

NLP Based Automated Persona Introducer

Dr. C.Bhanuprakash¹, Raksha Jain², Gayathri R³

^{1,2,3}Dept of Master of Computer Applications

^{1,2,3}Siddaganga Institute of Technology, Tumkur, India.

Abstract- The intention behind this paper is to present an article based on the newly developed application with a titled NLP based Automated Persona Introducer. This android based application simplifies the process of recruitment related assignments. Automated Persona Introducer is an application developed based on Natural Language Processing concept[NLP]. NLP is the technology used to guide the computers to understand the human's natural language. Currently, it is one of the hot areas in the field of artificial intelligence. It deals with interpretation of words and text patterns between the computers and human beings. This is possible in the form of series and sequence of interactions by using natural language processing concepts. The ultimate objective of NLP is to read, decipher, understand and make sense of the human language in a natural manner. The primary focus of persona introducer is to provide the complete information about the candidate to the machine to convert the given text pattern to audio form. When once the application starts functioning, it converts all the text patterns into audio visual format and displays the information about the person on the screen along with the background voice with a predefined speech. In this way, it improves the effectiveness of introduction of any person/fresh candidate with a minimum set of data and saves the time for a recruiter to read the resumes of the candidates during the recruitment process.

Keywords- Introducer, Resume builder, recruitment, interview, written test. Natural Language Processing, Artificial Intelligence, text patterns, machine learning, soft computing, text scanner.

I. INTRODUCTION

Persona Introducer is a mobile based android application which works on the basis of Natural Language Processing concepts. It has been developed purely on the basis of artificial intelligence and machine learning concepts. Here, it trains the computer machine by feeding series of predefined sequence of text patterns. To work with this application, a person has to login with a proper username and password credentials. Later, he needs to enter the basic information about himself in the form of attributes that include first name, middle name, last name, date of birth, address, educational data which covers his primary school, high school, pre-university school, graduation college, marks scored in SSLC,

PUC, Degree and Post graduation if any. After entering the data in to the machine, the app will start processing the entered text patterns into equivalent speech words with a sound. It will be like a person is able to introducing himself by speaking entered text words. This will definitely eliminate the efforts put to process the resume that include its reading, reading time and patience of reading.

This app is developed on the android platform (Android studio) by using Java JDK tool. Android Studio is an Integrated Development Environment for the development of an android application. It supports Gradle based build system, emulator, code, link tools, templates, proguard and Github integration. In addition to this, it has modalities that include android application modules, library modules and google app engine modules. It provides more features with advanced coding techniques which ensure more precise results. The user interface used here will provide user friendly features which ensure more customization options. It has a drag and drop features which facilitates the developers to go in an easier ways and consumes less developmental time. Even it is easy for the beginners to learn and implement.

Android studio works on the basis of IntelliJ IDEA. IntelliJ provides the facility for code analysis in a static state. It also assists during coding and helps in analysing the errors. It also facilitates developmental activities in a more effective ways by providing ergonomic design features.

Natural Language Processing [NLP], is defined as the automatic manipulation of natural language based on speech and text by using software. It is a collective term referring to automatic computational processing of human languages that include both algorithms that take human-produced text as input, and algorithms that produce natural looking text as outputs. It refers the way, we humans communicate with each other. Humans may speak to each other, as a species, more than they write. It may even be easier to learn to speak than to write. Taking the significance of this type of data patterns, there are methods to understand and reason about natural language, just like we do for other types of data. The amount of the efforts to put on this type of the research on this area is very hard. Because, the type of the language used by child is different, the language used by adults is different and type of the language used by senior members is different. It is very

difficult for the scientist who attempts to make a model the relevant scenario and it is hard for the engineer who attempts to build the systems that deal with natural language input / output.

Human language is highly ambiguous. It is always changing and evolving. Moreover, language has various meanings for the same word at different locations. People are crazy in producing the language, understanding the language, capable of expressing the language, perceiving and interpreting the language is totally different with nuanced meanings. At the same time, humans are great users of language, and also they are very poor at formally understanding and describing the rules that govern the language.

As machine learning practitioners and researchers, they are interested in working with text data and predefined speech words. Here, it is concerned with the tools and methods used to solve the requirements from the perspective point of Natural Language Processing. In general, higher-level NLP applications involve aspects that emulate intelligent behavior and apparent comprehension of natural language. In other words, the technical processes and operations of increasingly advanced aspects of cognitive behavior represent one of the developmental trajectories of NLP.

As per the long-standing trends in the field, it is possible to extrapolate future directions of NLP. As of now, three trends among the topics of the long-standing series are as follows.

- i) Cognitive aspects of natural language.
- ii) Multi-linguality, Potentiality and Modality.
- iii) Elimination of symbolic representations.

In this paper, it is trying to present a practical approach to demonstrate the concepts of natural language processing with predefined text patterns to generate personal introduction with spoken words along with background sound. These applications automatically sense the text patterns and translate into equivalent speaking words in a more meaningful manner.

II. RELATED WORK

There have been many approaches made by many researchers on NLP as well as AI based speech recognition concepts.

"Deep Learning For NLP-ACL 2012 Tutorial" [1]. Socher, Richard, made a research on Deep Learning along with skepticism. They got feedback from many research

scholars on this topic. They published a tutorial on this topic for the first time in the year 2013. Until 2015, deep learning had evolved into the major framework of NLP.

"Natural language grammar induction using a constituent-context model" [2], Klein, Dan; Manning, Christopher have presented a novel approach to the unsupervised learning of syntactic analyses of natural language text. This work has focused on optimizing the likelihood of generative PCFG models. In fact, they employ a simpler probabilistic model over trees based directly on constituent identity and linear context. This method produces much higher quality analyses, giving the best published results on the ATIS dataset.

In a titled "NLP Approaches to Computational Argumentation" [4], the authors have presented how humans will argue and debate on the intellectual activities of the human mind. They also explored the differences between multiple perspectives and conceptualizations. Their work involves a wide variety of complex and interwoven actions. They focused on wide range of conceptual capabilities and activities based on CL and NLP. The main aim of this article would be to introduce rapidly evolving field to the Computational Linguistics community.

Chomskyan linguistics [5] supports the investigation of corner cases that stress the limits of its theoretical models, typically created using theoretical experiments, rather than the systematic investigation of typical phenomena that occur in real-world data, as in the case of corpus linguistics. The processing and use of such corpora of realistic data is a basic part of machine-learning algorithms for natural language processing. In addition to this, theoretical aspects of Chomskyan linguistics such as "poverty of the stimulus" argument aids the general learning algorithms which are typically used in machine learning. But these algorithms cannot be successful in language processing. As a result of this, the Chomskyan paradigm discouraged the application of such models to language processing.

Another aspect of grammar induction where there has already been substantial success is the induction of parts-of-speech. Various distributional clustering approaches have been resulted in relatively high-quality clustering methods, though the clusters' resemblance to classical parts-of-speech varies substantially.

The compact grammar of MDL is close to some general linguistic argumentation which has argued for minimal grammars on grounds of analytical or cognitive economy. Perhaps, the basic disadvantage of MDL systems does not have

to do with the objective function. These systems finish their growing structures slowly, in a bottom-up manner. Therefore, their induction quality is specified by how do they are able to heuristically predict the local intermediate structures will fit into good final global solutions.

In this work, we take the part-of speech induction problem and worked with sequences of parts-of-speech rather than words. In some ways this ensures the problem easier, such as by reducing parsing of the text patterns, but in other ways it complicates the task of parsing.

III. METHODOLOGY

3.1 The implementation of this application requires the following software and hardware specifications

- Operating systems : Windows 10
- Android studio, XML, Java (JDK tool)
- Hardware : I5 processor and above
- RAM: 4GB and above
- Hard disk : 1GB and above

Installation ofJDK tool and Android studio will be required for the implementation of this application.

The flow of data in this application takesplace as shown in the figure-1. Data flow starts from the authorization of the user in the form of authorized username and password. After identifying that the user is authorized one, it will allow him/she into enter the application. He needs to enter the personal information by using user interfaces which are shown in the figure-3, figure-4 and figure-5.

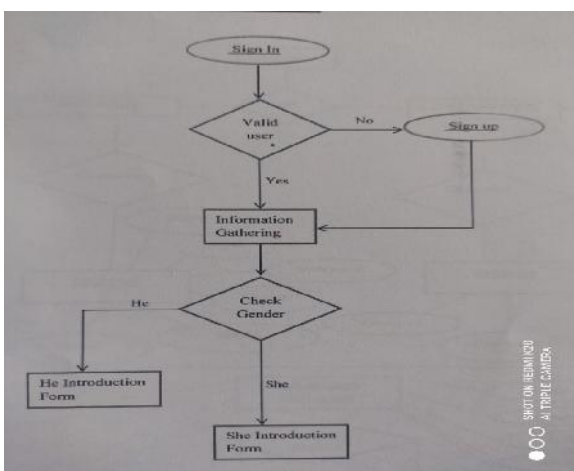


Figure-1, Data flow diagram for Persona Introducer.

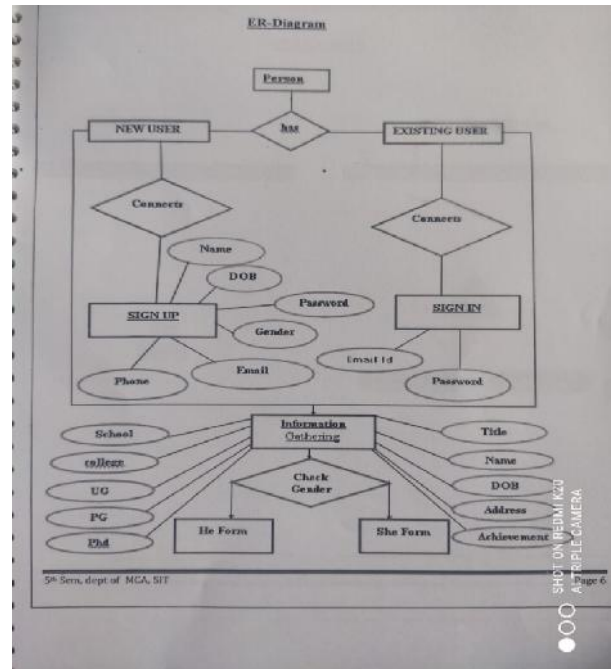


Figure-2, Entity-Relationship diagram for Persona Introducer.

The Entity-Relationship diagram for this application is as shown in the figure-2. This ER diagram has entities that include person, new user and existing user. The entity person has the attributes that include name, gender, date of birth, address, phone, email id,educational qualifications that include school, college, Undergraduate, Postgraduate, Doctorate and his / her achievements. The relationships exists in this ER diagram include connects, has and check gender.



Figure-3, User Interface to accept personal information.

The user interface to accept user’s personal information is as shown in the figure-3. This interface has many fields that include Mr/Mrs/Miss suffix, Name, date of birth, address, school, percentage of marks and college. The educational qualifications of the candidate can be entered into another user interface as shown in the figure-4. This interface has many fields that include name of the college studied for under graduation, name of the university, name of the course, percentage of marks obtained in degree. This form also consists of fields pertaining to post graduation course that include name of the college studied, name of the course, name of the university, percentage of marks obtained.

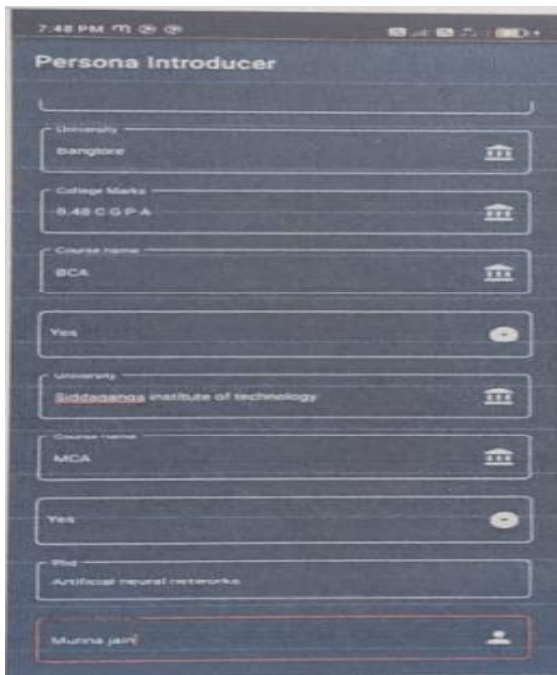


Figure-4 User Interface to accept under graduate information of a person.

The doctorate course (if any) details of the candidate and equivalent related information can be entered by using the interface as shown in the figure-5. This interface has fields that include name of the area of doctoral degree, area of specialization and additional attributes such as phone number, email id, contact address and permanent address. After entering the data for all the above fields in to the database, the result will be displayed by using the form as shown in the figure-6. This form will display the result by make use of pre defined key words along with the data accepted thru the above interfaces from the database. Here all the text patterns will be converted into sounded words. This was possible by make use of speech recognition technology that is implemented in the application by using proper coding and data structures.

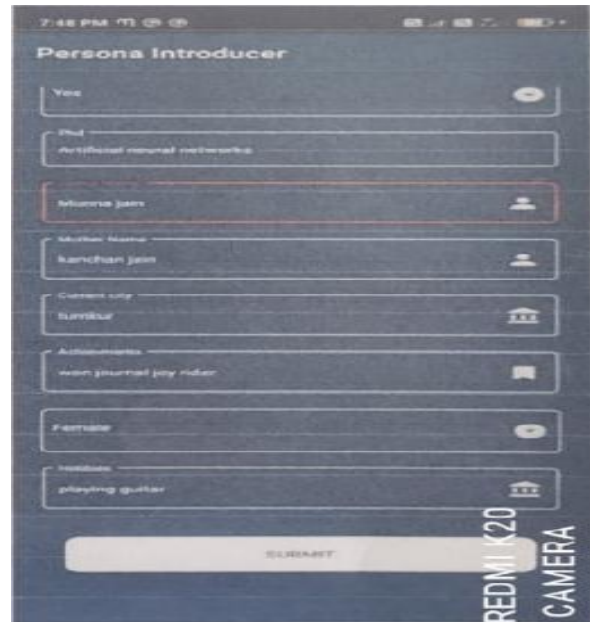


Figure-5, User Interface to accept Doctoral degree (optional) information of a person.

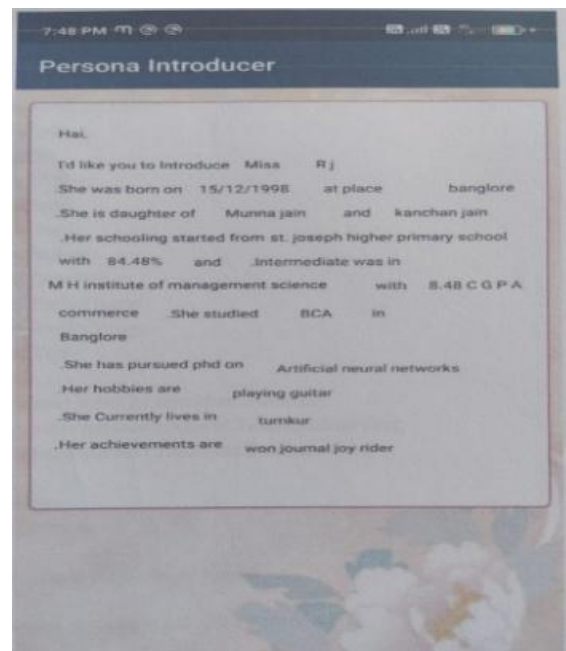


Figure-6, Interface to display overall information of a person.

IV. CONCLUSION

This is the sincere and small attempt to develop an application which works purely based on Natural Language Processing concepts. There are many challenging issues encountered during the development of this application. Because, the kind of the language being used, type of the language accent used, and the meaning for the same word is different from areas to areas and locality to locality. By considering all these issues, we have customized this

application to recognize only with few pre defined words. These words will be accepted thru fields provided in the user interfaces. The data accepted thru these fields are also restricted with only one type of standard language accent. Totally, this application is restricted with only few key words with only one accent. But still lot more scope is there to improve this application to work for more number of spoken words with few more language accents by considering more area locations. We would highly appreciate your valuable feedback, healthy comments and suggestions.

REFERENCES

- [1] Socher, Richard - "Deep Learning For NLP-ACL 2012 Tutorial" – www.socher.org. Retrieved 2020-08-17.
- [2] NASER S. BARGHOUTI AND GAIL E. KAISER - : "Concurrency Control in Advanced Database Applications" – Department of Computer Science, Columbia University, Newyork.
- [3] Guida, G.; Mauri, G. (July 1986). "Evaluation of natural language processing systems: Issues and approaches". Proceedings of the IEEE. **74** (7): 1026–1035. doi:10.1109/PROC.1986.13580. ISSN 1558-2256. S2CID 30688575.
- [4] "NLP Approaches to Computational Argumentation – ACL 2016, Berlin". Retrieved 2021-03-09.
- [5] Goldberg, Yoav (2016). "A Primer on Neural Network Models for Natural Language Processing". Journal of Artificial Intelligence Research. **57**: 345–420. arXiv:1807.10854. doi:10.1613/jair.4992. S2CID 8273530.
- [6] Turchin, Alexander; Florez Builes, Luisa F. (2021-03-19). "Using Natural Language Processing to Measure and Improve Quality of Diabetes Care: A Systematic Review". Journal of Diabetes Science and Technology. **15** (3): 553–560. doi:10.1177/19322968211000831. ISSN 1932-2968. PMID 33736486
- [7] Duan, Yucong; Cruz, Christophe (2011). "Formalizing Semantic of Natural Language through Conceptualization from Existence". International Journal of Innovation, Management and Technology. **2** (1): 37–42. Archived from the original on 2011-10-09

Authors Profiles:



C. Bhanuprakash, received his graduation B.E. in Mechanical Engineering and Master of Computer Applications degrees from Siddaganga Institute of Technology, Tumkur, of Bangalore University, in 1993 and 1998, respectively.

He received his doctoral degree in the discipline of Computer Science and Information Science Engineering in the year 2021 from Visveshvariah Technological University. He has 20 years of experience in Teaching, 4 Years of Industry experience and 8 years of research experience. Presently he is working as Assistant Professor in the department of Master of Computer Applications, Siddaganga Institute of Technology, Tumkur, Karnataka.

His areas of specialization include design and development of database applications, Operation Research, Optimization Techniques, Intelligent Data Analysis, Advanced Operating systems, Neural Networks, Data Mining and other Soft Computing Techniques.

Ms Raksha Jain, Ms Gayathri R, are pursuing their post graduation degree MCA at Department of Master of Computer Applications, Siddaganga Institute of Technology, Tumkur, under the University of VTU, Belgaum. Their areas of interests include Programming, framing the data structure for the innovative algorithms and developing the unique kind of applications.