EXAMPLE BASED MACHINE TRANSLATION SYSTEM

Project report submitted in partial fulfillment of the requirement for the degree of Bachelor of Technology

in

Computer Science and Engineering/Information Technology

By

AASTHA SHARMA(141238)
DIVYA JAIN(141237)

Under the supervision of

Ms. RUHI MAHAJAN

to

Department of Computer Science & Engineering and Information Technology

Jaypee University of Information Technology Waknaghat, Solan-173234, Himachal Pradesh

Candidate's Declaration

I hereby declare that the work presented in this report entitled **" EXAMPLE BASED MACHINE TRANSLATION SYSTEM"** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering/Information Technology** submitted in the department of Computer Science & Engineering and Information Technology**,** Jaypee University of Information Technology Waknaghat is an authentic record of my own work carried out over a period from August 2017 to May 2018 under the supervision of **Ms. Ruhi Mahajan** (Assistant Professor,CSE).
The matter embodied in the report has not been submitted for the award of any other degree or diploma.

Aastha Sharma,141238

Divya Jain,141237

This is to certify that the above statement made by the candidate is true to the best of my knowledge.

Ms. Ruhi Mahajan
Assistant Professor
CSE
Dated:

ACKNOWLEDGEMENT

We owe our significant appreciation to our project supervisor **Ms. Ruhi Mahajan**, who took distinct fascination and guided every one of us along in our undertaking work titled – **Example Based Machine Translation System**, till the fulfillment of our venture by giving all the essential data to building up the task. The task improvement helped us in research and we became more acquainted with a considerable measure of new things in our area. We are extremely appreciative to her.

ABSTRACT

Machine Translation (MT) is an undertaking in Natural Language Processing (NLP), where the programmed frameworks are utilized to interpret the content starting with one dialect then onto the next while safeguarding the significance of source dialect. In this work, we give our endeavors in creating Example based interpretation framework. A wide assortment of machine interpretation approaches have been created in past years. Example Based Machine Translation motor (EBMT) is an interpretation framework requiring basically no learning of the structure of a dialect, only an expansive parallel corpus of illustration sentences and a bilingual word reference. Info writings are portioned into successions of words happening in the corpus, for which interpretations are dictated by sub sentential arrangement of the sentence sets containing those groupings. These fractional interpretations are then joined with the consequences of other interpretation motors to shape the last interpretation delivered by the framework.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF GRAPHS

LIST OF TABLES

CHAPTER-1
INTRODUCTION

Example based machine translation (EBMT) is a conventional model of translation. Similar to Statistical MT, it depends upon huge corpora and tries to some degree to dismiss customary semantic thoughts (in spite of the fact that this does not confine them completely from utilizing the said ideas to enhance their yield). EBMT frameworks are appealing in that they require at least earlier learning and are thusly rapidly versatile to numerous dialect sets. We ask that creators take after some basic rules. Generally, we request that you influence your paper to look precisely like this record. Machine translation (MT) inquire about has made considerable progress since the plan to utilize PC to computerize the translation procedure and the significant approach is Statistical Machine Translation (SMT). An other option to SMT is Example-based machine translation (EBMT). The most imperative regular component amongst SMT and EBMT is to utilize a bilingual corpus (translation examples) for the translation of new data sources. The two techniques abuse translation learning certainly installed in translation examples, and make MT framework upkeep and change considerably simpler contrasted and Rule-Based Machine Translation**.**

Then again, EBMT is not quite the same as SMT in that SMT dithers to abuse rich phonetic assets, for example, a bilingual dictionary and parsers. EBMT does not consider such an imperative. SMT essentially joins words or expressions (generally little pieces) with high likelihood, EBMT tries to utilize bigger translation examples. At the point when EBMT tries to utilize bigger examples, it can better deal with examples which are spasmodic as a word-string, yet ceaseless fundamentally. In like manner, however it isn't inescapable, EBMT can normally deal with syntactic data. Other than that, the distinction in the middle of EBMT and SMT, EBMT isn't the trade for SMT. SMT is a characteristic approach when phonetic assets, for example, parsers and a bilingual dictionary are not accessible. Then again, on the off chance that that such phonetic assets are accessible, it is likewise normal to perceive how exact MT can be accomplished utilizing all the accessible assets. EBMT is a more practical,

1.1 PROBLEM STATEMENT

The world is turning into a worldwide town. There are hundred's of dialects being talked over the world. The official dialects of various states and countries are additionally unique as indicated by their social and topographical contrasts.

1.1.1 GAP ANALYSIS

A large portion of the substance accessible in computerized organize is in English dialect. The substance appeared in English must be introduced in a dialect which can be comprehended by the target group. There is huge segment of populace at nationwide in the whole world who can't grasp English dialect. It has achieved dialect hindrance in the side lines of computerized age. Machine Translation (MT), can defeat this hindrance. In this postulation, a proposed Statistical Based Machine Translation framework for making a translation of English content to Hindi dialect has been proposed. English is the source dialect and the Hindi is the objective dialect.

1.2 OBJECTIVES

The Objectives of the system are as follows:

1. To understand the matching, retrieval and transfer phases of EBMT.
2. Check for the presence of the source sentence.
3. Extraction of matched portion in the corpus.
4. Recombining the translated fragments.
5. To generate the target sentence.
6. Test and evaluate the system.

1.3 METHODOLOGY

Example based machine translation frameworks perform the translation of a given input s by performing the following functions:

Matching-Search for the entered input by the user in bilingual corpus.

Retrieval- Extract that part of the sentence which matches with the bilingual corpus.

Rearrangement- Combine all the translated fragments according to the alignment rules. EBMT for all intents and purposes in light of the recovery of source sentences like s in the bilingual corpus, thus EBMT is otherwise called source-similarity based translation. Example based machine translation concentrated on different natural issues like the span of Parallel Corpora, Granularity of Examples, Amount of Examples and Suitability of Examples .

1.4 MACHINE TRANSLATION

Machine Translation is a domain that scrutinizes the use of Natural Language Processing to translate the text entered in one language to another. At the basic level Machine Translation does the work of simple substitution in the source language to get the required output in the target language. Existing softwares which performs the work of machine translation allows for customization by domain , improving output by restraining the range within which we can substitute the text.

1.4.1 NEED FOR MACHINE TRANSLATION

Machine Translation softwares are expected to change the given arcticles or any documents of text to the target language with atmost accuracy. Also it is very important that it preserves the meaning of the actual sentences in the output. These kind of MT frameworks can remove the dialect hindrances by making accessible work affluent wellsprings of writing accessible to individuals over the globe.

Machine Translation likewise conquers the mechanical boundaries. The greater part of the data accessible is in English which is comprehended by just 4% of the populace. This has let to computerized isolate in which just little segment of civilization can comprehend the substance displayed in advanced configuration. MT can assist in such manner to conquer the computerized separate.

1.4.2 PROBLEMS IN MT

There are a few basic and expressive contrasts among dialects, which make programmed translation a troublesome assignment. A portion of these issues are as per the following:

1.4.2.1 WORD ORDER

Word arrange in dialects contrasts. Some Alignment should be possible by naming the run of the mill request of subject (S), verb (V) and protest (O) in a sentence . A few dialects have word arranges as SOV. The objective dialect may have an alternate word arrange. In these kind of examples, word to word translation is troublesome. English dialect has SVO and Hindi dialect has SOV sentence structure.

1.4.2.2 WORD SENSE

A similar word may have distinctive faculties while being meant another dialect. The determination of right word particular to the setting is critical.

1.4.2.3 PRONOUN DECLARATION

The issue of not settling the pronominal references is imperative for machine translation. Uncertain references can prompt mistaken translation.

1.4.2.4 IDIOMS

An informal articulation may pass on an alternate significance, that what is obvious from its words. For instance, a figure of speech in English dialect '*JACK OF ALL TRADES*', would not pass on the plan meaning when converted into Hindi dialect.

1.4.2.5 AMBIGUITY

In computational linguistics, Word Sense disambiguation (WSD) is an open issue of quality dialect preparing, which administers the way toward distinguishing which feeling of a word (i.e. significance) is utilized as a part of a sentence, when the word has various implications.

1.4.3 TYPES OF MACHINE TRANSLATION

The following are four types of Machine Translation (MT) systems:

1.4.3.1 MT FOR WATCHERS

MT for watchers is proposed for per users who needed to access some data written in outside dialect who are additionally arranged to acknowledge conceivable terrible 'harsh' translation as opposed to nothing. This was the sort of MT imagined by the pioneers. This came in with the requirement to translate armed mechanical archives.

1.4.3.2 MT FOR REVISERS

MT for revisers goes for creating crude translation consequently with a value similar to that of the principal drafts delivered by human. The translation yield can be viewed as just as catch up on with the goal that the expert translator can be liberated from that exhausting and tedious errand.

1.4.3.3 MT FOR TRANSLATORS

MT for translator's goes for serving human translators carry out their activity by giving online word references, vocabulary and translation memory. This kind of machine translation framework is generally joined into the translation work stations and the software based translation instruments.

1.4.3.4 MT FOR AUTHORS

MT for writers goes for writers needing to have their writings converted into one or a few dialects and tolerating to compose under control of the framework or to enable the framework to disambiguate the articulation so attractive translation can be acquired with no modification.

1.5 APPROACHES TO MT

There are four approaches to machine translation. These are discussed as follows:

1.5.1 RULE-BASED MT

A Rule-based MT framework parses the source content and delivers a middle of the road portrayal, which might be a parse tree or some dynamic portrayal.

1.5.2 DIRECT-BASED MT

A direct-based MT framework completes word-by-word translation with the assistance of a bilingual lexicon, more often than not took after by some syntactic adjustment.

1.5.3 CORPUS-BASED MT

Corpus based MT frameworks require sentence-adjusted parallel content for every dialect combine. The corpus based approach is additionally ordered into measurable and example based machine translation approaches.

1.5.4 KNOWLEDGE-BASED MT

Early MT frameworks are described by the grammar. Semantic highlights are appended to the syntactic structures and semantic preparing happens simply after syntactic handling. Semantic-based ways to deal with dialect examination have been presented by AI scientists. The drew closer require a substantial information base that incorporates both ontological and lexical learning.

1.6 ORGANIZATION

This report is divided into FIVE sections. Chapter 2 talks about the survey of writing on Example based Machine Translation System. Chapter 3 subtle elements of system plan and usage of the proposed English to Hindi Example Based Machine Translation System. The result and discussion, based on the execution of the proposed framework, has been examined in Chapter 4. The conclusion and future extension, of the framework which has been produced, is talked about in Chapter 5.

CHAPTER-2
LITERATURE SURVEY

2.1 Merging Example-Based and Statistical Machine Translation: An Experiment (2002), [1]

In spite of the energizing work achieved over the previous decade in the field of Statistical Machine Translation (SMT), we are still a long way from the purpose of having the capacity to state that machine translation completely addresses the issues of genuine clients. In a past report [6], we have demonstrated how a SMT motor could profit by expressed assets, particularly while translating writings altogether different from those used to prepare the framework. In the present paper, we talk about the opening of SMT to examples consequently removed from a Translation Memory (TM). We report comes about on a reasonable estimated translation assignment utilizing the database of a business bilingual concordance.

2.2 Wrapper Syntax for Example Based Machine Translation System(2007)[2]

In this paper example-based machine translation enhances when we include a wrapper level that fuses syntactic data. TransBooster gains by the way that MT frameworks for the most part bargain better with shorter sentences, and uses syntactic comment to deteriorate source dialect sentences into shorter, less difficult pieces that have a higher shot of being effectively deciphered. The subsequent translations are recomposed into target dialect sentences.

2.3 Implementation of Example Based Machine Translation System(2007)[3]

The whole framework will change over the source dialect content into target dialect content utilizing characteristic dialect handling. It will utilize the machine translation procedure which is superior to the current devices accessible in the market. The calculation is with the end goal that, there is word reference/vocabulary of English and Hindi. The parsing will be legitimate. The mapping procedure will likewise be utilized. Every one of the Literals will be isolated utilizing dividing and stemming procedures. The root word will be recognized utilizing manmade brainpower and bilingual translation.

2.4 Quantum Neural Network Based Machine Translator for Hindi to English(2008), [4]

In this work it is exhibited that the quantum neural system approach for the issue of machine translation. It has shown the sensible precision on different scores. It might be noticed that BLEU score accomplished 0.7502, NIST score accomplished 6.5773, ROUGE-L score accomplished 0.9233, and METEOR score accomplished 0.5456 precision.

2.5 Hierarchical Phrase-Based Translation Representations(2011)[5]

This paper thinks about a few translation portrayals for a synchronous setting free syntax parse including CFGs/hypergraphs, limited state automata (FSA), and pushdown automata (PDA). The portrayal decision is appeared to decide the frame and many-sided quality of target LM crossing point and most brief way calculations that take after.

2.6 A Lightweight Evaluation Framework for Machine Translation Reordering(2011)[6]

Here a straightforward structure for assessing word arrange autonomously of lexical decision by looking at the framework's reordering of a source sentence to reference reordering information created from physically word-adjusted translations. At the point when used to assess a framework that performs reordering as a preprocessing step our system enables the parser and reordering tenets to be assessed amazingly rapidly without tedious end to-end machine translation tests.

2.7 Addressing the Rare Word Problem in Neural Machine Translation(2015), [7]

In this work, model of Sutskever et al. (2014), which utilizes a profound LSTM to encode the info succession and a different profound LSTM to yield the translation. The encoder peruses the source sentence, single word at once, and produces an expansive vector that speaks to the whole source sentence. The decoder is introduced with this vector and creates an translation, single word at once, until the point that it emanates the finish of-sentence image .

2.8 Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation(2016)[8].

This paper acquaints a basic technique with translate between numerous dialects utilizing a solitary model, exploiting multilingual information to enhance NMT for all dialects included. Our strategy requires no change to the conventional NMT display engineering. Rather, we add a simulated token to the information Alignment to demonstrate the required target dialect, a straightforward change to the information as it were.

2.9 Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation(2016),[9]

This work displays the outline and execution of GNMT, a generation NMT framework at Google, that means to give answers for the above issues. In our usage, the repetitive systems are Long Short-Term Memory RNNs. Our LSTM RNNs have 8 layers, with lingering associations between layers to support inclination stream. For parallelism, we interface the consideration from the base layer of the decoder system to the best layer of the encoder arrange. To enhance derivation time, we utilize low-accuracy number-crunching for induction, which is additionally quickened by exceptional equipment. To viably manage uncommon words, we utilize sub-word units (otherwise called "wordpieces") for sources of info and yields in our framework.

2.10 Use of Fuzzy Tool for Example Based Machine Translation(2016) [10]

This examination paper proposes work conveyed in machine translation. It demonstrates the examination of different methods like SMT, SSER, EBMT, RBMT and so on utilized for machine translation. In the wake of experiencing the aftereffect of different systems it infer that still better outcome can be picked up . In this manner proposes a thought of utilizing Fuzzy rationale apparatus to enhance the regular learning process and show signs of improvement result around 81.7 rate. This work demonstrates the usage of the instrument for refining the outcome in example based machine translation utilizing fluffy rationale.

2.11 Massive Exploration of Neural Machine Translation Architectures(2017) [11]

This work shows the principal thorough examination of engineering hyper parameters for Neural Machine Translation frameworks. Utilizing a sum of in excess of 250,000 GPU hours, we investigate regular varieties of NMT structures and give knowledge into which building decisions matter most. We report BLEU scores, perplexities, demonstrate sizes, and meeting time for all trials, including fluctuation numbers ascertained over a few keeps running of each trial. Moreover, we discharge to the general population another product system that was utilized to run the analyses.

2.12 Hybrid Approach for English-Hindi Machine Translation(2017)[12]

In this paper, an expanded consolidated approach of expression based measurable machine translation (SMT), example based MT (EBMT) and run based MT (RBMT) is proposed to build up a novel crossover information driven MT framework fit for beating the standard SMT, EBMT and RBMT frameworks from which it is determined. To put it plainly, the proposed cross breed MT process is guided by the control based MT subsequent to getting an Alignment of incomplete applicant translations gave by EBMT and SMT subsystems. Past works have demonstrated that EBMT frameworks are equipped for beating the expression based SMT frameworks and RBMT approach has the quality of producing basically and morphologically more exact outcomes. This half breed approach builds the familiarity, exactness and syntactic accuracy which enhance the nature of a machine translation framework. An examination of the proposed half breed machine translation (HTM) show with eminent translators i.e. Google, BING and Babylonian is additionally displayed which demonstrates that the proposed show works better on sentences with vagueness and in addition included expressions than others.

2.13 Experiments with matching algorithms in example based machine translation(2017),[13].

This paper displays a few coordinating calculations utilized for an example based machine translation framework amongst English and Spanish. The translation database was removed from the Web and changed as needs be for the motivations behind the framework. We will depict how a string-based coordinating calculation can be enhanced using morphological and semantic data.

2.14 Machine Translation Using Semantic Web Technologies: A Survey (2017)[14].

This paper presents an expansive number of machine interpretation approaches have been produced as of late with the point of moving substance effortlessly over dialects. Not withstanding, the writing proposes that numerous hindrances must be managed to accomplish better programmed interpretations. A focal issue that machine interpretation frameworks must deal with is vagueness. A promising method for beating this issue is utilizing semantic web advances. This article introduces the aftereffects of an efficient audit of methodologies that depend on semantic web advances inside machine interpretation approaches for deciphering writings. In general, our review proposes that while semantic web advances can improve the nature of machine interpretation yields for different issues, the mix of both is still in its earliest stages.

2.15 Machine Translation Journal (2018),[15].

Machine Translation is changing and widening its extent important to envelop all branches of Computational Linguistics and Language Engineering wherever they join a MULTILINGUAL angle. We in this way welcome entries to the diary on THEORETICAL DESCRIPTIVE OR COMPUTATIONAL ASPECTS of any of the accompanying themes: machine interpretation and machine-supported interpretation human interpretation hypothesis and practice multilingual content creation and age multilingual data recovery multilingual regular dialect interfaces multilingual exchange frameworks multilingual message understanding frameworks corpus-based and measurable dialect demonstrating connectionist ways to deal with interpretation aggregation and utilization of bi-and multilingual corpora talk wonders and their treatment in (human or machine) interpretation learning building contrastive etymology morphology sentence structure semantics pragmatics PC helped dialect direction and learning programming restriction and internationalization discourse preparing particularly for discourse interpretation phonetics phonology computational ramifications of non-Roman character sets multilingual word-handling the multilingual data society (sociological and lawful and in addition etymological perspectives) minority dialects history of machine interpretation. We would likewise welcome your proposals about different highlights you might want to find in this diary for instance uncommon issues squibs topical remark.

CHAPTER-3
SYSTEM DEVELOPMENT

3.1 TYPES OF MACHINE TRANSLATION

Since most recent a very long while, individuals have built up a number of translation ways to deal with change one dialect substance to another going from basic word-to-word translation frameworks to corpus based measurable models.

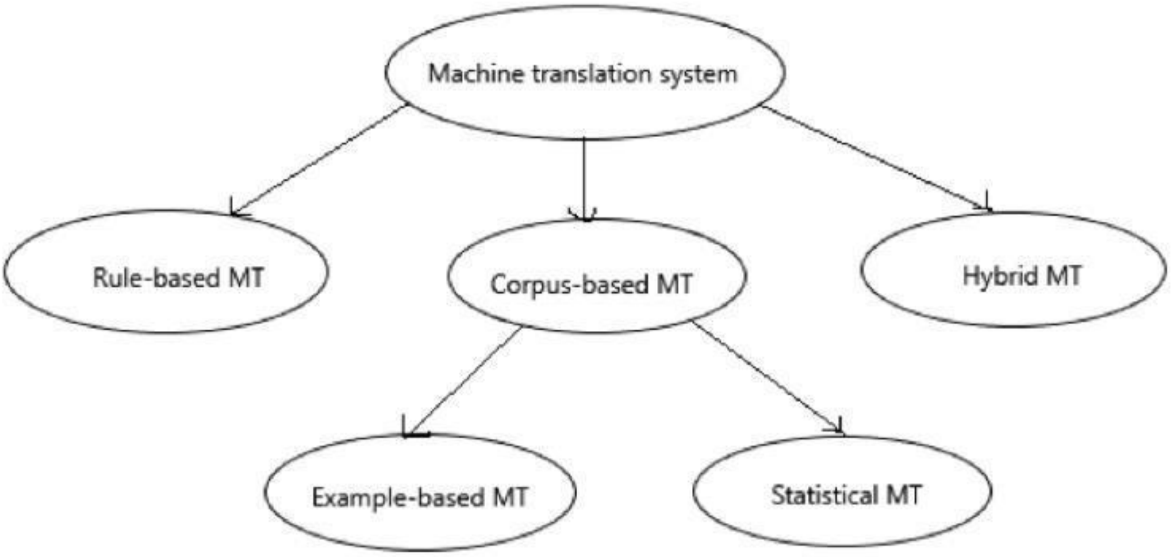


Figure3.1: Types of MT

3.1.1 RULE BASED MACHINE TRANSLATION

Rule-based machine translation framework is completed by design coordinating of the principles. The accomplishment lies in dodging the example coordinating of unproductive standards. Information and ideas are utilized for dialect understanding. General globe information is needed for tackling understanding issues, for example, disambiguation. Setting particular education can be utilized to decide the referent of thing phrases and disambiguating word faculties in view of what bodes well in the present circumstance. A learning portrayal consists of knowledge base and surmising techniques. Derivation strategies apply inference guidelines to get new sentences from the information base.

3.1.2 STATISTICAL BASED MACHINE TRANSLATION

A statistical approach based English-to-Hindi machine translation framework is produced comprising of three preparing units Language Model, Translation Model and Decoder. Dialect show ascertains the likelihood of a sentence in target dialect. Translation show intended to figure the target sentence likelihood for the given source sentence. Decoder's activity is to choose the objective sentence which amplifies the likelihood. The SMT display is prepared on the parallel dataset of 6000 sentences sets. Google translator which is an overall famous and for the most part utilized bilingual translator, is additionally in view of the SMT approach. Google translator learns the SMT parameters from their gigantic corpus gathered from all over the web. The SMT exactness relies upon the corpus quality and the parameter estimation required be to learn translator learns the SMT parameters from their huge corpus collected from all over the web. The SMT accuracy depends on the corpus quality and the parameter estimation needed be to learn.

3.1.3 EXAMPLE BASED MACHINE TRANSLATION SYSTEM

Example based machine translation frameworks (EBMT) perform the translation of a given information sentence s in three continuous stages (I) (matching) check for presence of the given info s in the bilingual corpus (ii) (retrieval) extraction of valuable fragments from the sentence that match in the bilingual corpus and (iii) (transfer) recombining the translated fragments. EBMT for all intents and purposes in light of the recovery of source sentences like s in the bilingual corpus, subsequently EBMT is otherwise called source-closeness based Translation. Example based machine translation concentrated on different instinctive issues of the EBMT like the measure of Parallel Corpora, Granularity of Examples, Amount of Examples and Suitability of Examples.

Table 3.1: Difference between example and statistical MT

EBMT	SMT
Example based MT frameworks utilize assortment of etymological assets, for example, lexicons and thesauri, and so forth., to translate content.	Statistical-based MT utilizes simply statistical based strategies in adjusting the words and age of writings.

3.1.4 HYBRID MACHINE TRANSLATION

Hybrid machine Translation is a strategy for machine Translation that consolidates attributes of different machine Translation approaches inside a solitary machine Translation framework . A multi-motor Hybrid way to deal with MT, using the measurable models to create the most ideal yield from numerous machine Translation frameworks. Promising outcomes for English-Hindi machine Translation on applying a choice tree technique to choose the best conceivable theory got from numerous RBMT, EBMT furthermore, SMT decoders has been discovered. The advantages of cross breed MT approaches as coupled different MT frameworks have the priority over using every MT independently . Obviously, it is been plainly noticeable that multi-motor MT approaches are equipped for outperforming the current person MT frameworks.

3.2. STAGES OF PROPOSED APPROACH

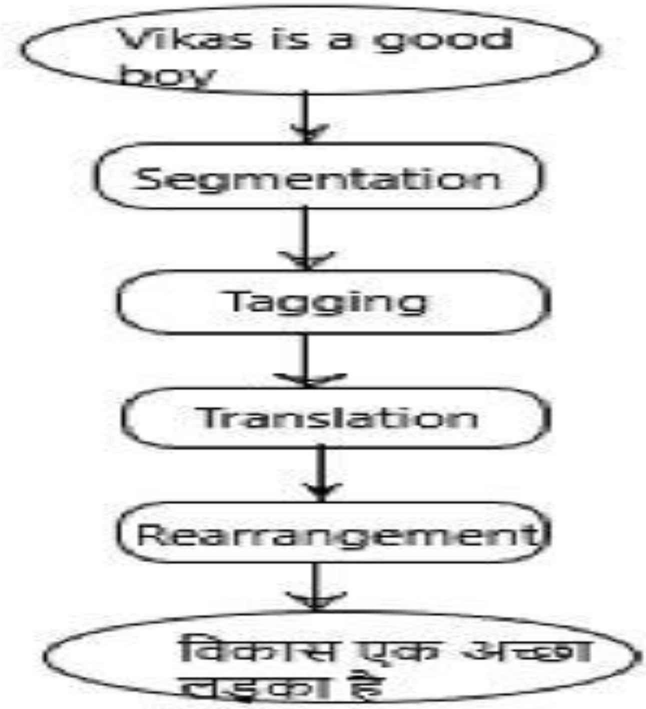


Figure 3.2 Stepwise procedure of the proposed approach

3.3 PROPOSED APPROACH FOR EXAMPLE BASED MACHINE TRANSLATION SYSTEM

In this chapter the design and implementation of the system has been discussed.
The essential thought of Example-Based Machine Translation (EBMT) is to use again examples of officially accessible translations as the reason for new translation. The procedure of EBMT is
separated into three phases:
3.3.1 Matching
3.3.2 Alignment
3.3.3 Recombination

3.3.1 MATCHING STAGE

The coordinating stage in example based machine translation discovers examples that will add to the translation based on their similitude with the info. The way coordinating stage ought to be actualized depends on how the examples are put away. In old frameworks, examples were put away as clarified tree structures and the constituents in the two dialects were associated by unequivocal connections. The contribution to be coordinated is parsed utilizing the syntax that was utilized to fabricate the example database and the tree is shaped. This tree is contrasted and trees in the example database.

The info and examples can be coordinated by looking at character by character. This procedure is called succession correlation. Alignment and recombination will be troublesome if this approach is utilized. Examples might be explained with Parts-Of-Speech labels. A few basic examples might be consolidated into a more broad single example containing factors.

The examples ought to be dissected to check whether they are appropriate for additionally handling. Covering or conflicting examples ought to be appropriately managed.

3.3.2 ALIGNMENT

Alignment is utilized to distinguish which parts of the comparing translation are to be reused. Alignment is finished by utilizing bilingual word reference or contrasting and different examples. The procedure of Alignment in example based machine translation must be robotized.

3.3.3 RECOMBINATION

Recombination is the last stage in example based machine translation approach. Recombination ensures that the reusable parts in example recognized amid alignment are assembling legitimacy. It takes source dialect sentences and an arrangement of translation patters as data sources and creates target dialect sentences as yields. The outline of recombination procedure relies upon past coordinating and arrangement stages.

3.4 DEVELOPMENT OF CORPUS

Example based Machine Translation framework makes utilization of a parallel corpus of source and target dialect sets. This parallel corpus is essential prerequisite before embraced preparing in Example based Machine Translation.The proposed framework has utilized parallel corpus of English and Hindi sentences. A parallel corpus of in excess of 5000 sentences has been created from which comprise of little sentences and the life history of opportunity contenders with reference to their trail in courts.

The specific EBMT framework that we are analyzing works in the following way. Given a broad corpus of adjusted source-dialect and target-dialect sentences, and a source-dialect sentence to translate:

1. It recognizes correct substrings of the sentence to be translated inside the source-dialect corpus, in this way restoring a progression of source-dialect sentences.
2. It takes the comparing sentences in the objective dialect corpus as the translations of the source-dialect corpus.
3. At that point for each combine of sentences:
3.1 It endeavors to adjust the source-and target-dialect sentences.
3.2 It recovers the part of the objective dialect sentence set apart as lined up with the corpus source-dialect sentence's substring and returns it as the translation of the information source-dialect lump.

The above framework is a specialization of summed up EBMT frameworks. Other particular frameworks may work on parse trees or just on whole sentences. The framework requires the following:
1. Sentence based alignment for source and target corpora.
2. Source- to target- corpus
3. Stemming algorithm

3.5 INDEXING

With a specific end goal to encourage the look for sentence substrings, we have to make a reversed record into the source-dialect corpus. To do this we circle through every one of the expressions of the corpus, including the present area (as characterized by sentence file in corpus and word file in sentence) into a hash table keyed by the fitting word. With a specific end goal to spare time in future runs we spare this to a file document.

3.6 CHUNK SCRUTINIZING AND COMPREHENDING

Keep two arrangements of chunks: current and completed. Circling through all words in the objective sentence: See whether areas for the present word broaden any lumps on the present rundown On the off chance that they do, broaden the chunk. Discard any chunks that are 1-word. These are rejected. Move to the finished rundown those chunks that were not able proceed. Begin another present chunk for every area. Toward the end, dump everything into finished. At that point, to prune, run each lump against each other: On the off chance that a lump appropriately subsumes another, evacuate the littler one. On the off chance that two lumps are e equivalent and we have excessively numerous of them, evacuate one.

3.7 POS TAGGING

The way toward relegating one of the parts of discourse to the given word is called Parts Of Speech labeling. It is ordinarily alluded to as POS labeling. Parts of discourse incorporate things, verbs, intensifiers, descriptors, pronouns, conjunction and their sub-classes.

Example:
Word: Wood,
Tag: Noun
Word: Run, Tag:
Verb
Word: Attractive, Tag: Adjective

Parts Of Speech tagger or POS tagger is a program that does the activity of tagging. Taggers utilize a few sorts of data: word references, dictionaries, principles, et cetera. Lexicons have class or classifications of a specific word. That is a word may have a place with in excess of one classification. For instance, run is both thing and verb. Taggers utilize probabilistic data to comprehend this equivocalness.

There are principally two sort of taggers: control based and stochastic. Run based taggers utilize manually written principles to recognize the label equivocalness. Stochastic taggers are either HMM based, picking the label arrangement which augments the result of word probability and label succession likelihood, or signal based, utilizing choice trees or most extreme entropy models to consolidate probabilistic highlights.

Preferably a run of the mill tagger ought to be powerful, proficient, precise, tunable and reusable. In all actuality taggers either unquestionably distinguish the tag for the given word or make the best figure in light of the accessible data. As the characteristic dialect is perplexing it is here and there troublesome for the taggers to settle on precise choices about labels. So infrequent mistakes in labeling isn't taken as a noteworthy barricade to inquire about.

3.8 ALIGNMENT

The alignment algorithm is as follows:
1. Stem the expressions of determined source sentence
2. Look up those words inxa translation corpus.
3. Stem the expressions of the previously defined target sentence.
4. Try to coordinate the objective words with the source words—wherever they coordinate, stamp the correspondence table.
5. Prune the table to expel far-fetched word correspondences.
6. Take just as much target message as is fundamental keeping in mind the end goal to cover all the remaining correspondences for the source dialect chunk.

The pruning calculation depends on the way that solitary words are not frequently savagely uprooted from their unique position. This presumption is valid amongst English and others; nonetheless, striking exemptions may (yet not really) incorporate the oft-referred to non-SVO dialects Hindi, Korean, Japanese, and Arabic. Likewise, the pruning calculation works best when most word correspondences are 1-to-1.

3.9 FLOW CHART OF THE PROPOSED ALGORITHM

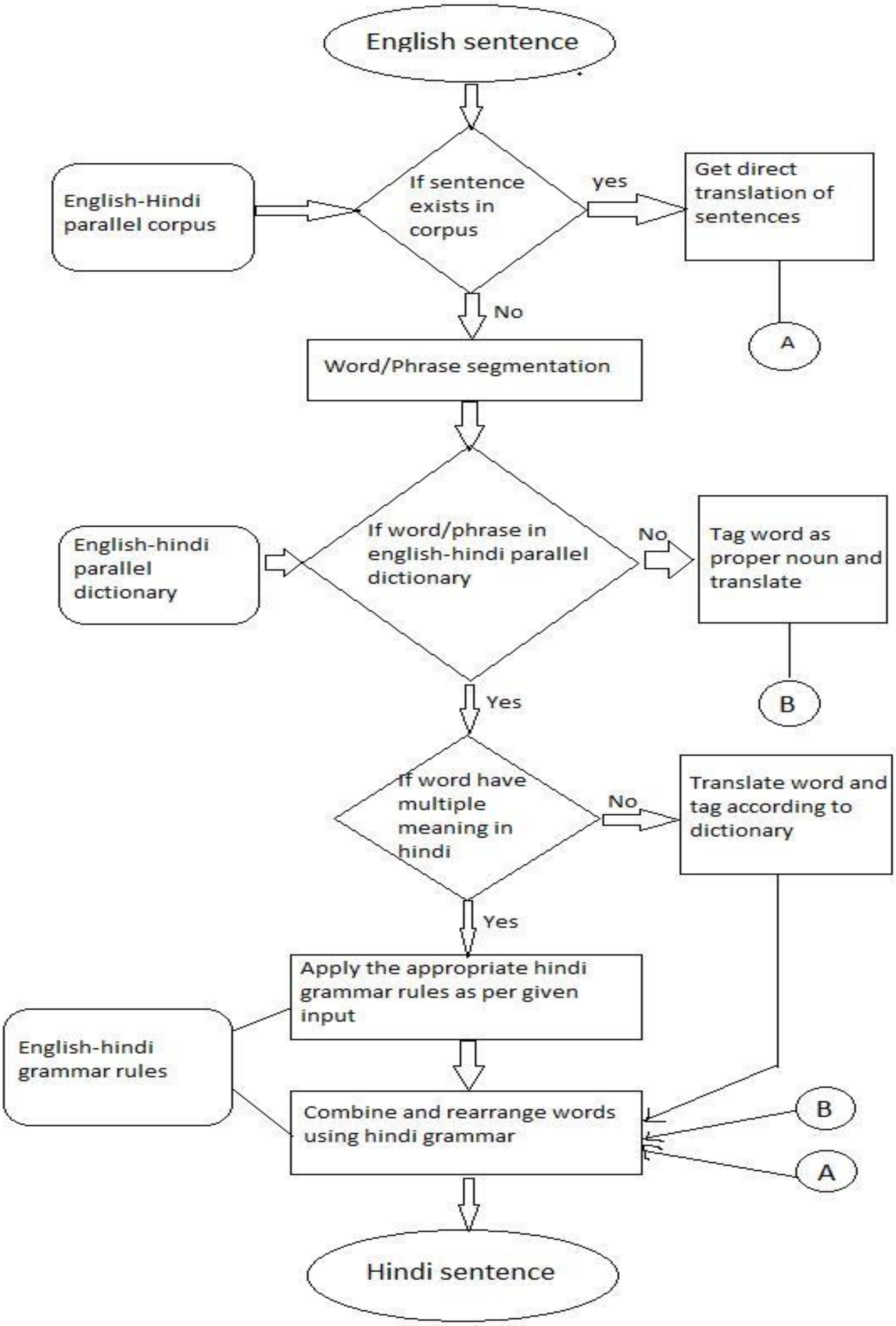

English sentence
English-Hindi parallel corpus
If sentence exists in corpus
yes
Get direct translation of sentences
A
No
Word/Phrase segmentation
English-hindi parallel dictionary
If word/phrase in english-hindi parallel dictionary
No
Tag word as proper noun and translate
B
Yes
If word have multiple meaning in hindi
No
Translate word and tag according to dictionary
Yes
Apply the appropriate hindi grammar rules as per given input
English-hindi grammar rules
Combine and rearrange words using hindi grammar
B
A
Hindi sentence

Figure 3.3 Proposed Algorithm

3.10 MACHINE TRANSLATION OF IDIOMS FROM ENGLISH TO HINDI

Machine translation (MT) is characterized as the utilization of a PC to decipher a content from one regular dialect, the source dialect (SL), into another, the objective dialect (TL), Machine translation is a PC application to the errand of dissecting the source message in one human dialect and creating a proportional content called „translated text" or „target text" in the other dialect with or without human help as it might require a pre-altering and a post-altering stage. Each dialect has its own colloquialisms, a unique sort of set articulations that include created inside a dialect. English and Hindi are inexhaustible in colloquialisms. A standout amongst the most critical parts of English is colloquialisms. They are much of the time utilized as a part of a wide assortment of circumstances, from well disposed discussions and conferences to more formal and composed settings. A colloquialism is a gathering of words which has, all in all, an alternate significance from the importance of its constituents. In different words, the significance of the informal articulation isn't the entirety of the words taken exclusively. Sayings are genuinely disputable. There is nobody set meaning of what a colloquialism is. The word itself comes either from Latin idiom, where it indicates uncommon property, or from Greek idiom, which means extraordinary element, unique expressing. Thus, the rationale forces relationship with components of dialect expresses that are run of the mill for a given dialect and, in this way, difficult to convert into another dialect. An informal articulation may pass on an alternate significance, that what is apparent from its words. For instance English: It's down-pouring like crazy Hindi translation By Google: यह पागल की तरह नीचे डालना है Clearly, the yield does not pass on the planned significance in target dialect.

3.10.1 ENGLISH LANGUAGE

English is presently the most broadly utilized dialect on the planet it is the third most regular local dialect on the planet, with in excess of 380 million local speakers. English Language is composed in Roman content. It is a West Germanic dialect that emerged in the Anglo-Saxon kingdoms of England. It is one of six authority dialects of the United Nations. India is one of the nations where English is talked as a moment dialect.

3.10.2 HINDI LANGUAGE

Hindi is one of the real dialects of India. It is the fifth most talked dialect on the planet with in excess of 180 million local speakers. It is composed in the Devanagari content. It is the national dialect of India and is the world second most talked dialect.

3.11 PROBLEMS IN IDIOM TRANSLATION

The translation issue is any kind of trouble in the source dialect (SL) message that obliges the translator to quit deciphering. This trouble is principally because of linguistic, social or lexical issues.

3.11.1 GRAMMATICAL PROBLEMS

Syntactic issues are the aftereffect of confused SL syntax, diverse TL language structure or distinctive TL word arrange. For instance, the word request of English and Hindi isn't same. English takes after SVO conspire while Hindi Follows SOV plot. Think about after figure of speech in English: "Add fuel to flame" Corresponding Hindi sentence is आग में घी का काम करना. Here in English expression, "fire" is finally position though in Hindi its partner आग is at first position of the expression.

3.11.2 CULTURAL PROBLEMS

Various issues might be brought up in diverse translation. The more noteworthy the hole between the source and target culture, the more genuine trouble would be. Translation amongst English and Hindi which have a place with two distinct societies (the Western and the Indian societies), and which have an alternate foundation is a best Example of such issues. Social issues may incorporate geological, religious, social and semantic ones. Consequently, the articulation "summer’s day" in „Shall I contrast thee with a summer‟s day‟ will be best converted into Hindi as ग्रीष्मऋतुto pass on a similar importance.

3.11.3 WORD SENSE VAGUENESS

This issue happens when there are various elucidation of words or sentence. Among these issues we have:

3.11.3 PHRASE LEVEL VAGUENESS

Phrase level vagueness happens when an expression can be translated in excess of one ways. For instance the articulation ‘spill the beans’ may allude to the beans that are really spilled or colloquially the expression may allude to spill out mystery data.

3.11.4 WORD LEVEL VAGUENESS

The word vagueness passes on the various understandings of words. For instance to hold up under the lion in his nook As bear have the various implications भालू,कष्टउठाना, फऱदेना, उत्पन्नकरना, रीछ, ऱेजाना

3.11.5 DIFFERENT METHODOLIGIES OF IDIOMS IN MACHINE TRANSLATION

The expression "methodology" alludes to a technique utilized to decipher a given component unit making utilization of at least one systems chose based on important parameters. presents a progression of techniques utilized by proficient translators.

3.11.6 USING AN IDIOM OF SAME MEANING AND STRUCTURE

It includes utilizing a colloquialism in the objective dialect which passes on generally an indistinguishable importance from that of the source-dialect figure of speech and, also comprises of proportionate lexical things. Example: to rub salt in wounds जले पर नमक छिड़कना

3.11.7 USING AN IDIOM OF SAME MEANING BUT DISSIMILAR STRUCTURE

Figures of speech of comparable importance however disparate frame allude to those having totally extraordinary implications and the events in which the colloquialisms are utilized are not alike also. Example: To rest like a log घोड़े बेच कर सोना

3.11.8 USING AN IDIOM TRANSLATION BY PARAPHRASE

Where the articulation is regularly revised utilizing different words to improve it and afterward translate. Example:

The suspension framework has been completely up rated to take unpleasant landscape in its walk.

निलंबन ढांचे को पूरी तरह से अपने चलने में अप्रिय परिदृश्य लेने के लिए रेट किया गया है।

Furthermore, The limit of the suspension framework has been raised in order to conquer the unpleasantness of the territory.

क्षेत्र की अप्रियता को जीतने के लिए निलंबन ढांचे की सीमा को उठाया गया है।

Second Example is more coherent in target dialect.

3.11.9 USING AN IDIOM TRANSLATION BY OMISSION

On the off chance that the saying has no nearby match, the framework can just exclude the figure of speech in target dialect. The importance won't be hurt, if this method is utilized when the words which will be overlooked are not essential to the advancement of the content. Translators can just exclude deciphering the word or the articulation being referred to.

3.12 MOTIVATION

Most of the current English to Hindi translation framework which makes the translation of English content into Hindi content does not extricate expressions from the information content amid translation. However, this is exceptionally uncommon that Idioms are available in the information content for MT System yet there is a need to remove Idioms from input message and decipher them accurately. So we built up a calculation for finding and translating English Idioms show in the information message and make an translation of them into Hindi content.

Example:
Sentence in English: He has settled his record with me
Output for this: वह मेरे साथ अपने खाते में बसे है
Clearly, the yield isn't coherent. In any Example, in the event that we by one means or another, find and supplant the figures of speech in above sentence as take after
He has चुकता किया हुआ his हिसाब किताब _with me and translate it with goggle decipher framework we get: वह चुकता किया हुआ उसके हिसाब किताब मेरे साथ है which is very comprehensible and much superior to anything past yield and subsequently spur us to work toward this path.

3.13 IMPLEMENTATION OF IDIOMS USING EBMT

Here, the point is to plan a framework for recognizing figures of speech and process them. This handled sentence will be then utilized as contribution by translation framework. The framework engineering is as take after

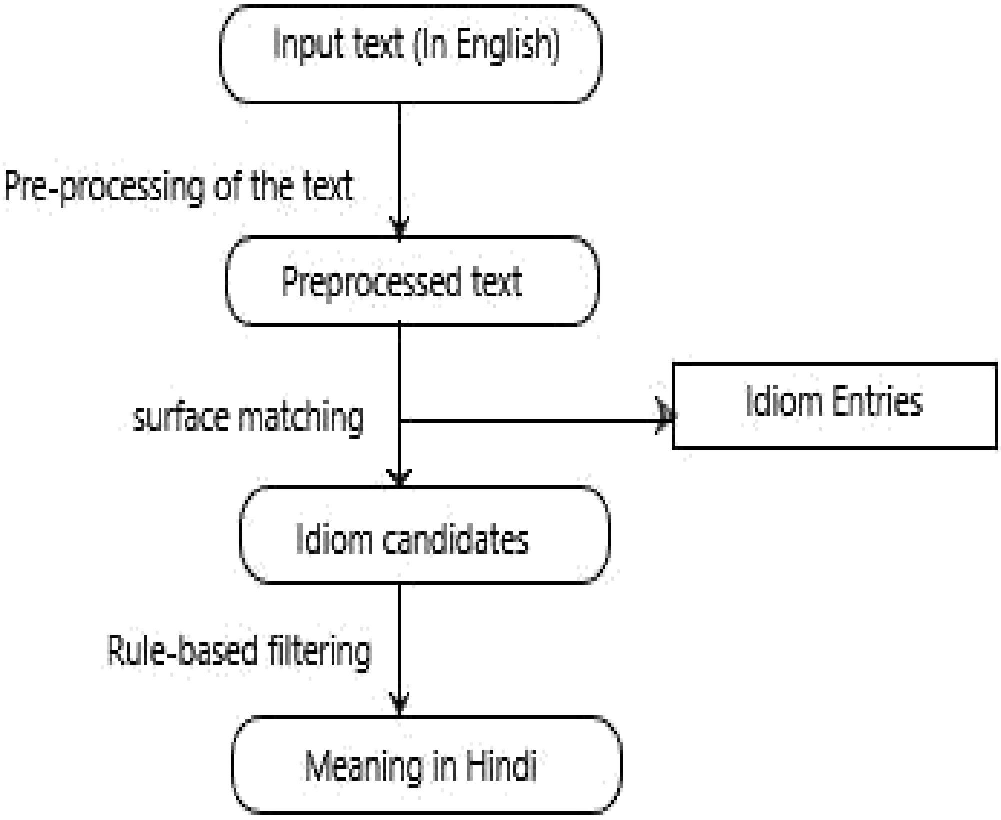


Figure 3.4 Flowchart for idioms

The framework comprises of three modules which incorporates Preprocessing (Paradigmatic substitution, syntagmatic increase, cancellation, Replacing arched type of verbs, Replacing Plural types of Nouns, articles, individual pronouns), Surface coordinating (FilteringPart-of -discourse labeling and lumping designs, distinguishing colloquialism applicants) and Post preparing module.

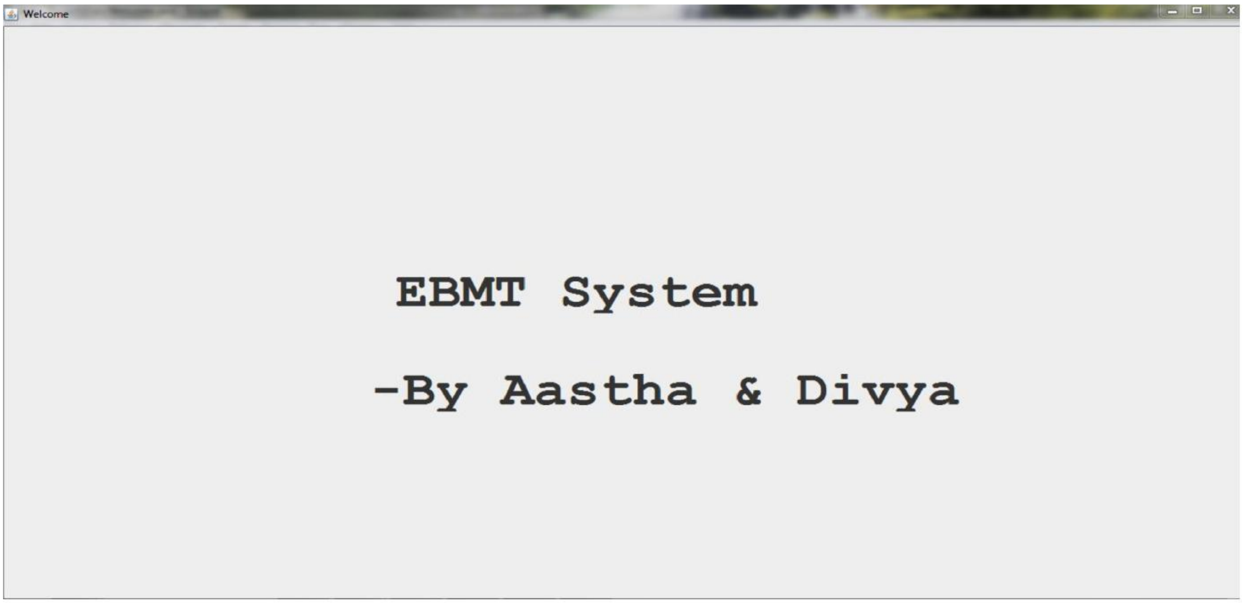
An English to Hindi Idiom Translator System Interface is made whose question will acknowledge a string in English dialect and returns its relating Hindi string. At the point when the client input an English content ,and taps the "hunt" catch ,the framework yields the figure of speech hopefuls with their significance (in Hindi).Here, we introduce the utilization of standard articulations in an translation framework for removing and deciphering the English sayings.

(a)Idiom Extraction utilizing Pattern coordinating: If there is an expression in sentence, it is extricated from the sentence by utilizing Pattern coordinating. Pursuit design coordinates the Idiom in the sentence and concentrates the Idioms from the sentence. Translating

(b)English maxims to Hindi idioms. Now, the removed Idioms is supplanted with the identical Hindi importance of that Idiom. It implies English figure of speech is converted into Hindi expressions.

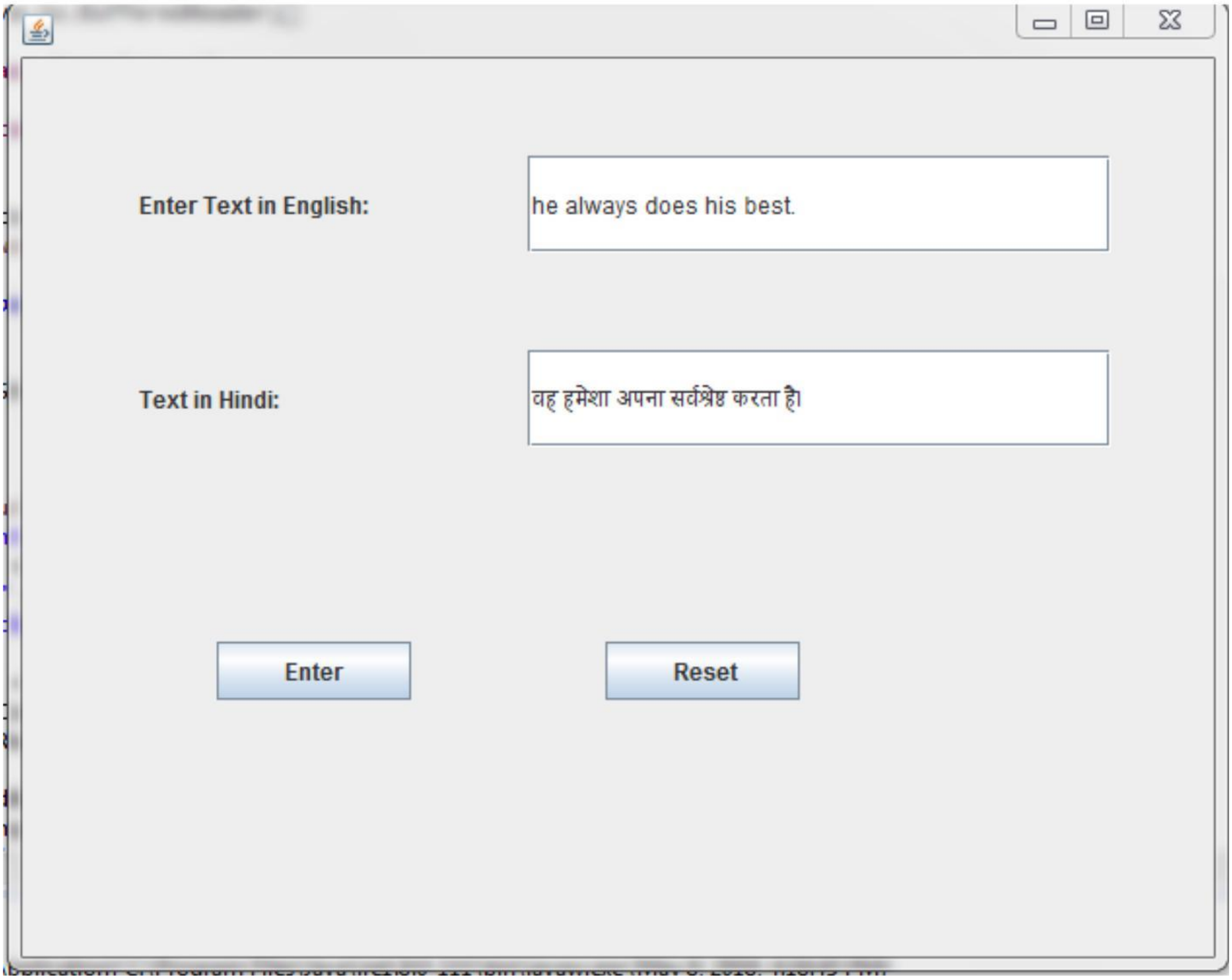
CHAPTER-4
PERFORMANCE ANALYIS

4.1 RESULTS:



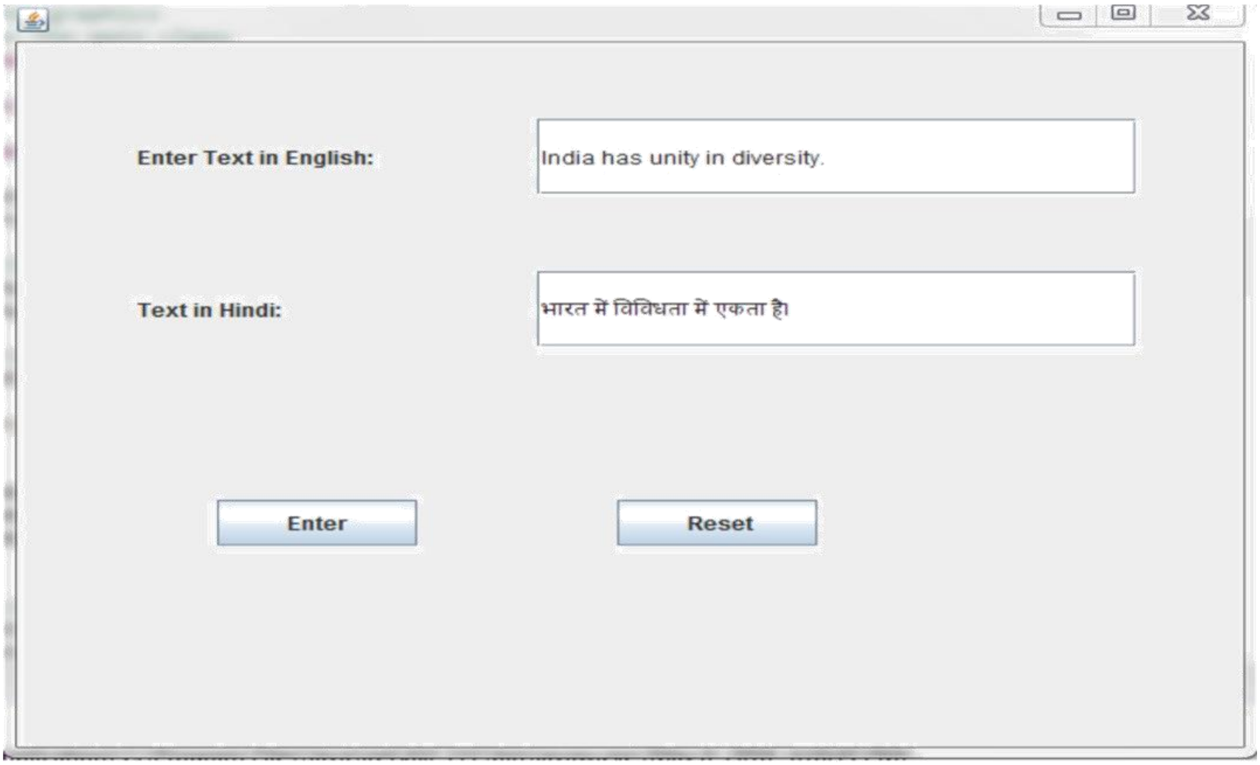
EBMT System

-By Aastha & Divya



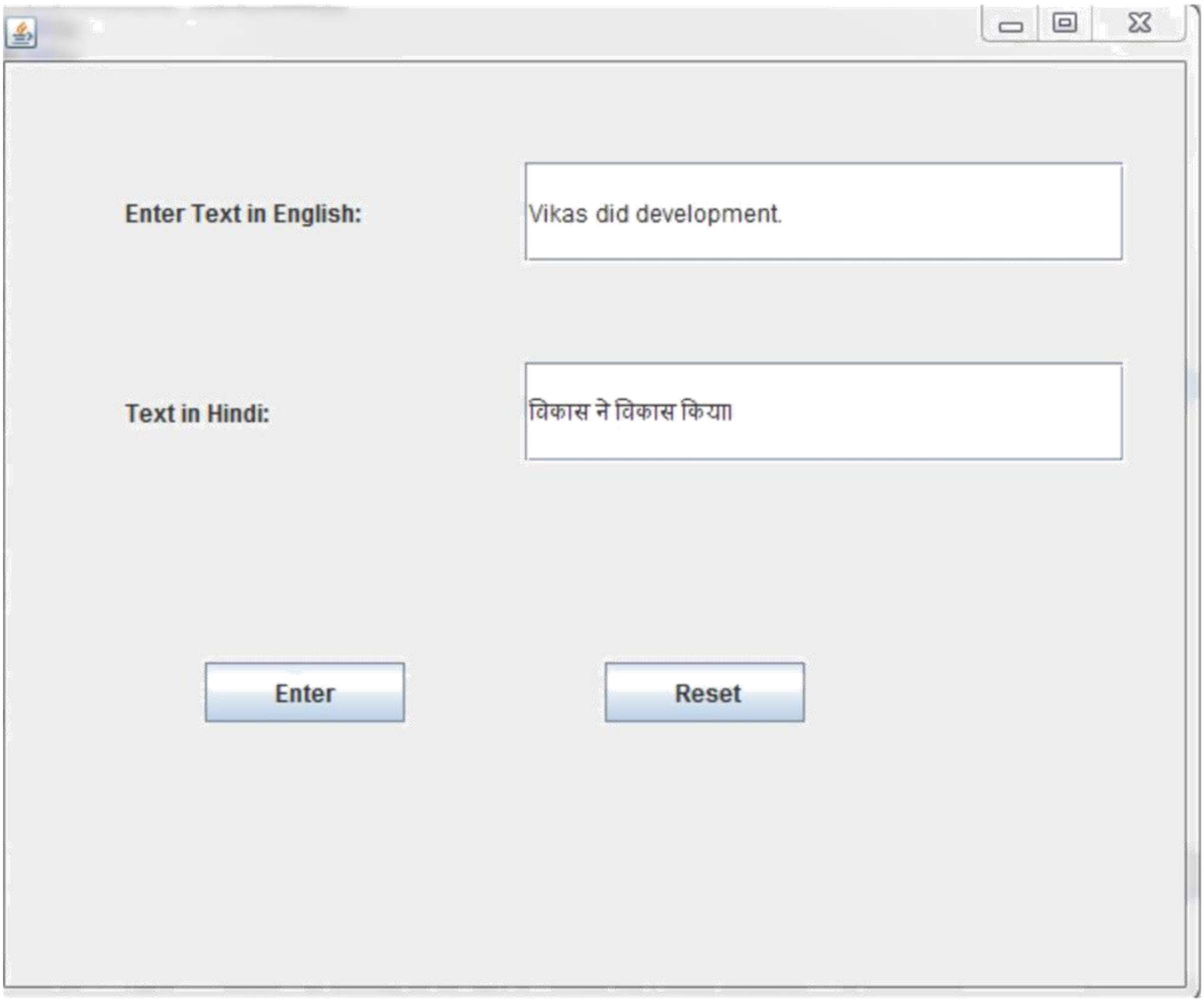
Enter Text in English: he always does his best.

Text in Hindi: वह हमेशा अपना सर्वश्रेष्ठ करता है।



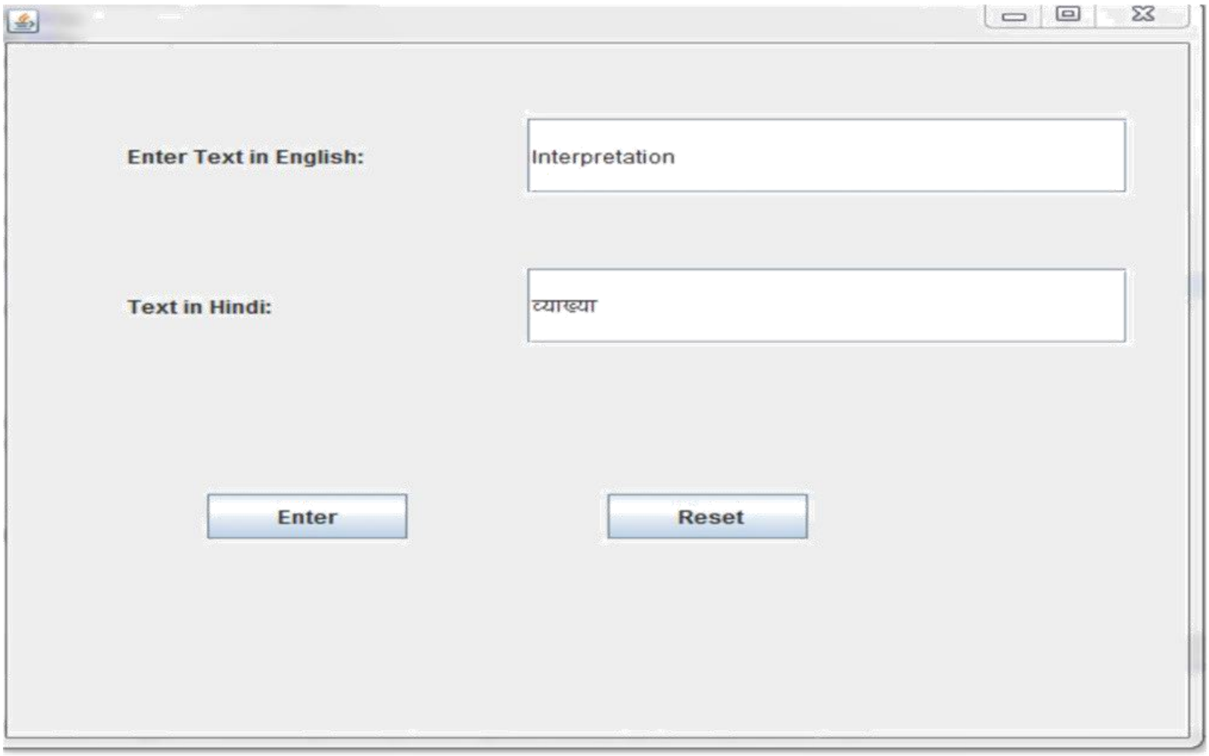
Enter Text in English: India has unity in diversity.

Text in Hindi: भारत में विविधता में एकता है।



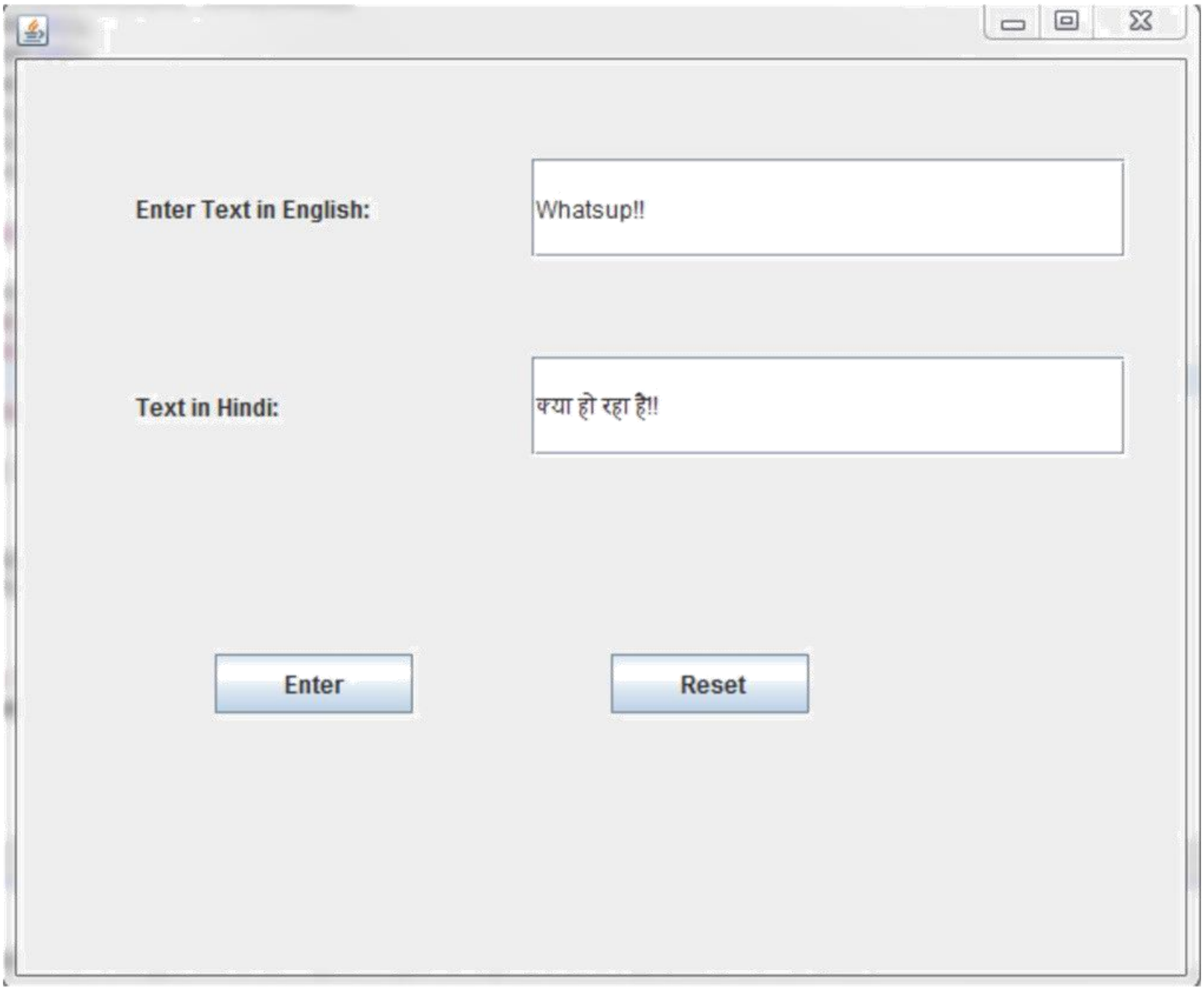
Enter Text in English: Vikas did development.

Text in Hindi: विकास ने विकास किया।



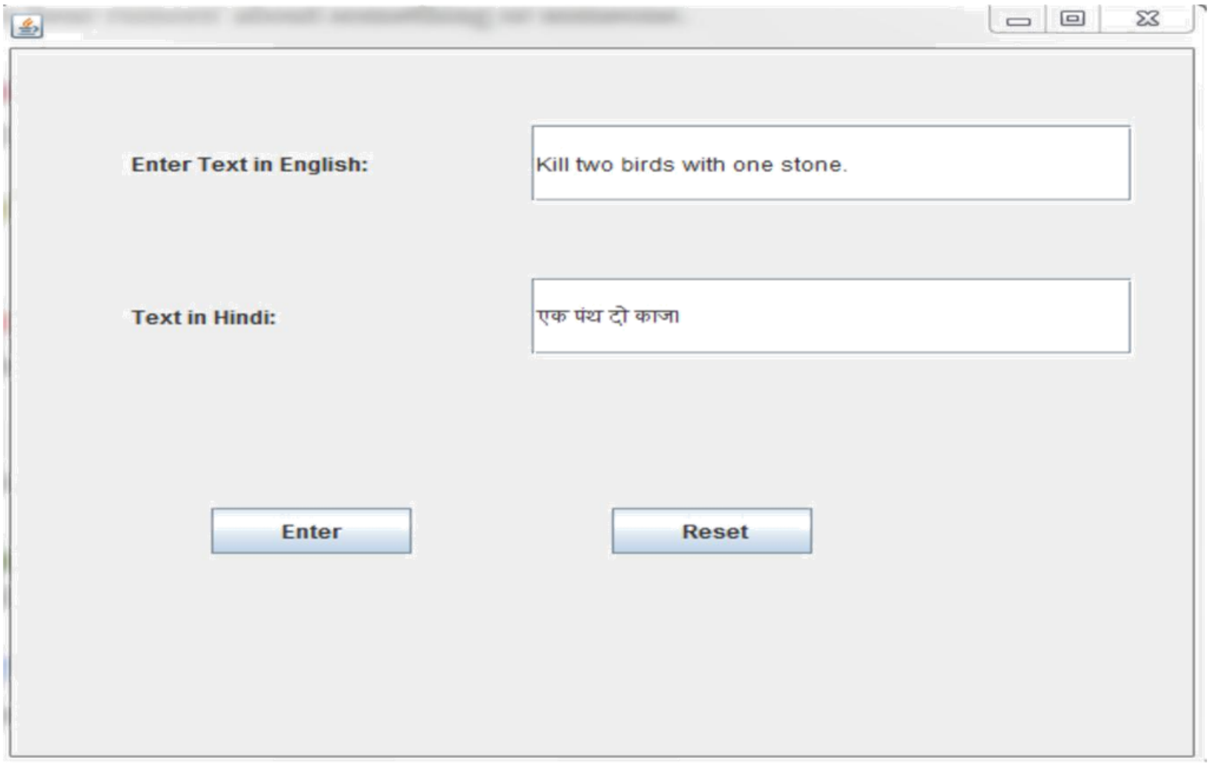
Enter Text in English: Interpretation

Text in Hindi: व्याख्या



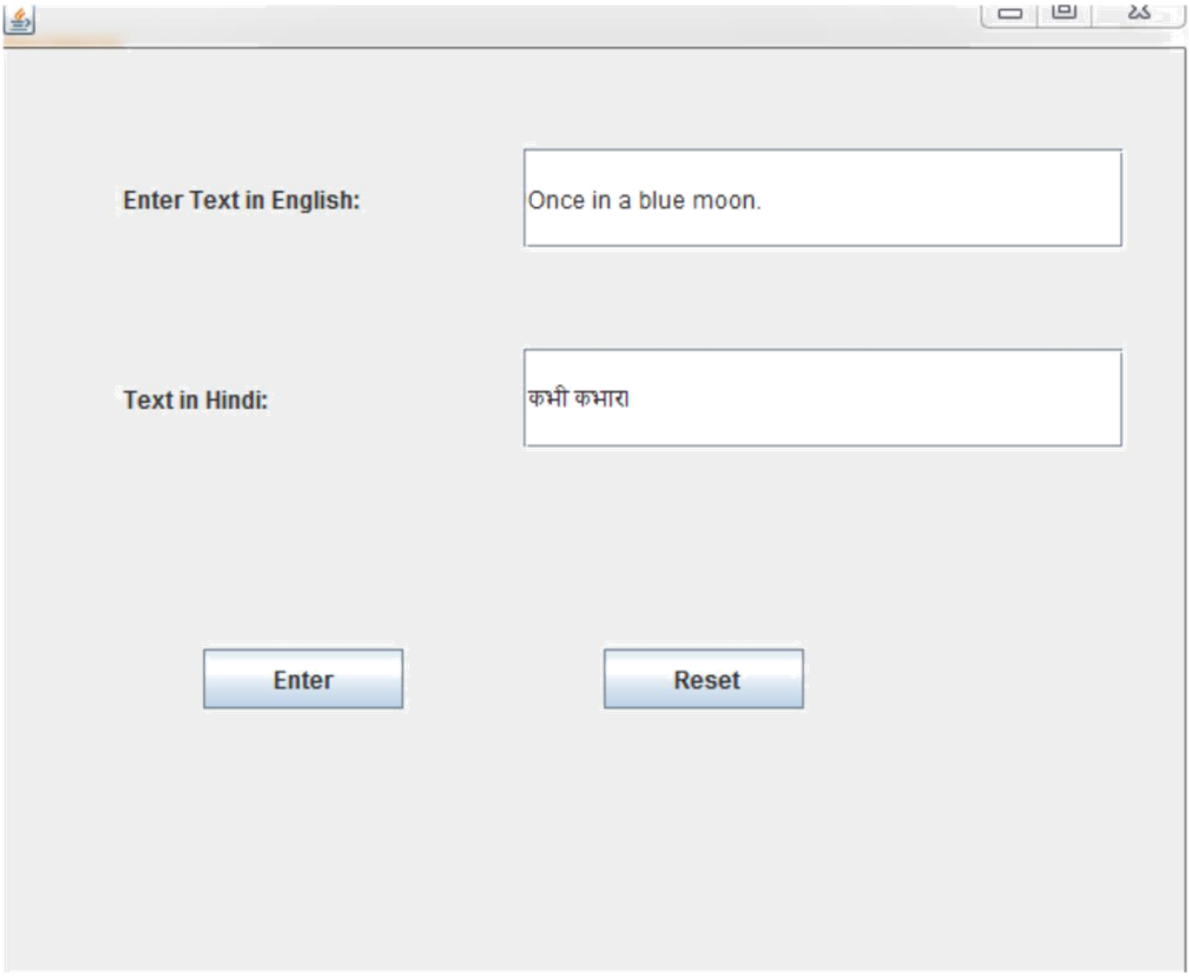
Enter Text in English: Whatsup!!

Text in Hindi: क्या हो रहा है!!



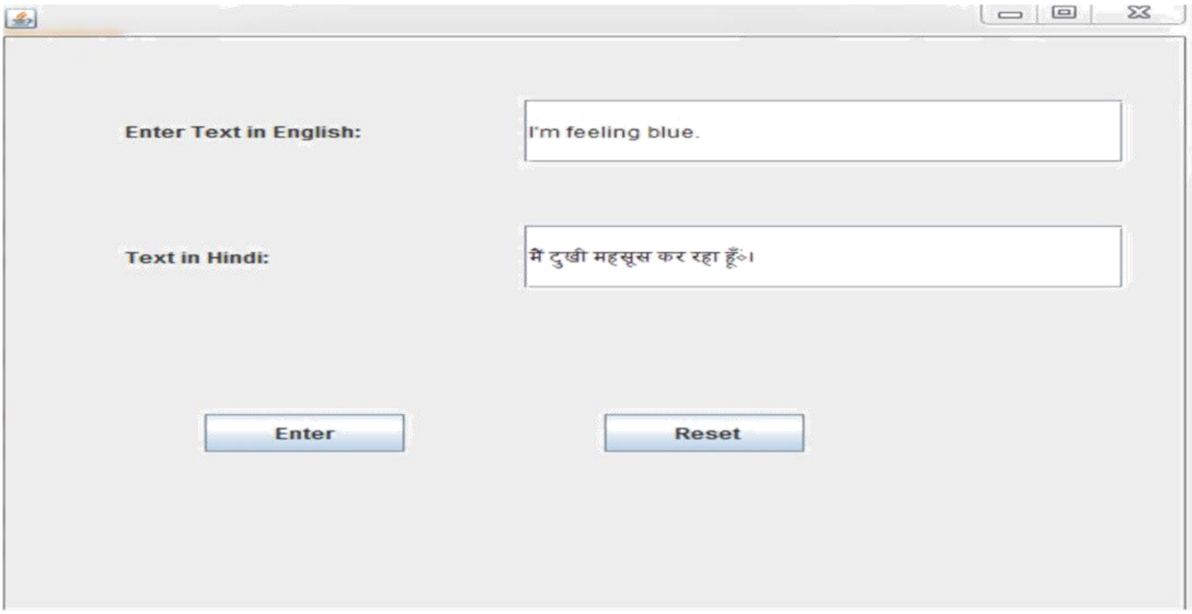
Enter Text in English: Kill two birds with one stone.

Text in Hindi: एक पंथ दो काज।



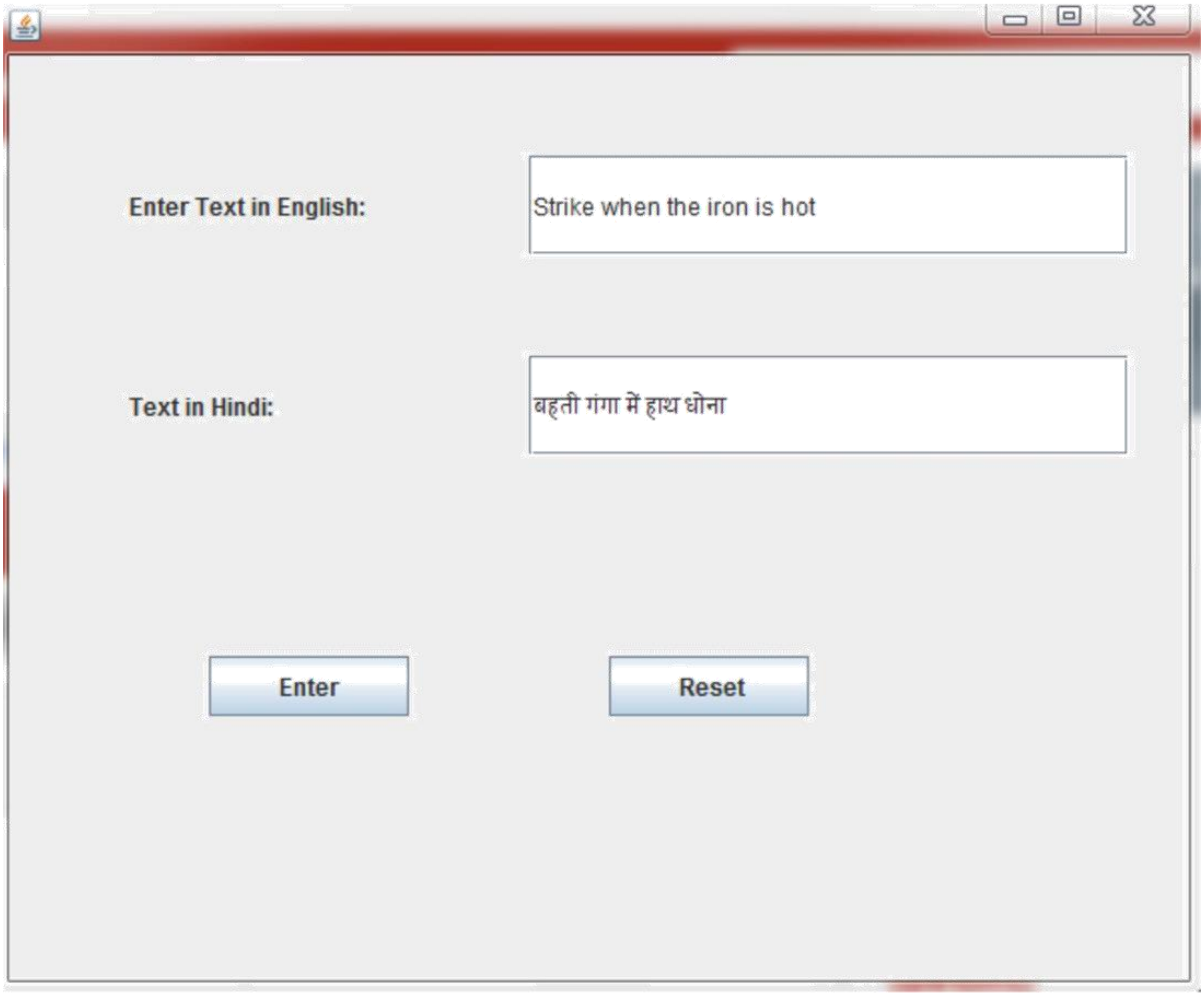
Enter Text in English: Once in a blue moon.

Text in Hindi: कभी कभार।



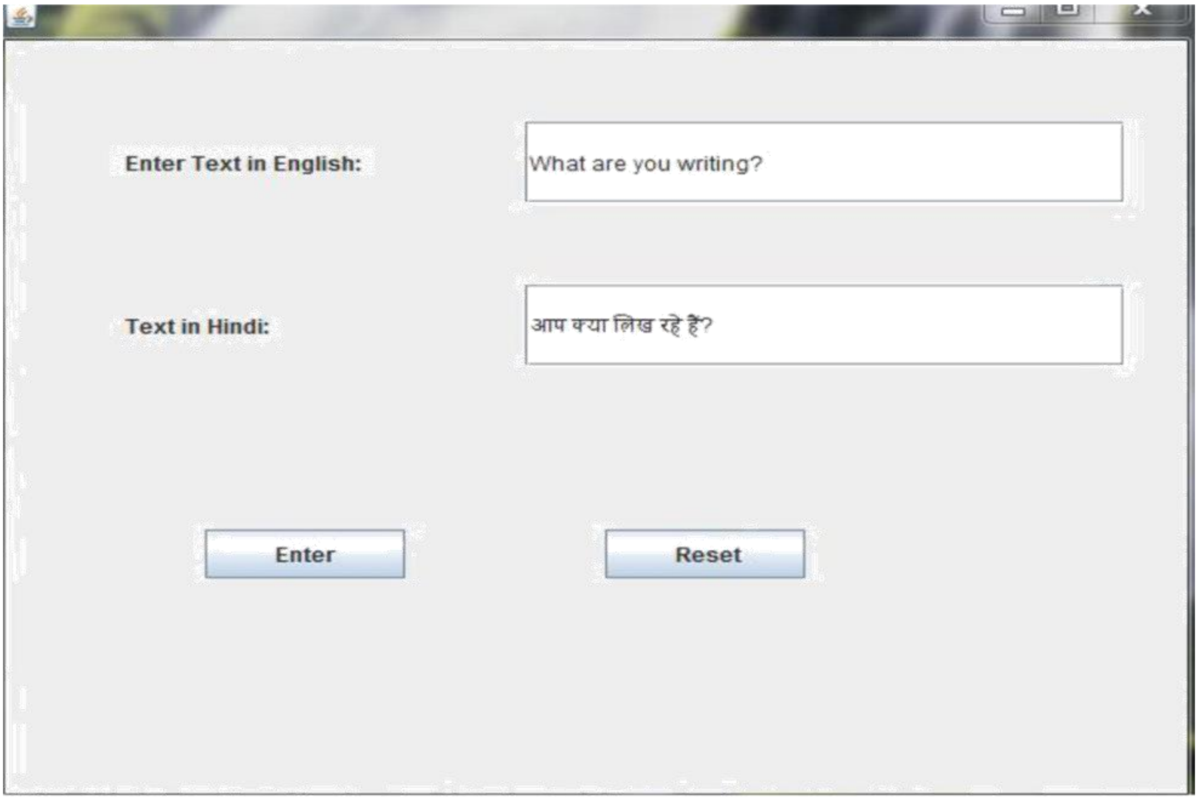
Enter Text in English: I'm feeling blue.

Text in Hindi: मैं दुखी महसूस कर रहा हूँ◌।



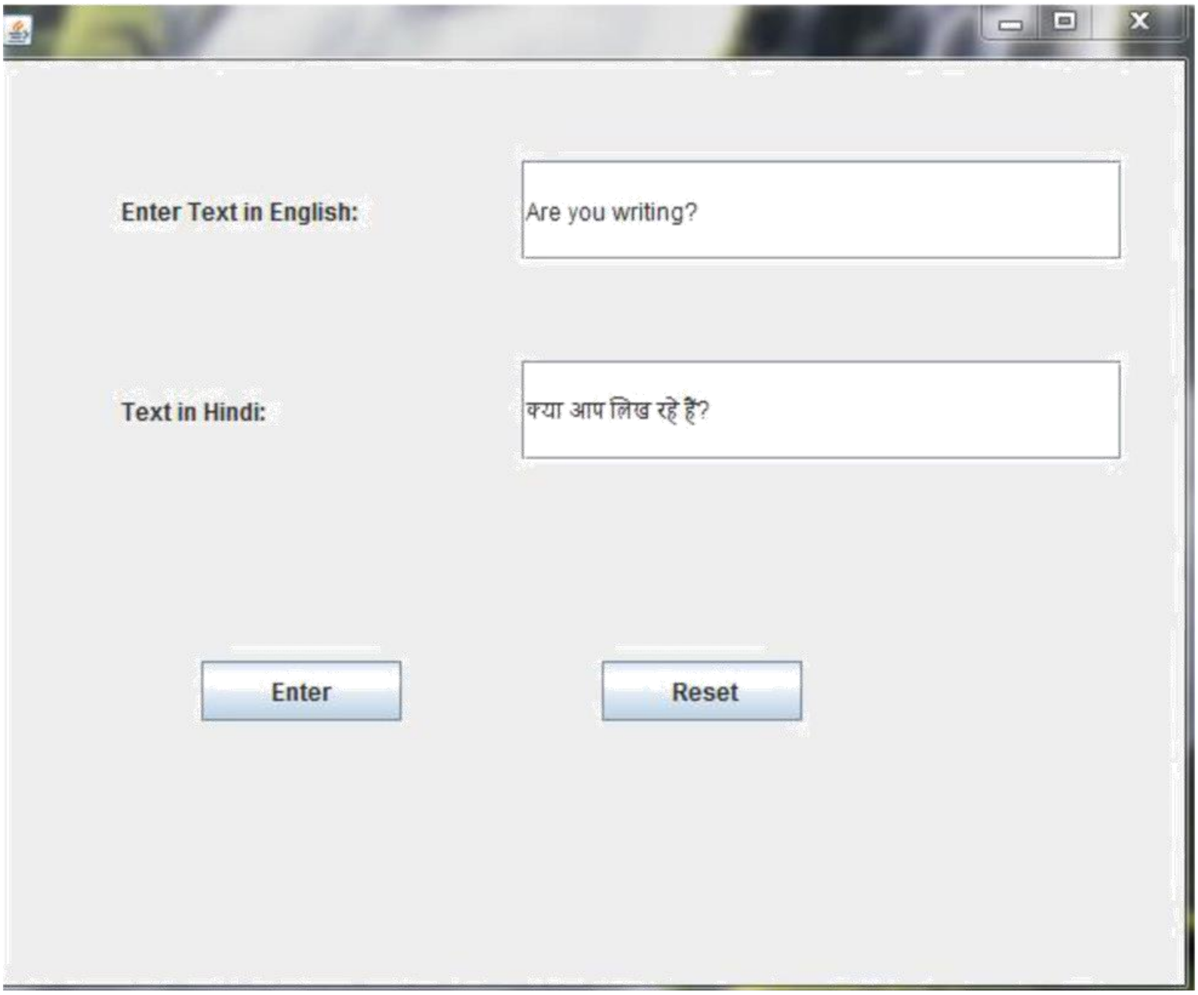
Enter Text in English: Strike when the iron is hot

Text in Hindi: बहती गंगा में हाथ धोना



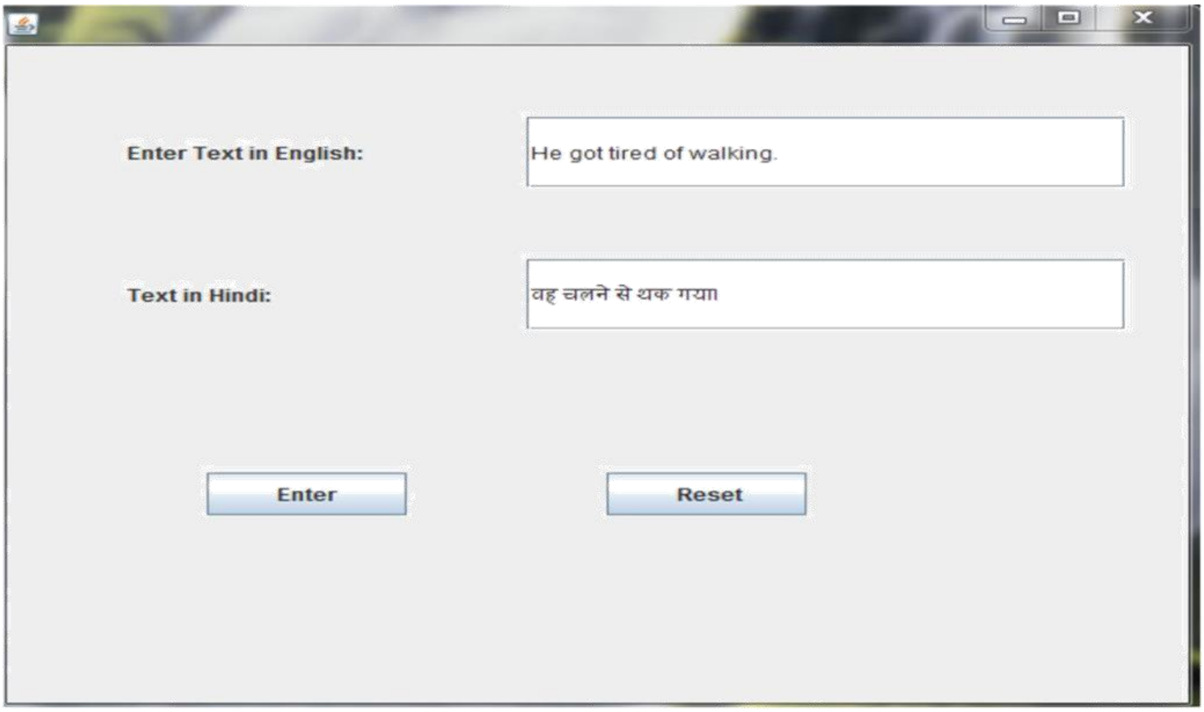
Enter Text in English: What are you writing?

Text in Hindi: आप क्या लिख रहे हैं?



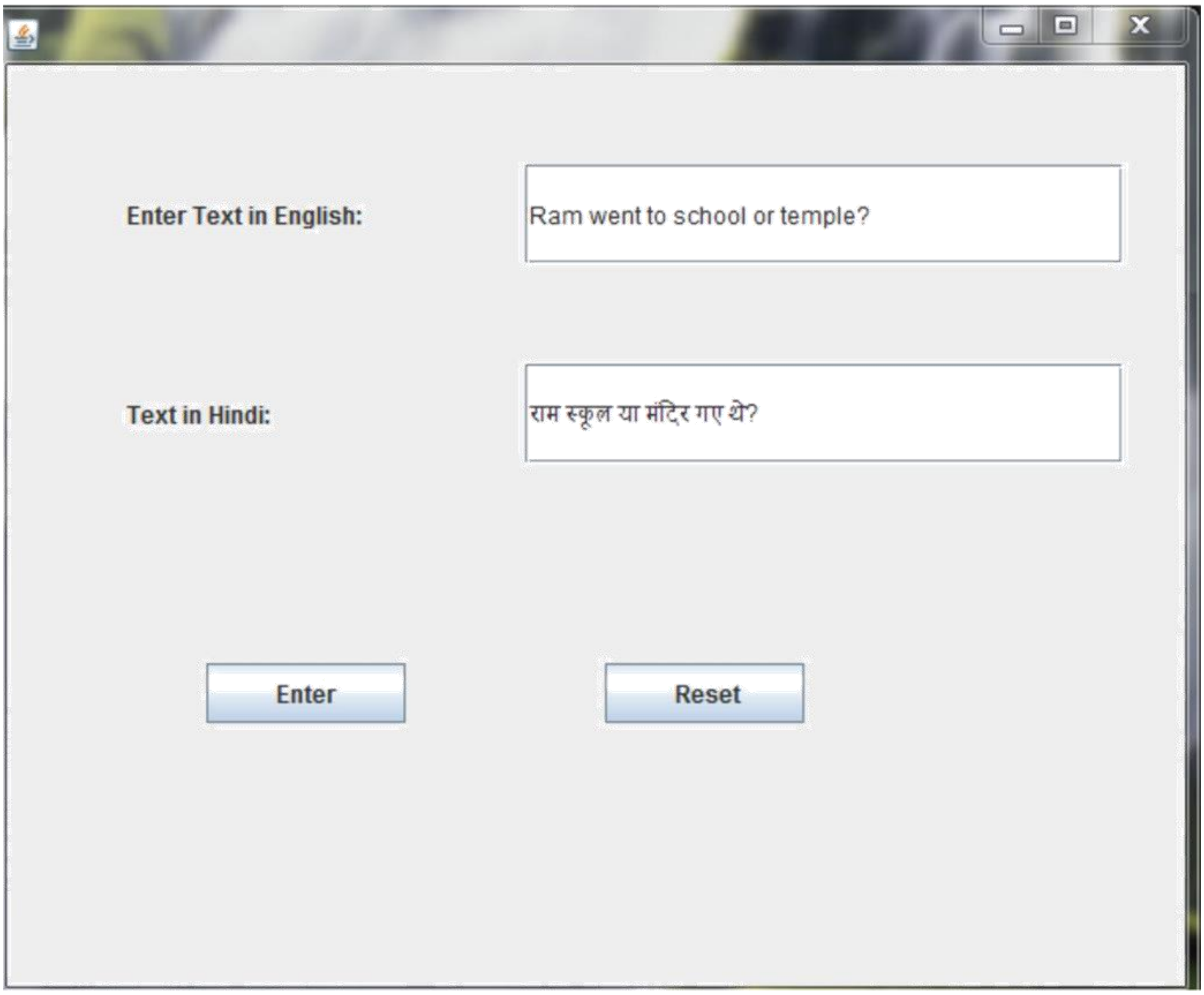
Enter Text in English: Are you writing?

Text in Hindi: क्या आप लिख रहे हैं?



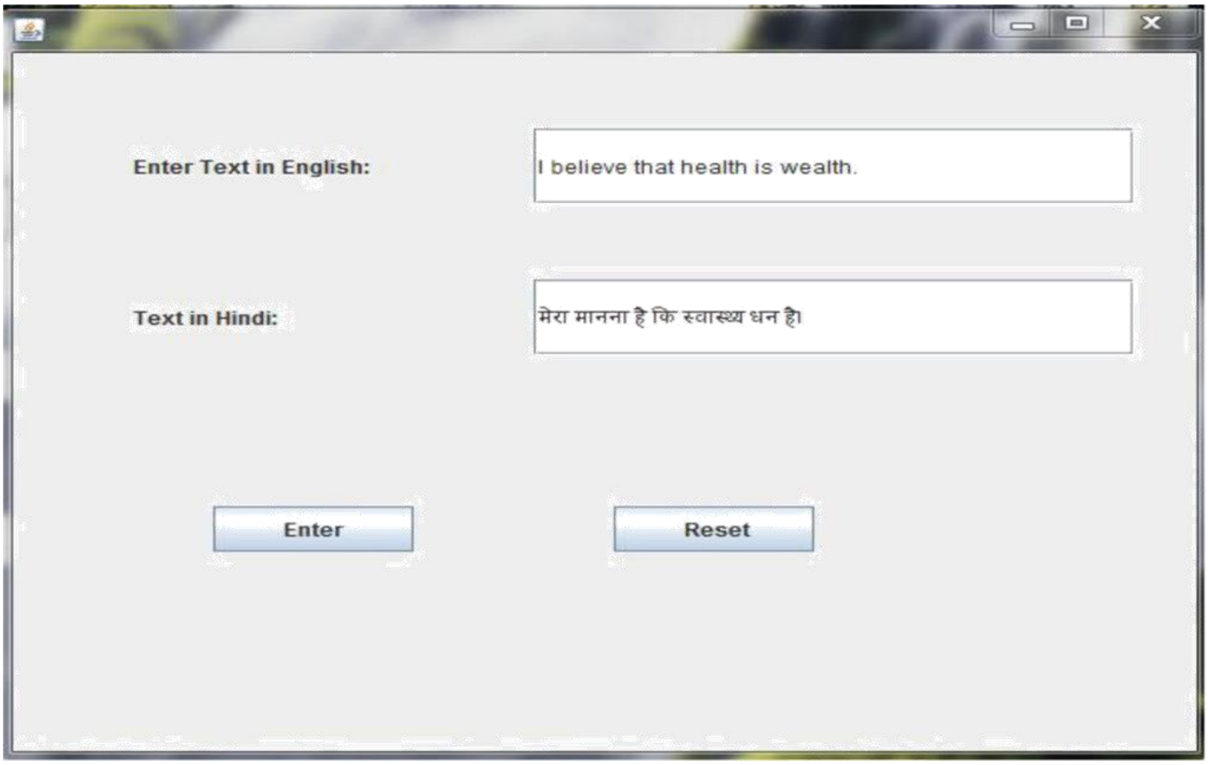
Enter Text in English: He got tired of walking.

Text in Hindi: वह चलने से थक गया।



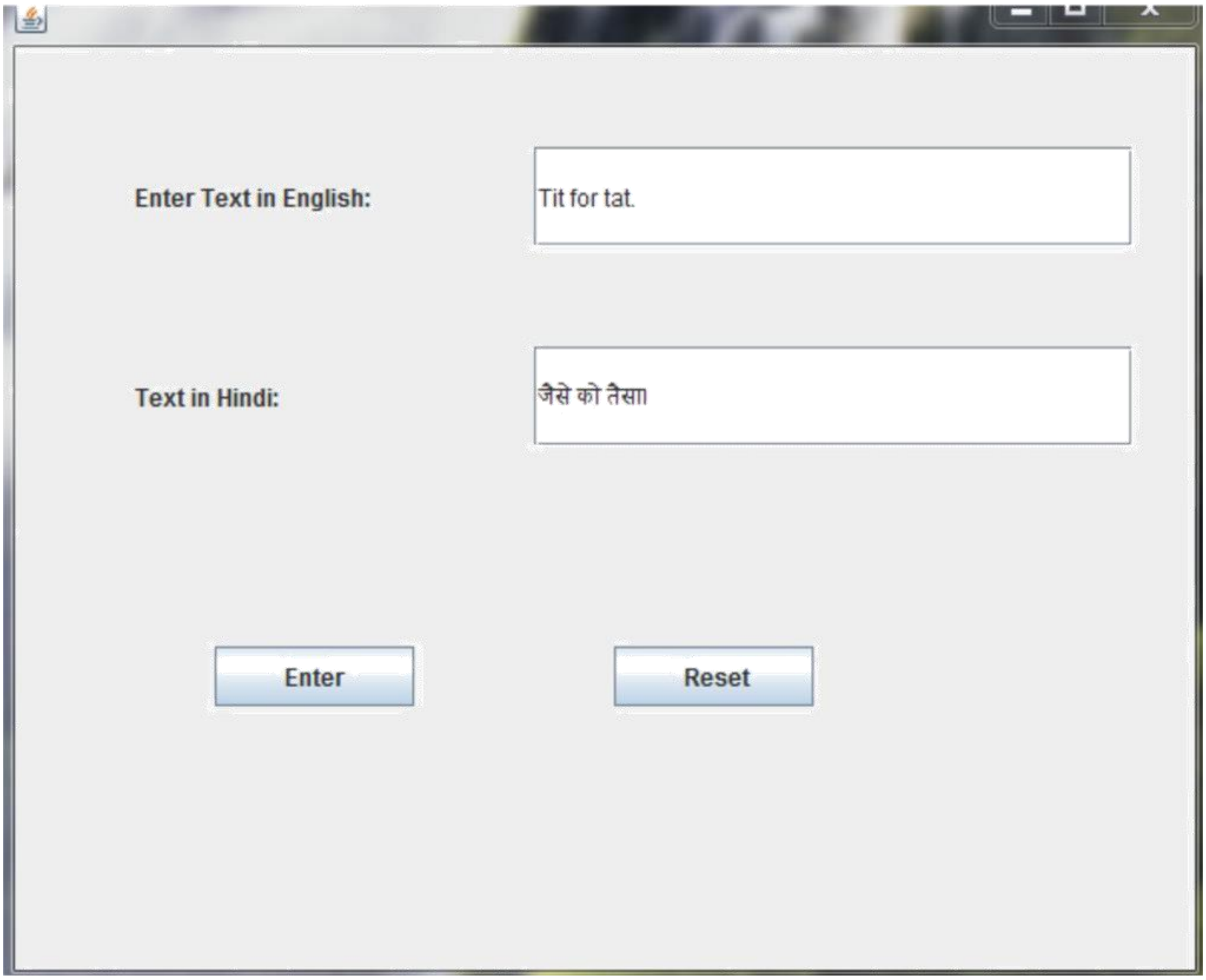
Enter Text in English: Ram went to school or temple?

Text in Hindi: राम स्कूल या मंदिर गए थे?



Enter Text in English: I believe that health is wealth.

Text in Hindi: मेरा मानना है कि स्वास्थ्य धन है।



Enter Text in English: Tit for tat.

Text in Hindi: जैसे को तैसा।

4.2 ANALYSIS

In this thesis, we have utilized the corpus comprising of 1000 basic and compound sentences , to do the trials of machine translation. The execution measurements utilized for assessment of translation are unigram precision, unigram recall, F-measure , BLEU, NIST, MWER and SSER. We have endeavored to look at the consequences of translation for three procedures : RBMT , SMT and EBMT .

4.2.1 UNIGRAM PRECISION: As specified previously, we have taken into account just correct coordinated matches between words. Exactness is computed as takes after:

Whereas is the quantity of words in the translation that match words in the reference translation, and q is the quantity of words in the translation. This might be deciphered as the portion of the words in the translation that are available in the reference translation.

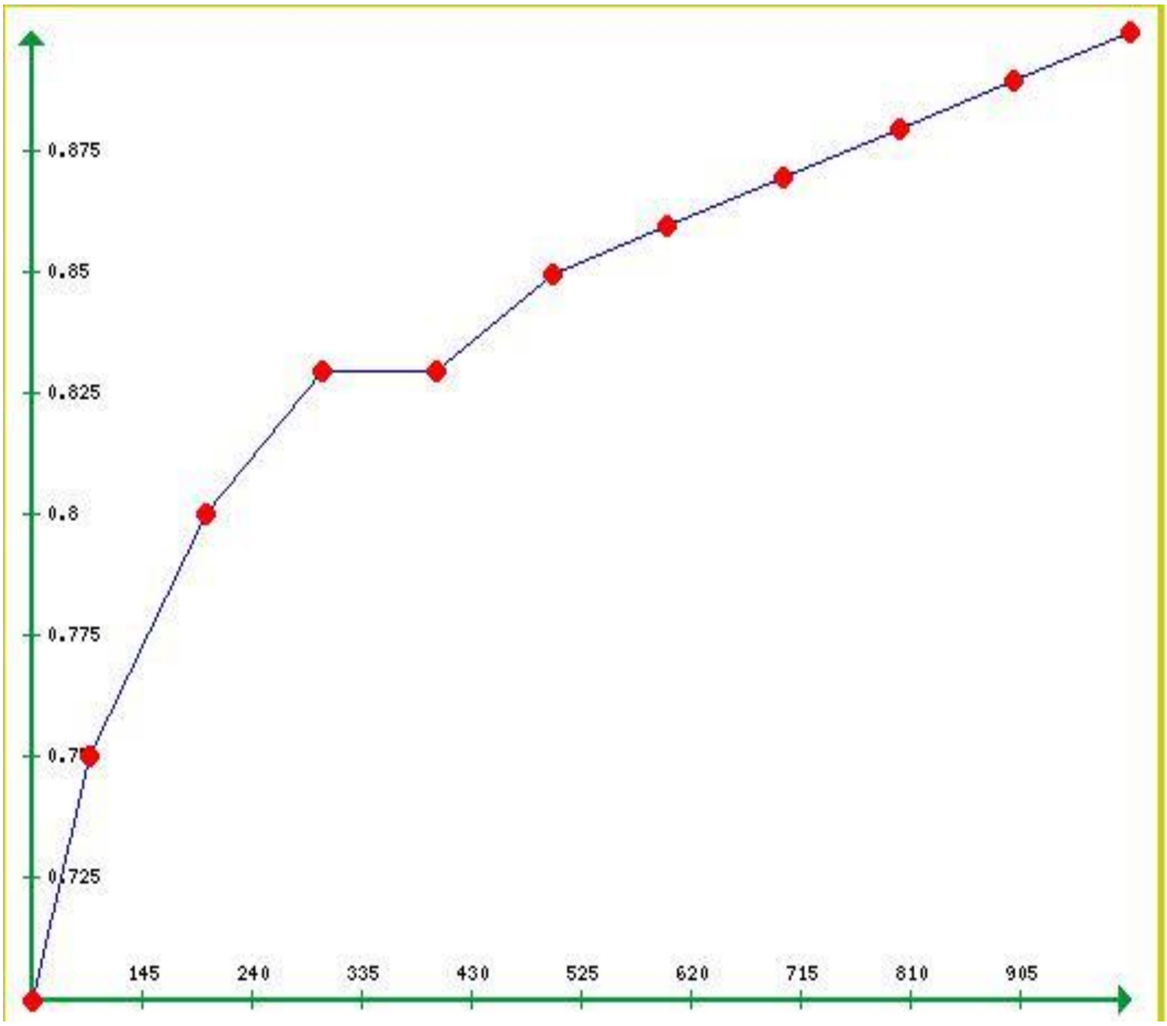


Figure 4.1: Unigram Precision

In this graph, corpus estimate is shifted from 0 to 1000 along x-pivot and comparing unigram exactness is plotted along y-hub.

4.2.2 UNIGRAM RECALL: Similarly as with exactness, just correct coordinated word matches are considered. Review is ascertained as takes after:

Where s is the quantity of coordinating words, and t is the quantity of words in the reference translation. This might be deciphered as the part of words in the reference that show up in the translation.

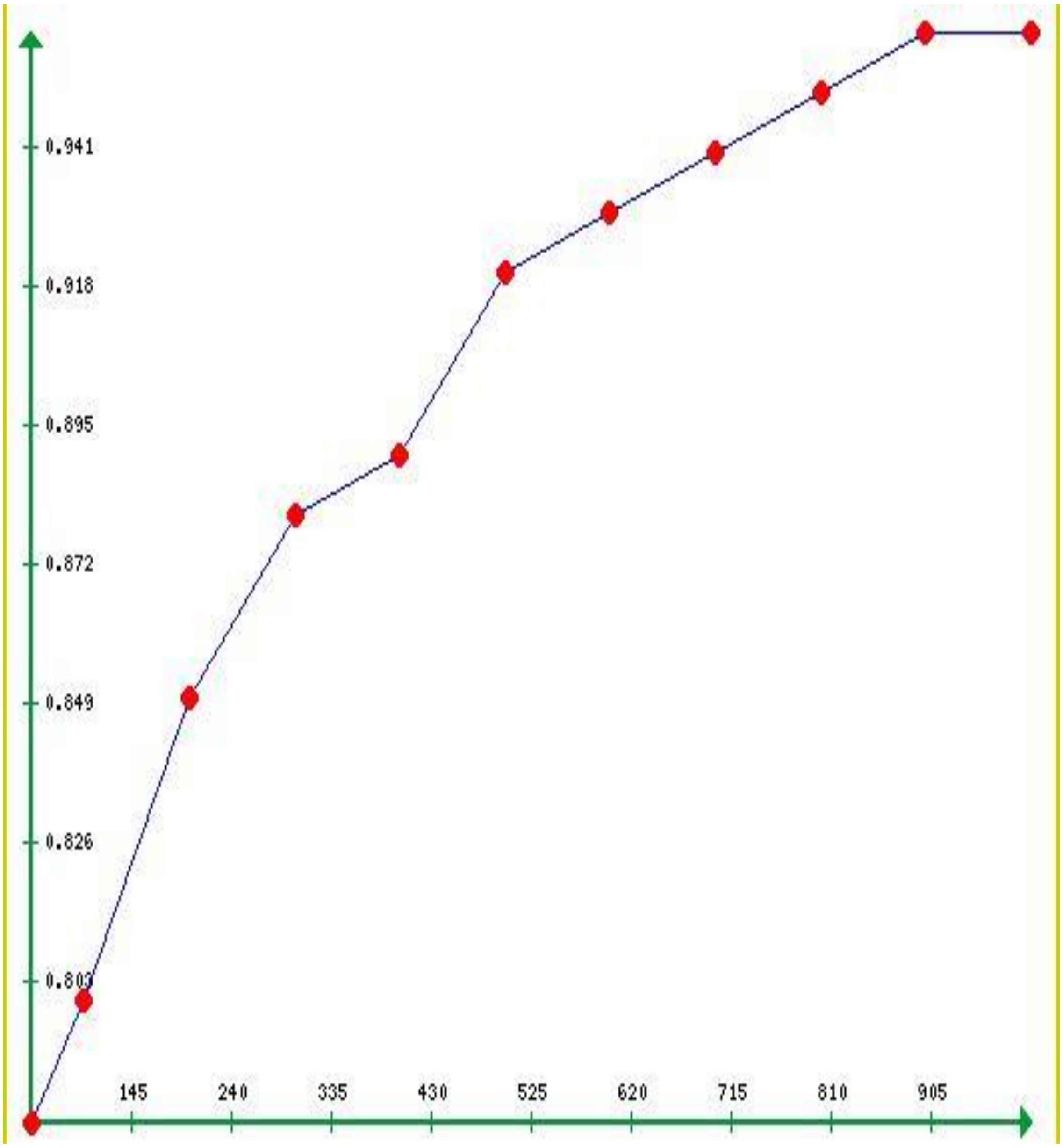


Figure 4.2: Unigram Recall

In this graph, corpus estimate is changes from 0 to 1000 along x-hub and comparing unigram review is plotted along y-pivot.

4.2.3 F-MEASURE : The F-measure of exactness and review is registered as takes after:

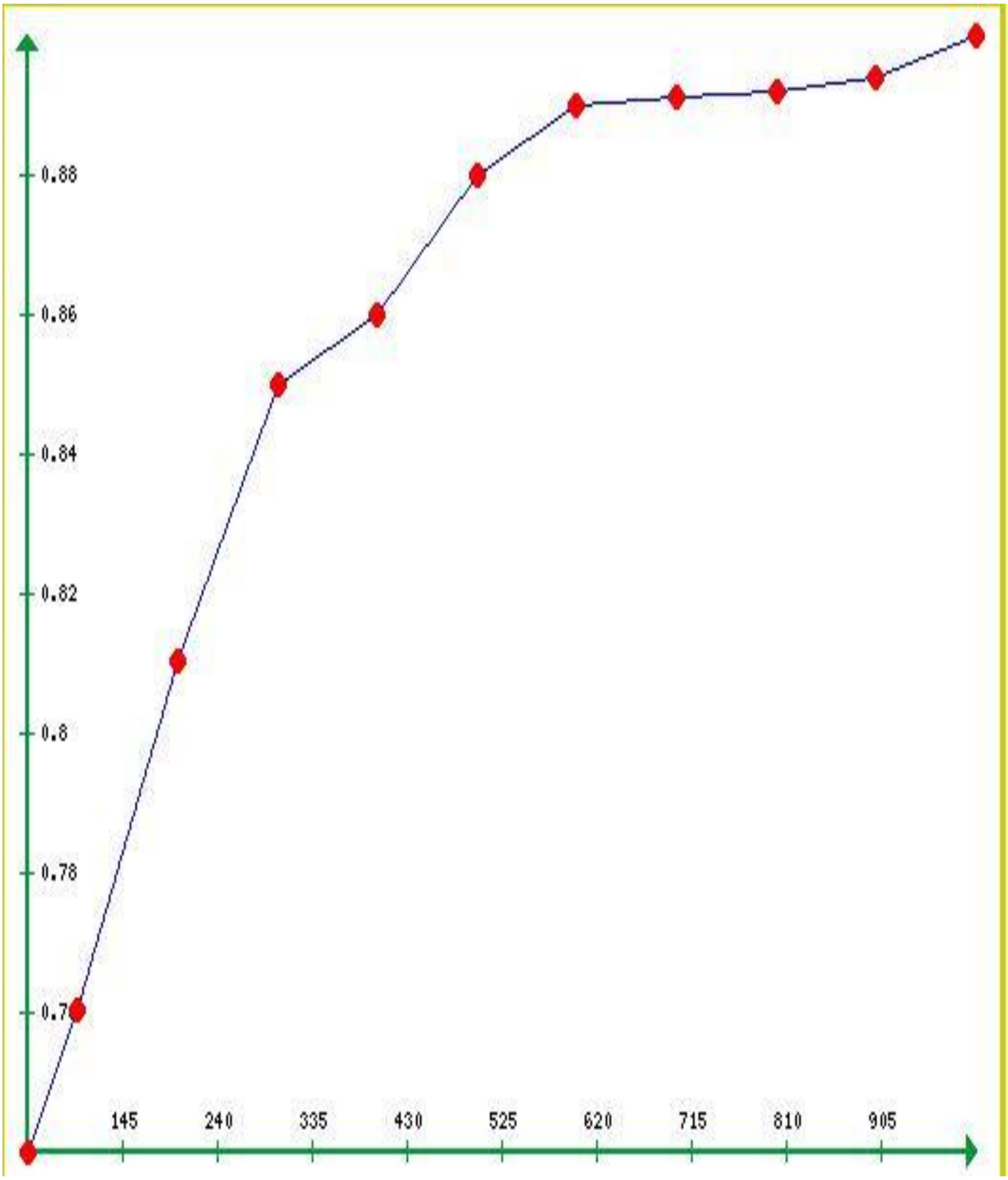


Figure 4.3 : F-measure

In this graph, corpus estimate is changed from 0 to 1000 along x-hub and comparing F-measure is plotted along y-pivot.

4.2.4 BLEU : This measures the exactness of n-grams as for the reference translations, with a quickness punishment. A higher BLEU score shows better translation.

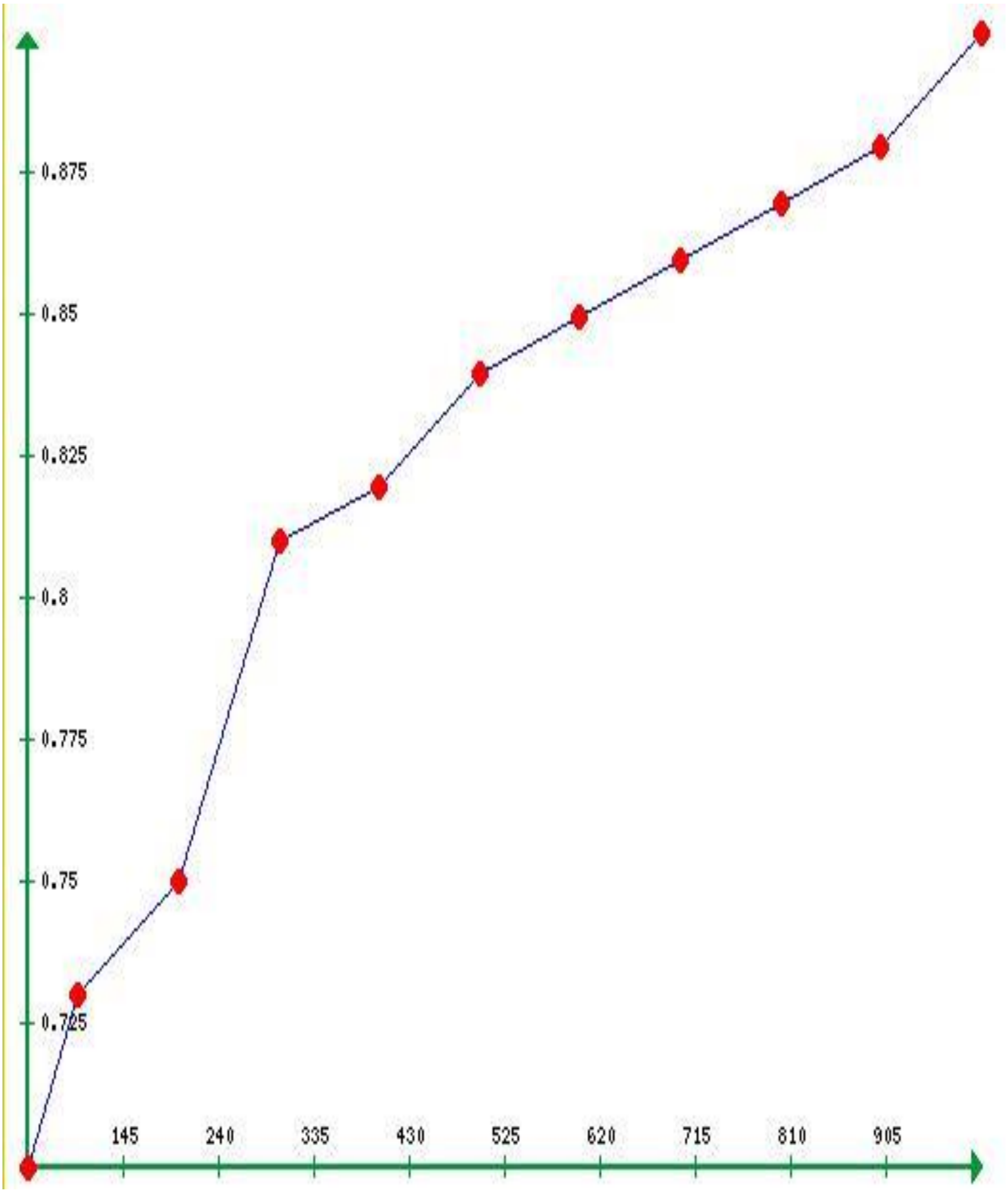


Figure 4.4 : BLEU

In this graph, corpus measure is changed from 0 to 1000 along x-pivot and relating BLEU is plotted along y-hub.

4.2.5 NIST : Translation ampleness is caught by NIST score. An translation utilizing similar words (1-grams) as in the references has a tendency to fulfill ampleness.

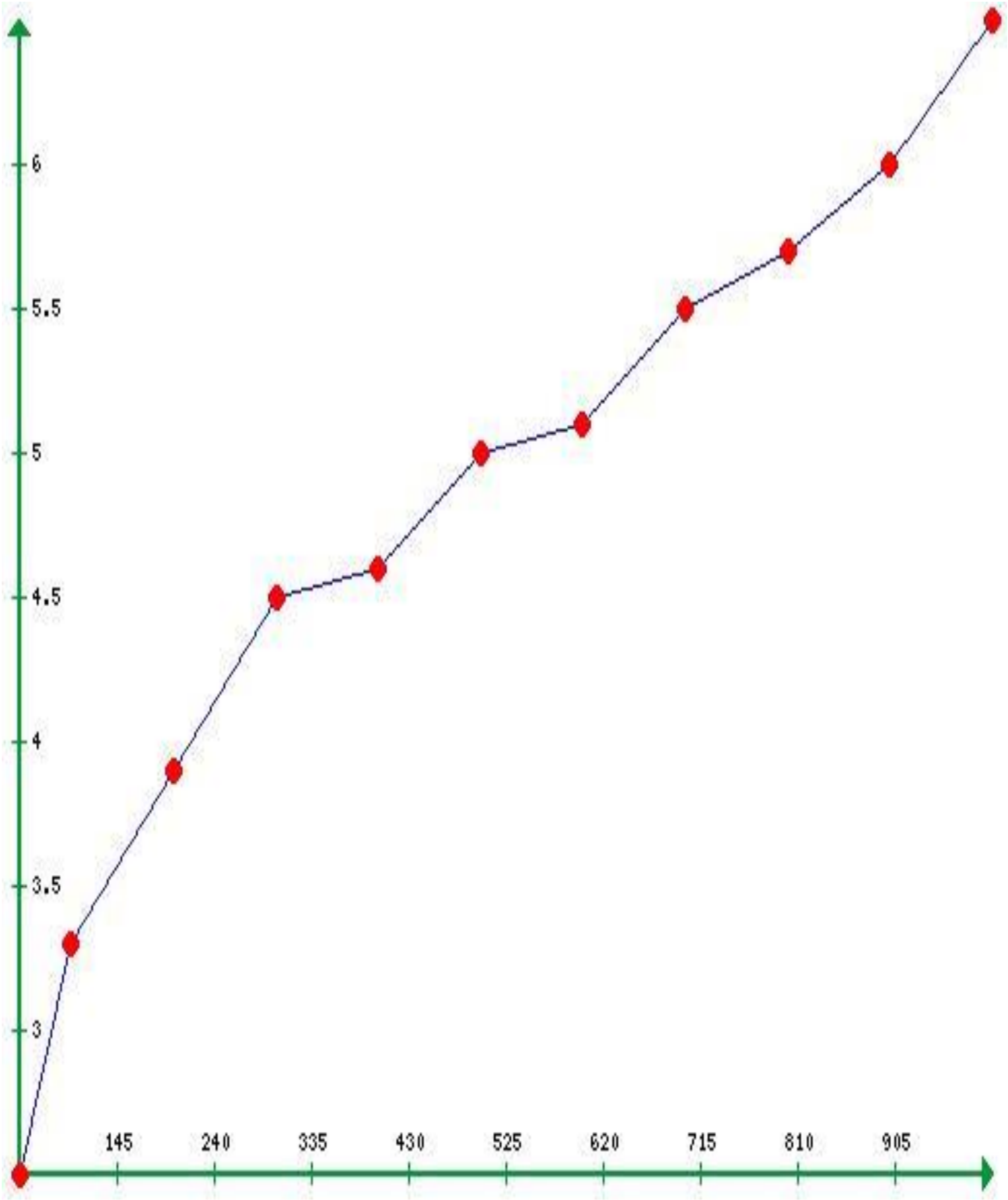


Figure 4.5: NIST

Execution assessment of EBMT for NIST is shown in above graph.

4.2.6 MWER : This measures the alter separate with the most comparable reference translation. In this way, a lower MWER score is alluring.

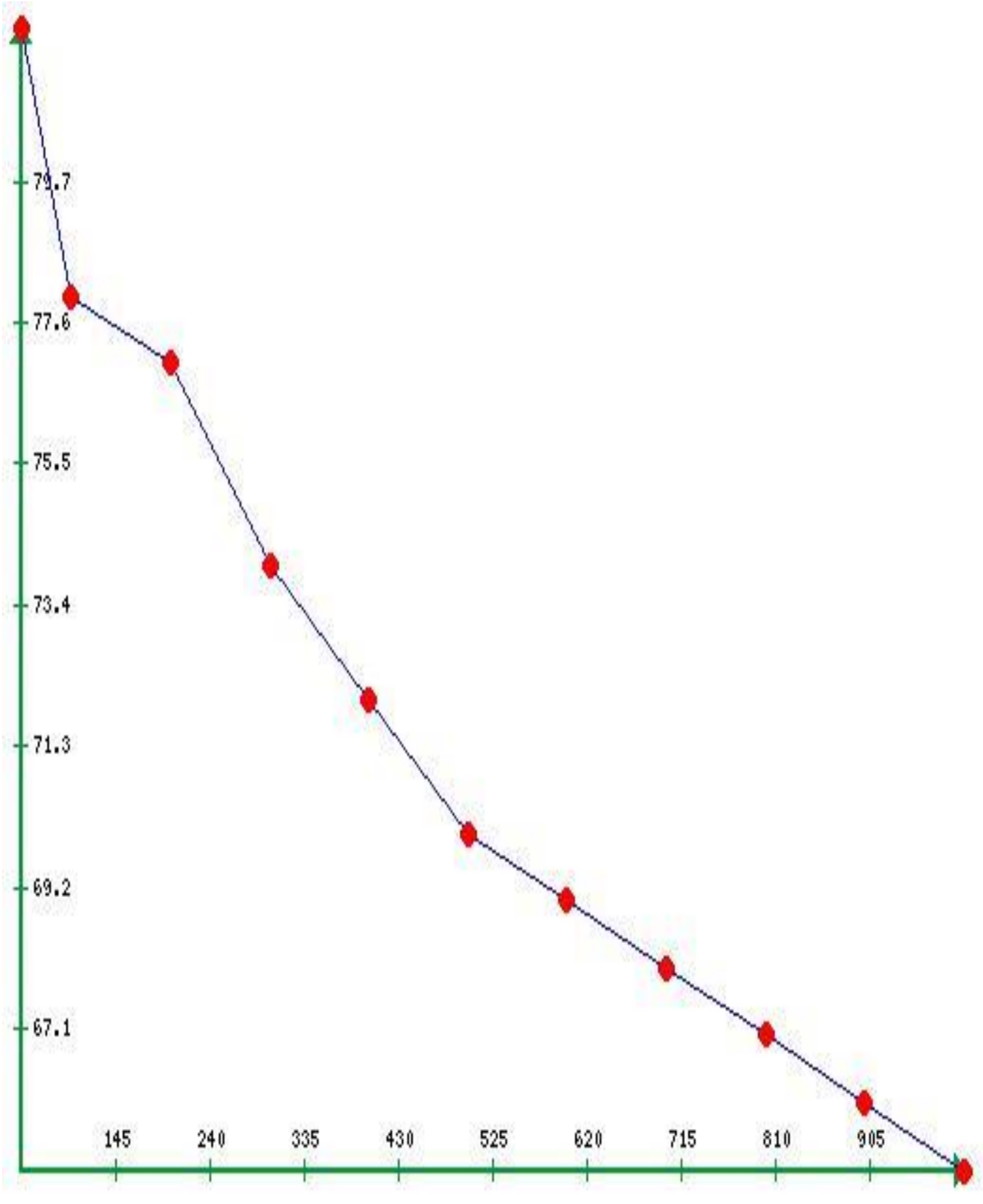


Figure4.6 : MWER

Execution assessment of EBMT for MWER is shown in above graph.

4.2.7 SSER : This is computed utilizing human judgments. Each sentence was judged by a human evaluator on the accompanying five-point scale, and the SSER was ascertained.

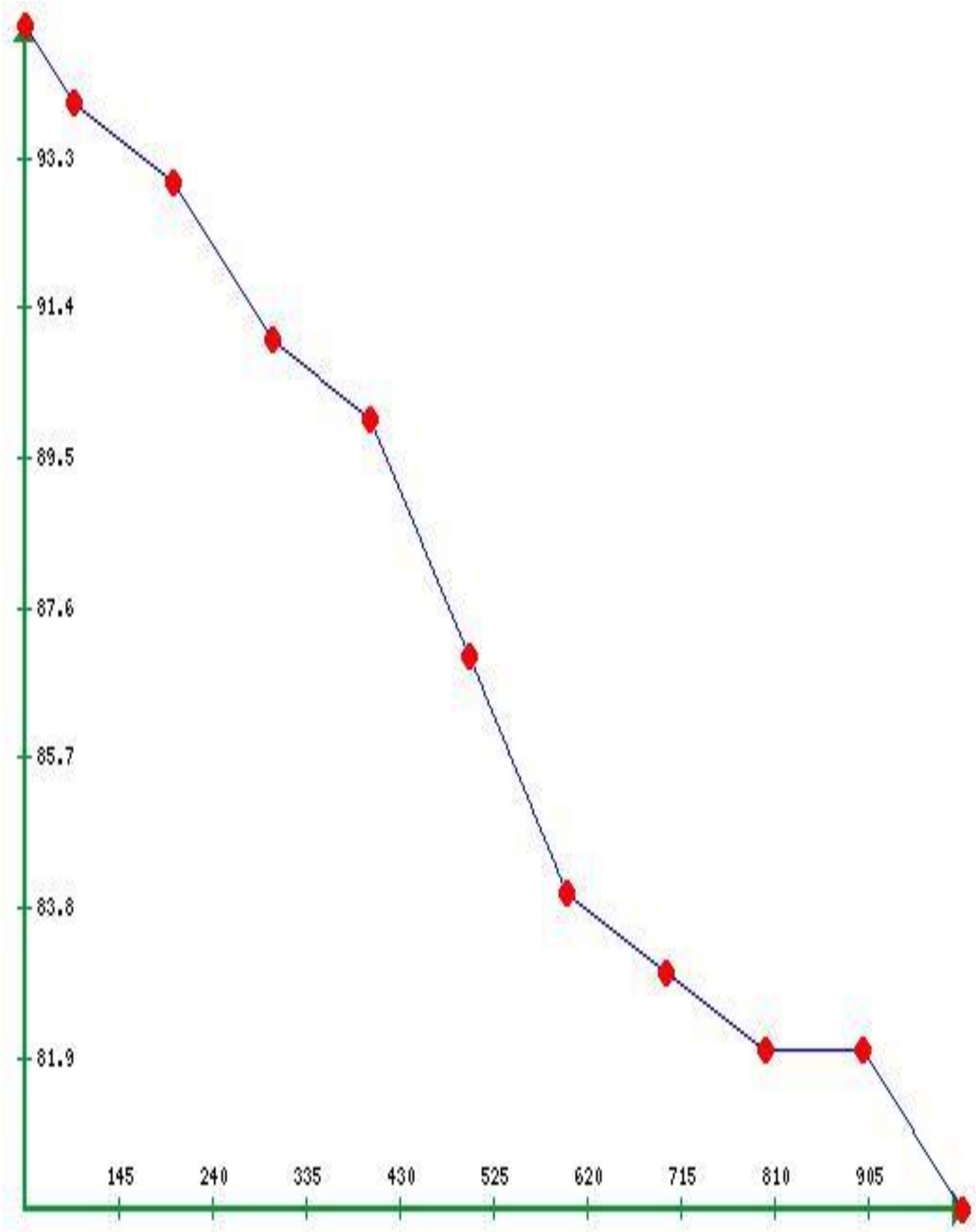


Figure 4.7: SSER

Execution assessment of EBMT for MWER is shown in above graph.

CHAPTER-5
CONCLUSIONS

In this theory we have exhibited the key methodologies of building up a machine translation framework. We have exhibited in detail our approach of building up a manage based framework for English-Hindi.

5.1 CONCLUSIONS

Various algorithms for machine translation have been examined previously. Since every calculation has its advantages and disadvantages, we proposed a multi-display approach where we progressively chose the best sentence from different translations. This is guided by a relapse display which devours the features from the source sentence and its corresponding translation from multiple models and selects the best sentence from the anticipated score.

Advancement of Example Based Machine Translation (EBMT) framework utilizes Java on Linux stage for translation starting with one dialect then onto the next. In this specific example, we will make an translation of English sentences to Hindi. The standard of translating in EBMT is straightforward: a framework chooses a fitting translation of an information sentence by breaking down the pre-deciphered sentences in the database. In this way, the bigger the database of pre-deciphered sentences, more noteworthy will be the exactness of the EBMT framework.

Example based translation is basically translation by similarity. This implies if an EBMT framework is given an arrangement of sentences in the source dialect (from which one is translating) and their comparing translations in the objective dialect, the framework can utilize these examples to decipher other such comparative source dialect sentences into target dialect sentences. The fundamental start is that, if a formerly deciphered sentence happens once more, a similar translation is probably going to be right once more.

The trial comes about demonstrate an expansion in execution over the pattern framework. In future, we intend to coordinate a couple of more phonetic and other statistical highlights, removed at the deciphering stage, which can be considered to enhance the determination criteria. Expectation of the quality score utilizing dynamic learning is an intriguing zone to be investigated. Successively running both the expression and various leveled framework may bring about increment in time of calculation as parse tree and other component calculation add to unraveling time.

5.2 FUTURE SCOPE

Following are the future bearings for English to Hindi EBMT system.

1. The work can be reached out to incorporate multilingual corpus of various dialects in the source-target combine. The objective and source dialects can be expanded from show one dialect

2. The framework can likewise be placed in the web-based to translate substance of one website page in English to Hindi.

3. A mobile application can likewise be created in which message containing English content is sent to the customer in Hindi dialect.

4. The corpus can be preprocessed to change its provision structure for enhancing the nature of translation.

5. The deciphered content can be reordered and handled to beat linguistic slip-ups which will be a piece of post-preparing. This will enhance score of human assessment.

REFERENCES

[1] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., and Zheng, X. Tensorflow: A system for large-scale machine learning. arXiv preprint arXiv:1605.08695 (2016).

[2] Bahdanau, D., Cho, K., and Bengio, Y. Neural machine translation by jointly learning to align and translate. In International Conference on Learning Representations (2015).

[3] Caglayan, O., Aransa, W., Wang, Y., Masana, M., García-Martínez, M., Bougares, F., Barrault, L., and van de Weijer, J. Does multimodality help human and machine for translation and image captioning? In Proceedings of the First Conference on Machine Translation (Berlin, Germany, August 2016), Association for Computational Linguistics, pp. 627–633.

[4] Caruana, R. Multitask learning. In Learning to learn. Springer, 1998, pp. 95–133.

[5] Cho, K., van Merrienboer, B., Gülçehre, Ç., Bougares, F., Schwenk, H., and Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In Conference on Empirical Methods in Natural Language Processing (2014).

[6] Crego, J., Kim, J., Klein, G., Rebollo, A., Yang, K., Senellart, J., Akhanov, E., Brunelle, P., Coquard, A., Deng, Y., Enoue, S., Geiss, C., Johanson, J., Khalsa, A., Khiari, R., Ko, B., Kobus, C., Lorieux, J., Martins, L., Nguyen, D.-C., Priori, A., Riccardi, T., Segal, N., Servan, C., Tiquet, C., Wang, B., Yang, J., Zhang, D., Zhou, J., and Zoldan, P. Systran's pure neural machine translation systems. arXiv preprint arXiv:1610.05540 (2016). 3The Korean translation does not contain spaces and uses '。' as punctuation symbol, and these are all artifacts of applying a Japanese postprocessor.

[7] Dong, D., Wu, H., He, W., Yu, D., and Wang, H. Multi-task learning for multiple language translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (2015), pp. 1723–1732.

[8] Firat, O., Cho, K., and Bengio, Y. Multi-way, multilingual neural machine translation with a shared attention mechanism. In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016 (2016), pp. 866–875.

[9] Firat, O., Cho, K., Sankaran, B., Yarman Vural, F., and Bengio, Y. Multi-way, multilingual neural machine translation. Computer Speech and Language (4 2016).

[10] Firat, O., Sankaran, B., Al-Onaizan, Y., Yarman-Vural, F. T., and Cho, K.

Zero-resource translation with multi-lingual neural machine translation. In EMNLP (2016).

[11] French, R. M. Catastrophic forgetting in connectionist networks. Trends in cognitive sciences 3, 4 (1999), 128–135.

[12] Gage, P. A new algorithm for data compression. C Users J. 12, 2 (Feb. 1994), 23–38.

[13] Gillick, D., Brunk, C., Vinyals, O., and Subramanya, A. Multilingual language processing from bytes. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (San Diego, California, June 2016), Association for Computational Linguistics, pp. 1296–1306.

[14] Hutchins, W. J., and Somers, H. L. An introduction to machine translation, vol. 362. Academic Press London, 1992.

[15] Kalchbrenner, N., and Blunsom, P. Recurrent continuous translation models. In Conference on Empirical Methods in Natural Language Processing (2013).

[16] Lee, J., Cho, K., and Hofmann, T. Fully character-level neural machine translation without explicit segmentation. arXiv preprint arXiv:1610.03017 (2016).

[17] Luong, M.-T., Le, Q. V., Sutskever, I., Vinyals, O., and Kaiser, L. Multi-task sequence to sequence learning. In International Conference on Learning Representations (2015).

[18] Luong, M.-T., Pham, H., and Manning, C. D. Effective approaches to attention-based neural machine translation. In Conference on Empirical Methods in Natural Language Processing (2015).

[19] Luong, M.-T., Sutskever, I., Le, Q. V., Vinyals, O., and Zaremba, W. Addressing the rare word problem in neural machine translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (2015).

[20] Maaten, L. V. D., and Hinton, G. Visualizing Data using t-SNE. Journal of Machine Learning Research 9 (2008).

[21] Richens, R. H. Interlingual machine translation. The Computer Journal 1, 3 (1958), 144–147.

[22] Schultz, T., and Kirchhoff, K. Multilingual speech processing. Elsevier

Academic Press, Amsterdam, Boston, Paris, 2006.

[23] Schuster, M., and Nakajima, K. Japanese and Korean voice search. 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (2012).

[24] Sébastien, J., Kyunghyun, C., Memisevic, R., and Bengio, Y. On using very large target vocabulary for neural machine translation. In Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (2015).

[25] Sennrich, R., Haddow, B., and Birch, A. Controlling politeness in neural machine translation via side constraints. In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016 (2016), pp. 35–40.

[26] Sennrich, R., Haddow, B., and Birch, A. Neural machine translation of rare words with subword units. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (2016).

[27] Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems (2014), pp. 3104–3112.

[28] Tsvetkov, Y., Sitaram, S., Faruqui, M., Lample, G., Littell, P., Mortensen, D., Black, A. W., Levin, L., and Dyer, C. Polyglot neural language models: A Example study in cross-lingual phonetic representation learning. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (San Diego, California, June 2016), Association for Computational Linguistics, pp. 1357–1366.

[29] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Łukasz Kaiser, Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M., and Dean, J. Google's neural machine translation system: Bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144 (2016).

[30] Yamagishi, H., Kanouchi, S., and Komachi, M. Controlling the voice of a sentence in japanese- toenglish neural machine translation. In Proceedings of the 3rd Workshop on Asian Translation (Osaka, Japan, December 2016), pp. 203–210.

[31] Zhou, J., Cao, Y., Wang, X., Li, P., and Xu, W. Deep recurrent models with fast-forward connections for neural machine translation. Transactions of the Association for Computational Linguistics 4 (2016), 371–383.

[32] Zoph, B., and Knight, K. Multi-source neural translation. In NAACL HLT 2016, The 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego California, USA, June 12-17, 2016 (2016), pp. 30–34.