

CHAPTER ONE

INTRODUCTION



1. Introduction

A natural language is human spoken language, such as English. NLP is a subfield of Artificial Intelligence and linguistics. It studies the Problems of automated generation and understanding of natural human languages. Natural language generation systems convert information from computer databases into normal-sounding human language, and natural language understanding systems Convert sample of human language into more formal representations that are easier for computer programs to manipulate.[1]

Generally Five Processing Stages in a NLP System:-

(Phonological Analysis, Morphological Analysis (lexical), Syntactic Analysis, Semantic Analysis and Pragmatic Analysis).

Artificial intelligence is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.[2]

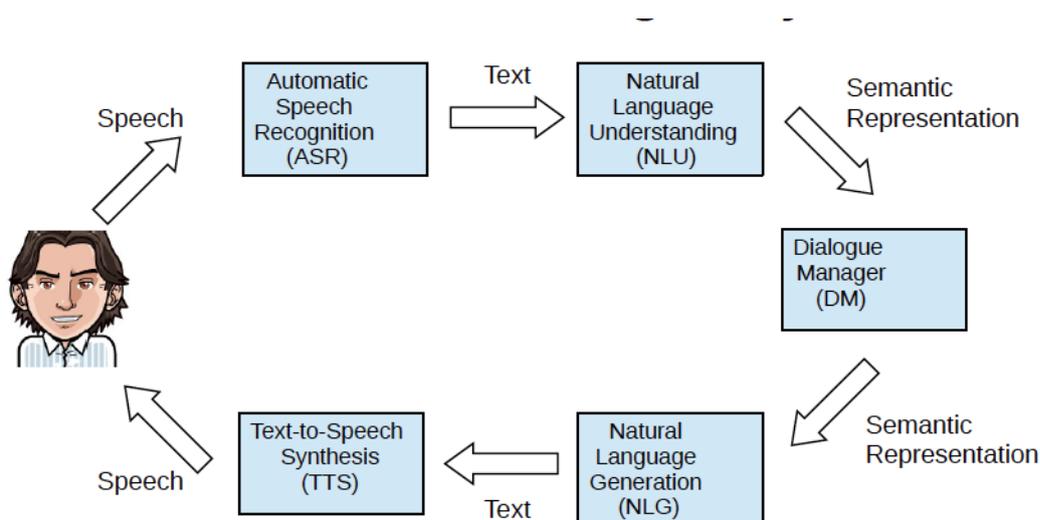
Morphological analysis algorithm based on the analysis of the root word and the word is made up of three or four characters and compare them with a list and in the case of matches is extracted prefix and suffixes and root, But in the case of mismatches extract morphological weight *and compares with the word* , *and then we extract the root.*[3]

2. Artificial intelligence

Artificial Intelligence is a branch of Science which deals with helping machines find solutions to complex problems in a more human-like fashion. This generally involves borrowing characteristics from human intelligence, and applying them as algorithms in a computer friendly way. A more or less flexible or efficient approach can be taken depending on the requirements established, which influences how artificial the intelligent behavior appears.[4]

3. Natural Language Processing

Short for natural language processing, a branch of artificial intelligence that deals with analyzing, understanding and generating the languages that humans use naturally in order to interface with computers in both written and spoken context using natural human languages instead of computer languages.[5]



3.1 morphology

Morphology is the study of how things are put together, like the make-up of animals and plants, or the branch of linguistics that studies the structure of word. In morphology, the word part morph- means "form" and -ology means "the study of." So, those who study how something is made or formed are engaged in morphology . In language morphology, you might study how prefixes and suffixes added to a word change its meaning.[6]

3.2 parser

In computer technology, a parser is a program, usually part of a compiler, that receives input in the form of sequential source program instructions, interactive online commands, markup tags, or some other defined interface and breaks them up into parts (for example, the nouns (objects), verbs (methods), and their attributes or options) that can then be managed by other programming (for example, other components in a compiler parser may also check to see that all input has been provided that is necessary).[7]

3.3 semantic

Semantics is the study of meaning in language. It can be applied to entire texts or to single words. For example, "destination" and "last stop" technically mean the same thing, but students of semantics analyze their subtle shades of meaning.

4. aim of project(morphology analysis)

The purpose of the project is to analyze the words to their roots and extracting Prefix, suffix and root. sometimes extract weights for the purpose of comparison with the word for root.

Example for word analysis:-

Unaccepted

Prefix: un

Suffix: ed

Stem: accept

CHAPTER TWO

NATURAL LANGUAGE PROCESSING



Natural Language Processing

2.1 introduction

The term Natural Language Processing encompasses a broad set of techniques for automated generation, manipulation and analysis of natural or human languages. Although most NLP techniques inherit largely from Linguistics and Artificial Intelligence, they are also influenced by relatively newer areas such as Machine Learning, Computational Statistics and Cognitive Science.[8]

2.2 The basic goal of (NLP) is to enable a person to communicate with a computer in a Language that they use in their everyday life.

2.3 understand of NLP

To understand something is transform it from one representation into another, where this second representation has been chosen to correspond to a set of available actions that could be performed and where the mapping has been designed. So that for each event, an appropriate action will be done.

2.4 Components of natural language analysis:

1. Phonological Analysis
2. Morphological Analysis (lexical Analysis)
3. Syntactic Analysis
4. Semantic Analysis
5. Discourse analysis
6. Pragmatic Analysis

The phonological analysis: Analysis of speech sounds of the world's languages, with a focus on both their articulator and acoustic properties. An introduction to phonetic alphabets, including practice in transcribing a variety of language samples. Analysis of the systematic organization of speech sounds, with reference to features and suprasegmentals.

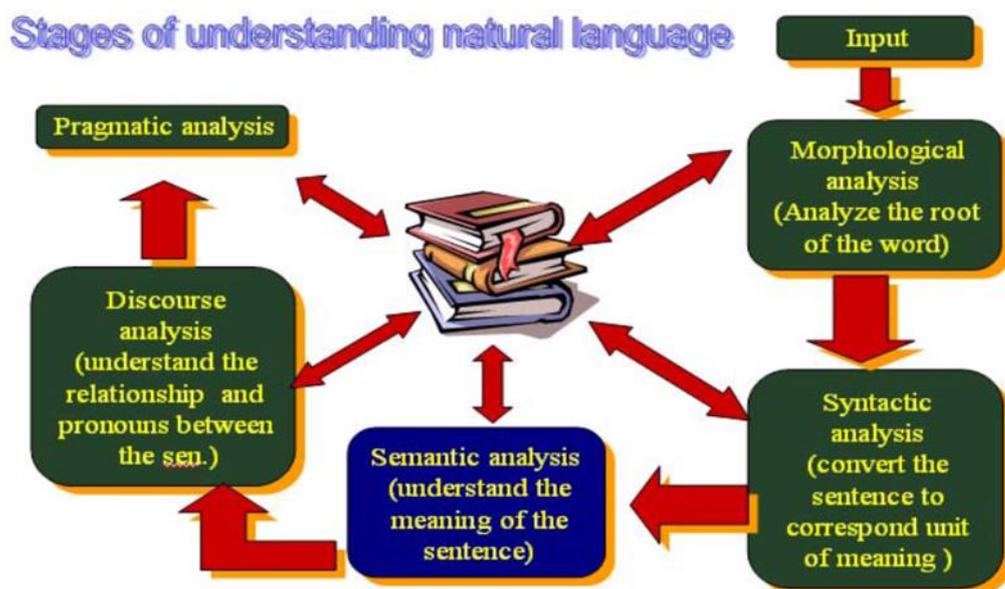
The lexical analyzer : reads The input text of source language character by character and produces tokens such as (names ,keywords ,punctuation marks ,discards white space and comments), which are the basic lexical units of the language. The process of breaking-up a text into its constituent tokens is known as tokenization. Tokenization occurs at a number of different levels: a text could be broken up into paragraphs, sentences, words, syllables, or phonemes.

Syntactic analysis: is concerned with the construction of sentences. Syntactic structure indicates how the words are related to each other. Syntax tree is assigned by a grammar and a lexicon.

Semantic analysis: Which produce a representation of the meaning of the text, the representation that commonly used include conceptual dependency, frames, and logic base representation. and it uses the knowledge about the meaning of the word and linguist structure, such as case roles of nouns or the transitivity.

Discourse Analysis: Analyze the effect of previous sentence on the next sentence.

Pragmatic Analysis: Interpret the sentence according to its meaning.



Written text processing:

1- Formal method.

2. Informal method.

Informal method:

Example:

Computers milk drinks.

Computer drinks milk.

Computers use data.

Formal method:

1. lexical analysis (word).
2. syntactical analysis (grammars).
3. semantic analysis (meanings).

Intelligent robot:

The robot would have to know:

1. The meaning of the words.
2. Relationship of one word to another.
3. Knowledge of grammars.
4. Associate descriptions and objects.
5. Analyze sentence in relation to another sentences.

e.g

-John drank milk

-he then put on his coat.

Natural Language and Computer Language

- Natural language are those that we use for communicating with each other, eg. Arabic, English, French, Japanese, etc.
- Natural language are expressive and easy for us to use.
- Computer languages are those that we use for controlling the operations of a computer, eg. Prolog, C, C++, C# ,java, python,...etc.
- Computer languages are easy for a computer to understand, but they are not expressive.

Applications for Natural Language Processing

- Machine Translation:** Machine translation (MT) is automated translation. It is the process by which computer software is used to translate a text from one natural language (such as English) to another (such as Spanish).
- Information Retrieval:(IR)** is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on or on full-text (or other content-based) indexing.
- Question Answering Systems:** Question answering (QA) is a computer science discipline within the fields of information retrieval and natural language processing (NLP), which is concerned with building systems that automatically answer questions posed by humans in a natural language.
- Speech Recognition :** include voice user interfaces such as voice dialing (e.g. "Call home"), call routing (e.g. "I would like to make a collect call").
- Intelligent Tutoring Systems:** in a tutoring system to teach elementary physics, such as mechanics, the system may present the theory and worked-out examples.

3. morphology

- Morphology is the study of the way words are built up from smaller meaning bearing units, morphemes. A morpheme is often defined as the minimal meaning-bearing unit in a language. So for example the word fox consists of a single morpheme (the morpheme fox) while the word cats consists of two: the morpheme cat and the morpheme –s.

- The morpheme is meaningful linguistic unit consisting of a words (such as dog) or a word element (such as the -s at the end of dogs) that can't be divided into smaller meaningful parts. Adjective: morphemic.

- The word is a speech sound or a combination of sounds, or its representation in writing, that symbolizes and communicates a meaning and may consist of a single morpheme or a combination of morphemes.

- In linguistics, morphology is the identification, analysis and description of the structure of a given language's morphemes and other linguistic units, such as root words, affixes, parts of speech, intonation/stress, or implied context (words in a lexicon are the subject matter of lexicology).

- The word is a speech sound or a combination of sounds, or its representation in writing, that symbolizes and communicates a meaning and may consist of a single morpheme or a combination of morphemes.

- In linguistics, morphology is the identification, analysis and description of the structure of a given language's morphemes and other linguistic units, such as root words, affixes, parts of speech, intonation/stress, or implied context (words in a lexicon are the subject matter of lexicology).

3.1 Types of morphemes

Free morphemes, like town and dog, can appear with other lexemes (as in town hall or dog house) or they can stand alone, i.e., "free".

Bound morphemes like "un-" appear only together with other morphemes to form a lexeme. Bound morphemes in general tend to be prefixes and suffixes.

Derivational morphemes can be added to a word to create (derive) another word: the addition of "-ness" to "happy," for example, to give "happiness." They carry semantic information.

Inflectional morphemes modify a word's tense, number aspect, and so on, without deriving a new word or a word in a new grammatical category (as in the "dog" morpheme if written with the plural marker morpheme "-s" becomes "dogs"). They carry grammatical information.

Zero Morphemes

A zero morpheme is a morpheme that is realized by a phonologically null affix (an empty string of phonological segments). In simpler terms, a zero invisible affix. It is also called a null morpheme, the process of adding a null morpheme is called zero affixation, null derivation or zero derivation.

Inflectional and Derivational

all affixes are bound and they are of two types:

derivational morphemes and inflectional morphemes. Morphology “is concerned with two quite different phenomena: derivation and inflection. Derivation has to do with the way morphemes get attached as affixes to existing lexical forms or stems in the process of word formation”. While inflectional morphology “does not create new words but adapts existing words so that they operate effectively in sentences. It is not a process of lexical innovation but of grammatical function. The differences between them are many, yet the most important ones are as follows:

- Many natural languages mainly uses prefixes and suffixes to express inflectional and derivational morphology.
- English has a relatively simple inflectional system: only nouns, verbs, and sometimes adjectives can be inflected,
- English Inflectional morphology is relatively simple and includes suffixes like (-s, -ed and -ing.)
- Arabic inflectional and derivational morphology are very Complexes.

A prefix is a letter or a group of letters attached to the beginning of a word that partly indicates its meaning. For example, the word prefix itself begins with a prefix—pre-, which generally means "before" Understanding the meanings of the common prefixes can help us deduce the meanings of new words that we encounter.

Prefix	Suffix
Co	Copilot
Dis	Disappear
Macro	Macroeconomics
Un	Unfinished
Tri	Tricycle
Uni	Unicorn

A suffix is a letter or a group of letters attached to the end of a word to form a new word or to alter the grammatical function of the original word. For example, the verb read can be made into the noun reader by adding the suffix -er; read can be made into the adjective readable by adding the suffix -able. Understanding the meanings of the common suffixes can help us deduce the meanings of new words that we encounter.

Suffix	Example
Al	Refusal
Dom	Freedom, kingdom

CHAPTER THREE

SYSTEM DESIGN AND IMPLEMENTATION



The dictionary of the morphology

The dictionary :- contains all the available word in a language with its classification such as : nouns, pronouns, verb, adjectives, ,adverb,...etc.

The morphology:- is the basic unit in NLP, and it is the process of breaking a word into its smallest meaningful components, or morphemes, is concerned with breaking a word into its root and suffix, or prefix .

Morphology analysis with English language

Dictionary words can be define as facts such as:

Word("boy", name, "boys", human or non- human, move).

Word("go", verb, "went", "going", "guan", "goes").

Word("tall", adj, type, human or non-human).

In morphology analysis the prefix or suffix either removed from the word or added to in order to extract the root of the word.

Prefix

suffix

In

ing

Im

ed

Un

es

Consider the following rules:-

1-Add ing when we delete e.

2-add ing to the verbs converts to nouns.

3-if y before her vowel, turn and add ing.

4-add ed to verbs and turn to past.

5-some word added to s when at combine or es.

Morphology analysis with Arabic language

Stages automatic language processing:-

- 1-Morphological Analysis:- Analyzing every word to their parts.
- 2- Composition analysis:- convert sentence to certain structures.
- 3- meaning analysis:- Inputs are converted into an internal representation of reserves sense.

Analyzer morphological:- Is a program studying the construction of the word and converted into the original root.

Morphological analyzer applications:-

- Spell Checker:- Analytic.
- Search:- analysis+ generation.
- Interpreter vocabulary.

Derivatives and solids:-

Derivative:- It is branching from the root, such as: مكتبة

Root(ك ت ب).

Solid:- It is the one who does not have a root ,such as: انسان, نعيم.

Root:- Length of 3-4Characters, Root is the original letters of the word and root is not the word, but rather a combination of letters.

the weight:- the weight + Letters Redundant.

Prefix:- Excess characters at the beginning of the word.

Suffix:- Excess characters at the end of the word.

Work requirements analyst:-

- 1- A list of all the prefix.
- 2- A list of all suffix.
- 3- A list of all weight.
- 4- A database of all triple and quadruple roots.
- 5- A database of all solid words.

Steps Action analyst:-

- 1) Search for the word in the list of solid words

If you find a finished;

If doesn't find, go to step second.

- 2) Using a list of prefix and suffixes, Delete all the prefix and suffixes related word.

- 3) If doesn't find prefix and suffix in word, go to step fourth.

- 4) Choose the appropriate weight of the word.

- 5) Extracted the root from the word.

Example:-

- Let's take the word (المرابطون).

First:- Looking for a word in the list of solid words.

ن	و	ط	ب	ا	ر	م	ل	ا	المدخلات
									الوزن
									الجذر

Second:- We are looking for all the prefix and suffixes related word.

ن	و	ط	ب	ا	ر	م	ل	ا	المدخلات
									الوزن
									الجذر

Third:- Choose from the list of weights appropriate weight to the word:

-The same number of characters.

-Match characters except - ف ع ل -

ن	و	ط	ب	ا	ر	م	ل	ا	المدخلات
		ل	ع	ا	ف	م			الوزن
									الجذر

Fourth:- Extract the root of the word:

-Selecting the corresponding characters to – ف ع ل -

ن	و	ط	ب	ا	ر	م	ل	ا	المدخلات
		ل	ع	ا	ف	م			الوزن
		ط	ب		ر				الجذر

The last:- Look for the root extracted (ربط) in the root database.

If found → the word is true.

If non found → There is a spelling error.

The algorithm

1. input Arabic word.

2. check if the word is invariable:

2.1 using the prefix table remove the largest prefix associated with the word.

2.2 if the remaining part of the word is in the list of invariable words then:

2.2.1 if the prefix is compatible with the word and the variable-invariable flag is false then return the type; value; grammatical rule

2.2.2 else go to step 2 below.

3. Analyze the word assuming that is variable.

3.1 Using the suffix table remove the largest suffix associated with the word.

3.2 Using the morphology pattern list, locate patterns that match the word .

A pattern is said to match a word if it satisfies the following conditions:

- * the pattern is equal to the word in length.
- * the root identified by the pattern is in the root table.
- * the removed suffix and prefix is compatible with the selected pattern.
- * the root is compatible with the pattern.

3.3 For each matching pattern return the type of the corresponding word, the root, the suffix and prefix.

3.4 Add the removed suffix back to the word letter by letter, starting with the first letter in the suffix. For each letter added repeat steps 3.2 and 3.3.

3.5 Add the removed prefix back to the word letter by letter, starting with the first letter in the prefix. For each letter added repeat steps 1 and 2.

examples

word	prefix	suffix	stem	weight	Root
الاولاد	ال		اولاد	افعال	ولد
يكتبون	ي	ون			كتب
يقاتلون	ي	ون	قاتل	فاعل	قتل
وباستخراجهم	وب	هم	استخراج	استفعال	خرج
كاتب				فاعل	كتب
يذهبون	ي	ون			ذهب

CHAPTER FOUR

USER INTERFACE



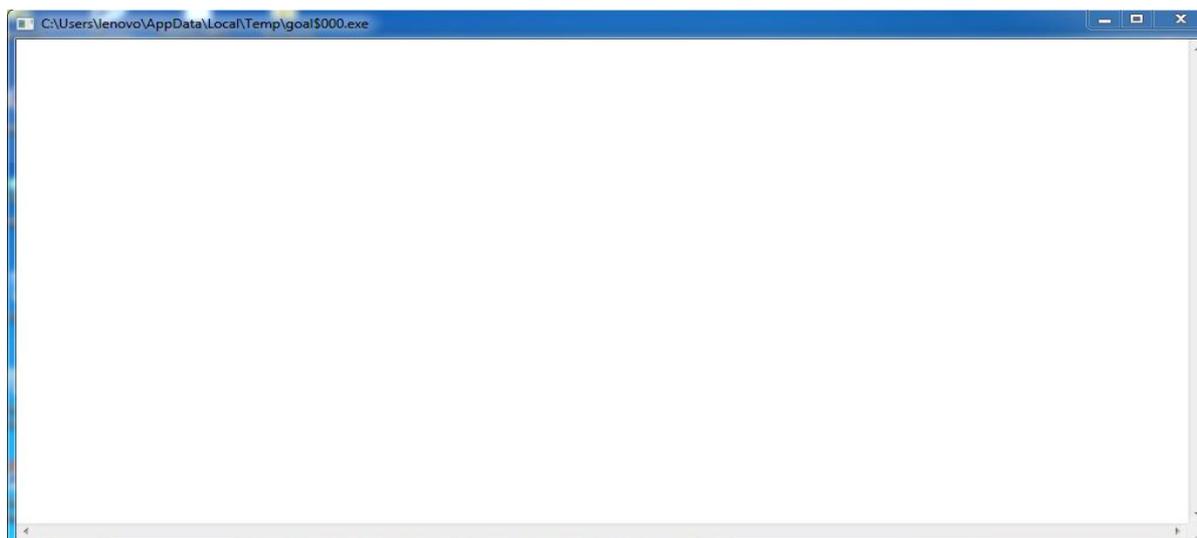
Beginning of the program:-

This section provides the user interaction with the system is very simple,

This part of the system is important not only to provide easy user interface interaction , but also allows the user to interact with the system and equipped with the necessary information to the process of analyzing words.

we'll take the way open user interface, and the method of Execution of the program (visual prolog).

The program interface



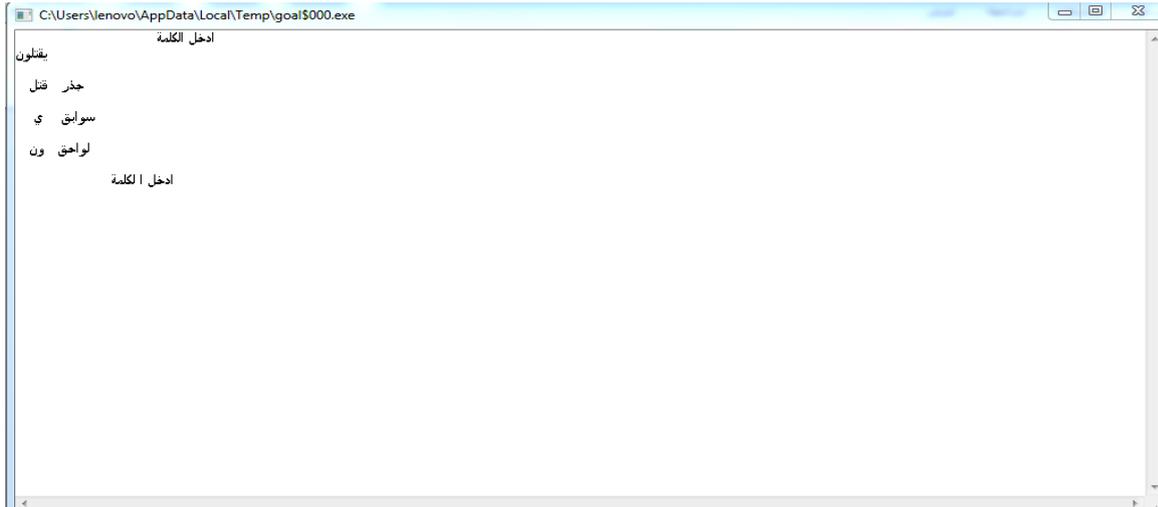
Case1: When you input the word into the program

الكلمة: يقتلون

السوابق: ي

اللواحق: ون

الجذر: قتل

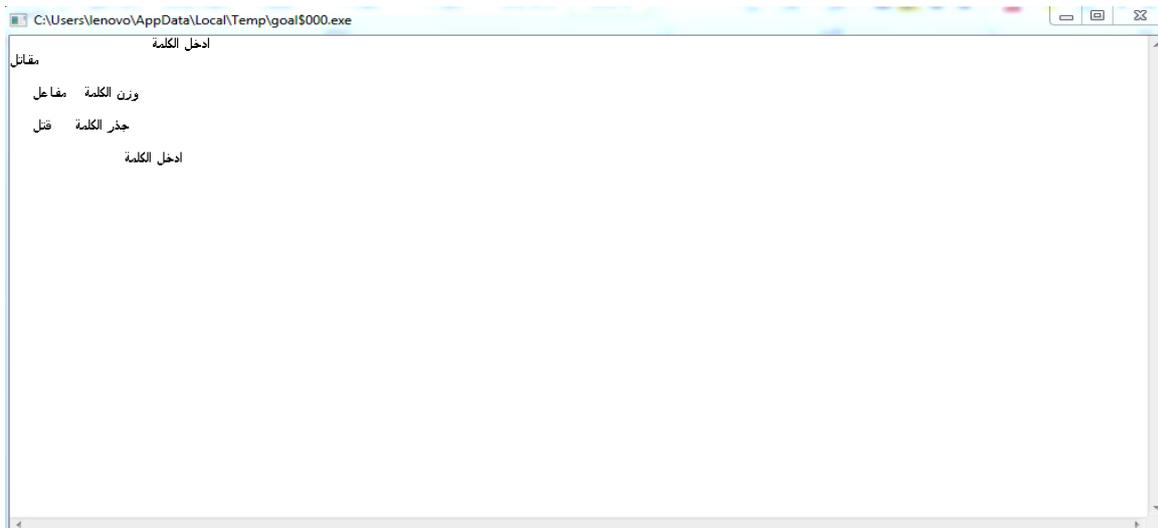


Case2: In this case you should enter the word and compare them with the corresponding weight .

الكلمة: مقاتل

الوزن المطابق: مفاعل

الجذر: قتل



Case3: In this case, combines case1 with case 2.

الكلمة: يقاتلون

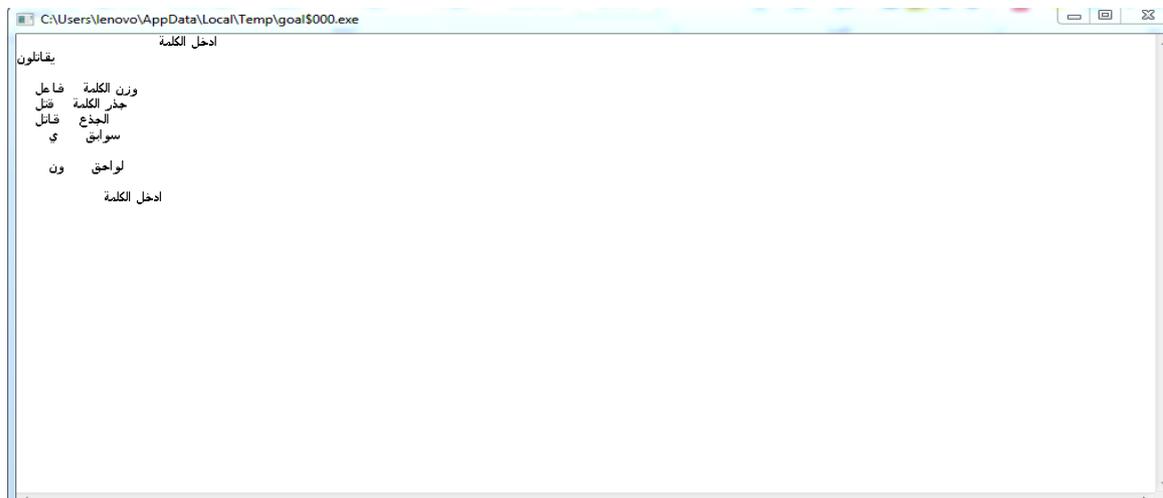
وزن الكلمة: فاعل

الجزر: قتل

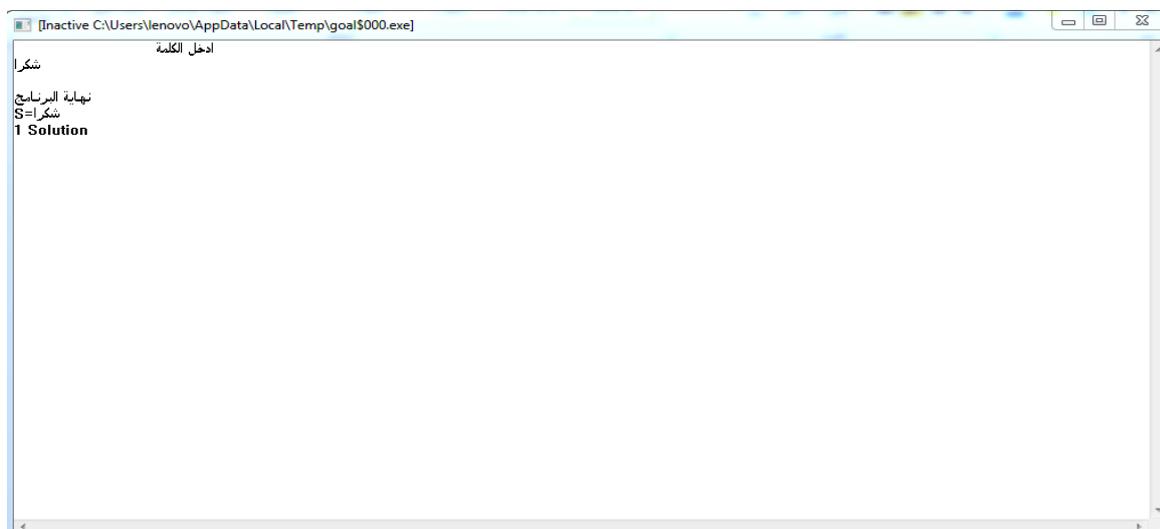
الجزع : قاتل

السوابق: ي

اللواحق: ون



When we want to get out of the interface implementation, input (شكرا)



CHAPTER FIVE

CONCLUSION



Conclusion:-

The Arabic word characterized by flexibility morphological g may be composed of " prefix, suffix, index", and be processed through these possibilities to deal with the root word and morphological weights in the Arabic language.

Project development:-

Vowels and substitution , such as letters that have added prefix and suffixes
And the need to deal with the weights be sized Long.