs in the list generated from RxNorm.
Actamin has the Generic Name Acetaminophen. Drugs like Tylenol, Dolphin, Hydrocodone, etc. might behaving the same base component generic name Acetaminophen. Then all these drugs are searched for their total market share from CMS to return the brand that has the best share in the market and could also be filtered for an area like New York “NY”.There were problems of exact string matches between the multiple databases, so NLP was proposed to be used to solve the problem

THE IIOAB JOURNAL
ISSN: 0976-3104

100 Results in 22.594 s | Warnings

prefLabel	tradename_of	gen_rx_name	has_tradename	tradename	drug_name
"Actamin"	161	"Acetaminophen"	1052413	"Pamprin Max Formula"	"ABILIFY"
"Actamin Oral Product"	1152842	"Acetaminophen Oral Product"	1437472	"Dolofin Infantil Oral Product"	"ABILIFY"
"Actamin Pill"	1152843	"Acetaminophen Pill"	1187315	"Tylenol Pill"	"ABILIFY"
"Acetaminophen 325 MG Oral Tablet [Actamin]"	313782	"Acetaminophen 325 MG Oral Tablet"	209384	"Acetaminophen 325 MG Oral Tablet [Tycolene]"	"ABILIFY"
"Acetaminophen 325 MG [Actamin]"	315263	"Acetaminophen 325 MG"	570402	"Acetaminophen 325 MG / butalbital 50 MG [Marten-Tab]"	"ABILIFY"
"Acetaminophen 500 MG [Actamin]"	315266	"Acetaminophen 500 MG"	1101750	"Acetaminophen 500 MG / pamabrom 25 MG [Midol Teen]"	"ABILIFY"

100 Results in 22.594 s | Warnings

drug_name	gen_cms_name	provider_state	provider_city	specialty_description	total_drug_cost
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"
"ABILIFY"	"ARIPIPRAZOLE"	"NY"	"GENESEO"	"Internal Medicine"	"6.48193E3"

Fig: 5. Shows result drug name and total drug cost of Geneseo city

……..

Drug-Drug interaction

It provides how one drug reacts in the presence of other drug.
Figure 6. Shows how drug Lipitor react in the presence of drug aliskiren (Atorvastatin may increase the serum concentration of Aliskiren.)

THE IIOAB JOURNAL
ISSN: 0976-3104

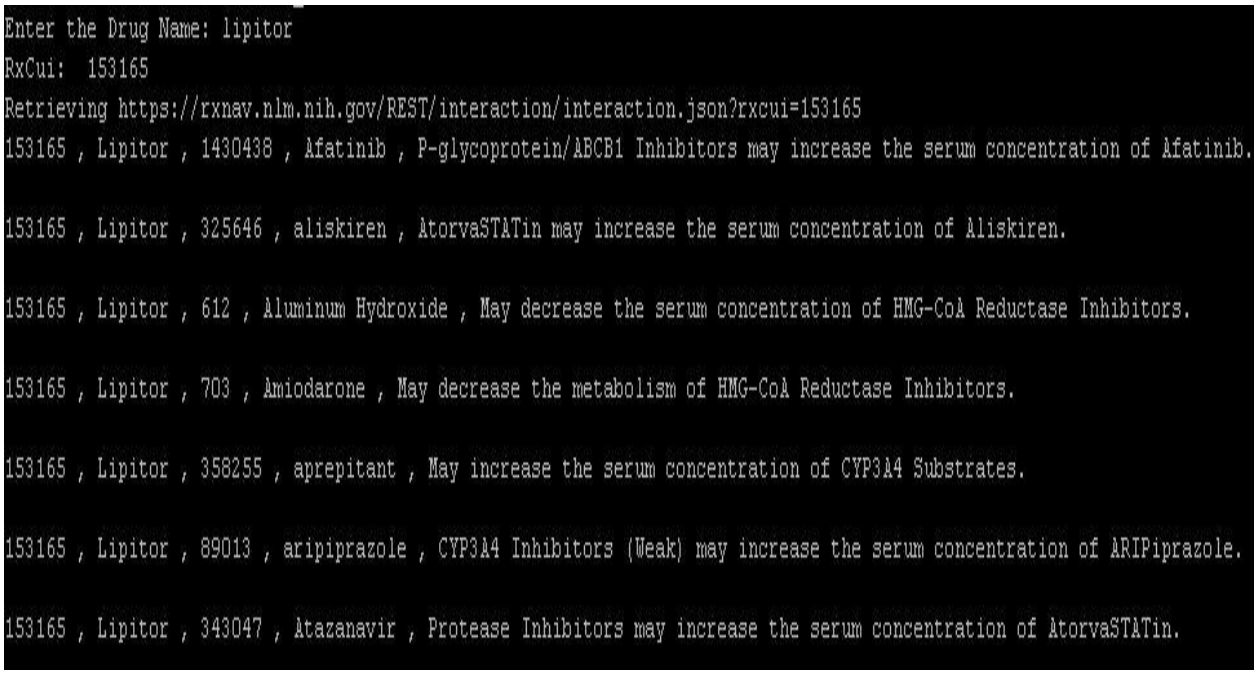

Fig: 6. Shows how Lipitor interact with other drug.

Here one drug interact with other drug and causes some adverse reaction to the patient.

Figure- 7, shows when one patient take drug named "6-(3'-5' Dimethylbenyl)-1-ethoxmethyl-5-isopropyluracil" with other drug named "Dabrafenib" may decrease the excretion of amephetamines.

CONCLUSION

The Real World Evidence in the pharmaceutical domain is an ultimate goal towards achieving better healthcare as per the Obama-Care and a platform that would showcase the implementation of this would be the first step in at least visualizing the RWE. RDF data model combined with Semantic Integration (instance mapping using NLP) was effective in answering questioning Competitive Intelligence. Ontologies provide a powerful framework in providing dictionaries and taxonomical relations that help to reason and inference the data for knowledge discovery. Manual curation is a tedious, error prone and labor intensive task. A semi-automated intelligent computer based solution that utilizes Ontologies, Semantic Integration and NLP could drastically reduce manual curation process and maintain high quality information.

THE IIOAB JOURNAL
ISSN: 0976-3104

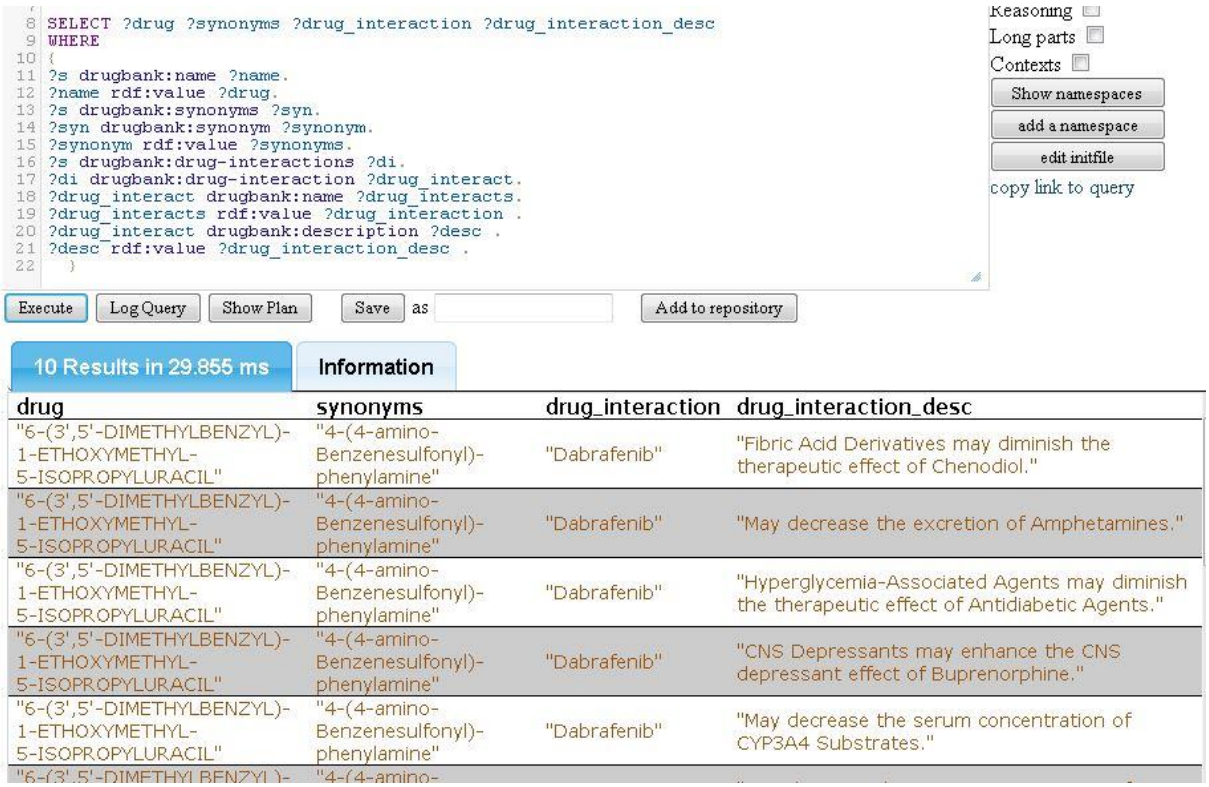

Fig: 7.Drug-Drug Interaction

……..

CONFLICT OF INTEREST

Authors declare no conflict of interest.

ACKNOWLEDGEMENT

None.

FINANCIAL DISCLOSURE

No financial support was received to carry out this project.

REFERENCES

[1] Andreas M. Pleil "Using Real World Data in Pharmacoeconomic Evaluations: Challenges, Opportunities, and Approaches" Worldwide Medical and Outcomes Research Pfizer, Inc., La Jolla, CA

[2] Mack C and Lang K. Using Real-World Data for Outcomes Research and Comparative Effectiveness Studies. Drug Discovery & Development. Nov. 4, 2014. Accessed at: http://www.dddmag.com/articles/2014/11/using-real-world-data-outcomes-research-and-comparative-effectiveness-studies.

[3] Christel M. "Evidence" Trail Elusive. Applied Clinical Trials. June 30, 2014. Accessed at: http://www.appliedclinicaltrialsonline.com/evidence-trail- elusive-0.

[4] John Davis, Rudi Studer, Paul Warren "Semantic Web Technologies: trends and research in ontology based system". USA, Willey;2006 ISBN: 978-0470025963

[5] Hyunsik Choi , Jihoon Son , YongHyun Cho , Min Kyoung Sung , Yon Dohn Chung, SPIDER: a system for scalable, parallel / distributed evaluation of large-scale RDF data, Proceedings of the 18th ACM conference on Information and knowledge management, November 02-06, 2009, Hong Kong, China [doi>10.1145/1645953.1646315].

[6] Resource Description Framework(RDF): Concepts and Abstract Syntax, http://www.w3.org/TR/rdf-concepts/

[7] Framework (RDF): Concepts and Abstract Syntax. Technical report, W3C, G. Klyne and J. J. Carroll. Resource Description 2004.

[8] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D.L. McGuinness, P.F. Patel-Schneider, and L.A. Stein. OWL Web Ontology Language Reference. Technical report, W3C, 2004.

[9] SPARQL Query Language for RDF, http://www.w3.org/TR/rdf-sparql- query/.

[10] Bernstein, M. Stocker, and C. Kiefer. SPARQL Query Optimization Using Selectivity Estimation. In Poster Proceedings of the 6th International Semantic Web Conference (ISWC), Lecture Notes in Computer Science. Springer, 2007.
[11] Olaf Hartig , Ralf Heese, The SPARQL Query Graph Model for Query Optimization, Proceedings of the 4th European conference on The Semantic Web: Research and Applications, June 03-07, 2007, Innsbruck, Austria [doi>10.1007/978-3- 540-72667- 8_40]
[12] Making a Semantic Web Business Case at Pfizer
[13] National Library of Medicine. Collection development manual: medical informatics. Available from: http://www.nlm.nih.gov/tsd/acquisitions/cdm/CDMBook .pdf. Accessed on June 5, 2015.
[14] Proceedings of the Pharmacy Practice Model Summit: An invitational consensus conference conducted by ASHP and the ASHP Research and Education Foundation, November 7–9, 2010, Dallas, Texas. Available from http://www.ashp.org/DocLibrary/PPMI/PPMISummit/Proceedings.aspx
[15] Erickson AK, editor, Pharmacy Today. Pharmacy:Harnessing the power of big data. American Pharmacists Association. http://www.pharmacist.com/pharmacy. Accessed March 20, 2015.
[16] Groves P, Kayyali B, Knott D, Van Kuiken S. The “big data” revolution in healthcare: accelerating value and innovation. New York, NY:Center for US Health System Reform, Business Technology Office, McKinsey and Company; 2013.
[17] RxNORM, http://bioportal.bioontology.org/ontologies/

ABOUT AUTHORS

Pragya Agrawal pursued her Master degree from VIT University, Vellore in Computer Science and Engineering.She has worked as a Project Trainee In Wipro Technologies for duration of 10 months. She has keen interest on Big Data Analytics and had attended short term training program on Big Data Analytics organized on 26th, 28th and 30th September 2014 conducted by School of Computing Science and Engineering and Centre for Ambient Intelligence and Advanced Networking Research at VIT University, Vellore. She got All India Rank-97 in Nationwide Interactive Science Olympiad 2007 in class 12th. She also has keen interest on Semantic Web technologies, Data Mining, Artificial Intelligence and Image Processing.

Swarnalatha Purushotham is an Associate Professor, in the School of Computer Science and Engineering, VIT University, at Vellore, India. She pursued her Ph.D degree in Image Processing and Intelligent Systems. She has published more than 50 papers in International Journals/International Conference Proceedings/National Conferences. She is having 14+ years of teaching experiences. She is a member of IACSIT, CSI, ACM, IACSIT, IEEE (WIE), ACEEE. She is an Editorial board member/reviewer of International/ National Journals and Conferences. Her current research interest includes Image Processing, Remote Sensing, Artificial Intelligence and Software Engineering.