Asynchronous Learning of Personal Assistant from User Histories Stored in Cloud

Sumathi.V.P, Vidyasagr V, Vanitha.V

Abstract— ***A personal assistant always provides us with a daily schedule, remembers all the events and intimates us at right time. Most of the times, our routine jobs are forgotten and user may get to know new information about our most obsessed things later only. Not everyone could afford a personal assistant to help with the daily activities or help the executive in remembering him/her about the activities. Due to heavy industrious works, a human mind naturally tend lose its integrity to forget most important tasks which may lead in missing flight timings or losing a relationship or even millions of amount. People most of the time tend to know the newer updates or information about the things they have been searching for only when they go looking for it again or when they see in others post. The proposed personal assistant system called Naturally Artificial Sanguine and Tangled Youngster (N.A.S.T.Y) does not violate any of the privacy issues and helps the user to remember their search topics and mobile applications usage statistics. The system has asynchronous method run continuously in the cloud server and finds the user's recent favourites and activities and stores that in separate PostgreSQL database table. To find the activities and favourites of the user, an android application is developed using Android Studio which gathers the usage statistics of the user's phone. The user wants to search any content over the web the proposed system helps to do it efficiently. The summarized text is shown as output in user mobile phone.***

Keywords — Named Entity Recognition, Personal Assistant, Machine Learning, Searching Assistant, Pattern Recognition

I. INTRODUCTION

A personal assistant is a system or an application that enables the user to commit the simple works easier and faster. A software application that is intelligent enough to understand the user's query and perform the reliable action. A good personal assistant answers the user's question to the point and doesn't answer irrelevant. If the AI system gets over trained or become biased at a situation it starts to make vague responses and it will start to lose its quality. Using these assistants, not only for performing simple tasks but also making simple suggestions based on the activities performed by the users. The primary purpose for developing the assistant is to increase the accuracy of suggestions given to the users by examining the activity patterns of the user. Many of the personal assistants suggests the users only when an action is performed by the user like booking a ticket to a movie, suggest booking a cab for the user. Upcoming technological world is throwing tones of gadgets and applications making the human life easier to live which can be used to enhance personal assistant.

The proposed work is emphasized the different tasks like to get all possible suggestions that can be given to the user even when the user is not aware of it. An asynchronous method helps in finding all such suggestions before the user gets aware of it, runs continuous analysis based on the user's activities like application usage statistics and queries made in a search engine. To run a task as background process and finds the favorites of the user to post them as a notification to the user via the android application. The user mobile usage details collected from the android application is sent to the cloud server. This enables the user to remember the tasks as the application will be sending out tasks as notifications to the user's phone. The asynchronous method tries to find newer details of his queries made in Google search engine. This is processed with Natural Language Tool Kit (NLTK) and the whole text of the page is summarized using Named Entity Recognition (NER) algorithm.

Objective of the proposed work

- To develop personal assistant to help the user to remember their activities
- To explore the mobile application usage pattern to the user to understand their activities in mobile phone on daily basis
- To provide summarized text as output to understand the searched content very effectively and quickly

The rest of the paper is organized into five sections. Section 2 discusses the existing related literatures. Section 3 discusses the materials and methods used for developing personal assistant. Section 4 provides results and discussions. Finally section 5 concludes the work.

II. REVIEW OF LITERATURES

Bayu Setiaji and Ferry Wahyu Wibowo[1] proposed a system where decisions are made with the knowledge, similarities formed between pre-processed data using bigram algorithm and the decisions and knowledge are stored in the database. Bigram algorithm is the process of separating the words adjacently. The accuracy 67.55% was obtained. In future chatbots incorporates computer reasoning systems where the system is trained with more knowledge extracted from informational database.

Casteleiro et.al.[2] deals with two different types of models in this work namely generative and selective model. Generative models use Recurrent Neural Network where hidden state is designed based on the sense of the context.

Revised Manuscript Received on July 18, 2019.

Sumathi.V.P, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India. (E-mail: sumathi.vp.cse@kct.ac.in)

Vidyasagr V, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India. (E-mail: sagar.vidhya816@gmail.com)

Vanitha.V, Department of Computer Science and Engineering, Kumaraguru College of Technology, Coimbatore, Tamilnadu, India. (E-mail: vanitha.v.cse@kct.ac.in)

More layers in bidirectional encoder plays a major role for improving the response accuracy. The issue in generative model is no proper responses for queries in generic contexts, which involves a wide range of vocabularies. Inconsistent responses generated by 4-layer SeQ2SeQ model trained with 25 million Twitter snippets.

Chen, Yan & Zhang, Yan-Qing[3] Queries submitted to a Web search engine are usually short and ambiguous. Currently, most search engines respond to a user's query by using the bag-of-words model. This paper proposes a personalized query suggestion agent that uses the Query-Concept bipartite graphs and Concept Relation Trees for query suggestion. The accuracy obtained is around 86% and topic based text similarity computation that constructs query text similarity graph is anticipated as a future work.

Karolina Owczarzak et.al.[4] proposed a system to perform analysis of underlying relations indicated one particular domain. The developed system was struggled to find the required ambiguity results and lexical relations, but the performance on adjuncts was poor. A number of classes of adjuncts, such as temporal ones, are fairly easy to identify once information about the object of the preposition is taken into account. The system developed gives out optimized performance 79.56% leaving out the dumbest queries made by the user. As a part of future work is to integrate an interface that comprises a compression model with sentence extraction.

The chatbot [5] proposed in this paper used to learn the sports news. The user can acquire sports knowledge by interacting with this bot. Bank chatbot [6] is trained to do customer service. The queries posted by the customer handled by this chatbot within a short period.

Sameera A. Abdul-Kader and Dr. John Woods [7] proposed a system with key factors like Human-Computer Speech interaction and Natural Language Toolkit (NLTK). Text Parsing, Pattern matching, AIML, Chat Script, SQL and relational database, Markov Chain, Language tricks, Canned responses, Typing errors and simulating keystrokes, Model of personal history are used to evaluate the response. The future works concentrate on designing Chatbots with more comprehensive knowledge bases in order to cover general topics by using the latest techniques. The technologies used in various chatbots are discussed in this paper[8]. The authors [9] were used N-gram model to predict word from large text. Statistical algorithm was used to find associated word.

In distributed approach[10] the relative strength of the words is computed efficiently. The authors [11] proposed a method without using knowledge base, the acquisition of lexical relation was computed. The proposed algorithm has wider applicability. The document compression technique was discussed in [12]. It didn't consider surrounding context. The local coherence theory within integer linear programming was used for text compression.

Personal Assistant Systems Available in the Web

- **Google Assistant:** .In the outside world, the best and accurate assistant till date is developed by Google. It is user friendly. The user feels best assistant and does the tasks so promptly. This system does all the tasks more efficiently using the data fed by the users.
- **Amazon Echo:** Amazon's Echo aka Alexa is a developer friendly assistant where developers in the outside world can develop models and feed data to it using the APIs given by Amazon for the developers.
- **Apple Siri:** Apple Siri is more like a developing assistant that fetched one of the top position due to its brand name. It answers the user query well and it is relevant. Compared to other assistants mentioned above Apple's Siri is not able to answer most of the user queries in general domain.
- Microsoft **Cortana:** Cortana again an assistant that only does when it is asked to do some tasks and the tasks are very limited. It is currently now in beta stage and updates are released very soon.
- **Samsung Bixby:** An assistant that is developed only for the users of Samsung smartphones. Users of samsung smartphones enables them to control their phone using this assistant and the tasks are limited to the samsung only .
- **DataBot Personal Assistant:** Databot is a funky AI assistant that is very similar to iOS devices and is ready to answer any user questions, as well as provide an overview of the latest world news
- **Andy Assistant:** It suggests the user to respond to messages and performs outgoing calls. It reads through the emails and clearly understands client questions and messages the user about it. Converting currencies and report the results to the user.

2.1 Motivations to develop Personal Assistants

- All the systems mentioned above, are only performing the tasks when the user asks to do, except for Google Assistant that show notifications for all the calendar entries using google assistant
- The above said assistants are responded only when the queries are posted in its API.
- When the assistants failed to answer the user queries at the time of submission, they don't have provision to answer the same queries in future.

The above said issues strongly motivate to develop the personal assistant with latest technologies like AI and cloud storage.

III. MATERIALS AND METHODS

The proposed system adds feature for studying on its own when the user explicitly needs it to. When the user asks the assistant for a newer information and if the assistant doesn't find the answer to it, then it tries to evolve and tries to find answer by studying through the answers given in the internet. Soon after it finds the answer to the query asked by its user, the assistant sends this out as a notification and as message to the android application thereby leaving out none of the queries given by the users.

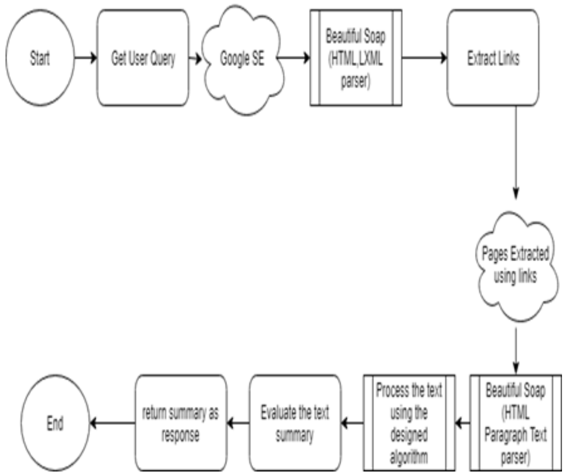

Start
Get User Query
Google SE
Beautiful Soap (HTML,LXML parser)
Extract Links
Pages Extracted using links
Beautiful Soap (HTML Paragraph Text parser)
Process the text using the designed algorithm
Evaluate the text summary
return summary as response
End

Figure 1 N.A.S.T.Y system structure

The proposed system has the following components:

1. NLTK an open source library used basically for processing the text data obtained from the web pages.
2. An android application helps us in serving the required data to make the application tailored made for the user's personal use.
3. A statistical model developed for this purpose to analyze the user frequent query and identify the user's recent favorites and what the user wishes for.
4. The asynchronous model gets the stored data from the database and runs continuously processing each and every query of the user and maps it with the mode value.
5. Mode value exceeds threshold value the respective query is sent and the system studies the recent news of that query or the application.
6. Assistant finds any newer information relevant to the query, it sends out the summary of the information to the user as message and a notification through N.A.S.T.Y API.

The proposed personal assistant N.A.S.T.Y processes are shown in Figure 1. This assistant is developed for the purpose that runs automatically and gives out suggestion at possible situations. A statistical model with almost no speculative response generating algorithm for suggesting the user. In this system, the user queries and his/her daily activities are obtained which acts as the primary data to convert the application, a tailored made one for the user. The data is nothing more than that of text values which are processed with NLTK to summarize the whole content. Before processing the text with the NLTK, first step is to measure the cosine similarities with each and every query. A threshold value is compared to evaluate similarity between two queries given by same user.

Developing a system that is to be deployed at one cloud space whose APIs are then shared to be consumed by multiple users, to create unique identification to each mobile user was the first challenge in this work. To achieve this, a hashing technique was developed in the android application which uses the parameters such as the date and time of installation, user's mobile name and the android phone username. This hash is then stored in a separate SQLite DB inside the mobile, which runs only once at the time of installation. Maintaining this value and passing this value for all API requests hit in the server, the user uniqueness is determined. This solves the first challenge without any compromises.

The primary data that should be obtained to make the application tailored fit to the users was the next challenge which was solved using an android application. Since android phones became one of the biggest part and a gadget that stays at every step of our life, this posed a source of truth for the data reliability purposes. Installation of this application requires a space of 12 MB in the user's android phone and using this application is very easy. This application fetches the user's data for each and every unlocks and send that to the server consuming on an average data usage of 9 MB per day. This personal assistant does not post unwanted notifications but works efficiently to change the user's life in a better way or to increase the user's productivity.

3.1 Methodology

Initially, a redis server ran within the Django server is implemented in separate thread to run all asynchronous methods. NLTK package is installed with the required stopwords, punkt, acronyms, abbreviations. API URLs are exposed to the internet for the applications to consume them. One of the API is to get all the application usage statistics data and then trigger the asynchronous method. This asynchronous method gets the post request data and then stores in the SQLite DB for later processing. On a daily basis, another process is performed again asynchronously to analyze the obtained application usage tracking data. This system will provide information about top, least 5 applications used, query relevant new information, suggestions about new applications are found. To get the user's complete activity, the application usage details, the queries posted by the users in search tab are to be obtained. In order to collect those data, web browsers histories are going to be analyzed.

NLTK Text Summarization Algorithm

Input : Html pages relevant to given search strings
Output : Summary of the all web page content

1. Using the BeautifulSoap library the html page is scraped and obtained with the html tags
2. The html tags are removed with a special regular expression algorithm which gets rid of all the html tags and retrieves the text alone.
3. Using the texts concatenated to one variable, it is sent to the NLTK library for ranking, vector formation and then extracting the summary out of it .
4. The concatenated text is first split into sentences and then compared with each other for the cosine similarity.
5. Cosine Similarity results in a floating value that shows how similar the two sentences are. This ranges from 0 to 1.
6. Before processing the sentences with each other, to do a two step optimization control NER algorithm is used to get a simple understanding of the two sentences.

NER says whether the sentences which are going to be compared are in same subject and tense. Subject of the sentence plays a major role, it is always resembling the core of the sentence.

Named Entity Recognition (NER) Algorithm

Input: Text document with multiple sentences in same concept

Output: Summary of the text document

1. The text is obtained and it is processed with RegexpTokenizer for tokenizing the text values.
2. After tokenizing the sentences, the stop words obtained from NLTK library used to evaluate the sentence for its integrity. Though the sentences are tokenized and integrity is checked, it is not be very sure of its meaning and grammar of the sentences.
3. To solve this, again the NLTK library provides us with WordNetLemmatizer. Processing the sentences in WordNetLemmatizer helps to ignore the obvious the words for processing.
4. PorterStemmer is used to clean the tokens and then join all the sentences. The resultant string is returned to source calling method.

IV. RESULTS AND DISCUSSIONS

The proposed system is implemented using Amazon Linux 2 AMI (HVM), 8GB SSD Volume Type, DDR 16GB RAM, Linux kernel 4.14 tuned for optimal performance on Amazon EC2, systemd 219, GCC 7.3, Glibc 2.26, Binutils 2.29.1.and software's like Open Source Library NLTK (Natural Language ToolKit), Open Source Framework Django, Redis, RQ , Python language interpreter, Android studio IDE.

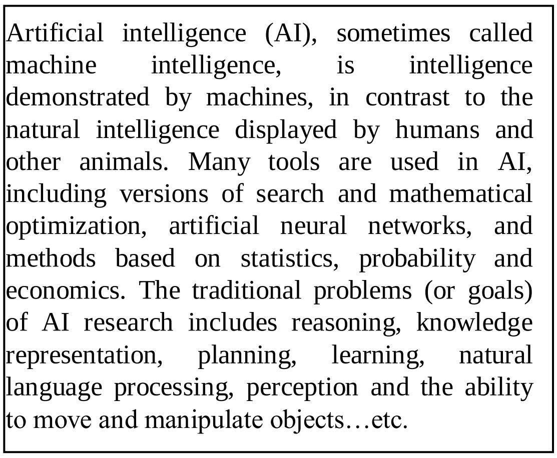
The N.A.S.T.Y assistant is chat application. First it creates the initial view for the assistant to be built. Using this, the adapter gets the value and adds it to the view. The three methods in main program which are OnTypeListener, OnMessageSentListener, OnStoppedTypingListener helped in stopping the assistant to perform and carry out a newer task. The asynchronous methods are handled using the redis queue framed by the redis server runs the analysis on a daily basis. Notification Holder, a table that holds all the analyzing messages. Once this table gets updated the user gets notified about it. The top and least 5 applications used by the user and user searched data are found by this method. Also, suggestion to use newer applications to the user and gathering new information about the applications the user has been using. On the whole, setting up the environment to be used by the user is made easy now with single click and converted this application into an 'On the Go' application. The user queries are found with the cosine similarities and then the information for their most searched subject is suggested for the user to read it. Thus the N.A.S.T.Y assistant ensures it is always keeping connected to the user and provide valuable insights to the user and evolve when the internet updates newer information.

4.1 Redis Queue Worker and Redis Server

In bash shell, after installing the required packages the Redis server is separate shell that runs continuously picking the asynchronous method calls. This starts a newer thread where it waits for any new requests that should be processed in the background without the user intervention.

- The RQ worker starts its thread listening in the default port.
- The queue is enabled to add processes in to it and process the tasks one by one.

The main and integral part of the code is to get the application usage statistics at all times from the user to the server to run analysis based on the data stored. This ensures to call a universal resource locator for every unlocks made by the user. This POST request made by the user has the unique hash value and application usage statistics in a dictionary. The server gets this request and creates new user if not created already for the hash value and stores the statistics of the application usage in the asynchronous method. Basically, once a call is made a true instance is returned after making a trigger to the asynchronous method.



Artificial intelligence (AI), sometimes called machine intelligence, is intelligence demonstrated by machines, in contrast to the natural intelligence displayed by humans and other animals. Many tools are used in AI, including versions of search and mathematical optimization, artificial neural networks, and methods based on statistics, probability and economics. The traditional problems (or goals) of AI research includes reasoning, knowledge representation, planning, learning, natural language processing, perception and the ability to move and manipulate objects…etc.

Figure 2 N.A.S.T.Y Summarized Text Output

The input text content is obtained from WikiPedia page. The total word count in the sample input is 13453. After summarization the resultant page word count is 174. Figure 3 shows the output of the proposed system N.A.S.T.Y . The generated sentence has proper meaning that conveys the same meaning when it is standing alone.

4.2 Real Time execution and Output Images

Table 1 shows the database structure used to mobile user details with unique identification in cloud server. The cloud entries are used to monitor the user activities.

Table 1 SQLite Database User Table

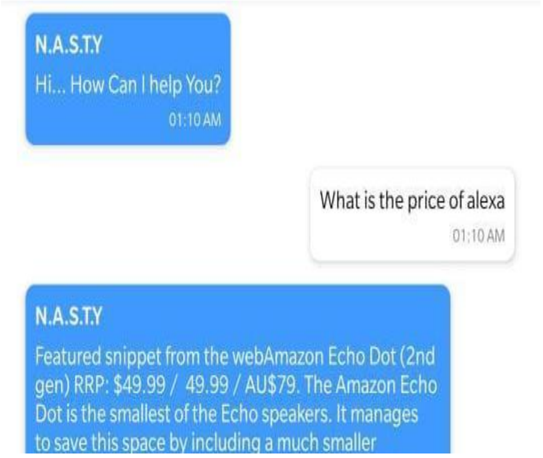
Field No.	Field name	Description
1	Id	User unique identification number
2	Password	Security code
3	Last login	Date and time of lost login
4	Is super user	Values 0 or 1 based on the nature of the user
5	Username	Name of the assistant user
6	Email	Email id of respective user

Table 2 shows the list of applications which is least frequently used applications by the user. The notification regarding applications usage pattern indicated to the user. Which help the user to realize their time spending pattern in various applications.

Table 2 User applications usage histories stored structure

S.No.	Field name	Description
1	Id	Application Identification Number
2	Application Name	Name of the Application
3	Created at	Date and time of installation
4	User id	Mobile user identification
5	Last login	Latest Application activated date and time

Figure 3 and 4 shows the summarized text of two keyword searched by the user. First query relevant to price of alexa and second query is related to information about black hole by NASA. If the N.A.S.T.Y assistant not able to answer the query at the time of submission it will continuously search the result as the background process. The asynchronous search starts automatically by the system by itself using time stamp.

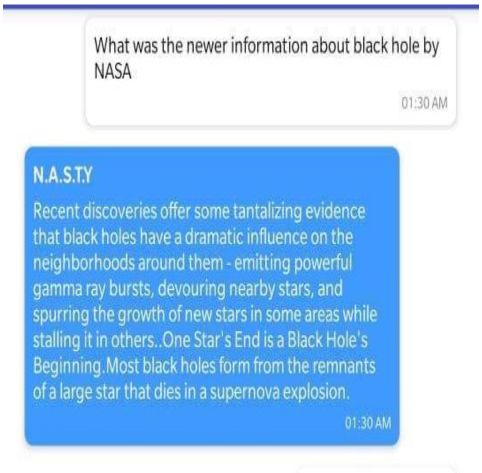


N.A.S.T.Y
Hi... How Can I help You?
01:10 AM

What is the price of alexa
01:10 AM

N.A.S.T.Y
Featured snippet from the webAmazon Echo Dot (2nd gen) RRP: $49.99 / 49.99 / AU$79. The Amazon Echo Dot is the smallest of the Echo speakers. It manages to save this space by including a much smaller

Figure 3 N.A.S.T.Y output for the keyword "Price of Amazon Alexa"



What was the newer information about black hole by NASA
01:30 AM

N.A.S.T.Y
Recent discoveries offer some tantalizing evidence that black holes have a dramatic influence on the neighborhoods around them - emitting powerful gamma ray bursts, devouring nearby stars, and spurring the growth of new stars in some areas while stalling it in others..One Star's End is a Black Hole's Beginning.Most black holes form from the remnants of a large star that dies in a supernova explosion.
01:30 AM

Figure 4 N.A.S.T.Y output for the keyword "Black hole by NASA"

V. CONCLUSIONS

N.A.S.T.Y a cloud based personal assistant has been developed with two features at the end of the day. The user can treat it as their own answer finder and it acts as their personal assistant giving out the possible suggestions whenever possible. This assistant always keeps track of mobile user activities and sends time stamp based notification and messages regarding their applications usage pattern. The assistant not only gets the long passages of information and summarize it to shorter ones but also answers should be short and clear to the point of query. The future scope of this work is to perform predefined user actions. Actions like booking a ticket for your favorite movie at the center of the auditorium, Ordering your favorite food from the restaurants using the food delivery applications like Swiggy, Zomato. Since the assistant deals with unsupervised learning method, this doesn't require models to be trained before releasing for use. Also this can be used as a personal trainer, which fetches data from gadgets like smart watches. N.A.S.T.Y assistant ensures it is always keeping connected to the user and provides valuable insights to the user and evolve when the internet updates newer information.

REFERENCES

1. Bayu Setiaji and Ferry Wahyu Wibowo, "Chatbot Using a Knowledge in Database: Human-to-Machine Conversation Modeling", 7th International Conference on Intelligent Systems, Modelling and Simulation, pp 72-77, 2016.
2. Casteleiro M.A., Tsarkov D., Parsia B., Sattler U., "Using Semantic Web Technologies to Underpin the SNOMED CT Query Language", In: Bramer M., Petridis M. (eds) Artificial Intelligence XXXIV. SGAI 2017. Lecture Notes in Computer Science, vol 10630. Springer, Cham
3. Chen, Yan & Zhang, Yan-Qing, "A personalized query suggestion agent based on query-concept bipartite graphs and Concept Relation Trees", International Journal of Advanced Intelligence Paradigms, 1(4), 2009.
4. Karolina Owczarzak, Ferdinand de Haan, George Krupka, Don Hindle, "Words you don't know: Evaluation of lexicon-based decompounding with unknown handling" , Proceedings of the First Workshop on Computational Approaches to Compound Analysis, pages 63–71, Dublin, Ireland, August 24 2014.
5. Ameya vichare, Ankur Gyani, Yashika Shirkhande, Nilesh Rathod, " A chatbot system demonstrating Intelligent Behavious using NLP", in IJARCET Vol 4 Issue 10 , October 2015.
6. Chaitrali S. Kulkarni, Amruta U. Bhavsar. Savita R. Pingale, Prof. Satish S. Kumbhar, " BANK CHAT BOT – An Intelligent Assistant System Using NLP and Machine Learning", IRJET, Volume 4 Issue 05, May 2017
7. Abdul-Kader, Sameera A., and J. C. Woods. "Survey on chatbot design techniques in speech conversation systems." International Journal of Advanced Computer Science and Applications 6, no. 7 (2015).
8. Luka Bradesko, Dunja Maldenic " A Survey of Chatbot System through a Loebner Prize Competition" International Conference January 2012.

9. Peter F. Brown, Peter V. deSouza, Robert L. Mercer, Vincent J. Della Pietra, Jenifer C. Lai " Class-Based n-gram Models of Natural Language" in IBM T.J Watson Research Center in Association for Computational Linguistics 1992.
10. Hindle, Donald, and Mats Rooth. "Structural ambiguity and lexical relations." Computational linguistics 19, no. 1 (1993).
11. Marti A. Hearst, Xerox Palo Alto Research Center, " Automatic Acquisition of Hyponyms from Large Text Corpora" in ACTI~S DE COLING-92, NANTES, 23-28 1992
12. James Clarke, Mirella Lapata "Modelling Compression with Discourse Constraints" in 2007 Association for Computational Linguistics.

WEB SITES

1. https://play.google.com/store/apps/details?id=com.google.android.apps.googleassistant
2. https://www.amazon.com/Amazon-com-Amazon-Alexa/
3. https://www.apple.com/in/siri/
4. https://www.microsoft.com/en-us/cortana
5. https://www.samsung.com/global/galaxy/apps/bixby/
6. http://www.databot-app.com/
7. http://www.andyforandroid.com/

Retrieval Number: I10920789S219/19©BEIESP
DOI : 10.35940/ijitee.I1092.0789S219

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication