An Anatomization of Language Detection and Translation using NLP Techniques

Bhagyashree P Pujeri, Jagadeesh Sai D

Abstract: ***The issue with identifying language relates to process of determining natural language in which specific text is written. This is one of the big difficulties in the processing of natural languages. Still, they also pose a problem in improving multiclass classification in this area. Language detection and translation a significant Language Identification task are required. The language analysis method may be carried out according to tools available in a particular language if the source language is known. A successful language detection algorithm determines the achievement of the sentiment analysis task and other identification tasks. Processing natural language and machine learning techniques involve knowledge that is annotated with its language. Algorithms for natural language processing must be updated according to language's grammar.This paper proposes a secure language detection and translation technique to solve the security in natural language processing problems. Language detection algorithm based on char n-gram based statistical detector and translation Yandex API is used.While translating, there should be encryption and decryption for that we are using AES Algorithm.***

Index Term: Language Identification, Natural Language Processing (NLP), AES, N-gram,Language Detection, and Translation.

I. INTRODUCTION

Language Recognition (LR) is a method in which the substance of the document is written in a natural language. Language identification is a wide-ranging research field because it is mostly considered in applications for natural language (NL) identification, like machine translation, informationretrieval, summary as well as answering questions, etc. They need prior identification of the language before processing. Identification of language falls into 2 methods: 1) non-computational,and 2) computational. Non-computational methods require authors to havingenough knowledge of language to be recognized, diacritics& symbols, the most common words utilized characterscombination, & so on.

In contrast, computational methodsdepend on statistical techniques to solve related problems rather than linguistic knowledge. Fast growth of less well-known languages on Internet has generated requirements for LR for applications such asmachine translation, spell checking, multi-lingual information retrieval, etc.Three factors complicate this task: several sizes of character set utilized to encode diverse languages,use various character sets for single language, also more than one language sharing the same script[2]. Automatic treatments of these texts, for any purpose requiring Natural Language processing.

Revised Manuscript Received on November 26, 2020.
***** Correspondence Author
Bhagyashree P Pujeri, Information Science, RIT, Bangalore, India.
Email: bhagyapujeri24@gmail.com
Jagadeesh Sai D, Information Science, RIT, Bangalore, India.
Email: durgamsai@gmail.com

Such as WWW indexing and interrogation or providing reading aids necessitates a preliminary identification of the language used. For example, morphological based stemming has proven essential in improving information retrieval, and applying language-specific algorithms implies knowing the language used.In developing new digital goods and services, data is now a sort of capital, on a par with financial and human capital. Everyone is overloaded with information overload due to the proliferation of information innews, medical records, corporate files, court hearings, government papers,as well as social media. The bulk of this data is unstructured, i.e., free text, making it hard to have reasoning and interpretation.

A. Sentiment Analysis (SA) for Language Identification

Sentiment Analysis (SA) was one of the fields of quantitative study in the production of natural languages [3].SA usually performs the processing of knowledge relevant to emotions or beliefs. From a community for a given subject. Furthermore, opinions havebeen obtained at the document level from specific applications. SA has gained prominence in many fields, including politics [4], business, and marketing. It has believed that the records will contain views while undertaking SA. However, for so many cases, only factual information and evidence are set out in such papers (the news document is one such example). Even materials that are supposed to contain feelings (opinions) that often include descriptive sentences as a part of them. Hence, the most crucial aspect of SA is the recognition of the form and essence of sentences. Thus, the sentences have extracted, classified, and included in the given analysis, depending on the subjective or objective. Classification of subjectivity is the critical activity at SA, which provides for the classification of sentences as factual or subjective. Generally speaking, SA requires several complicated processes. The analysis has accompanied by a set of activities, including the designation of emotions, individualinterpretation, the extraction of opinion holders, and the extraction of aspects or objects[5]. The subjective research includes evaluating the same as subjective or objective as a document or a word to mark. Those documents or phrases classified as objective are automatically discarded after this stage because they are not very useful for the SA[6].

B. Language Processing

Artificial intelligence is now widely debated as a buzzword and is rapidly evolving. AI is computer programs that can do something smart like a person. It's merely a machine that mimics human beings to execute tasks in his absence and, generally speaking, often in improved and productive way. Machine learning is an AI subgroup of AI.

Using machine learning, machine intelligence is enhanced by learning algorithms and analyzing various types of data. Machine learning is a subset of Deep Learning and Neural Networks. According to the performance obtained, deep learning algorithms repeatedly analyze various data sets through algorithms and enhance machine intelligence. In the field of computer science, the analysis of natural languages is an important part of machine learning and computational linguistics. Natural Language Processing (NLP) field includes the development of computer systems with natural and human language to perform meaningful tasks. In the future, NLP is so important because it allows one to construct models and procedures, which embrace information as an input or as a voice or a word or both and exploit it on an algorithm in the machine.Input can then be voice, text, or picture where both speech and written output of an NLP device can be processed. Different algorithms produced to improve the effectiveness of the text type processing of the language that we will address here are:

- Sequence 2 Sequence framework
- The long short term memory
- User preference graph framework
- Named Entity Recognition framework
- Feature-based sentence extraction by fuzzy inference rules.
- A template-based approachbyan automatic text summarization
- Word Embedding model[8].

C. Language Identification

Identification of a language typicalmeans process that tries to categorizetext intoa predefined set of available languages in a language. It is a crucial technique for NLP, mainly in the operating text dependsupon language and classification. Excellent results have been obtained by several researchers[9][10][11][12] on language Recognition rely ona few European languages picked. But, most Asian &African languages also staynot tested. This highlights that the search engines have even less supporting for most African & Asian languages in their language-specific searchingcapability.For correct language categorization, all LSE properties are essential. Also, to defineaccurate tools for text processing atthe last level, LSE identification is necessary. To choose a good translator to translate source text into an additional language, a computer translation tool must first learn the script.

D. N-gram

An n-gram mayseem from longer sequences as a sub-sequence of N objects. A word, letter, syllable, or some logical data form specified by the application can be referred to as the item described. As it is simple to apply and determine next possible succession from the known sequence with great accuracy, the n-gram probability model is one of the very successful NLP methods for statistical results. The main principle of just using n-gram is that each language has its specific n-grams and that these n-grams are often used much more than other languages, offering a vocabulary hint. A monogram is named as n-gram order 1 (that is n=1); n-gram of order 2 is called bigram, n-gram of order 3 is called trigram. N-gram number 2 is sometimes called bigram. The "rest" is usually called "n-gram." "Use" Number 456 b As an example, correct lists (all distinguished via space) for the character-level trigrams and bigrams will be as follows:

Bigram: No o- -4 45 56

Trigram: No- o-4 -45 456

Numerousauthors [13][14] announced that the best language recognition result was obtained using the trigram model on preferred European languages. But, several African & Asian languages have not dependedupon Latin script which usesseveral European languages. Therefore, the analysis tests the efficacy of the n-gram orders (n=1, 2,3,..., 6) &unique n-gram mix framework for language orders. Recognition in preferred languages. The paper is structuredinthe following sections: Relevant history and related research are given in Section II. Section III addresses our problem statement, proposed system model, and algorithm. Section IV discusses the effects of the simulation and the analysis. The article is finalized in Section Vand Section VI, which determines the future scope.

II. LITERATURE SURVEY

A. NLP

M. Uma(2019) ProposedSuch a framework uses NLP to easily By offering standardized NL queries as input & receiving SQL queries as output, access to relevant information from the railway reservation database. Measures concerned in this method werelemmatization, tokenization, and PoS marking, mapping &parsing. For proposed system dataset, a list of 2880 structured NL queries on train tickets &seats is available. We achieved 98.89% precision. The paper explains use of NLP and the use of standard phrases to mapping queries to SQL in the English[15].

M. R. Hasan et al.(2019) PresentedTwitter data to study public opinions on a product. Firstly, to filter tweets, we have developed an NLP based pre-processed data system. Secondly, to evaluate sentiment, we integrate the model definition of Bag of Words and Term Frequency-Inverse Text Frequency (TF-IDF). It is an effort to utilize BoW & TFIDF together to distinguish positive & negative tweets accurately.We have originated that the accuracy of SA can be significantly improved using the TF-IDF vectorizer, and simulation results indicate the efficacy of our proposed method. Using NLP technology, they achieved 85.25 percent precision in sentiment analysis. One of the primary sources of knowledge is social networking sites such as Facebook, Twitter, Instagram, etc. A business organization may benefit from its product by collecting [16].

Y. A. Solangi(2018) Opinion mining (OM) recently in a diverse research area. NLP techniques for an opinion mining & interpretation of sentimentshave been studied in this paper. NLP is initially checked. In this work, mining is studied and evaluated at different levels. Problems are listed at the end, and some suggestions for OM and SA are proposed. Like most online networks on the Internet, OM has been a basic way to deal with such a vast amount of knowledge being investigated. In a wide range of modern fields, various applications pop up. In the meantime, views have different pronouncements that carry problems to be discussed[17].

J. Ding et al.(**2018)**Built an electronic model of a robot. In the lane prepared in the laboratory, the robot can move and precisely predict the direction. In the right path, the robot's motion is going, representing about 72 percent. Researchers have thought in the past that automating robots is a far-reaching method. Some scientists believe that better software is built for a robot that can automatically drive the robot. Several scientists have now built robots that can be automated and applied to several tasks[18].

B. Das and A. Kumar(2017)A nonlinear programming approach (NLP) was implemented to optimally plan the regular operating strategy of the connected energy storage system. To maximize the economic benefits of the hybrid energy system, the issue of optimization on an hourly basis is considered to be energy demand, market prices, and operational constraints. Using the NLP approach, the optimization problem is solved by General Algebraic Modeling Software (GAMS), an easy and quick comp solution framework.The optimization model is analyzed with wind farm data. The results show that the model is reliable and offers an ideal solution to the energy storage system's regular operating schedule, thus optimizing revenue. With rising global energy demand mostly met by fossil fuels with limited supplies and adverse environmental impacts, renewable energy sources abundant in nature with less to no ecological damage are the most promising choice.However, renewable energy sources are intermittent, causing grid imbalances and impacting the electric grid's stability. Therefore, the introduction of energy storage plays a keyfunction in ensuring the stability & effective incorporation of renewable energy sources into the grid [19].

B. Sentiment Analysis for NLP

Y. Peng and T. Chou (2019) Proposed a standard definition into color palettes automatically. A system has proposed integrating color imagery with sentiment analysis. The algorithm consists of four stages. Second, it describes terms affect as the basis for classifying texts. In this analysis, the conditions influencing the CIS picture are names. Second, it's collecting related to Google and Wikipedia text corpora. Third, by way of model training, word2vec is used to evaluate the linguistic affinity of words and colors. Also, a prototype program has been designed to illustrate its effectiveness for the automated design of color palettes. Designers have sometimes been forced to guess from vague and often conflicting reports of customers about color combinations and other specifications. Perhaps so much time is spent on collaboration and adjustment [20].

A. Razzaq et al.(2019) Presented a Sentiment Analysis and Prediction (ASAP) three-step Text Pattern solution by K-Nearest Neighbor (KNN). At first, sentences converted into tokens and words removed from the stop. Second, the sentence, paragraph, and text's polarityare determined by adding weighted terms, strength clauses, and shifters of feelings. The resulting characteristics extracted during this stage played a significant role in improving performance. Lastly, the input text pattern is predicted using the KNN classifier based on derived functions. The model's training and research carried out on publicly accessible twitter and film review datasets. The results of the experiments demonstrated a satisfactory change in contrast with current solutions. Additionally, the text analysis system based on Interface (Hello World) designed to perform text analytics. Sentiment Analysis, a non-diminishing research field in text mining, involves a computational method to extract useful text information. Numerous ways aimed at understanding the text's feelings, but they could not enter the emotions [21].

M. Wongkar and A. Angdresey(2019) presented python's use of python. A SA framework for twitter analysis (TA) was implemented. This SA is to accumulate data utilizing python libraries, process text, check training data, and categorize booksvia the NBapproach, carried out in so many phases. The Naïve Bayes approach helps classify communities or the degree of human feelings. This study's findings presented that the measure of Jokowi-Ma'ruf Amin pair's positive sentiment polarity was 45.45 percent and 54.55 percent negative[22].

J. Ding et al.(**2018)** Concerned With the conduct of sentiment analysis at the company level. First, we create a physically labeled dataset with 3 thousand selected comments from the 231,732 issues from ten GitHub OSS projects. After that,they have designed& construct SentiSW, an entity-leveled analysis system that includes classifying sentiments & identifying entities, which can categorize comments on issues into <; sentiment, entity > tuples.They test the identification of objects by manual annotation and achieve 75.15 percent[23].

K. S. Sabra et al.(2017)presented a new method for developing an Arabic-speaking sentiment lexicon WordNet learning and comparing it to an Arabic database. Sentiment Scrutiny is a method of determining sentiment from a text written in anNLP about the person to which it refers. This role has carried out using sentiment lexicons. There are several lexicons available for carrying out this function in English using WordNet [24].

C. AES

A. Alipour e al. **[2016]**Non-profiling DLSCA is used for secret encryption with the help of a correlated noise generation against a secret AES counter-measure. We demonstrate that AES can provide equal protection with correlated noise generation as a lightweight countermeasure about required traces of power to obtain the secret key under CPA and in non-profiling DLSCA attacks. The latest cryptography work has become one of the strongest attacks on standard encryption algorithms like AES [25].

T. N. Dang and H. M. Vo(2019) The sophisticated AES algorithm that produces dynamic keys. Here the sequential transmission and indexing of the 16-byte data frame in numerical order. The key is modified automatically with encrypted data as an example of IoT systems inside a car tracking framework. In an IoT environment, the protection of data is essential**.** AES is one of their most popular algorithms. Many phases, such as AddRoundkey, Sub Bytes, Move Rows, and Mix Colum, are included. However, the traditional AES algorithm produces fixed encryption keys that are exchanged between the sender and the recipient. The key is, therefore, easily stolen[26].

F. R. Nuradha et al.(2019) When the AES program worked, the proposal utilizedthe CW1173 board to recover trace. Investigator was able to discover all sub-key keys ofthe AES algorithm through CPA analysis with Hamming distance S-box output tool.

The results demonstrate that AES encryption requires more excellent protection so that it is not easy to attack.Encryption is a technology that operates to render unauthorized parties impossible to interpret it through scrambling data. The AES, also called Rijndael, defines electronic data encryptionATmega328p is the method used for implementing AES. ATmega328P is an 8-bit RISC architecture microcontroller depends upon AVR CMOS. Correlation Power Analysis (CPA) is an attack which helps to locate hidden encrypting key stored on the hardware computer [27]

S. K. R et al.(2017) Operating frequency can be 291.68 MHz and a sequential decry AES setup can offer a power of up to 37.21 Gbps. FPGA are three modules that form part of the AES Hardware implementation. Cryptography has an essential role in the protection of data transmission. The primary objective is to improve the efficiency of Advanced Encryption Standard algo by hardware technology by simultaneously processing several rounds. The key benefit of the AES algorithm is that both sofas are incorporated intohardware. Hardware implements advanced encryption requirements since hardware implementation has the benefits of increased performance and improved security. Field-Programmable Gate Arrays are one of the popular solutions for output gradation. Four transformation stages are included in AES function computation for encryption:sub bytes, shift rows, column Mix, add round key. Four transformation stages are included in AES function computation for Decryption: Invsubbytes, Invshiftrow, InvMixColumn, Invaddround key. In Spartan 6 FPGA & Development Board, XC6SL is simulated&advanced encryption standard (AES) developed[28].

U. Arom-oon(2017) Demonstrated AES small-scale n/w cryptosystem implements the execution of an AES algorithm, FIPS 197, on Real-Time Operating System (RTOS) microcontroller to secure data in small-level n/w, like wireless UAV communication. AES algorithm's ECB(Electronic Code Book) mode is often utilized as a cryptographic heart. RTOS has a pre-emptive scheduling algo scheduler in which every task to provide access [29].

III. PROPOSED METHODOLOGY

A. Problem Statement

The critical challenge is the overload of information, which presents a significant problem with accessing a particular, relevant piece of data from vast datasets. Due to consistency and usability problems, semantic and meaning comprehension is essential and challenging for summary systems. It is also crucial to identify the context of interaction b/wobjects&entities, mainly with high-dimensional, heterogeneous, complex,as well as poor-quality data.Because there is no aspect to encrypt and decrypt data in a secure format. To find a relationship between objects&entities, semantics are essential. Extraction of text and visual data by persons and objects could not offer reliable information unless the interaction's meaning & semantics were known. Often, instead of keyword-based search, search engines currently available will search for items (objects or entities). Semantic search engines are needed since user queries traditionally written in natural language are better understood.

B. Methodology

Methodologies used throughout comprise data extraction, NLP, and ml techniques that play an essential role in deciding language detection. Language detection algorithm based on char n-gram based statistical detector and translation Yandex API is used.The Yandex online machine translation tool can be accessed from this API. It can translate separate words or full texts in over 90 languages. The API allows Yandex to be implemented.While translating, there should be encryption and decryption for that we are using AES Algorithm.The main aim to get a secure detection and translation environment.

C. Material and Method

In recent years, NLI has focused the attention of many authors and researchers. With the influx of new researchers, the most substantive research in this area has contributed to the joint mission[30]. The role focuses on the recognition of a writer's native language based on his writing in another language. The second language, in this case, was English. The task was to predict a writer's native language from the provided text / XML file containing English-language Facebook comments.Four languages were proposed to include for this task. They were English, Hindi, Urdu, Gujarati.

a) Dataset

The task's training dataset was XML files, which contain FB comments in English by different native language speakers. XML files were annotated as EG HI, UR, GJ for English, Hindi, Urdu, Gujarati language, respectively [31].

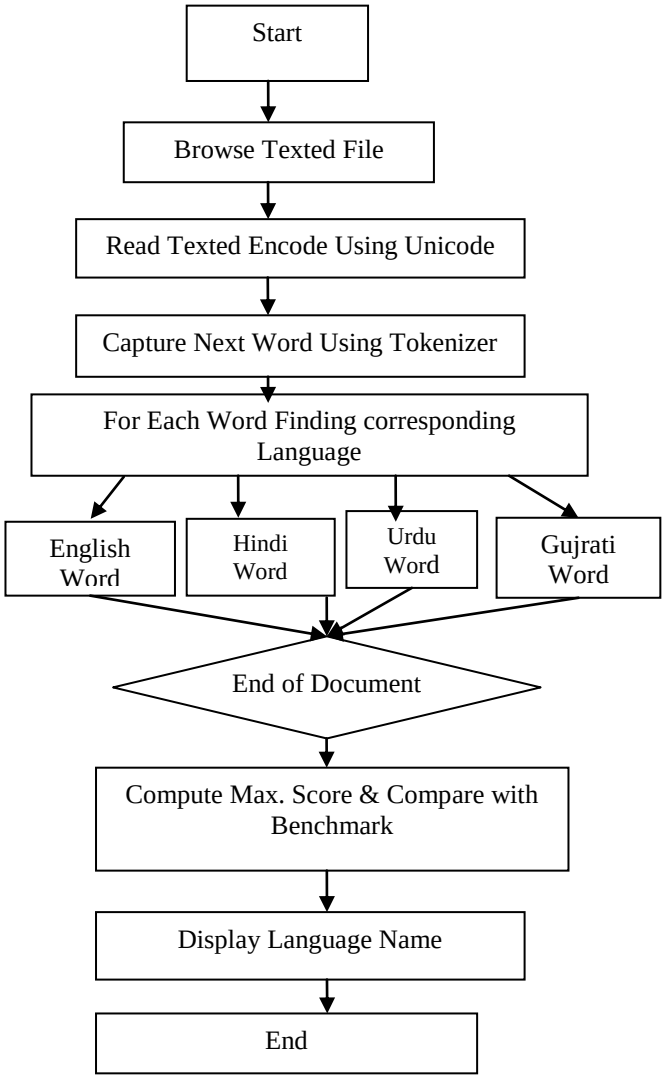

Fig.1 Flow Chart for Language Detection

The above Fig.1 describes the proper step by step approach for the Language Detection process. A language detection algorithm is used the determine the language of a given text. Some languages can be determined reliably from their script alone. A widely used approach is supervised machine learning algorithms based on character n-gram based statistical detector.

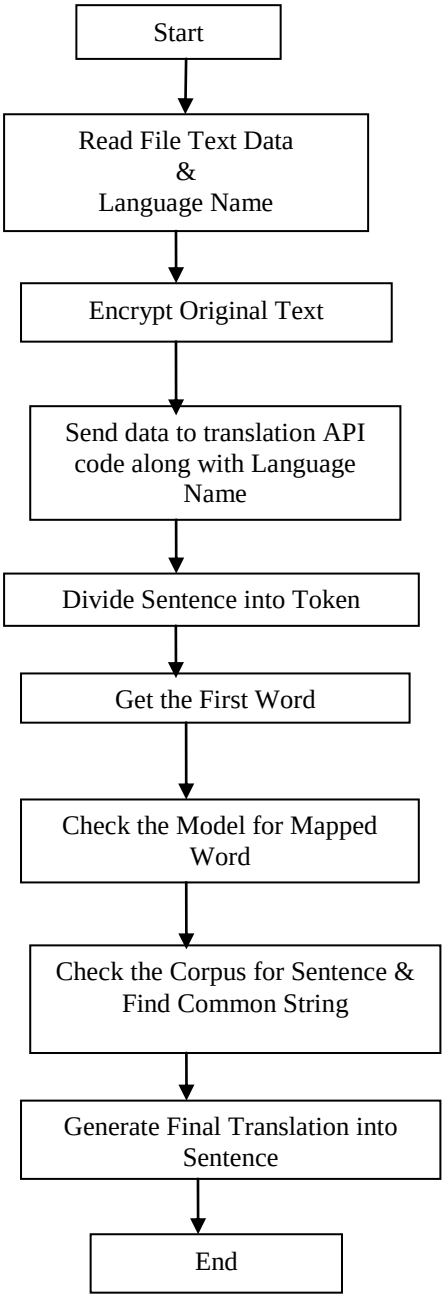

Fig.2 Flow Chart for Language Translation

The above Fig.2 describes the proper step by step approach tothe Language Translation process.For several NLP functions, such as text summarization, speech recognition, DNA sequence modeling, among others, Sequence-to-Sequence (seq2seq) models are used. Our goal is to translate those phrases from one language to another.

D. AES

AES is a symmetric block cipher selected to safeguard sensitive information by the U.S. government. To encrypt confidential data, AES is executed in hardware &software around the world. Cybersecurity and electronic data protection are essential for government information security.

AES was first developed by NISTin 1997 to substitute the Data Encryption Standard (DES)at that time when it realized the needtohappen tosusceptible to the brute-force attacks (BFAs). ES contains 3 block ciphers: AES-128, AES-192 & AES-256. To decrypt &encrypt messagesblock, AES-128, AES-192 and AES-256utilizes128-bit, 192-bit &256-bit key size to encrypt & decrypt messages, respectively. Every cipher encrypts & decrypts data into 128-bit blocks with 128, 192, and 256-bit encryption keys. Ciphers use the same encryption and decryption key, which is also known as the secret key, and both the sender & the receiver must know and also use the similar private key. Knowledge is categorized into three groups by the government: confidential, hidden, or top secret. To safeguard the Confidential and Secret stage, all key lengths can be used. [32].

E. Tokenization

Extracting words from the text may appear to be a simple task. The top-down method breaks book on whitespace characters such as space, Tab, or a punctuation character. Nonwhite spacenames are concatenated to form a word or token. The bottom-up method builds tokens one character from a text stream until a nontoken character is encountered. The simplest definition of a token is any consecutive string of alphanumeric characters. Between tokens, we find one or more nontoken characters.

F. Language Translation

Once the Language Identification task is completed, then the next task is to translate the document. Machine translation [33] [34] It is the translation method into the target language from the source language. The following is a list of problems when attempting to do machine translation that one has to face. Not all words have corresponding words in one language.

- In different tongues inspecific examples, a word in one language must be represented in another by a group of words. 2 provided languages may have different.
- Structures. English has an SVO structure, for instance, whilstTelugu orKannada havean SOV structure. There is also a shortage of one-to-one communication.
- Speech sections for two languages. Kannada / Telugu color words, for instance, are nouns, although they are adjectives in English. The forms in which sentences are placed together vary between languages, too. Words may have more than one meaning and sometimes a meaning.
- In a language, a group of words / an entire sentence can have above one meaning. Ambiguity is called this problem. Not all problems with translation can be explained by applying.
- Grammar'svalues. It's too tricky for software programs to forecast
- Meaning. Translation includes not only grammar &vocabulary however also information collected by previous experience.

Algorithm:

1: Start
2: Browse Text File
3: If detect language=yes, then goto step 5
4: Ifsee language=No, then goto step 1
5: Detect Language API
6: EncryptOriginal Text Using AES
7: If translationlanguage=yes then goto step 9
8: Otherwise, goto step 1
9: Translation language API
9: Decrypt
10: Display the Result

IV. RESULT AND DISCUSSION

It presents the results obtained with Java using the eclipse implemented methods in the outcome analysis. Firstly, the results of the classification assignment of orientation are presented. We submitted the output of the system for test data provided shared task workshop. A single run of each method for four different languages was submitted, and the results of native language classification for all the languages. We have to find out the secure language detection and translation environment.

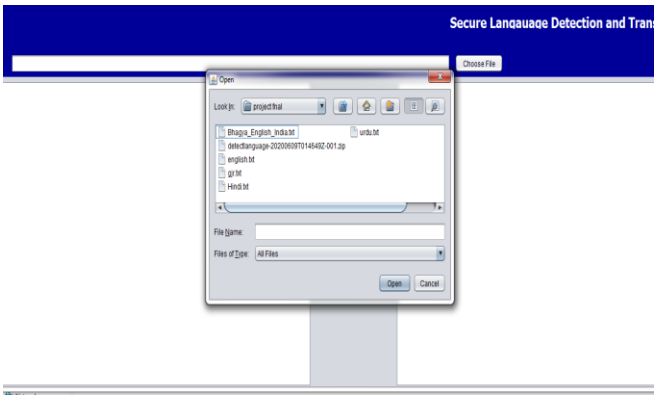

Fig.3 Browse texted file

The fig.3 above shows thebrowse texted file. Language Detector Model is used, and therefore the training step is not required.

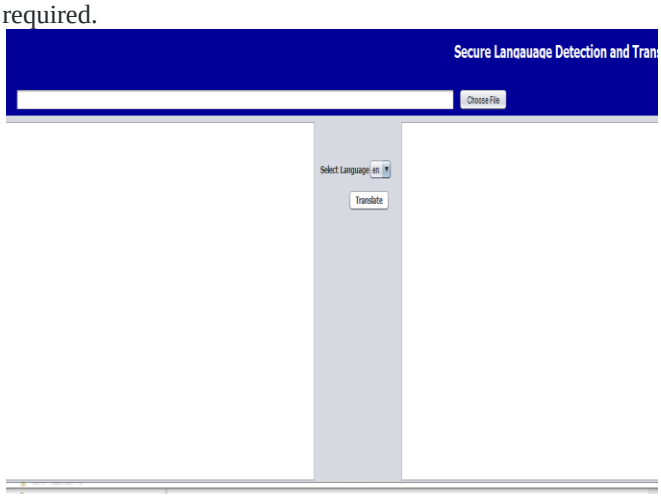

Fig. 4Selection of Language for Translation

The Fig.4 above shows the Selection of Language forTranslation. The main aim to translate given sentences from one language to another. Here, both the input and output are sentences.

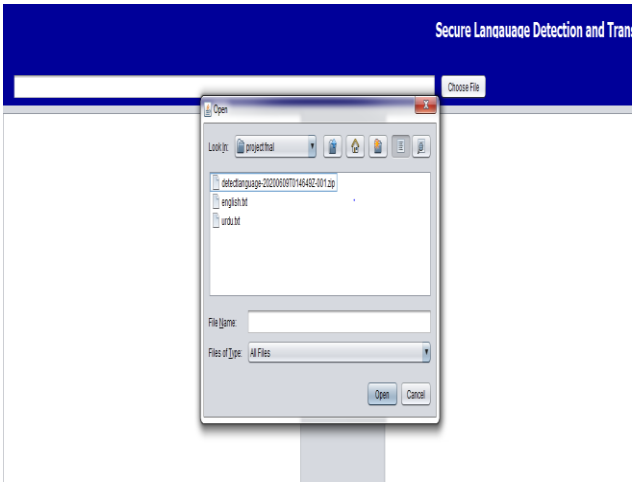

Fig.5Browse File for Translation

The fig.5 above shows the File for Translation. It is the task to convert one natural language automatically into another, maintain the essence of the input text, and produce fluent text in the output language.

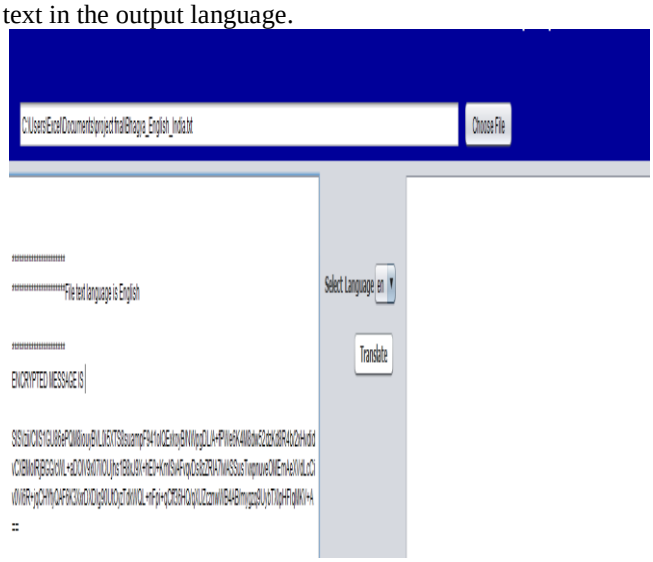

Fig. 6 Encryption of Original Text of File

The results have given in the above fig.6 based on the AES technique. Itis used to encrypt the original text of the selected file.

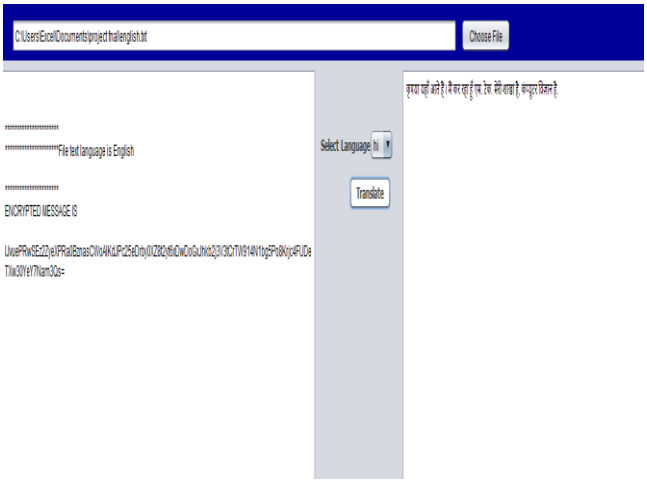

Fig.7 Translation the Encrypted Text for Single Sentence

The results given in the above fig.7have based on the translation methodology, which provides secure detection and translation environment.

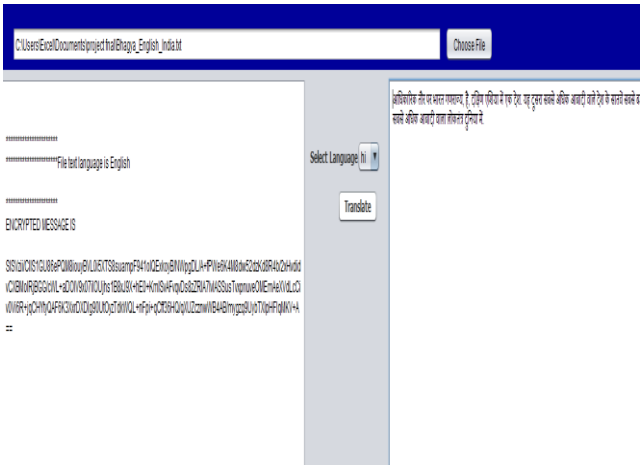

Fig. 8Translation the Encrypted Text for Multiple Sentence

The results given in the above fig.8 have based on the translation methodology, which provides secure detection and translation environment.

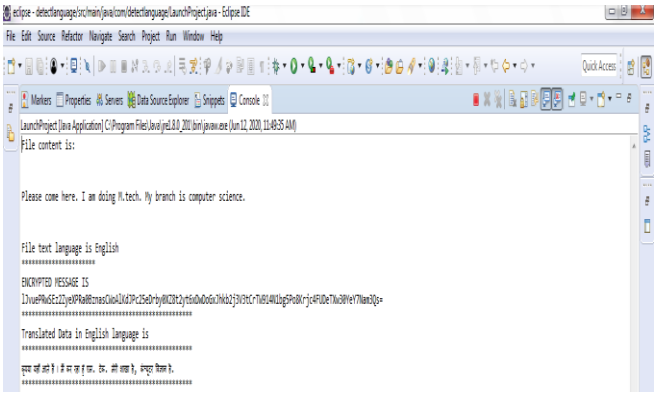

Fig.9 Display the Results(Console Screen)

V. CONCLUSION

For applications such as the fast growth of less-known languages on Internet have generated requirements for language recognition through machine translation, spell checking, multi-language information retrieval, etc. Three factors complicate this role: different sizes of character sets utilized to encode further languages, use of various character sets for single language, as well as more than one language sharing the same script. In this paper, particularly for English, Hindi, Urdu, and Gujrati, we present modules to identify and translate English into Indian languages or Indian Interlingual Languages. A language detection algorithm based on a statistical detector based on char n-gram and Yandex API is used for translation. There should be encryption and decryption when translating, so we use the AES algorithm to obtain a protected language detection and translation environment.One of the essential fields of natural language processing is language translation.

FUTURE SCOPE

We will also use this language identifier module for translation in the future. For bilingual computer translation from English into the Urdu / Hindi language, this will be very helpful.

One of the fundamental difficulties is that English has the structure of the Subject Verb Object (SVO), whereas Urdu has the design of the Subject Object Verb (SOV) in Machine Translation.This study's proposal is a contribution and breakthrough.Our future work also lies in improving the NLI system's performance by considering features, which can classify native languages in a better way, and Language recognition task n-grams may be used for multi-lingual sentiment analysis for fine-grained classification.

REFERENCES

1. M. PadróCirera, L. PadróCirera, Comparing methods for language identification, In Procesamiento del lenguaje natural, Barcelona: Sociedad Española para el Procesamiento del Lenguaje Natural, pp. 155-161, 2004.
2. Penelope Sibun, and Jeffry C Reynar. Language Identification: Examining the issues
3. R. Prabowo and M. Thelwall." Sentiment analysis: A combined approach." Journal of Informetrics , 3(2): 143-157,b2009.
4. J. Bollen, H. Mao, and X. Zeng "Twitter mood predicts the stock market." Journal of Computational Science, 2(1): 1-8, 2011.
5. G. Leech, R. Barnett, and P. Karel, Guidelines for the standardization of corpora's syntactic annotation, EAGLES Document EAG-TCWG-SASG/1.8, 1996.
6. C. Li, B. Xu, G. Wu, S. He, G. Tian, and Y. Zhou, Parallel recursive deep model for sentiment analysis, in Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, 2015, pp. 15–26.
7. Penelope Sibun, and Jeffry C Reynar. Language Identification: Examining the issues
8. Matthew Henderson, Ramial-Rfou, Brian Strope, etal."Efficient Natural Language Response Suggestion for Smart Reply."
9. T. Dunning, Statistical Identification of Language. Technical Report, New Mexico State University, 1994, MCCS 94-273.
10. W. B. Cavnar and J. M. Trenkle, N-Gram-Based Text Categorization. In Proceedings of SDAIR-94, 3rd Annual Symposium on Document Analysis and Information Retrieval, 1994, pp. 161–175.
11. I. Suzuki, Y. Mikami, A. Ohsato, A Language and Character Set Determination Method Based on N-gram Statistics. ACM Transaction on Asian Language Information Processing, 2002, Vol. 1. No. 3, pp. 270–279.
12. T. Ölvecký, N-Gram Based Statistics Aimed at Language Identification. In Mária Bieliková (Ed.), Proceedings of IIT.SRC 2005: Student Research Conference in Informatics and Information Technologies, Bratislava, 2005, pp. 1–7.
13. Grefenstette G., Comparing two language identification schemes. In 3rd International Conference on Statistical Analysis of Textual Data, Rome, Italy, 1995
14. Bashir Ahmed, Sung-Hyuk Cha, and Charles. Tappert "Language Identification from Text Using N-gram Based Cumulative Frequency Addition" published in Proceedings of Student/Faculty Research Day, CSIS, Pace University, May 7th, 2007
15. M. Uma, V. Sneha, G. Sneha, J. Bhuvana and B. Bharathi, "Formation of SQL from Natural Language Query using NLP," 2019 International Conference on Computational Intelligence in Data Science (ICCIDS), Chennai, India, 2019, pp. 1-5, doi: 10.1109/ICCIDS.2019.8862080
16. M. R. Hasan, M. Maliha and M. Arifuzzaman, "Sentiment Analysis with NLP on Twitter Data," 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), Rajshahi, Bangladesh, 2019, pp. 1-4, doi: 10.1109/IC4ME247184.2019.9036670.
17. Y. A. Solangi, Z. A. Solangi, S. Aarain, A. Abro, G. A. Mallah, and A. Shah, "Review on Natural Language Processing (NLP) and Its Toolkits for Opinion Mining and Sentiment Analysis," 2018 IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bangkok, Thailand, 2018, pp. 1-4, doi: 10.1109/ICETAS.2018.8629198.
18. M. Piyaneeranart and M. Ketcham, "Modeling for Robot-Driven Prototype Automation," 2018 International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), Pattaya, Thailand, 2018, pp. 1-6, doi: 10.1109/iSAI-NLP.2018.8692828.
19. B. Das and A. Kumar, "Optimal scheduling of energy storage system using NLP approach," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-5, doi: 10.1109/ICCCNT.2017.8204080.
20. Y. Peng and T. Chou, "Automatic Color Palette Design Using Color Image and Sentiment Analysis," 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 2019, pp. 389-392, doi: 10.1109/ICCCBDA.2019.8725717.
21. A. Razzaq et al., "Text sentiment analysis using frequency-based vigorous features," in China Communications, vol. 16, no. 12, pp. 145-153, Dec. 2019, doi: 10.23919/JCC.2019.12.011.
22. [22] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm Of The Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), Semarang, Indonesia, 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.
23. J. Ding, H. Sun, X. Wang and X. Liu, "Entity-Level Sentiment Analysis of Issue Comments," 2018 IEEE/ACM 3rd International Workshop on Emotion Awareness in Software Engineering (SEmotion), Gothenburg, 2018, pp. 7-13.
24. K. S. Sabra, R. N. Zantout, M. A. E. Abed and L. Hamandi, "Sentiment analysis: Arabic sentiment lexicons," 2017 Sensors Networks Smart and Emerging Technologies (SENSET), Beirut, 2017, pp. 1-4, doi: 10.1109/SENSET.2017.8125054.
25. A. Alipour, A. Papadimitriou, V. Beroulle, E. Arabi, and D. Hély, "On the Performance of Non-Profiled Differential Deep Learning Attacks against an AES Encryption Algorithm Protected using a Correlated Noise Generation based Hiding Countermeasure," 2020 Design, Automation & Test in Europe Conference & Exhibition (DATE), Grenoble, France, 2020, pp. 614-617, doi: 10.23919/DATE48585.2020.9116387
26. [T. N. Dang and H. M. Vo, "Advanced AES Algorithm Using Dynamic Key in the Internet of Things System," 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), Singapore, 2019, pp. 682-686, doi: 10.1109/CCOMS.2019.8821647.
27. F. R. Nuradha, S. D. Putra, Y. Kurniawan, and M. A. Rizqulloh, "Attack on AES Encryption Microcontroller Devices With CorrelationPower Analysis," 2019 International Symposium on Electronics and Smart Devices (ISESD), Badung-Bali, Indonesia, 2019, pp. 1-4, doi: 10.1109/ISESD.2019.8909447.
28. S. K. R, S. R, M. A. M, P. K. M. S, and R. M, "Design of High-Speed AES System for Efficient Data Encryption and Decryption System using FPGA," 2018 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT), Msyuru, India, 2018, pp. 1279-1282, doi: 10.1109/ICEECCOT43722.2018.9001535.
29. U. Arom-on, "An AES cryptosystem for small scale network," 2017 Third Asian Conference on Defence Technology (ACDT), Phuket, 2017, pp. 49-53, doi: 10.1109/ACDT.2017.7886156.
30. Anand Kumar M, Barathi Ganesh HB, Shivkaran S, Soman K P, Paolo Rosso. Overview of the INLI PAN at FIRE-2017 Track on Indian Native Language Identification. In: Notebook Papers of FIRE 2017, FIRE-2017, Bangalore, India, December 8-10, CEUR Workshop Proceedings
31. https://www.kaggle.com/rtatman/sentiment-lexicons-for-81-languages?select=correctedMetadata.csv
32. Abdullah, Ako. (, 2017). Advanced Encryption Standard (AES) Algorithm to Encrypt and Decrypt Data.
33. Machine Translation, Prof. Abdullah H. Homiedan
34. S. Kereto, C. Wongchaisuwat, Y. Poovarawan. 1993. Machine translation research and development. In Proceedings of the Symposium on Natural Language

AUTHORS PROFILE

Ms.Bhagyashree P Pujeri completed B.E in computer Science from S.G Balekundri Institute of Technology, Belgavi, India with First Class in year 2018. Currently she is pursuing her Master of Technology in Information Science from M S Ramaiah Institute of Technology, Bangalore, India. Her research interest areas are Natural Language Processing and Linguistics and Image Processing.

Mr.JagadeeshSai D Master of Technology from Jawaharlal Nehru TechnologyUniversity, Hyderabad, India in the year2013. Currently he is pursuing his PhD in CSE in M S Ramaiah Institute of Technology, Bangalore. He is serving as an assistant professor in IS&E department of MSRamaiah Institute of Technology, Bangalore. He has membership of professional societies like IEEE, ISTE, and ACM. He is interested in subjects related toProgramming languages, Scripting, Cryptography,and Distributed Systems Operating System.