STM JOURNALS

Journal of Artificial Intelligence Research & Advances
ISSN: 2395-6720 (Online)
Volume 4, Issue 2
www.stmjournals.com

Unstructured Data Analysis with the Use of Cognitive Capabilities of Watson Analytics

Sindhu B. Jigali**, Nirmala C.R.
Department of Computer Science and Engineering, Bapuji Institute of Engineering and Technology, Davanagere, Karnataka, India

Abstract

Intellectual science concentrates on how individuals investigate for data when connecting with the web. The exploratory hunt refines search exercises and updates the inward portrayals of the inquiry issues. Exploratory data recovery frequently includes badly characterized search objectives and advancing criteria for assessment of pertinence. The associations amongst people and the data framework include more intellectual movement, and frameworks that bolster exploratory pursuit along these lines, need to consider the subjective complexities required amid the dynamic data recovery prepare. Building adaptable psychological model to investigate unstructured information is testing process. IBM's Bulemix cloud Watson administrations help individuals to fabricate extreme intellectual space. It fabricates psychological, cloud-based investigation applications by opening significant bits of knowledge covered up in unstructured information, including restrictive information, and also open information. The proposed application uses the utilization of Watson's Discovery administration to concentrate later and important data of an organization by questioning in common dialect.

Keywords: *Cognitive computing, Watson analytics, discovery service, natural language processing*

****Author for Correspondence*** E-mail: sindhubjigali94@gmail.com

INTRODUCTION

Cognitive models of data recovery lay on the blend of zones, for example, intellectual science, human-PC communication, data recovery, and library science. These models portray the connection between a man's intellectual model of the data looked for and the association of this data in a data framework. These models endeavor to see how a man is hunting down data so that the database and the inquiry of this database can be outlined so as to best serve the client. Data recovery may consolidate numerous errands and psychological issues, especially on the grounds that diverse individuals may have distinctive strategies for endeavoring to discover this data and anticipate that the data will be in various structures [1]. Subjective models of data recovery may endeavor at something as evidently dull as enhancing indexed lists or might be something more intricate, for example, endeavoring to make a database which can be questioned with natural language search.

Features of Cognitive Systems (Figure 1)

Adaptive: Cognitive model learns as data changes happen, and as objectives and prerequisites develop. They may likewise resolve equivocalness and endure capriciousness. They can be designed to feast upon dynamic information continuously, or close constant.

Interactive: Cognitive models connect effortlessly with clients, so clients can characterize their requirements serenely. They likewise collaborate with different processors, gadgets, and cloud administrations, and with individuals.

Iterative and State Full: Cognitive models help in characterizing an issue by making inquiries or finding extra source input if an issue articulation is uncertain or fragmented. They may "recall" past communications in a procedure and return data that is reasonable for the particular application at that point in time.

Contextual: These models comprehend, distinguish and extricate relevant components; for example, mean language structure, time area, suitable space, directions, client's profile, process, errand and objective. They draw on different wellspring of data, including both organized and unstructured advanced data.

Cognitive computing depends on having the right data and structuring it so that the algorithms can operate effectively on it. The architecture of cognition and the theory of cognition outline the structure of the various parts of the mind and make commitments to the use of rules, associative networks, and other aspects [2]. The cognitive architecture implements the theory on computers. The software used to implement the cognitive architectures are also "cognitive architectures". Thus, cognitive architecture also refers to blueprint for intelligent agents. It proposes (artificial) computational processes that act like certain cognitive systems, most often, like a person, or acts intelligent under some definition. Cognitive architectures form a subset of general agent architectures. The term 'architecture' implies an approach that attempts to model behavior, along with structural properties of the modeled system [3].

Analytics is characterized as the disclosure, translation, and correspondence of significant examples in information. Particularly significant in territories which are rich with recorded information, analytics depends on the simultaneous application of measurements, PC programming and operations research to evaluate execution. Analytics regularly supports information representation to communicate insight. The analysis of unstructured information sorts is another test getting consideration in the business [4].

Watson Analytics: A New Era of Decision Making

Watson analytics is a savvy information disclosure device from IBM examination. Watson arrangements intend to upgrade, scale, and quicken human skill focusing on an extensive variety of complex arrangement. Three ways were distinguished to bring Watson into market; they are:

- Natural language processing.
- Hypothesis generation.
- Evidence based learning.

It uses the psychological capacities of Watson technology to quicken day by day logical exercises of businessmen in three ways.

- Provides semantic acknowledgment in information.
- Identifies the beginning stages for examination.
- Enables communication in common dialect.

Key elements of Watson investigation are:

- Natural language dialogue.
- Automated prescient examination.
- One-click examination.
- Smart information revelation.
- Simplified examination.
- Accessible progressed examination.

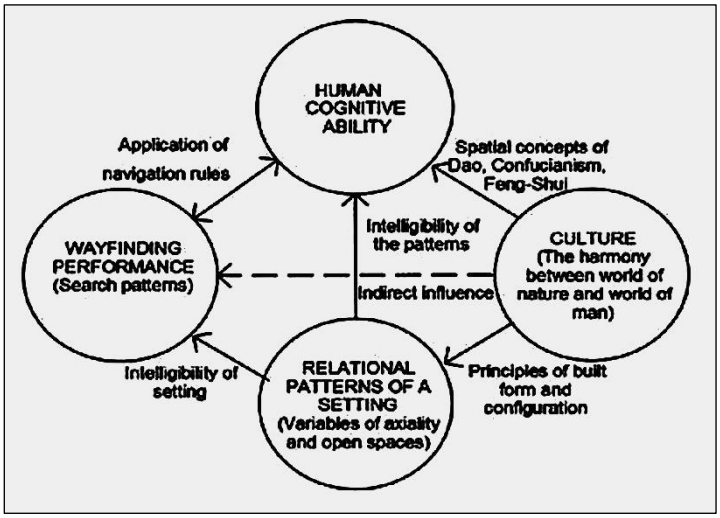

Fig. 1: *Cognitive Computing Reference Model.*

Watson uses IBM's DeepQA software and the Apache UIMA (Unstructured Information Management Architecture) system (Figure 2). The framework is composed in different languages, including Java, C++, and Prolog, and keeps running on the SUSE Linux Enterprise Server 11 operating system utilizing Apache Hadoop structure to give circulated registering [5].

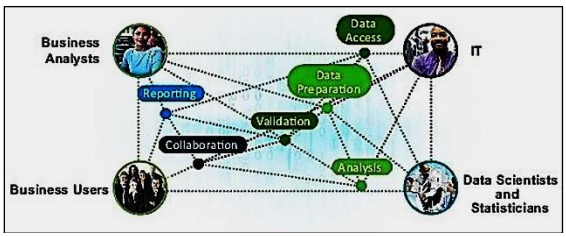

Fig. 2: *Analytics Made Easy by Watson Analytics.*

Watson services uses natural language processing and machine learning techniques**.** Many challenges in NLP involve: natural language understanding, enabling computers to derive meaning from human or natural language input; and others involve natural language generation [6]. NLP comprises all quantitative approaches to automated language processing, including probabilistic modeling, information theory, and linear algebra. The technology for statistical NLP involves learning from data.

The services provided by Watson analytics are: discovery, conversation, document conversion, language translator, natural language classifier, personality insights, retrieve and rank, speech to text, tone analyzer, visual recognition and trade off analytics [7]. The IBM Watson discovery service rapidly builds cognitive, cloud-based exploration applications that unlock the actionable insights hidden in unstructured data, including people's proprietary data, as well as public and third-party data.

It uses data analysis that is combined with cognitive intuition to take unstructured data and enrich it, so that the necessary information can be discovered. IBM Watson discovery brings together a functionally rich set of integrated, automated Watson APIs to:

- Crawl, convert, enrich and normalize data.
- Securely explore the proprietary content.
- Apply additional enrichments such as concepts, relations, and sentiment through natural language processing.

LITERATURE SURVEY

The earlier proposed methodologies and analytics-approaches, costs for the computing resources and the platforms, lag in drawing the semantics from the unstructured data. They do not provide visualization for the results of analytics which makes difficult in analyzing the outcome of analytics insights that have to be obtained from unstructured data. It is very difficult to build the ontology for analytics with the existing approaches because integration of various analytics processes and approaches is difficult with platform dependency and knowledge database cannot be easily built [8].

These drawbacks can be overcome by the use of Watson services for data analytics provided by IBM Bluemix cloud. Watson analytics is a collection of services which is built on Apache UIMA framework. It is a cloud based smart data discovery tool that learns from data using natural language processing, cognitive services, statistical algorithms and machine learning techniques etc. It provides services for cleaning, transforming, creation of conclusions and decision making on true meaning of the data [9]. It makes analysis easier with integrated services by accelerating the analytics in three ways:

1. Provides semantic recognition of concepts in source data.
2. Identifies starting points for analysis.
3. Enables interaction in natural language.

Watson analytics is able to analyze the data just in four steps:

1. Analysis of question.
2. Hypothesis generation.
3. Evidence scoring.
4. Final merging and ranking.

The services of Watson analytics can be easily used with the API_keys provided for each service. The services of the Watson analytics can be integrated with other services enable user to draw valuable insights from the data.

Problem Statement

The early proposed cognitive search engines were designed based on traditional library science theories related to retrieval basic facts and simple information through an interface. They lag in processing of queries that are posed in natural language. Query building to query the unstructured data is very difficult. They also lag in drawing sentiments form unstructured data.

Objective

The objective is to design an application that queries the unstructured data, by building and structuring queries using natural language (Figure 3).

ARCHITECTURE

The proposed application follows the architectural pattern of cognitive model (Figure 3). The application is associated with the credentials of discovery service and accepts the query in a natural language as input. The query is forwarded to the discovery service running on the Watson server through the interface. This query is embodied in request where the request consists of functionalities to be addressed. The discovery service running on Watson server prepares the unstructured data, enriches the query to make it point to information of interest and then insights are integrated to the response. The discovery service cleans and normalizes the data using NLP results and stores the indexes into a collection [10].

Further, the service builds queries based on the indexed data to draw actionable insights. This actionable insight would deliver better outcomes. This service can also be integrated with other Watson services like alchemy language to extract sentiments, entities like people, topic etc. and uses discovery news service, wherein it is a public data set that has been pre-enriched with cognitive insights.

This public, unstructured data set is also used to query for insights that can be integrated into applications. Watson discovery news is a dataset of primarily English language news sources that is updated continuously, with approximately 300,000 new articles and blogs added daily. The outputs are the actionable insights drawn by services for specific queries and are sent as response to the application. The output response includes entities, sentiments, recent and relevant information about the query [11].

IMPLEMENTATION AND PROTOTYPE

The following are the steps to be followed in designing the news application. It requires a service credential of the discovery service that is running on Bluemix cloud. Once the credential is integrated with the application, the application can make use of the service. The service processes the input content and provides the response to the query posed in natural language [12].

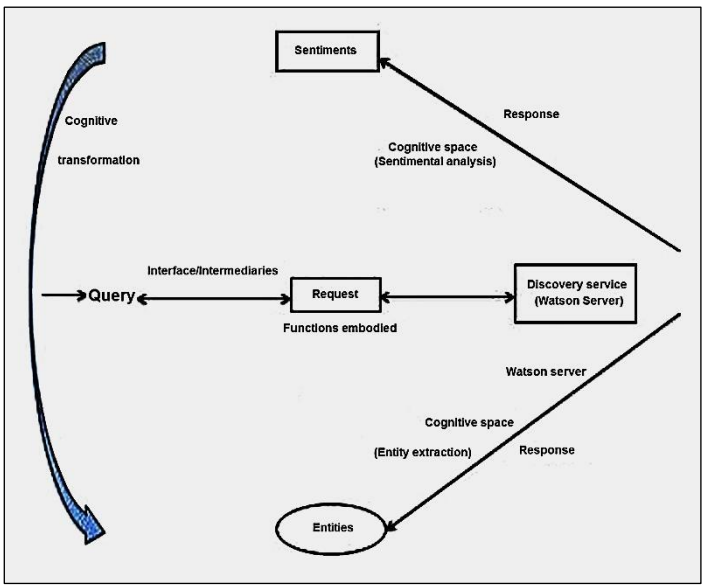

Fig. 3: *News Application Using Watson Analytics Discovery Service.*

Getting the Credential for Discovery Service

```
var express = require('express');
var router = express.Router();
var watson = require('watson-developer-
cloud');
var discovery = watson.Discovery({
api_key:
'81e32a4cc260b79a714c23f4f3e7b1a833332
e2f'});
```

The API key for the discovery service is added to the news application so that the application is able to use the service running on the Watson server.

Query Building

```
import React from 'react';
import Collapsed from './Collapsed.jsx';
import Expanded from './Expanded.jsx';
const Query=props=> (
(props.query && props.query.text) ?
<Collapsed query={props.query}
onQueryChange={props.onQueryChange} /> :
<Expanded
onQueryChange={props.onQueryChange} />
);
Query.propTypes={
query: React.PropTypes.shape({
text: React.PropTypes.string,
date: React.PropTypes.object,
dateRangeRestricted: React.PropTypes.bool,
}),
onQueryChange:
React.PropTypes.func.isRequired,
};export default Query;
```

Queries to Display Aggregations Used in the Application

```
const moment = require('moment');
const entities = [
'nested(enrichedTitle.entities).filter(enriched
Title.entities.type:
Company).term(enrichedTitle.entities.text)',
'nested(enrichedTitle.entities).filter(enriched
Title.entities.type:
Person).term(enrichedTitle.entities.text)','ter
m(enrichedTitle.concepts.text)',
];
const sentiments=[
'term(blekko.basedomain).term(docSentimen
t.type)',
'term(docSentiment.type)',
'min(docSentiment.score)',
'max(docSentiment.score)',
];const mentions=[
line'filter(enrichedTitle.entities.type::Compa
ny).term(enriched
Title.entities.text).timeslice(blekko.chrondate,
1day).term(doc Sentiment.type)'
];
module.exports={
aggregations: [].concat(entities, sentiments,
mentions),
entities,
sentiments,
mentions,
build(query, full){
const params={
count: 5,
return: 'title,enrichedTitle.text,url,host',
query: `"${query.text}",language:english`,
};
if (full){
params.aggregations=[].concat(entities,
sentiments, mentions);
}
if (query.date) {
params.filter=`blekko.hostrank>20,blekko.chr
ondate>${moment(query.date.
from).unix()},blekko.chrondate<${moment(q
uery.date.to).uni x()}`;
}       return params;
},
};
```

Query Processing

The query posed in natural language is sent to the service as a request. Discovery service employs natural language processing in order to perform ingestion where it crawls, converts and enriches the unstructured data, and it securely explores the proprietary content. This data is normalized and indexed to a collection, so that the query can understand the data faster, create hypothesis and deliver better insights. The service returns sentiments, entities of interest, relevant and recent information related to the posed query.

Response and Result

```
app.post('/api/query', (req, res, next)=>{
const params=queryBuilder.build(req.body);
discovery.query(params, (error, response) => {
if (error){
```

```
next(error);
} else{
res.json(response);
}
});
});
require('./config/error-handler')(app);
module.exports = app;
```

The response received by the application includes sentiments, entities of interest, relevant and recent information related to the posed query.

Software Requirements

1. Node.js
2. CLI-cloud foundry tool for Windows operating system.

RESULTS

Below are presented the results of the whole analysis in Figures 4–10.

Fig. 4: *Query in Natural Language.*

Fig. 5: *Querying for Recent News.*

Fig. 6: *Entities Related to the Query.*

Fig. 7: *Topics Related to the Queried Company.*

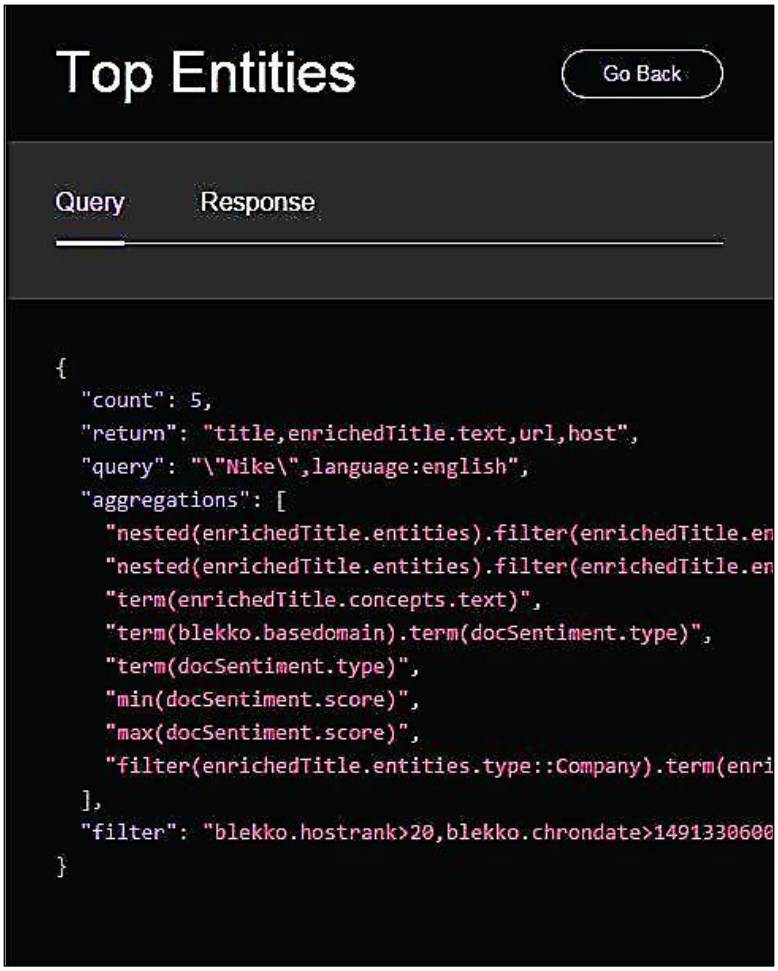

Fig. 8: *Automatic Query Building to Extract Entities as the Recent and Relevant Information Changes.*

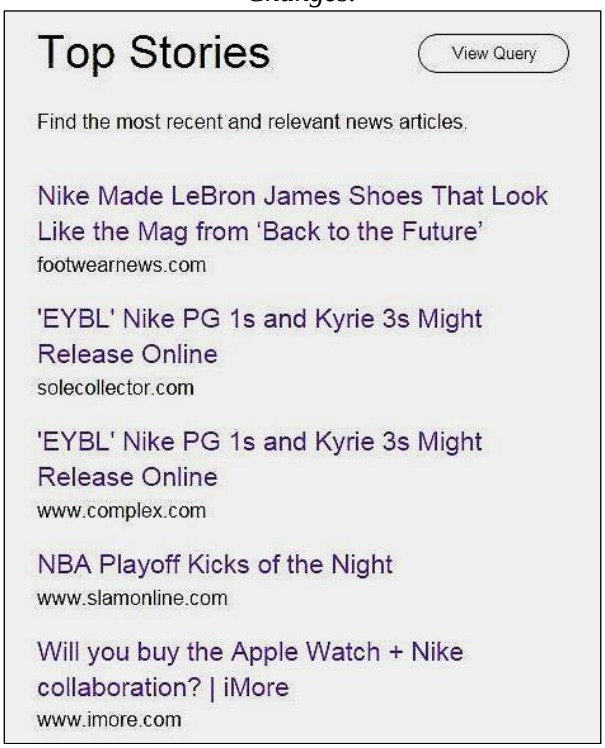

Fig. 9: *Recent and Relevant News about the Company Specified in Query.*

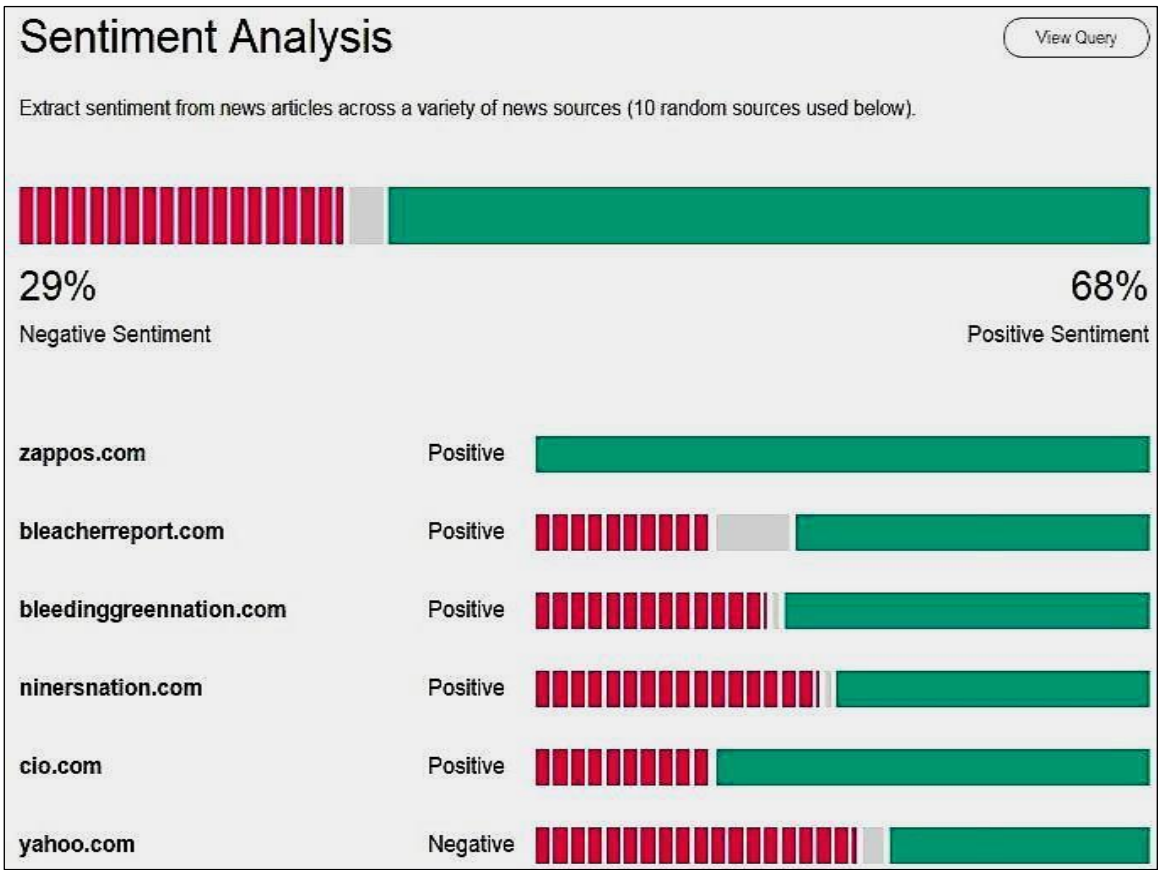

Fig. 10: *Extraction of Sentiments with the Recent Information in News Articles.*

CONCLUSION

Cognitive computing is the recreation of human points of view in a modernized model. Cognitive computing includes self-learning frameworks that utilize data mining, design acknowledgement and natural language processing to impersonate the way the human mind works. The objective of psychological registering is to make robotized IT frameworks that are fit for taking care of issues without needing human support. The IBM Watson discovery benefit serves to quickly fabricate psychological, cloud-based investigation applications that open noteworthy bits of knowledge covered up in unstructured information.

The proposed application builds a tough cognitive space using Watson discovery service. This application accepts the queries in natural language to process a large amount of unstructured data. It converts and enriches the data by using Watson API's to add Meta data to the content and makes easier to explore and discover insights. The future scope of the proposed work is, by integration of discovery news service, the application can add additional enrichments such as concepts and relations, through natural language processing.

REFERENCES

1. Wikipedia. Big Data. Available from https://en.wikipedia.org/wiki/Big_data
2. Wikipedia. Cognitive Search. Available from https://en.wikipedia.org/wiki/cognitive search
3. Wikipedia. Web Blogger. Available from https://en.wikipedia.org/wiki/Blog
4. IBM Cloud-Bluemix. Available from www.ibm.com/cloud-computing/bluemix/
5. Yunus Yetis, *et al*. Application of Big Data Analytics via Cloud Computing. IEEE. 2016.
6. Alisa Sotsenko, *et al*. Using a Rich Context Model for Real-Time Big Data Analytics in Twitter. IEEE. 2016.
7. Leung Carson K, *et al*. A Data Science Model for Big Data Analytics of Frequent Patterns. 2016.
8. Gloria Bordogna. Clustering Geo-Tagged Tweets for Advanced Big Data Analytics. 2016.

9. Omesh Kumar, *et al*. Visualization: A Novel Approach for Big Data Analytics. 2016.
10. Awais Ahmad. Defining Human Behaviors using Big Data Analytics in Social Internet of Things. *IEEE 30th International Conference on Advanced Information Networking and Application.* 2016.
11. Nural Mustafa V, *et al*. Using Semantics in Predictive Big Data Analytics. 2015.
12. Ravi Vatrapu, *et al*. Social Set Analysis: A Set Theoretical Approach to Big Data Analytics. 2010.

Cite this Article
Sindhu B. Jigali, Nirmala CR. Unstructured Data Analysis with the Use of Cognitive Capabilities of Watson Analytics. *Journal of Artificial Intelligence Research & Advances*. 2017; 4(2): 10–18p.