Cell Phone Mart

CSCI 586 Project Report

Adarsh Rajanikanth, Malatesha Somasundar Anantha, Neelima Vangipuram, Sahil Wadhwa

Group 1 Wednesday

University of Southern California

1. Background

"There is a revolution occurring and it's all about making the Web meaningful, understandable, and machine-readable, whether it is based on an intranet, extranet or, Internet. This is called Semantic Web and it will transition us towards a knowledge-centric viewpoint of everything." The Semantic Web provides an infrastructure that enables not just Web pages, but databases, services, programs, sensors, personal devices and even household appliances to both consume and produce data on the Web [1]. URIs are a fundamental component of the current Web and are in turn a foundation for the Semantic Web. URIs, a fundamental component of Semantic Web provides the ability for uniquely identifying resources, RDF leverages URIs and the Web to provide a powerful means of expressing and representing the relationships and their meaning and Ontologies can be used to power advanced services such as more accurate Web search, intelligent software agents and knowledge management.

In our project titled Cell Phone Mart, we apply the concepts of Semantic Web and, by using Ontologies, integrate data from various sources to provide a unified view to the user. The data can be queried and used to extract information required by the user. We have provided queries that analyses various cellphone brands.

2. Problem Statement

There are many sources on the Internet that provide diverse set of disjoint information about Mobile Phones. **Kaggle**, one of the most comprehensive platforms that provides public datasets. The datasets on Kaggle are consistently accessible, and the forums helps to understand the nuances of the data. The data that we acquired from Kaggle had the information about unlocked mobile phones reviews from Amazon with attributes like – product name, brand name, price, rating, reviews, and review votes.

Similarly, **DeviceSpecifications & GSM Arena,** a comprehensive source that provides information about full specifications and features of various mobile devices such as – standard mobile phones and smartphones with attributes such as – dimensions, weight, SoC, CPU, Cores, RAM, Storage, Display, Battery, OS, Camera, SIM Card, Wi-Fi, USB, Bluetooth and positioning. **Wikipedia**, the most common and extensively used online source for anything contributed to the third data source to extract information about the Brand with attributes such as – type, founded in, headquarters, area served, other products, number of employees and its website link.

With data scattered around at multiple data sources, the user can't get an integrated piece of information about a mobile phone of his choice at one place. For example, if a user wants to know which mobile phone had the highest rating with a certain screen size, he would have to sieve through all the sources. Likewise, identifying which mobile brand had the highest selling rate based on a key feature a user is interested in, becomes time-consuming otherwise.

In our project, we try to implement a few such key performance indices. We gather information about mobile phones from various data sources, unite and shape them, and provide a SPARQL endpoint. The results of the SPARQL queries are displayed on the browser using HTML and JavaScript. A provision for visualization is also provided which uses High charts to display charts and tables to display the query results.

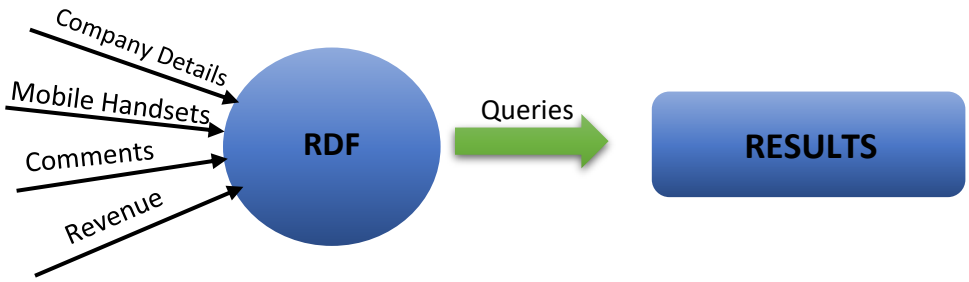


Figure 1. Workflow

3. Scope

In our project, we are focusing on data available from the following websites. The attributes crawled from each page are listed in Table 1:

1. Kaggle
2. Device Specifications
3. Wikipedia
4. GSM Arena

We have also extracted the yearly revenue of Brands starting from 2014 to 2016 by web scrapping Wikipedia and the scrapping individual Brand pages which were not present in Wikipedia's revenue data.

Website	Extracted Information
Kaggle	product name, brand name, price, rating, reviews, review votes
Device Specification, GSM Arena	dimensions, weight, SoC, CPU, Cores, RAM, Storage, Display, Battery, OS, Camera, SIM Card, Wi-Fi, USB, Bluetooth and positioning
Wikipedia	type, founded in, headquarters, area served, other products, number of employees and its website link
Revenue Dataset	Profit_Id, Profit_2014, Profit_2015, Profit_2016

Table 1. Information extracted from websites

4. Approach

In this section, we discuss different phases of the development and execution of our project. The project can be divided into four phases.

1. Data Acquisition
2. Data Cleaning
3. Data Modeling & Integration
4. Data Linking
5. Querying

4.1 Data Acquisition

We implemented Java-based crawlers to extract data from the above-mentioned data sources by

1. Using jsoup library that provides an API to extract and manipulate data from the web pages we used this to extract data from various websites like Device specifications, GSM Arena, Wikipedia to obtain Cell phone's technical specs, Detailed features, Brand info respectively.

2. For analysis of user reviews on individual models we got the users comments data set for cell phones on amazon from Kaggle which helped us to analyze each model based on user rating and comments in amazon.

Website	No. of records generated
Kaggle	4,13,000 records were generated for comments and review
Device Specification, GSM Arena	8,632 models were captured for various brands of mobile phones
Wikipedia	108 different brand information was captured
Revenue Data	25896 records pertaining to the profits through years 2014 to 2016

Table 2. Statistics about generated data

Data Acquisition Challenges: Web Scrapping may sound simple but is accompanied with its own challenges. We faced the following:

- 9,000 records are from various above-mentioned Data Sources.
- Non-uniform structure and hence it took long time for data refinement making it available for mapping.
- Navigation through elements and attributes in an XML document

4.2 **Data Cleaning**

- Organized scrapped data by removing the duplicates rows from the CSV file by writing generic Java code.
- Removed some special characters and supported some by using specific data structures.
- Used specific datatype for handling website links/image URLs.
- Enforced UTF-Encoding for proper conversion of data to RDF.

4.3 **Data Modeling & Integration**

We created an Ontology for our model as shown in Figure 2. There are total of 4 classes namely – Brand, Model, Comments, and Profits. A total of 4 Object Properties – has_brand, has_comment, has_model, profit_has_model. A total of 31 data properties. The object properties link different data sources.

has_brand links the model data source with the brand data source.

has_model links the comments data source with the model data source.

profit_has_model links the profit data source with the model data source.

There are 31 data properties from 4 data sources, each of these can be considered as the values associated with a particular data source. Figure 2 gives the complete details regarding both object and data properties.

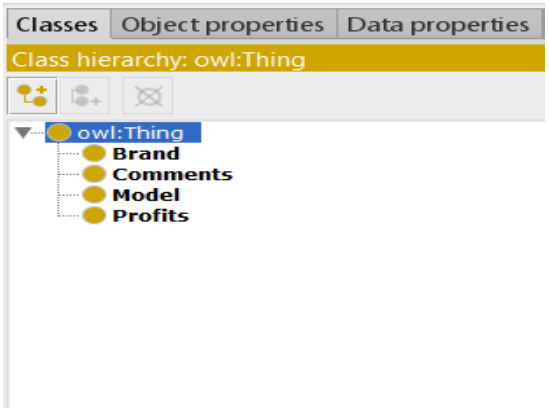


Figure 3. Classes

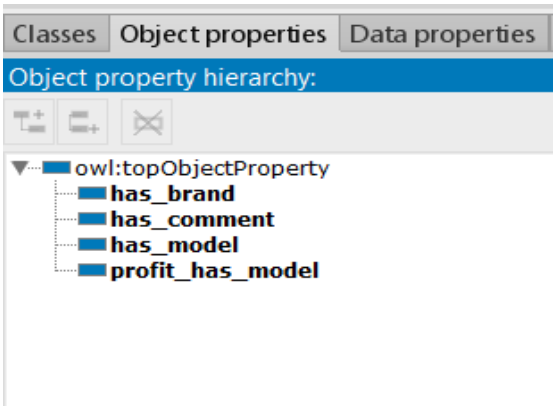


Figure 4. Object Properties

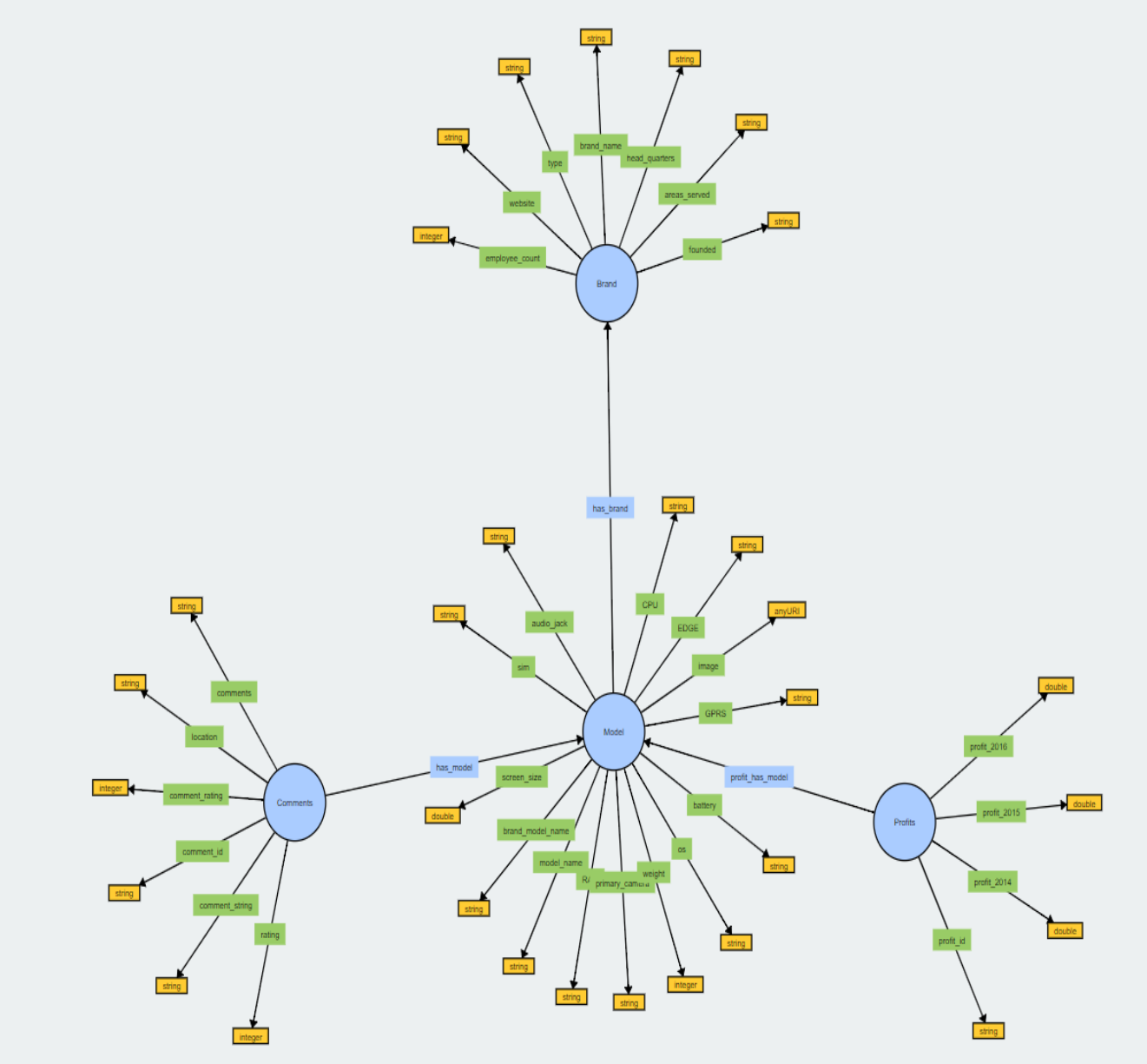


Figure 2. Ontology Graph

4.3 **Data Linking**

Data linking played a most prominent role in this application because the data was scrapped from various data resources and hence there was a requirement for linking or mapping of data in some way such that it makes sense. In order to do so we created a unique primary key in individual data set and then we linked them in owl using their URI so that we can obtain interlinked owl file containing mapped data. This was done by utilizing Protégé tool.

In below images we can notice that we have created unique id for individual data set and we have mapped them by utilizing their common attributes so that we can obtain N:1 and 1:N mapping between entities of mapped classes of different data sets.

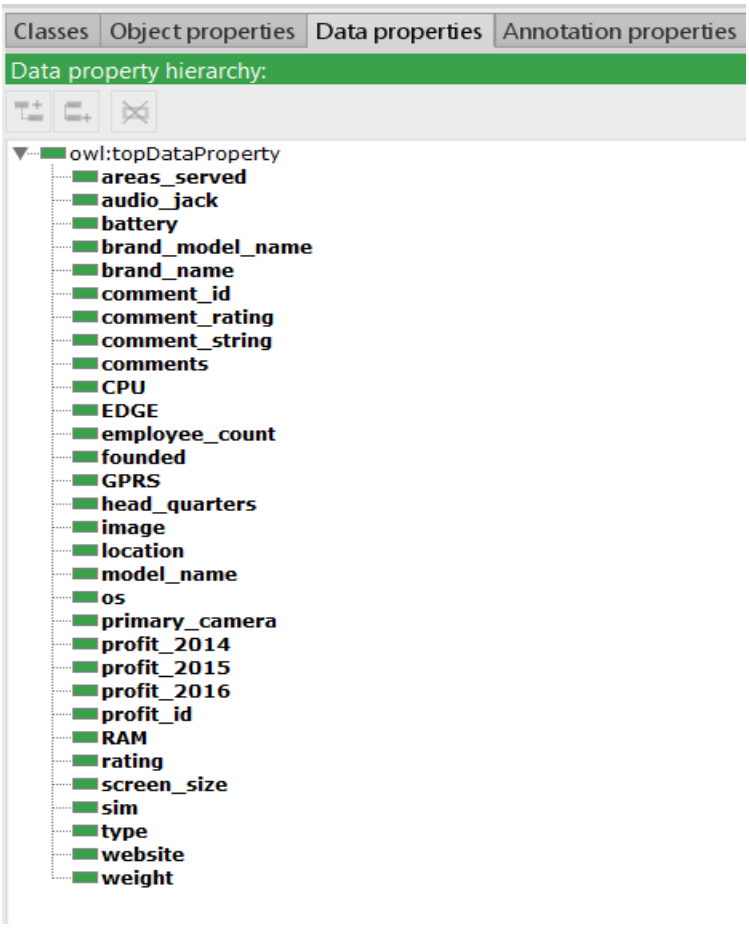


Figure 5. Data Properties

COMMENTS

	A	B	C	D	E	F	G
1	Brand Model comment 0	Brand Model	Brand	Model	Rating	Review	Location
2	Acer Liquid Jade Z comment 1	Acer Liquid Jade Z	Acer	Liquid Jade Z	1	The description says 16GB rom, it	CALIFORNIA
3	Acer Liquid Jade Z comment 2	Acer Liquid Jade Z	Acer	Liquid Jade Z	2	I had high hopes for this Acer phc	CALIFORNIA
4	Acer Liquid M220 comment 3	Acer Liquid M220	Acer	Liquid M220	1	First off, great service as usual. I	ARIZONA
5	Acer Liquid M220 comment 4	Acer Liquid M220	Acer	Liquid M220	1	I did not receive my order items.	CALIFORNIA
6	Acer Liquid M220 comment 5	Acer Liquid M220	Acer	Liquid M220	1	The phones were advertised as b	TEXAS
7	Acer Liquid M220 comment 6	Acer Liquid M220	Acer	Liquid M220	1	manure	FLORIDA
8	Acer Liquid M220 comment 7	Acer Liquid M220	Acer	Liquid M220	1	phone very poor quality. I recom	TEXAS
9	Acer Liquid M220 comment 8	Acer Liquid M220	Acer	Liquid M220	3	It'sn a powerfull phone.	FLORIDA
10	Acer Liquid M220 comment 9	Acer Liquid M220	Acer	Liquid M220	3	this phone was a replacement fo	ARIZONA
11	Acer Liquid M220 comment 10	Acer Liquid M220	Acer	Liquid M220	4	It's work well	CALIFORNIA
12	Acer Liquid M220 comment 11	Acer Liquid M220	Acer	Liquid M220	4	Item is good. The only thing is AT	CALIFORNIA

REVENUE

	A	B	C	D	E	F	G
1	profit id	model_id	Brand	model	Revenue 2016	Revenue 2015	Revenue 2014
2	Motorola E1000 1	Motorola E1000	Motorola	E1000	3.995884152	2.487442254	2.936621763
3	Motorola RAZR V XT889 1	Motorola RAZR V XT889	Motorola	RAZR V XT889	3.994552982	2.960712931	2.557519431
4	Motorola Moto 360 . 1	Motorola Moto 360 .	Motorola	Moto 360 .	3.991374214	3.304798991	2.779012273
5	HTC Desire 620 1	HTC Desire 620	HTC	Desire 620	3.990173801	3.169891124	3.363771433
6	Motorola V690 1	Motorola V690	Motorola	V690	3.986258922	3.462777975	3.392175545
7	HTC Dream 1	HTC Dream	HTC	Dream	3.985789737	3.613500601	3.554501269
8	Motorola Moto C Plus 1	Motorola Moto C Plus	Motorola	Moto C Plus	3.981485045	2.9331731	2.591083731
9	Motorola Moto 360 46m	Motorola Moto 360 46m	Motorola	Moto 360 46mm .	3.980722458	2.564511098	2.921502369
10	Motorola W360 1	Motorola W360	Motorola	W360	3.978904883	2.538550885	2.238986419
11	HTC Desire 820q dual sim	HTC Desire 820q dual sin	HTC	Desire 820q dual sim	3.978792933	3.143284196	3.869683061
12	Motorola E380 1	Motorola E380	Motorola	E380	3.978414749	2.198565592	3.275390667
13	HTC Touch 1	HTC Touch	HTC	Touch	3.977423143	2.411044322	2.689163358
14	Motorola RAZR V3 1	Motorola RAZR V3	Motorola	RAZR V3	3.973519073	2.682289476	3.111203281
15	Motorola MPx100 1	Motorola MPx100	Motorola	MPx100	3.973232893	2.955173778	2.907872705
16	Motorola M3688 1	Motorola M3688	Motorola	M3688	3.966866738	3.216377745	3.524824391
17	HTC DROID Incredible 4G	HTC DROID Incredible 4C	HTC	DROID Incredible 4G LTI	3.960629571	2.653048523	3.381693967
18	Motorola Droid Turbo 2 1	Motorola Droid Turbo 2	Motorola	Droid Turbo 2	3.959711795	2.288929778	3.094122462

SPECIFICATIONS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	brand	model	network_t	2G_bands	3G_bands	4G_bands	network_s	GPRS	EDGE	announce	status	dimentior	weight_g	weight_oz	SIM	display_ty	display_re	display_s	OS	CPU	Chipset	GPU	memory_c	internal_r	RAM
2	Acer	Iconia Tal	GSM / HSF	GSM 850 ,	HSDPA 85	LTE band :	HSPA 42.2	Yes	Yes	2016	Aug	Available.	191.7 x 1(	260	9.17	Dual SIM	IPS LCD ca	7.0 inches	720 x 128	Android 6	Quad-cor	Mediatek	Mali-T72(	microSD	16/32 GB	2 GB RAM
3	Acer	Liquid Z6	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA 42.2	Yes	Yes	2016	Aug	Available.	153.8 x 7!	169	5.96	Single SIM	IPS LCD ca	5.5 inches	1080 x 19	Android 6	Octa-core	Mediatek	Mali-T72(	microSD	32 GB	3 GB RAM
4	Acer	Liquid Z6	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2016	Aug	Available.	-			Single SIM	IPS LCD ca	5.0 inches	720 x 128	Android 6	Quad-cor	Mediatek	Mali-T72(	microSD	8 GB	1 GB RAM
5	Acer	Iconia Tal	No cellula	N/A				No	No	2016	Apr	Available.	259 x 167 x 8.9 mm (10.20 x 6.!			No	IPS LCD ca	10.1 inche	1920 x 12	Android 6	Quad-cor	Mediatek	Mali-T72(	microSD	16/32/64	2 GB RAM
6	Acer	Liquid X2	GSM / HSF	GSM 850 ,	HSDPA 90	LTE 800 /	HSPA 42.2	Yes	Yes	2015	Apr	Available.	153.3 x 7!	166	5.86	Triple SIM	IPS LCD ca	5.5 inches	720 x 128	Android 5	Octa-core	Mediatek	Mali-T72(	microSD	32 GB	3 GB RAM
7	Acer	Liquid Jad	GSM / HSF	GSM 850 ,	HSDPA 90	LTE	HSPA 42.2	Yes	Yes	2016	Feb	Available.	-			Dual SIM	Super AMI	5.5 inches	1080 x 19	Android 6	Hexa-core	Qualcomm	MSM899	microSD	32 GB	3 GB RAM
8	Acer	Liquid Zes	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2016	Apr	Available.	154 x 77 x 10 mm (6.06 x 3.03 :			Single SIM	IPS LCD ca	5.5 inches	720 x 128	Android 6	Quad-cor	Mediatek	Mali-T72(	microSD	16 GB	2 GB RAM
9	Acer	Liquid Zes	GSM / HSF	GSM 850 ,	HSDPA 85	LTE	HSPA LTE	Yes	Yes	2016	Feb	Available.	145.7 x 7!	125	4.41	Single SIM	IPS LCD ca	5.0 inches	720 x 128	Android 6	Quad-cor	Mediatek	Mali-400(	microSD	8 GB	1 GB RAM
10	Acer	Predator (	No cellula	N/A				No	No	2015	Sep	Available.	217.9 x 1!	353.8	12.49	No	IPS LCD ca	8.0 inches	1920 x 12	Android 5	Quad-cor	Intel Atom	x7-Z8700	microSD	32/64 GB	2 GB RAM
11	Acer	Liquid Jad	GSM / HSF	GSM 850 ,	HSDPA 85	LTE 800 /	HSPA 42.2	Yes	Yes	2015	Sep	Available.	156.5 x 7!	150	5.29	Dual SIM	AMOLED c	5.5 inches	1080 x 19	Microsoft	Hexa-core	Qualcomr	Adreno 41	microSD	32 GB	3 GB RAM
12	Acer	Liquid Z3(	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	136 x 66.5	142	5.01	Single SIM	IPS LCD ca	4.5 inches	480 x 854	Android 5	Quad-cor	Qualcomr	Adreno 3(	microSD	8 GB	1 GB RAM
13	Acer	Liquid Z3(	GSM / HSF	GSM 850 ,	HSDPA		HSPA 42.2	Yes	Yes	2015	Sep	Available.	136 x 66.5	142	5.01	Single SIM	IPS LCD ca	4.5 inches	480 x 854	Android 5	Quad-cor	Qualcomr	Adreno 3(	microSD	8 GB	1 GB RAM
14	Acer	Liquid Z6(	GSM / HSF	GSM 850 ,	HSDPA 90	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	156.3 x 7!	165	5.82	Single SIM	IPS LCD ca	5.5 inches	720 x 128	Android 5	Octa-core	Mediatek	Mali-T72(	microSD	32 GB	3 GB RAM
15	Acer	Liquid Z6(	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	156.3 x 7!	165	5.82	Single SIM	IPS LCD ca	5.5 inches	720 x 128	Android 5	Quad-cor	Mediatek	Mali-T72(	microSD	8 GB	1 GB RAM
16	Acer	Liquid Z5(	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	144 x 70.3	145	5.11	Single SIM	IPS LCD ca	5.0 inches	720 x 128	Android 5	Octa-core	Mediatek	Mali-T72(	microSD	32 GB	3 GB RAM
17	Acer	Liquid Z5(	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	144 x 70.3	145	5.11	Single SIM	IPS LCD ca	5.0 inches	720 x 128	Android 5	Quad-cor	Mediatek	Mali-T72(	microSD	8 GB	1 GB RAM
18	Acer	Liquid M3	GSM / HSF	GSM 850 ,	HSDPA	LTE	HSPA LTE	Yes	Yes	2015	Sep	Available.	136 x 66.5	142	5.01	Single SIM	IPS LCD ca	4.5 inches	480 x 854	Microsoft	Quad-cor	Qualcomr	Adreno 3(	microSD	8 GB	1 GB RAM
19	Acer	Liquid M3	GSM / HSF	GSM 850 ,	HSDPA		HSPA 42.2	Yes	Yes	2015	Sep	Available.	136 x 66.5	142	5.01	Single SIM	IPS LCD ca	4.5 inches	480 x 854	Microsoft	Quad-cor	Qualcomr	Adreno 3(	microSD	8 GB	1 GB RAM
20	Acer	Iconia Tal	No cellula	N/A				No	No	2015	Apr	Available.	260 x 176	540	1.19	No	LED-backl	10.1 inche	1920 x 12	Android 5	Quad-cor	Intel Atom	Z3735F	microSD	16/32/64	2 GB RAM
21	Acer	Iconia On	No cellula	N/A				No	No	2015	Apr	Available.	213.9 x 1!	355	12.52	No	IPS LCD ca	8.0 inches	800 x 128	Android 5	Quad-cor	Intel Atom	Z3735G	microSD	16/32 GB	1 GB RAM
22	Acer	Iconia Tal	No cellula	N/A				No	No	2014	Oct	Available.	256.5 x 17	508	1.12	No	LED-backl	10.1 inche	1280 x 80	Android 4	Quad-cor	Mediatek	Mali-450(	microSD	16 GB	1 GB RAM
23	Acer	Iconia Tal	No cellula	N/A				No	No	2014	Oct	Available.	256.5 x 17	508	1.12	No	LED-backl	10.1 inche	1920 x 12	Android 4	Quad-cor	Mediatek	Mali-450(	microSD	32 GB	2 GB RAM
24	Acer	Liquid Jad	GSM / HSF	GSM 850 ,	HSDPA 90	LTE band :	HSPA 21.1	Yes	Yes	2015	Mai	Available.	143.5 x 69	110	3.88	Dual SIM	IPS LCD ca	5.0 inches	720 x 128	Android 5	Quad-cor	Mediatek	Mali-T76(	microSD	8 GB	1 GB RAM
25	Acer	Liquid Z5	GSM / HSF	GSM 850 ,	HSDPA 900 / 2100		HSPA	Yes	Yes	2015	Mai	Available.	154 x 86 x	118	4.16	Single SIM	TFT capac	5.0 inches	480 x 854	Android 4	Quad-cor	Mediatek	Mali-400(	microSD	8 GB	1 GB RAM
26	Acer	Liquid Z2	GSM / HSF	GSM 850 ,	HSDPA 900 / 2100		HSPA	Yes	Yes	2015	Mai	Available.	125.3 x 64	120	4.23	Single SIM	Capacitiv	4.0 inches	480 x 800	Android 5	Dual-core	Qualcomr	Adreno 3(	microSD	8 GB	1 GB RAM
27	Acer	Liquid M2	GSM / HSF	GSM 850 ,	HSDPA 900 / 2100		HSPA 21.1	Yes	Yes	2015	Mai	Available.	124.9 x 64	119	4.2	Single SIM	TFT capac	4.0 inches	480 x 800	Microsoft	Dual-core	Qualcomr	Adreno 3(	microSD	4 GB	512 MB R
28	Acer	Liquid Z4	GSM / HSF	GSM 850 ,	HSDPA 90	LTE band :	HSPA LTE	Yes	Yes	2015	Jan	Available.	136 x 66.7	145	5.11	Single SIM	IPS LCD ca	4.5 inches	540 x 960	Android 4	Quad-cor	Mediatek	Mali-T76(	microSD	8 GB	1 GB RAM
29	Acer	Liquid Jad	GSM / HSF	GSM 850 ,	HSDPA 90	LTE band :	HSPA 21.1	Yes	Yes	2014	Dec	Available.	143 x 69 x	116	4.09	Dual SIM	IPS LCD ca	5.0 inches	720 x 128	Android 4	Octa-core	Mediatek	Mali-T76(	microSD	16 GB	2 GB RAM
30	Acer	Liquid Z5(	GSM / HSF	GSM 850 ,	HSDPA 900 / 2100		HSPA	Yes	Yes	2014	Sep	Available.	145 x 73 x	150	5.29	Single SIM	IPS LCD ca	5.0 inches	720 x 128	Android 4	Quad-cor	Mediatek	Mali-400(	microSD	4 GB	1 GB RAM
31	Acer	Liquid X1	GSM / HSF	GSM 850 ,	HSDPA 90	LTE band :	HSPA LTE	Yes	Yes	2014	Jun	Available.	153.3 x 8(	164	5.78	Micro-SIN	IPS LCD ca	5.7 inches	720 x 128	Android 4	Octa-core	Mediatek	Mali-450(	microSD	16 GB	2 GB RAM
32	Acer	Liquid Jad	GSM / HSF	GSM 850 ,	HSDPA 900 / 2100		HSPA	Yes	Yes	2014	Jun	Available.	140.5 x 69	110	3.88	Dual SIM	IPS LCD ca	5.0 inches	720 x 128	Android 4	Quad-cor	Mediatek	Mali-400(	microSD	8 GB	1 GB RAM

BRANDS

	A	B	C	D	E	F	G	H	I
1	Brand	Type	Founded_In	Headquarters	Area_Served	Number	Website_URL	Other_Products	Country
2	Acer	Public	1976	Xizhi, New Taipei, Taiwan	Worldwide	7967	https://www.acer.com/ac/en/US/content/home	Desktops, laptops, netbooks, servers	Taiwan
3	alcatel	Trademark	2004	Paris, France	Worldwide	52673	https://networks.nokia.com	Hardware, software and services to te	France
4	Allview	Public	1999	NULL	Worldwide	0	NULL	none	unknown
5	Amazon	Public	1994	Seattle, Washington, U.S	Worldwide	541900	https://www.amazon.com	Amazon Appstore, Amazon Echo, Ar	USA
6	Amoi	Public	1997	NULL	Worldwide	0	NULL	LCD Television	unknown
7	Apple	Public	1976	Cupertino, California, USA	Worldwide	123000	https://www.apple.com	Macintosh, iPod, iPhone, iPad, Appli	USA
8	Archos	Public	1988	Igny, France	Worldwide	200	http://www.archos.com	Portable media players, tablet compu	France
9	Asus	Public	1989	Beitou District, Taipei, Taiwan	Worldwide	17000	https://www.asus.com/us/	Desktops, laptops, netbooks, mobile	Taiwan
10	AT&T	Public	1983	Dallas, Texas, United States	Worldwide	273000	https://www.att.com	Satellite television, Fixed line telephc	USA
11	Benefon	Public	2011	Salo, Finland	Worldwide	200	http://www.twigworld.com	Personal safety and GPS tracking pr	Finland
12	BenQ	Subsidiary	2001	Taipei, Taiwan	Worldwide	100000	http://www.benq.us	TFT LCD monitors, digital projectors	Taiwan
13	BenQ-Siemens	Subsidiary	2005	Taoyuan, Taiwan	Worldwide	3000	http://www.benqmobile.com	none	Taiwan
14	Bird	Public	1992	Fenghua, China	Worldwide	0	http://www.chinabird.com/en/about/about.asp	none	China
15	BlackBerry	Public	1984	Waterloo, Canada	Worldwide	4044	https://us.blackberry.com	BlackBerry Messenger, BlackBerry, [	Canada
16	BLU	Public	2009	Miami, Florida, USA	Worldwide	500	http://bluproducts.com/android-phones/	none	USA
17	Bosch	Private	1886	Gerlingen, Germany	Worldwide	390000	https://www.bosch.com	Automotive parts, power tools, securi	Germany
18	BQ	Self-owned	2010	Madrid, Spain	Spain	1000	https://www.bq.com/en/	multimedia devices, 3D printers and	Spain
19	Casio	Public	1946	Shibuya, Tokyo, Japan	Worldwide	12287	https://world.casio.com	Watches (includes G-Shock and Wa	Japan
20	Cat	Public	1993	NULL	Worldwide	0	https://www.catphones.com/shop/cat-s60-smartphor	none	unknown
21	Celkon	Private	2009	Hyderabad, India	Worldwide	200	http://www.celkonmobiles.com	N/A	India
22	Chea	Public	1991	NULL	Worldwide	0	NULL	none	unknown
23	Coolpad	Public	1993	Shenzhen, China	Worldwide	5634	http://www.coolpad.com.hk/eng/global/home.php	Yulong Computers	China
24	Dell	Public	2016	One Dell Way, Round Rock, Tex	Worldwide	138000	https://www.delltechnologies.com/en-us/index.htm	Personal computers, Servers, SAN s	unknown
25	Emporia	Private	1992	Austria	Worldwide	50	http://www.emporia.eu	none	Austria
26	Energizer	Public	2002	NULL	Worldwide	0	NULL	none	unknown

Figure 6. Generating a consolidated dataset for Protégé

By integrating all the above data sets we got a consolidated data set with mapped key value pairs, now we can use single query to get the resquired results from the consolidated dataset.

Challenges in Data Linking – Due to the huge heterogeneity in data, we faced a few challenges while linking these datasets. The dataset acquired for Comments and reviews about mobile phones had redundant entries for comments, we resolved this ambiguity by deleting the duplicates. Then we had to assign a unique id for each of the rows of different datasets for linking. We carefully chose the brand name as the unique key for the brand xls sheet, the concatenation of brand name and the model name was the unique for the model sheet. For revenue and comments we assigned a unique id by hashing all the column values of each row. We also observed that a brand can have multiple models, for instance Samsung could produce hundreds of models under their brand. This problem of data linking with cardinality of n:1 was resolved by applying transformation rules(Figure 8). The below screenshot shows the process of data linking.

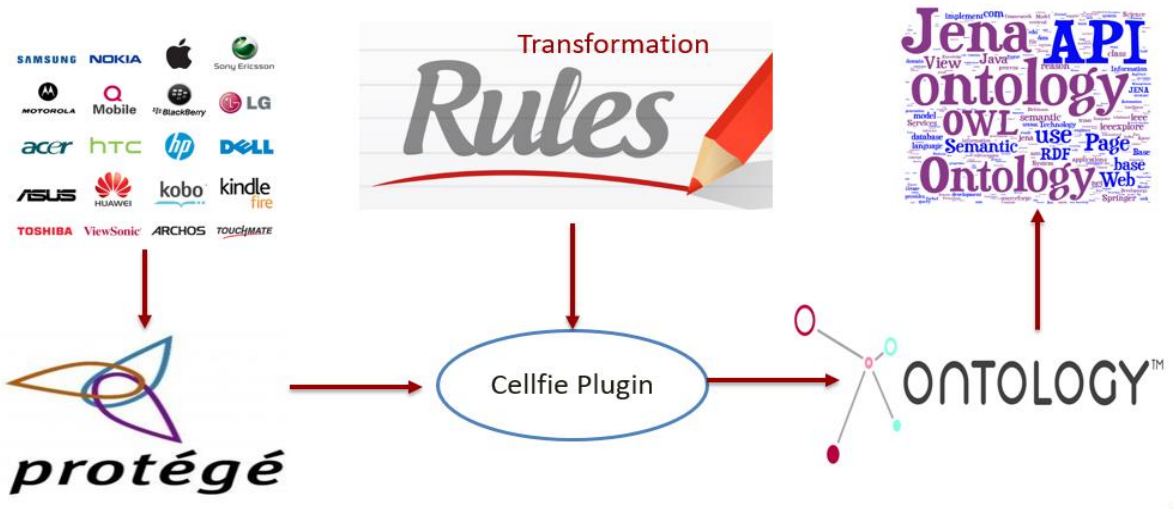


Figure 7. Data Linking Steps

In the figure 6, we can see that data was collected from heterogenous data sources and a consolidated dataset was generated to be fed to Protégé accompanied by Cellfie plugin to which we issued transformation rules to create a resulting OWL file. This OWL file was then given as an input to Apache Fuseki Jena that was our application's SPARQL endpoint.

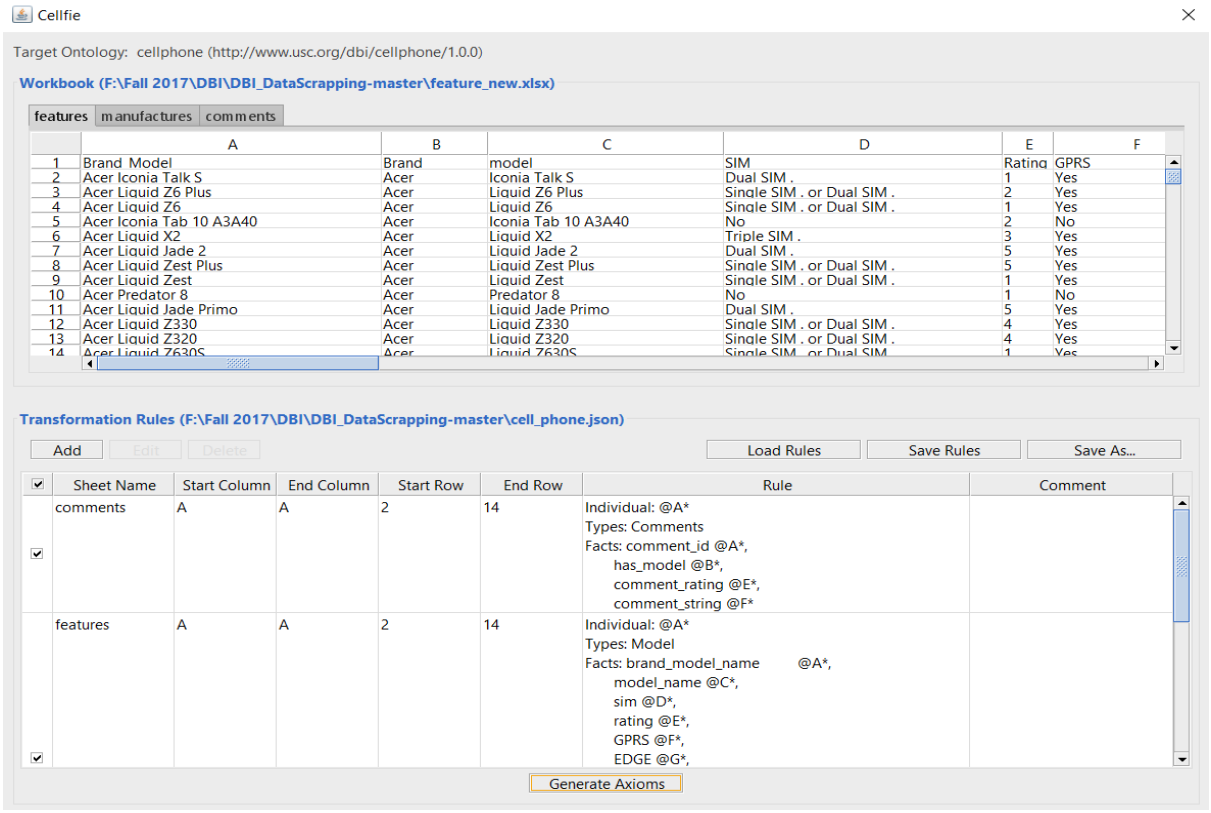


Figure 8. Adding data from CSV to RDF/OWL file by applying transformation rules

4.4 **Querying**

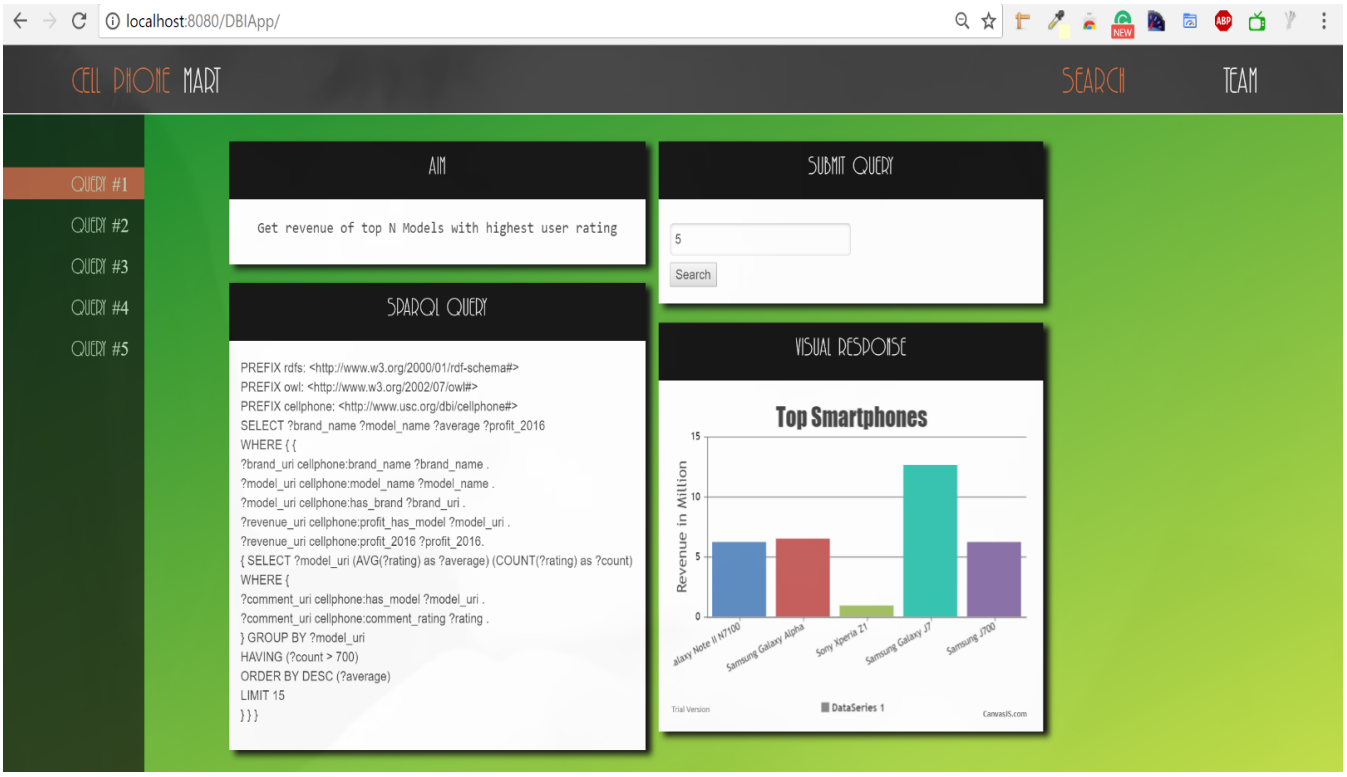
After data integration and linking, the RDF dataset had roughly 14,00,000 triples. We hosted the RDF dataset in Apache Jena Fuseki server. This server acted as a SPARQL end point. The repository was then queried using Apache Jena API. A user-interface was developed using HTML, jQuery and JavaScript to display the results of the query. We also used High charts, a JavaScript library to implement bar charts and histograms for the results of the queries. The user interface is shown in Figure 9. The queries are pre-defined and provisions for users to change the parameters of the query is provided in the UI, the functionality of the user issuing their own queries from the UI is in scope for future work. When user chooses to run a pre-configured query, the actual SPARQL query is also shown in the text area, which can be used as a template for custom queries. The list of pre-configured queries and their descriptions is given in Table 3.

Query	**Query Description**
Query #1	**Get the revenue of top N models with highest user rating**: In this query, we are finding the cell phone models which has the best reviews according to Amazon customers, then we get the profits obtained by the companies by selling each of these phone models. This helps in understanding the general trend of the revenue generated by the best rated cell phones. We have also given the bar chart representation for visualizing the trend.

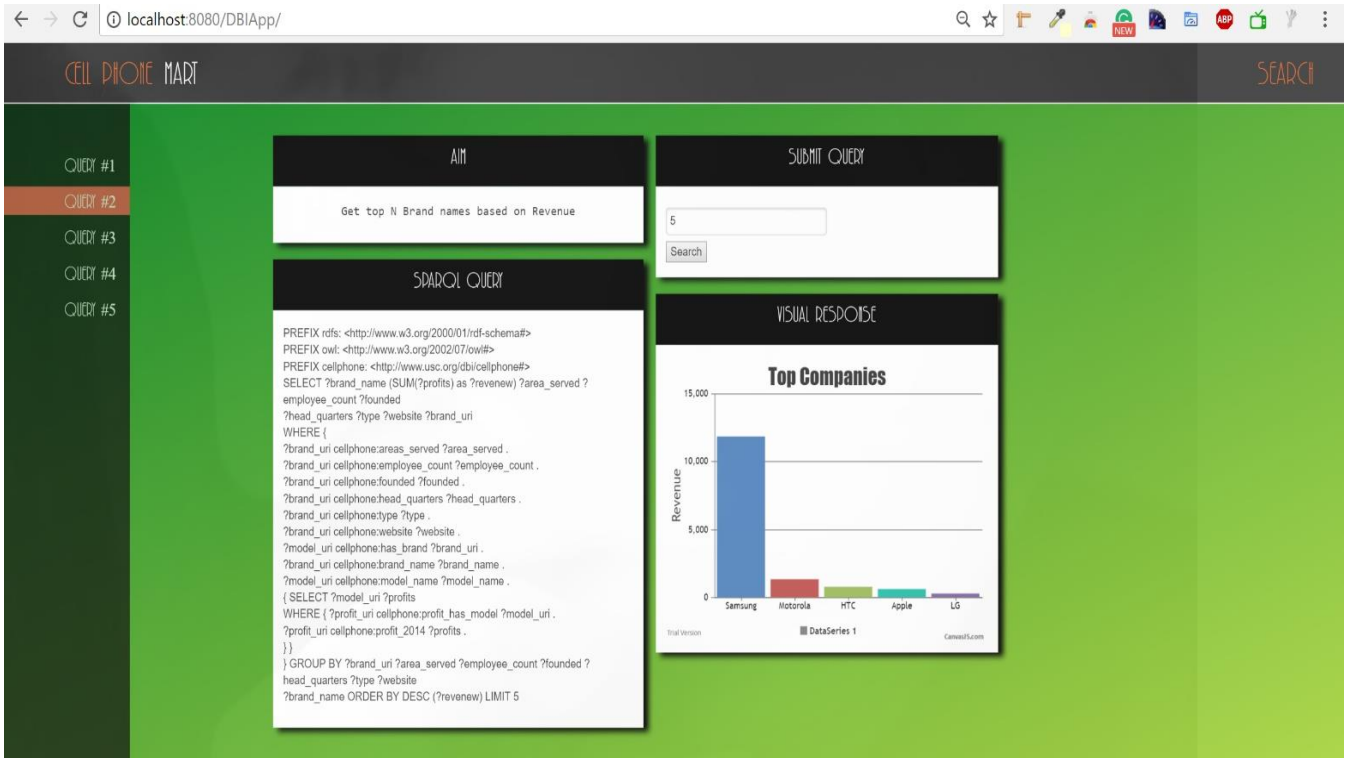
Query #2	**Get top N models based on their revenue:** In this query we are retrieving top N models, which are having highest gross profit among all the available brands. This helps user in make better choice of a model while buying a cell phone.
Query #3	**Get all the details of a cell phone based on brand and model:** Helps user to get all the details of a cell phone by entering the brand name and model name. This query returns feature data, comments, ratings and profits.
Query #4	**Get all the phones manufactured by reputed brands whose weight is less than a given weight:** Here users are asked to enter the maximum weight of a cell phone which they wish to buy. The query returns all the cell phones whose weights is below the user entered weights.
Query #5	**Get all the cell phones with screen size greater than X and has best user rating:** In this query users enters the size of cell phone in inches, then we will retrieve all the top-rated cell phones greater than or equal to the user specified size.

Table 3. Query Description

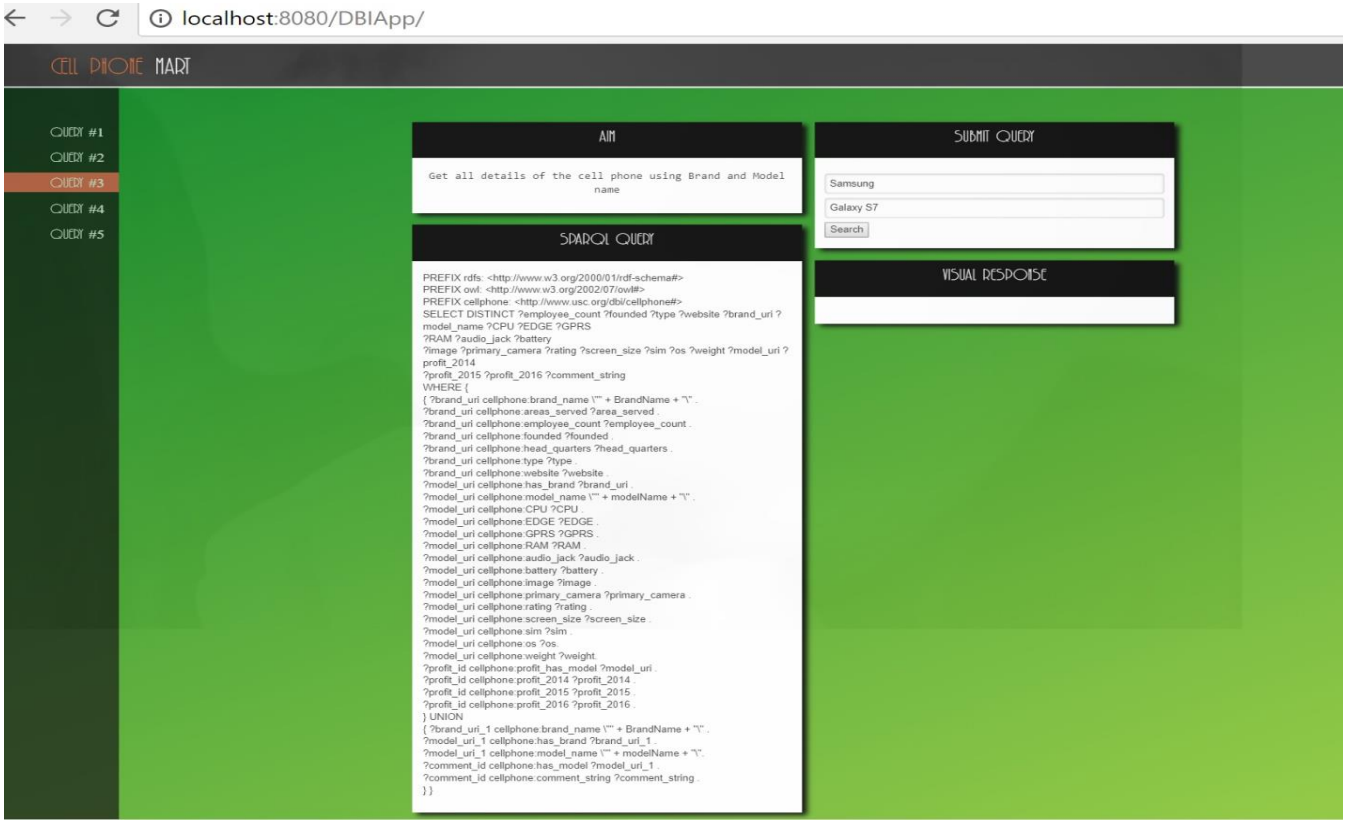
1.Get the revenue of top N models with highest user rating:



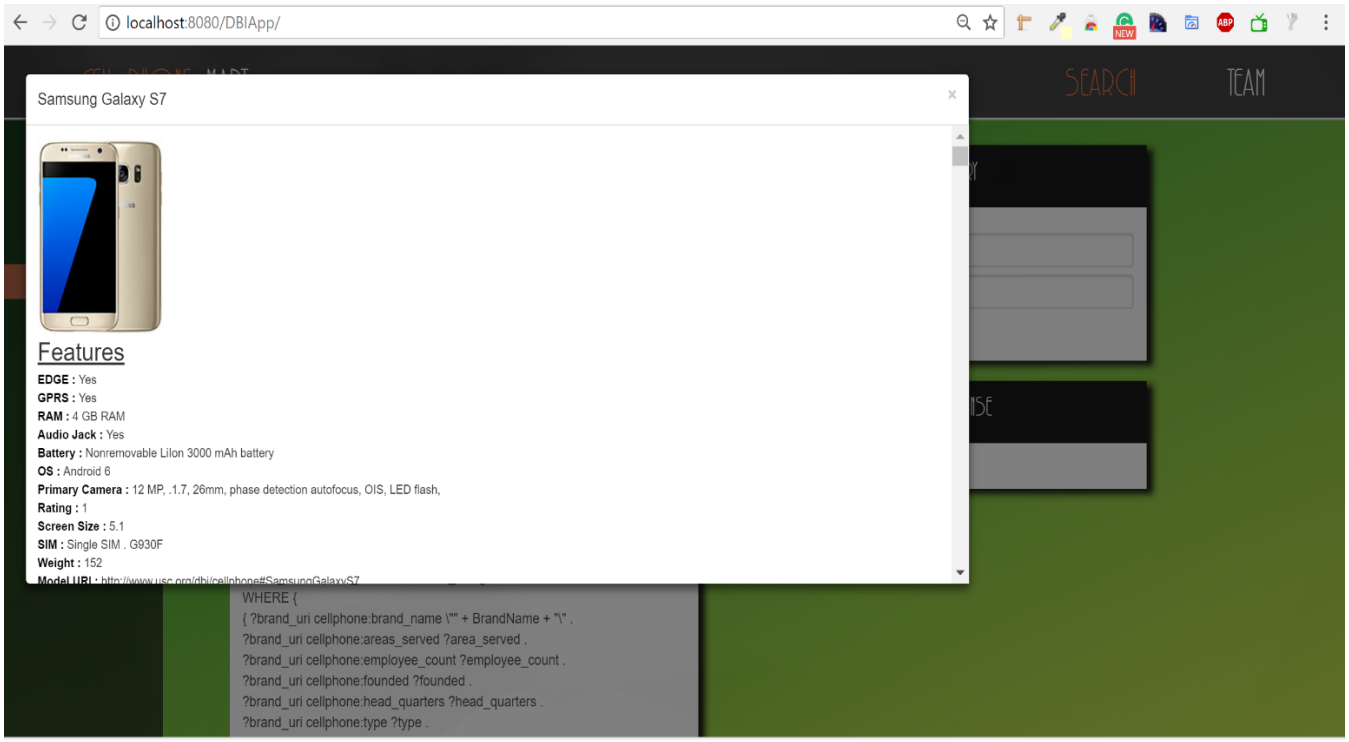
2. Get top N models based on their revenue:



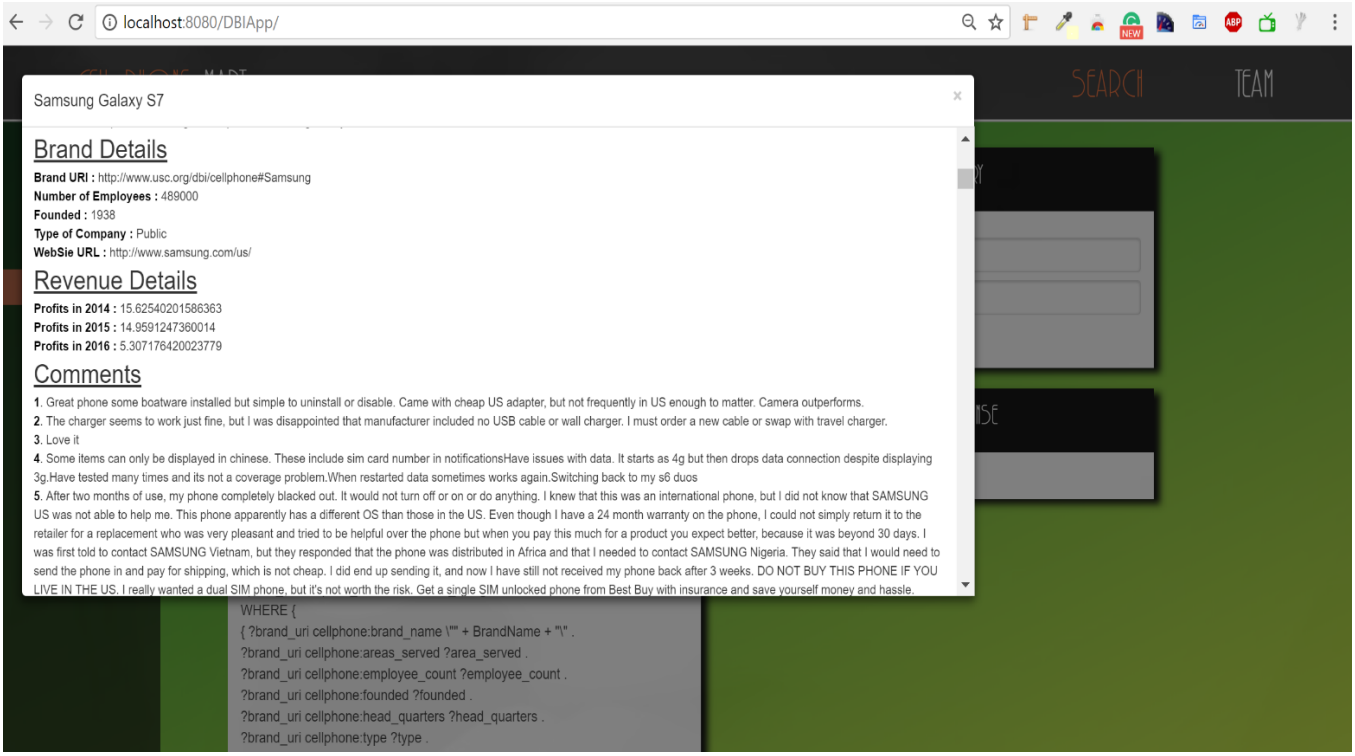
3. Get all the details of a cell phone based given brand and model:



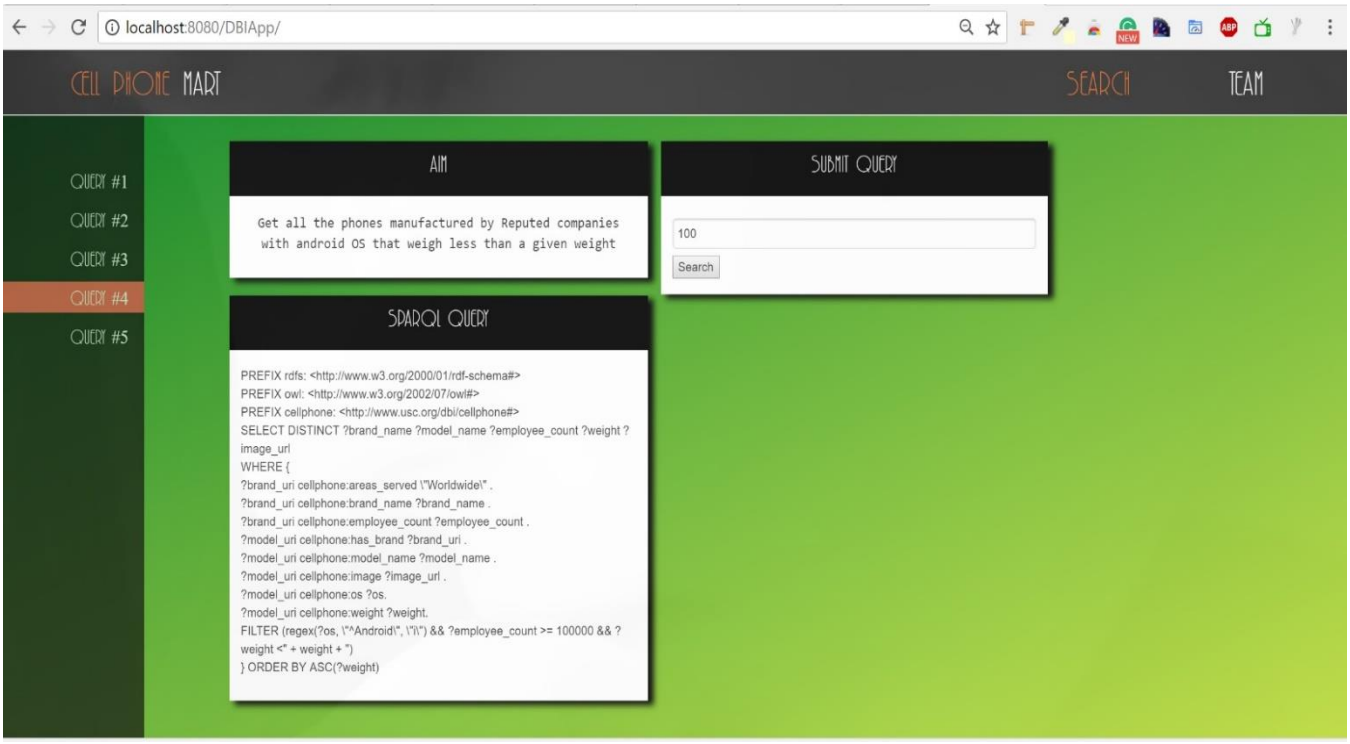
Device Specs as part of result:



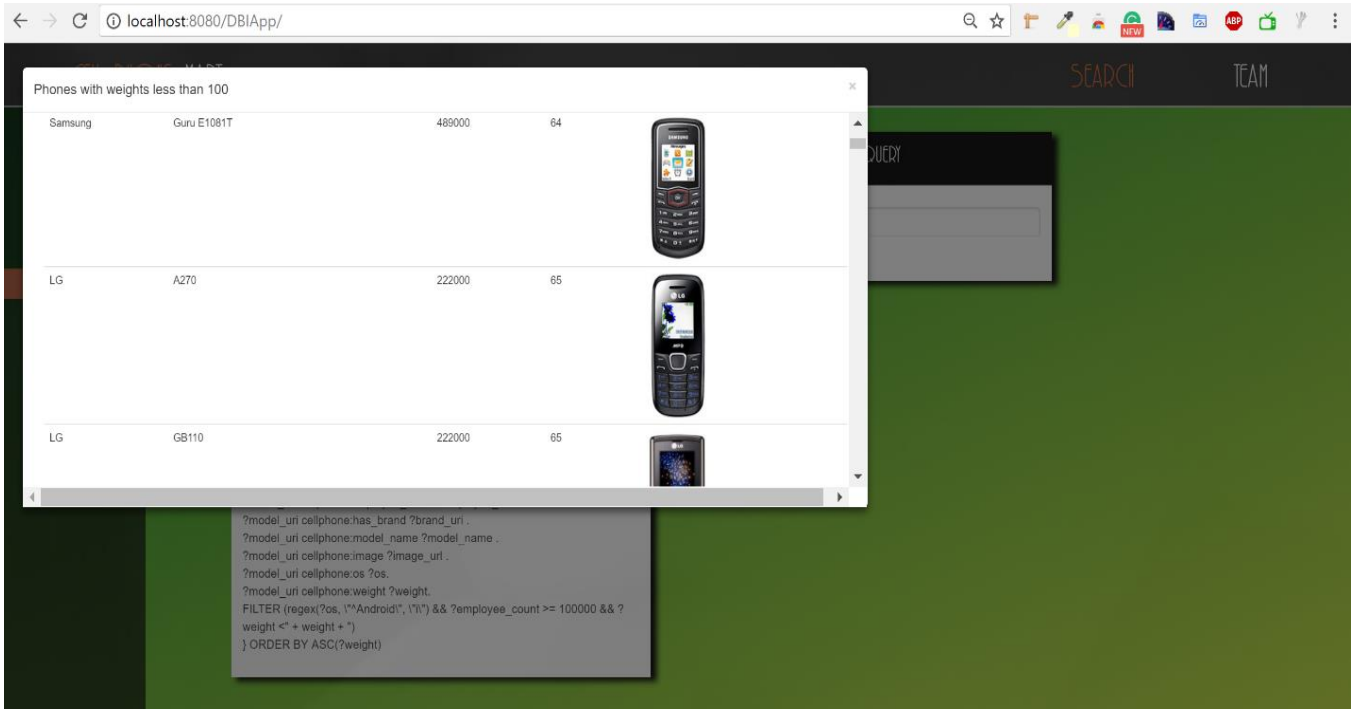
Brand, revenue, rating as part of results:



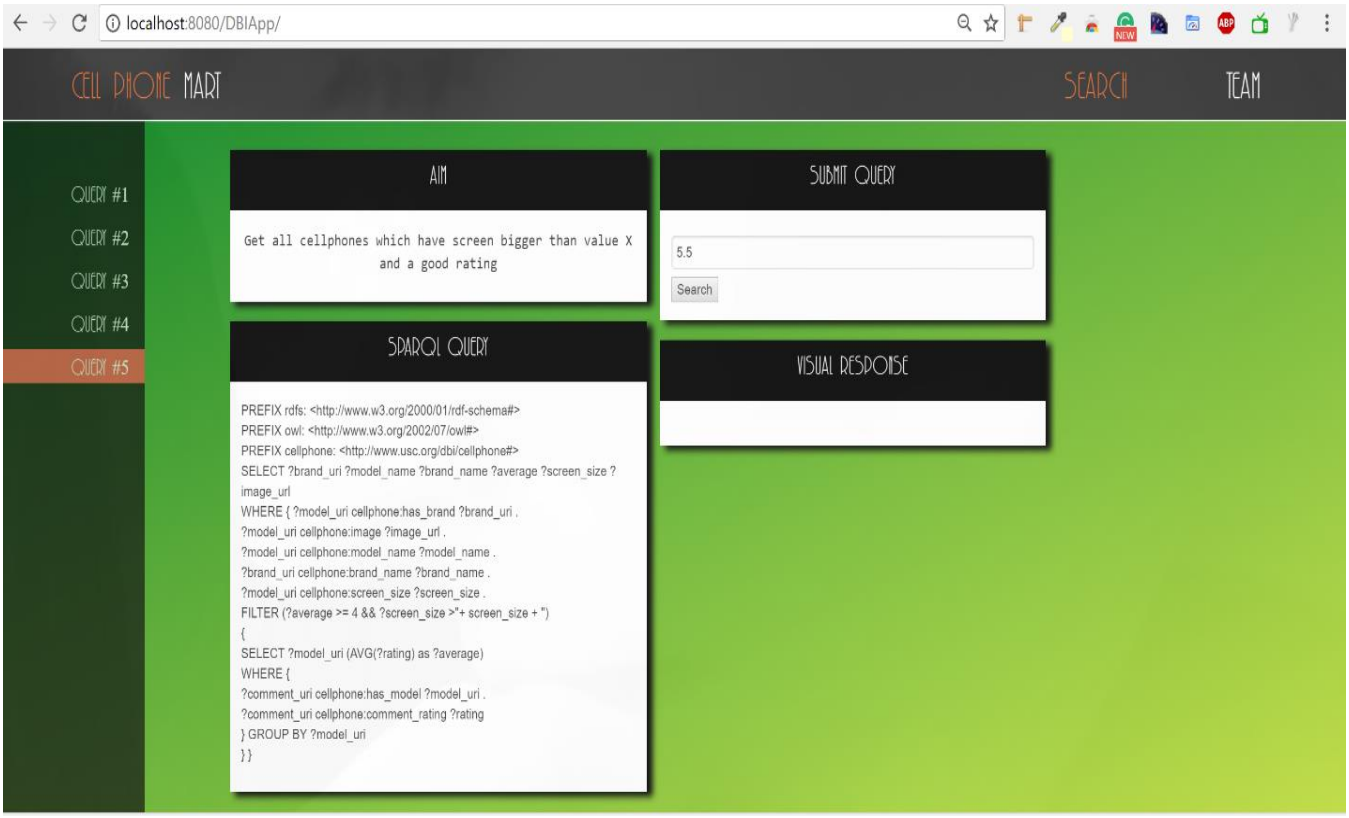
4. Get all the phones manufactured by reputed brands whose weight is less than a given weight:



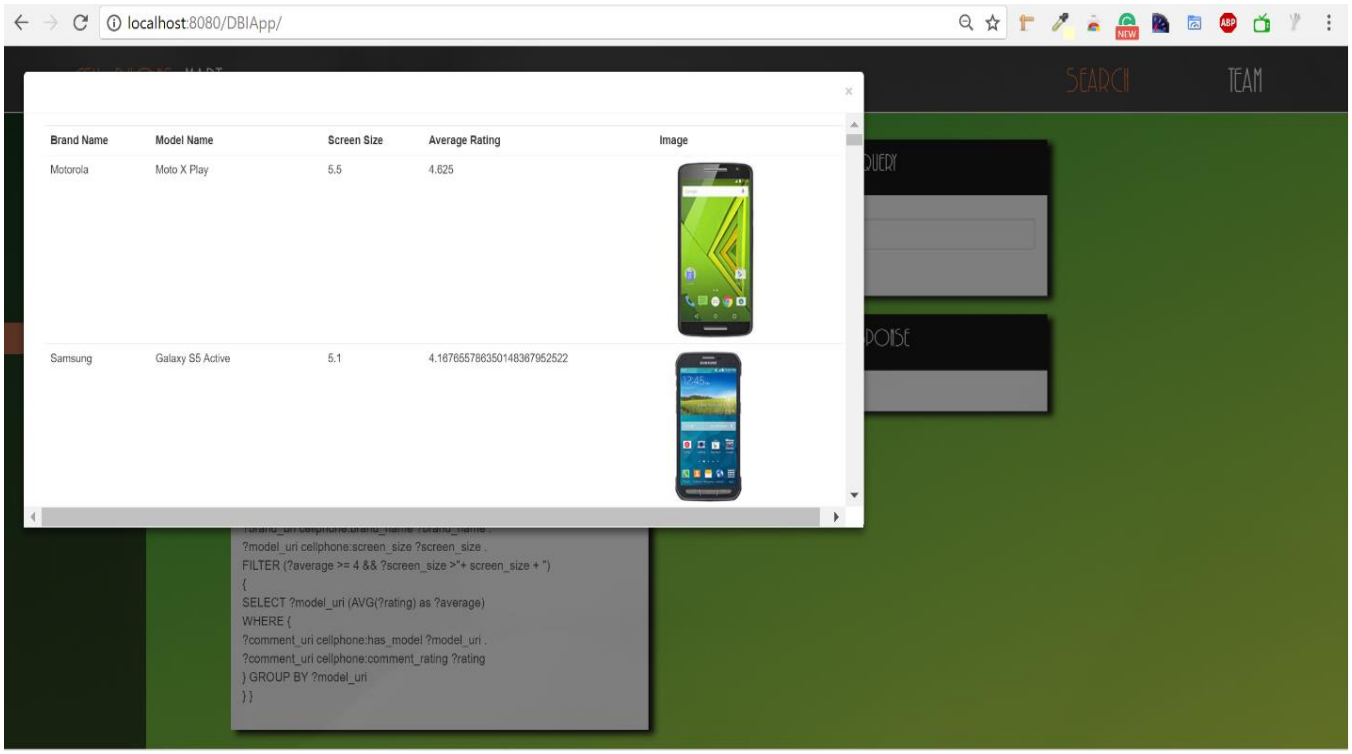
Result with all models along with weight:



5. Get all the cell phones with screen size greater than X and has best user rating:



Result has all the phones whose screen size is above X (5.5 inches) and with their average rating:



5. Conclusion & Future Work

Use of Ontology is progressively gaining popularity to unify data in various domains. In Cell Phone Mart, we showed one of its many uses in mobile phone industry by answering queries which would have required exploring multiple web pages to answer.

For future work, we could have an endpoint for more generalized queries for the user to key in. More datasets could be added to extend the current schema and the ontology for a better data management and more complicated queries. For instance, data related to stocks of a company. The current scope of Cell Phone Mart included a set of predefined queries covering a few key performance indices a user would be interested in; however, this could be extended with various analytical tools, and statistical concepts. We used high charts, a JavaScript charting library, in future high stocks could be used to show trends of various company stocks.

6. References

1. Integrating Applications on the Semantic Web, Hendler, James, Berners-Lee, Tim and Miller, Eric "Integrating Applications on the Semantic Web," Journal of the Institute of Electrical Engineers of Japan, Vol 122(10), October 2002
2. Integrated Movie Database - CSCI 586 Project Report by Muhammad Rizwan Saeed, Santhoshi Priyanka Gooty Agraharam, Ran Ao
3. https://www.youtube.com/watch?v=R9ERlUgvgwM
4. https://github.com/protegeproject/cellfie-plugin/wiki/Grocery-Tutorial
5. https://jena.apache.org/documentation/serving_data/
6. https://protegewiki.stanford.edu/wiki/Excel_Import