In computer science, lexical analysis, lexing or tokenization is the process of converting a sequence of characters (such as in a computer program or web page) into a sequence of tokens (strings with an assigned and thus identified meaning). A program that performs lexical analysis may be termed a lexer, tokenizer,[1] or scanner, although scanner is also a term for the first stage of a lexer. A lexer is generally combined with a parser, which together analyze the syntax of programming languages, web pages, and so forth

Token: A token is a group of characters having collective meaning: typically a word or punctuation mark, separated by a lexical analyzer and passed to a parser. A lexeme is an actual character sequence forming a specific instance of a token, such as num. The pattern matches each string in the set.

In this project, we give the sentence as the input to the chat bot that we created and what happens is the keyword that the chatbot extracts gets analysed and after that it will give the preferred output to us. The compiler tokenizes the text we enter and does the token matching.

When you split a large unit (long string) into a group of sub-units (smaller strings), each of the sub-units (smaller strings) is referred to as a "token". If there are no more sub-units, then you are done parsing.