ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Amplified Semantic Parsing of Open Street Map using KNN

Poonam Devi [1], Rachna [2]

M.Tech. Student, Department of Computer Science & Engineering, N.G.F College of Engineering and Technology at Palwal, Haryana, India[1]

Assistant Professor, Department of Computer Science & Engineering, N.G.F College of Engineering and Technology at Palwal, Haryana, India[2]

ABSTRACT: The significance and inspirations driving Geographic Information Systems (GIS) extraction has been changing and creating well ordered, invigorated greater responsibility by the OSM. This improvement gave dynamic strategies for sharing besides, preparing data by swarm sourcing, for instance, Open Street Map, moreover called "the Wikipedia of maps" by a couple of experts. Right when swarm sourcing assembles massive data which withhold in corpus information, with help of general regional locale with fluctuating level of mapping foundation, the grouping of this plan should be on separating the data instead of social event it, utilizing semantic parser we give the plan to question Open Street Map data by taking a gander at it with prohibitive data or data of regulatory guide workplaces and overviews the investigation work for evaluation of Open Street Map and moreover discusses the future orientation utilizing Machine Discernible Dialect. In this manner, in this plan we propose to utilize the corpus information more than 500 MB from maps to separate an exact semantic structure that will construct the premise of a characteristic dialect interface to Open Street Map. Moreover, we utilize reaction construct learning with respect to parser results to adjust a measurable machine interpretation framework for social database access to Open Street Map. Our structure permits to outline common dialect articulations, for example, "adjacent", "north of", or "in strolling separation" to spatial polygons on an intuitive guide. Besides, it joins syntactic many-sided quality and compositionality with a sensible lexical fluctuation of questions, making it a fascinating new freely accessible dataset for look into on semantic parsing.

KEYWORDS: Open Street Maps (OSM), Geographic Information Systems (GIS), Machine Readable Language, Semantic Parsing, K Nearest Neighbours (KNN).

I. INTRODUCTION

Open Street Map (OSM) is a group constructed database of geographic information, containing client contributed neighbourhood and state-of-the-art data about historic points everywhere throughout the world. While the primary Programming interface is improved for altering map information, there exists a Programming interface that permit to channel delineate in view of inquiry criteria, for example, area, sort of articles, or highlights with which objects are labelled. In any case, issuing a question that is executable against the OSM database still requires nitty gritty information of database internals. something that can't be normal from a layman client. The objective of our work is the improvement of an interface to OSM that gives a client a chance to make an inquiry in normal dialect, which is then parsed into a database question that is executable against an online separating apparatus and profits OSM information for an intelligent guide. A client with less information of OSM ought to have the capacity to make inquiries that grasp the "fluffiness" of normal dialect, for instance, "What are the areas, names and phone quantities of lodgings in Paris with wheelchair get to that are near the station Gare du Nord?". To discover such data one would need to issue a question that requires point by point information of the database and the inquiry dialect:

“location[place=‘Bombay’]→.a;node(area.a)
[vide=‘cinema’]→.b;node(around.b:2000)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

[entertainment='cinema '][wheelchair='yes'];out;".

As a beginning stage for a characteristic dialect interface we assembled a corpus of 2,380 common dialect inquiries combined with Machine Readable Language (MRL) formulae that we used to separate data utilizing a xml parser. We decide to physically making a corpus of OSMs from which structure and weights of a xml parser can be educated for a few reasons. We will give the OSM people group an arrangement of test addresses that can be executed and whose database question portrayal can be examined, beneath delineates similar information of OSM corpus.

```
<?xml version="1.0" encoding="UTF-8"?>
 <osm version="0.6"/>
        <node id="1" lat="21.4219827" lon="39.8336534">
        <tag k="traffic" v="light"/>
        </node>
        <node id="2" lat="21.4221823" lon="39.8331833">
                <tag k="highway" v="motorway_junction"/>
        </node>
        . . .
        <way id="6">
                <nd ref="1"/>
                <nd ref="2"/>
                <tag k="highway" v="service"/>
        </way>
        <way id="8">
                <nd ref="4"/>
                <nd ref="5"/>
        <tag k="type" v="multipolygon"/>
        </way>
        . . .
        <relation id="2">
                <member type="relation" ref="1" role="inner"/>
                <member type="way" ref="6" role="inner"/>
        <tag k="highway" v="primary"/>
        </relation>
        . . .
</osm>
```

Figure 1: Extensible Mark-up Language OSM Corpus data.

However, the determination and facilitation of Open Street Map (OSM) clients and engineers to perceive how the complex geological actualities can be issued as basic normal dialect questions that are parsed into executable channels on Open Street Maps objects.

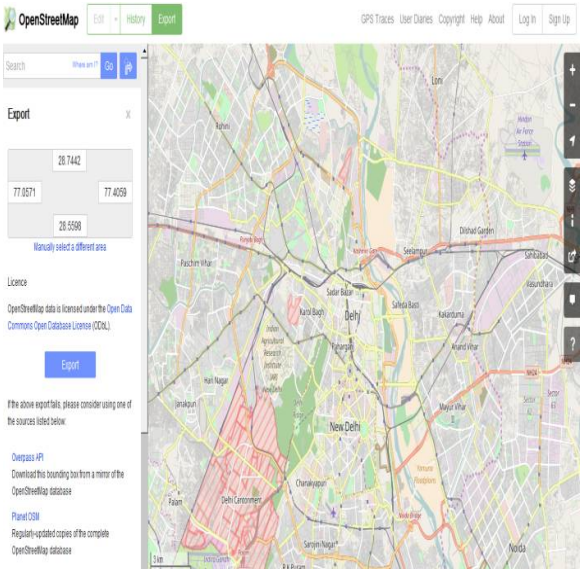

Figure 2: Example OSM Map from Mapzen.com

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

Under the scheme, we inculcate the plan and usage of a parsing and translation system for the Extensible Markup Language and the Rich Data Formats. These incorporate high flexibility in understanding of markup and meta-information, from one perspective, and proficient semantic content parsing is the place of area to extract and espionage the accurate information, on the other hand parsing of Extensible Markup Language content does not should be exceedingly adaptable, but rather it must be productive and dependable. A few quick parsers for Extensible Markup Language exists which can be reused as off-the-rack parts. We chose to re-utilize the semantic parser via KNN technique. In the structure exhibited here, a question situated Extensible Markup Language parser wraps parser the generic xml parsers. It shapes a reflection and incorporates procedural parsers executed in different machine learning models. We are utilizing Semantic Parse using KNN to join the interface of Expat parsers used earlier. The strategy is to arrange the sets and configuration alternative, parse summons the parsing of Extensible Markup Language content, and reset tidies up the semantic parsers. Extensible Markup Language has a comparable interface to parse semantically and can be connected to the framework by indicating a solitary configurable parameters.

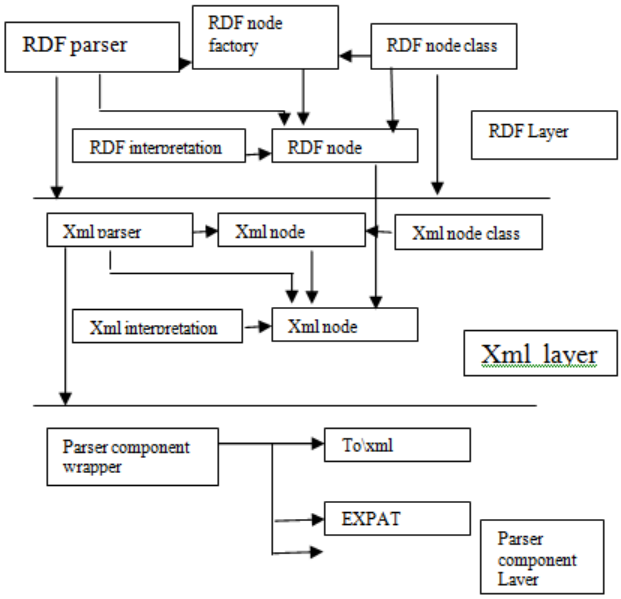

Figure 3: Rich Data Format and Extensible Mark-up Parsers

Open Street Map entire world dataset is free and available as a XML 700 GB document called MapZen.osm, refreshed on week by week premise. Figure 1 gives a bit of the MapZen.osm Extensible Mark-up Language document. The document comprises of the accompanying three rudimentary information writes: Open Street Map entire world dataset is free and available as a XML 700 GB document called MapZen.osm, refreshed on week by week premise. Figure 1 gives a bit of the MapZen.osm Extensible Mark-up Language document. The document comprises of the accompanying three rudimentary information writes:

(1) Element-Node, which is distinct because the position in the space related with a hub identifier, scope and longitude with latitude organizes the navigational model.

(2) Element-Way, which speaks to a line between two hubs, and related with the way identifier and the two hubs identifiers of the two end purposes of the line. The line could be basically a street, a shopping center, a stopping, city/nation limit, or part of a lake shape etc.

(3) Connection, which speaks to the relations between element-nodes, element-ways, or even other scalar relations and is utilized to express KNN measures. For instance, to express the limits of a specific lake, the element-nodes should be characterized, at that point the ways that interface nodes to each other, at that point a connection or relation that associates the routes together to express the lake limit with the flow direction. As the dataset is contributed by various volunteers, one lake might be communicated in different relations that could be isolated or settled (e.g., a connection inside a connection), where every connection is made out of either element-ways or element- nodes. For

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

instance, one lake is made out of two separate relations A and B. Connection C incorporates an arrangement of ways that frame some portion of the lake, while connection A is formed on one way and a connection B. At that point, connection C is made out of an arrangement of ways that shape other piece of the lake. Together, relations A, B , and C frame the entire state of the lake. Each of the three crude information writes element-node, element-way, and connection is related with an arrangement of labels. A tag is essentially a (key, esteem) combine that gives additional data about the crude information compose. Not at all like the three essential natives, labels are not predefined, where volunteers can include new labels and adjust existing one. This makes OSM information boisterous with non-standard labels. In conspire we have encountered a few labels that are confounded by volunteers. For instance, a few volunteers included a skyway as street arrange edge and the other way around.

II. RELATED WORK

Graham et al.,[1] Slowly but surely OSM is gaining popularity in these countries. Perhaps OpenStreetMap is helping address the participation inequality that is strongly represented in many types of User Generated Content on the Internet today. This “Digital Divide” indicates that very small groups with specific demographic and geographical characteristics are responsible for production of most of the UGC we see on the Internet today. However, these map visualizations indicate that OSM is reaching into countries and regions which heretofore would have felt the consequences of the digital divide. Improvements in ICT infrastructures and Information and Technology education for socially deprived groups such as women and children coupled with more ubiquitous access to smart phone technology has provided an environment where participation in OSM can increase. Research will need to be undertaken to gain a better understanding of the social processes involved in these changes.

[2] OSM has its own geology crosswise over time and space. At the end of the day, seldom observe indistinguishable examples of commitments in two distinct locales/nations. When talking about OSM quality and commitments arranges the significance of concentrate different contextual investigations has been featured. Thus, in this area, two distinct maps are produced from the Open Street Map insights, which exhibit the heterogeneity of Open Street Map in various nations. This guide shows a topical arrangement of made hubs, which is one of the key components in estimating Open Street Map commitments. It ought to be noticed that in this similar report, the span of the nation, populace, total national output Gross Domestic Product and various other physical qualities of the nations are not contemplated. In any case, they are of incredible significance in performing further top to bottom investigation.

[3] This paper thinks about the combined total closest neighbor(MANN) inquiry. Weiwei sun build up a calculation for handling this question, the Quick Pruning calculation. It utilizes the Euclidean total separation between an objective point and the question set as the pruning separation to prune away superfluous target focuses, the analysis comes about demonstrate that it can dispose of a significant piece of target focuses which thusly spare the execution time and I/O cost., The ideal area inquiry issue in view of street systems. unambiguous to a street arrange on which a few customers and servers are found. Every customer find out the server that is nearest to her for administration and her cost of getting served is equivalent to the(network) separate between the customer and the server serving her duplicated by her weight or significance. The ideal area inquiry issue is to find an area for setting up another server with the end goal that the greatest cost of customers being served by the servers (including the new server) is limited. This issue has been contemplated some time recently, however the cutting edge is as yet not sufficiently proficient. In this paper, creator propose an effective calculation for the ideal area question issue, which depends on an original thought of closest area part. They additionally talk about three augmentations of the ideal area question issue, in particular the ideal different area inquiry issue, the ideal area inquiry issue on 3D street systems, and the ideal area question issue with another target. Broad examinations were directed which demonstrated that our calculations are speedier than the best in class by no less than a request of extent on vast genuine benchmark datasets. For instance, on our biggest genuine datasets, the cutting edge kept running for over 10 hours however our calculation kept running inside 3 minutes just (i.e., >=400 times quicker).

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

III. PROPOSED SYSTEM

As discussed and depicted in section 1 the proposed scheme architecture is elaborated below where as the Open Street Module file is downloaded from map-zen or open street map for the experiment and results however, the architecture or workflow is depicted below:-

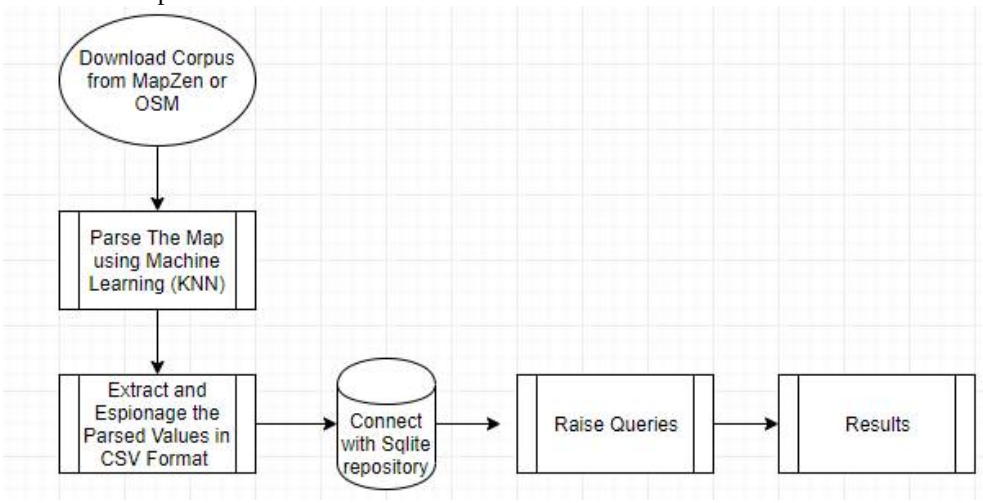

Figure 4: Proposed Architecture of the Amplified Semantic Parsing of Open Street Map using KNN

Algorithm using python herewith for explanation :-

```
import xml.etree.ElementTree as ET  # Use cElementTree or lxml if too slow
OSM_FILE = "/ML/new-delhi_india.osm"
SAMPLE_FILE = "/ML/new-delhi_sample.osm"
k = 100 # Parameter: take every k-th top level element
def get_element(osm_file, tags=('node', 'way', 'relation')):
    context = iter(ET.iterparse(osm_file, events=('start', 'end')))
    _, root = next(context)
    for event, elem in context:
        if event == 'end' and elem.tag in tags:
            yield elem
            root.clear()
with open(SAMPLE_FILE, 'wb') as output:
    output.write('<?xml version="1.0" encoding="UTF-8"?>\n')
    output.write('<osm>\n  ')
    # Write every kth top level element
    for i, element in enumerate(get_element(OSM_FILE)):
        if i % k == 0:
            output.write(ET.tostring(element, encoding='utf-8'))
    output.write('</osm>')
def count_tags(filename):
    tags = {}
    for event, elem in ET.iterparse(filename):
        if elem.tag in tags:
            tags[elem.tag] += 1
        else:
            tags[elem.tag] = 1
```

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

```
return tags
pprint.pprint(count_tags(OSMFILE))
```

IV. PROPOSED EXPERIMENTAL WORK

Open Street Map is presented as another information base that has not, to the best of our insight, been utilized for question replying, and offer another corpus to the exploration group. Work assembles the premise of a characteristic dialect interface to Open Street Map that will empower for fascinating bearings of future research, e.g., reaction based figuring out how to enhance parsing and multilingual database get to with efficient and accurate results.

1. Information Espionage: Separating information from OSM isn't a minor assignment. The whole OSM dataset is kept back to back in one monster volume go into a semi-organized Extensible Mark-up Language arrangement. The Extensible Mark-up Language record begins with root-nodes(hubs), at that point routes, and finally, relations, while labels are settled in everything about data sorts. A principle challenge in extricating information from OSM is recognizing the comments (i.e., labels or Elements) that infer arranged geo-spatial data. The Information Extraction module take the uniform asset locator of the packed mapzen.osm record as partner degree info, and yields many arranged documents. Each computer document contains a steady spatial arrangement of information, e.g., silver screen corridor detail and stopping office. We will talk about each extraction work exclusively in light of circular segment extraction, beneath calculation delineates the usual way of doing things for increase.

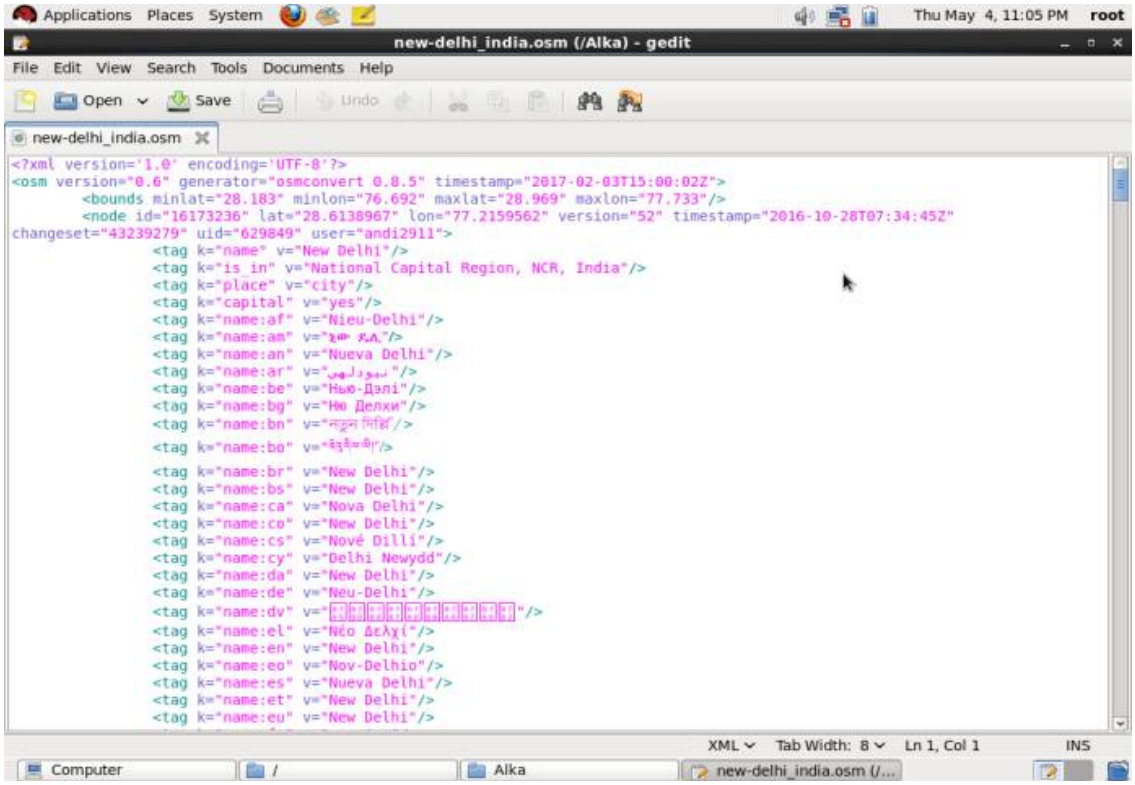

Figure 5: OSM Raw Information

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

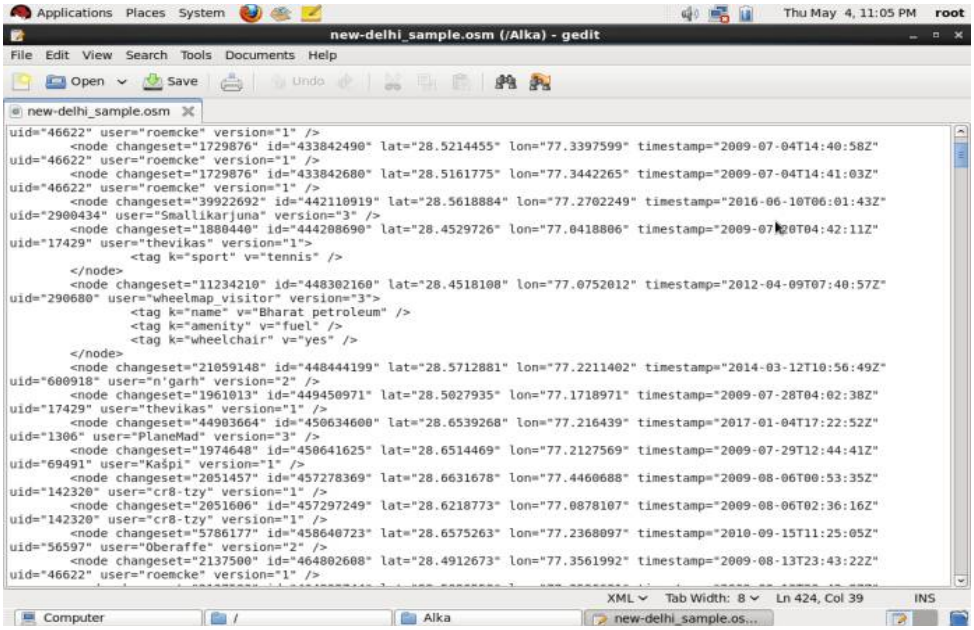

Figure 6: Espionage Information

2. Processing or Parsing: The geo-spatial corpus creation process was guided by the objective to match an assorted scope of inquiries with MRL formulae. These ought to incorporate the most vital Open Street Map labels or elements with the goal that the parser can take in a mapping between these labels and the distinctive comparing normal dialect articulations. It is unambiguously characterized by means of a setting free language structure so one can simply determine regardless of whether a recipe is substantial. While the composed type of the Machine Readable Language is a section structure, this structure can without much of a stretch be encoded as a tree by taking a pre-arrange traversal which makes it simple and productive to work with.

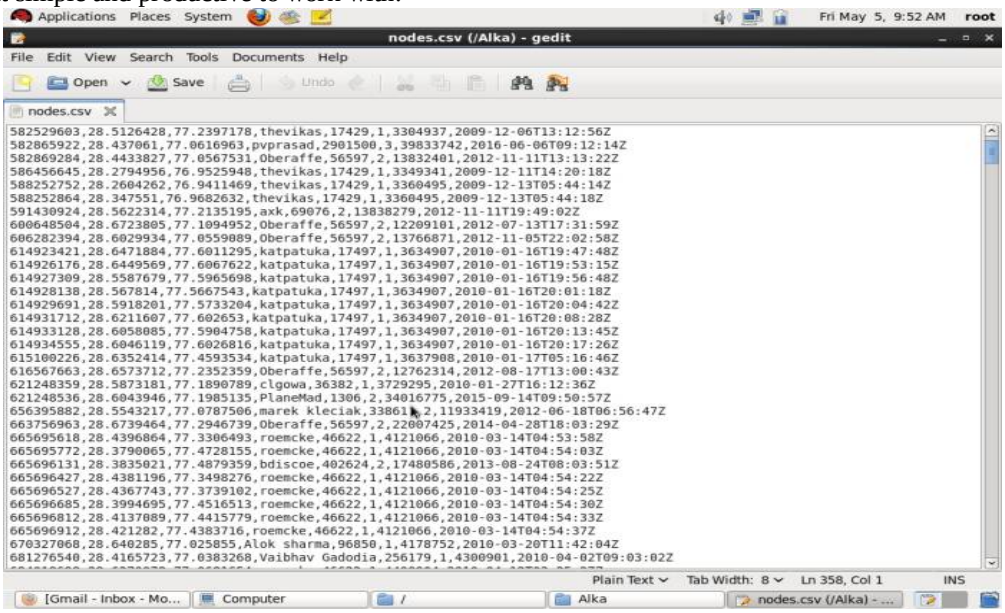

Figure 7: Information vide Nodes and Ways

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

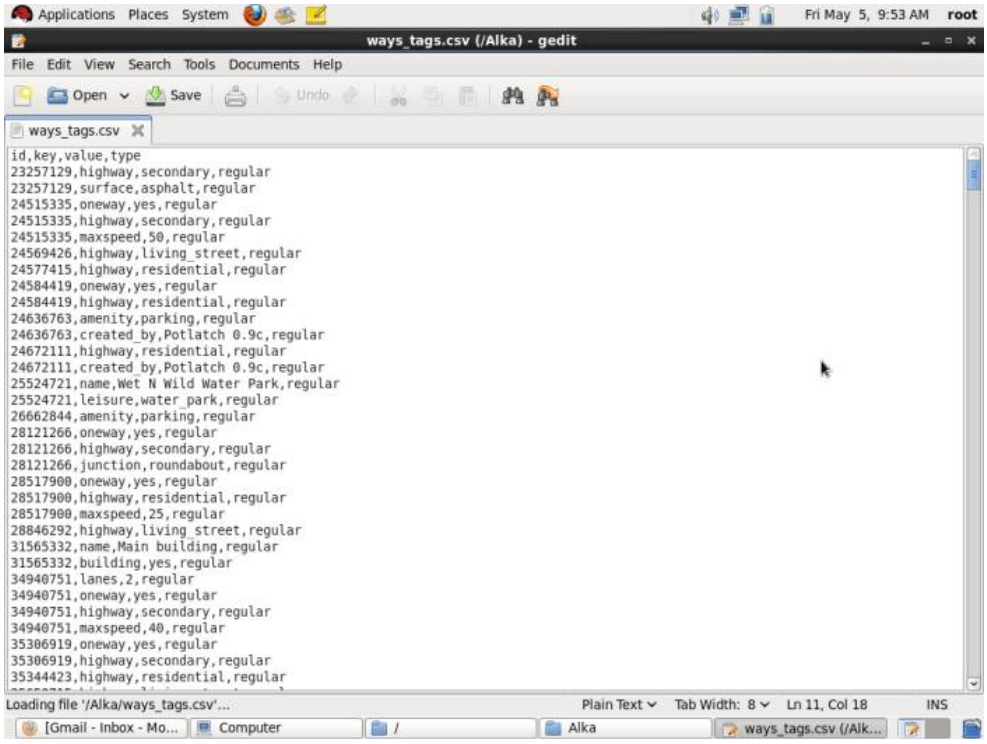

Figure 8: Information from Tags, Ways & Their Relations

3. Data Extraction & Establishment: Since even the Bridge Inquiry Wizard expects clients to be comfortable with the label set of Open Street Map key-esteem sets, it unusable for clients with just easygoing or no learning of Open Street Maps inside structure. In any case, we could utilize parts of the client inquiry log to detail common dialect questions. For instance clients would enter the inquiry "Which shopping center in delhi have wheelchair access?" with Machine Readable Language :-

"query(area(keyval(name,'bombay')),nwr(keyval(entertainment,'cinema'),keyval(utilitychair,'yes')), findkey(location))" can be exhibited as a tree. A preorder traversal gives: "query@3 area@1 keyval@2 name@0bombay nwr@2 keyval@2 tourism@0 mall@s keyval@2 wheelchair@0 yes@s findkey@1 name@0"

4. Question Administration: A solitary database inquiry is encoded in the administrator questioning() which will hold the Bridge question and in addition facilitate determinations about what sort of answer ought to be recovered. A couple of administrators are straightforwardly gotten from Bridge, just re-composed as a tree structure. In that capacity Open Street Map key-esteem sets are encoded utilizing the administrator keyval() which takes two contentions, the first being the key and the second the esteem. The region administrator from Bridge specifically means the administrator region. Hubs or nodes, ways and relations are gathered together under the nwr_() administrator which will supply the union of the inquiry keep running with the 3 composes thus. This is fundamental in light of the fact that regularly structures, e.g. schools, might be spoken to as any of the 3 writes relying upon how particular the annotator needed to be. Both region_() and nwr_() at that point take one mineral more keyval_() contentions. In the event that territory_() and nwr() show up as kin in the tree (for a case see Figure 2), at that point just the items that exist in zone() will be sought to decide whether they satisfy the nwr()_ imperatives.

5. Various pattern Operators**:** : Some further pattern based operators were expected to show the Machine Readable Language equation for complex inquiries. also, as it would utilized when the client requests two distinct pieces of data ("Whereas he closest shopping center and the nearest stopping?" or "Give me the site and name of ..."). or then again() is utilized to make unions, as, required in a sentence, for example, "Give me the nearest shopping center or eatery." "*"

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

can be utilized as a trump card in an esteem position, e.g. ['parking'=*] will restored any noteworthy items, be it a stronghold, a landmark or something different. nodup() restores a set without any copies.

V. PSEUDO CODE & RESULTS

Amplified Semantic parsing algorithm using KNN :

Step 1. Download Open Street Maps vide MapZen repository.
Step 2. Enticing the Mammoth Raw Information termed as corpus
Step 3. Traversing inside the OSM Data using semantic model.
Step 4..Referencing the OSM Corpus using semantic parsing
Step 5. Figure out elements, ways, lays, attributes and relations.
Step 6. Integrating and KNN Algorithm
Step 7. Refurbish OSM data to CSV raw data.
Step 8. Binding CSV with Object Relational Models
Step 9. Forming and Integrating the Query Plan
Step 10. Espionage Information as per Query Plan
Step 11. Display Results.

Pseudo Code:

Believe k as the preferred number of adjoining neighbors and S:=$p_1,...,p_n$ be the set of preparation samples in the form $p_1=(x_i,c_i)$, where x_i is the d-dimensional feature vector of the point *pi* and *ci* is the class that *pi* belongs to calculate the distance d(x′,xi) between p′ and all pi belonging to S
Sort all points pi according to the key d(x′,xi)
Select the first k points from the sorted list, those are the k closest training samples to p′
Assign a class to p′ based on majority vote: c′=argmaxy∑(xi,ci) belonging to S, I(y=ci)

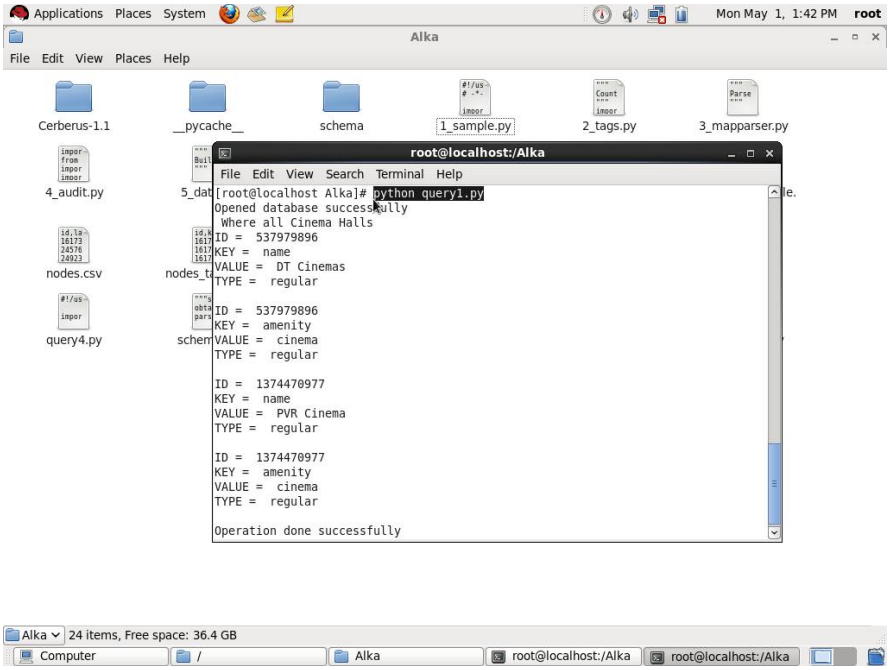

Figure:9 Query (Entertainment Units which provides wheelchair for differently able peoples)

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 5, May 2018

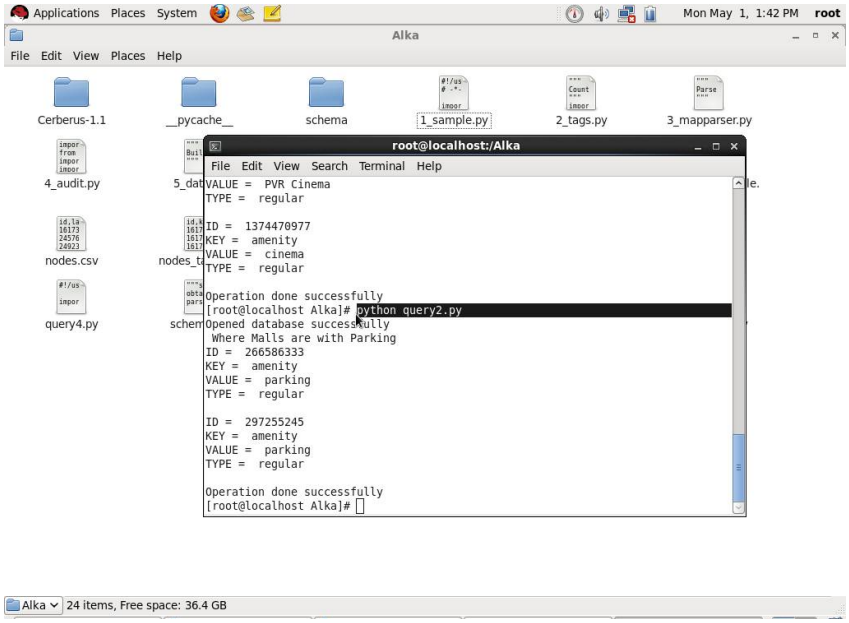

Figure: 10 Query (Facility that provides the vehicle parking in the premises)

VI. CONCLUSION AND FUTURE WORK

Under the scheme we developed an scenario to question the Open Street Maps catalog for multifaceted geological particulars through innate idiom incorporated using semantic models developed by KNN. The explanation expertise is a semantic models and parsing that is educated in administer approach from an extensive arrangement of inquiries commented on with executable Machine Readable Language. Our OSM query model is bigger than past clarified question-answer corpora, while including a wide assortment of testing questions. Terms, for example, "close-by", "in the south of ", "inside x miles" are especially appropriate for a characteristic dialect inquiry interface that permits to delineate fluffiness of normal dialect to adaptable geo-spatial information with KNN technique.

REFERENCES

1. Graham M, Hale S, Stephens M (2012) Digital divide: the geography of internet access. Environ Plan A 44:1009–1010
2. Bakillah M, Liang S, Mobasheri A, Jokar Arsanjani J, Zipf A (2014a) Fine-resolution population mapping using OpenStreetMap points-of-interest. Int J Geogr Inf Sci 28:1940–1963
3. Weiwei Zhang. *. , Anping Liao, Shu Peng, Xinyan Zheng, Ming Li. National Geomatics Center of China, 28 Lianhuachi West Road, Beijing 100830, PR China;.
4. Bakillah M, Andrés Domínguez J, Zipf A, Liang SHL, Mostafavi MA (2013) Multi-agent evacuation simulation data model with social considerations for disaster management context. In: Zlatanova S, Peters R, Dilo A, Scholten H (eds) Intelligent systems for crisis management, SE-1. Springer, Berlin, Heidelberg, pp 3–16. doi:10.1007/978-3-642-33218-0_1
5. Bakillah M, Liang S, Mobasheri A, Jokar Arsanjani J, Zipf A (2014a) Fine-resolution population mapping using OpenStreetMap points-of-interest. Int J Geogr Inf Sci 28:1940–1963
6. Bakillah M, Lauer J, Liang S, Zipf A, Jokar Arsanjani J, Loos L, Mobasheri A (2014b) Exploiting big VGI to improve routing and navigation services. In: Big data techniques and technologies in geoinformatics, pp 177–192
7. Fan H, Zipf A, Fu Q, Neis P (2014) Quality assessment for building footprints data onOpenStreetMap. Int J Geogr Inf Sci 28:700–719. doi: 10.1080/13658816.2013.867495
8. Goetz M (2013) Towards generating highly detailed 3D CityGML models from OpenStreetMap.Int J Geogr Inf Sci 27:845–865
9. Goetz M, Zipf A (2012) Using crowdsourced indoor geodata for agent-based indoor evacuation simulations. ISPRS Int J Geo-Inf 1(2):186–208. doi:10.3390/ijgi1020186
10. Hagenauer J, Helbich M (2012) Mining Urban land-use patterns from volunteered geographic information by means of genetic algorithms and artificial neural networks. Int J Geogr Inf Sci 26:963–982
11. Helbich M, Amelunxen C, Neis P (2012) Comparative spatial analysis of positional accuracy of OpenStreetMap and proprietary geodata. In: Jekel T, Car A, Strobl J, Griesebner G (eds) Geospatial crossroads GI_Forum 2012, Wichmann, Heidelberg, Salzburg, pp 24–33