Differentiate between lexemes and tokens. Give examples to support your discussion.

Answers

A lexeme is a sequence of characters in the source program that matches the pattern for a token and is identified by the lexical analyzer as an instance of that token. The term lexeme is used both in the study of language and in lexical analysis of computer program compilation. In computer programming, lexemes are part of the input stream used to identify tokens. Lexemes are one of the building blocks of language. Examples include; IF, while, <, <=, =, etc.

A token is a pair consisting of a token name and an optional attribute value. The token name is an abstract symbol representing a kind of lexical unit, e.g., a particular keyword, or sequence of input characters denoting an identifier. The token names are the input symbols that the parser processes. A token can also be said to be a syntactic category that forms a class of lexemes I.e. which class the lexeme belong is it a keyword or identifier or anything else. An invalid or illegal token produces an error. Examples include; IF, ID, number, relation, number etc.