

Link to the github repo:

https://github.com/arbazkh4an/lexical_analyser.git

Regex Definition Table

Token Type	Regex Pattern	Description / Matches
NEWLINE	\n	Increments line number counter (yylineno++).
WHITESPACE	[\t\r]+	Ignores spaces, tabs, and carriage returns.
COMMENT (Single-line)	\/\/"[^\\n]*	Matches // followed by all characters until newline.
COMMENT (Multi-line)	\/\/*\" (handled in C loop)	Matches start of a block comment /* ... */ and counts lines.
STRING_LITERAL	\"([^\\"\\] \\\\.)*\"	Matches double-quoted strings like "hello" with escaped characters allowed.
CHAR_LITERAL	\\'([^\\"\\] \\\\.)\'	Matches single character literals like 'a' or '\n'.
NUMBER	[0-9]+(\.[0-9]+)?([eE][+-]?[0-9]+)?	Matches integers, floats, and scientific notation (e.g., 12, 3.14, 2.5e10).
IDENTIFIER / KEYWORD	[A-Za-z_][A-Za-z0-9_]*	Matches variable names and keywords, differentiated via isKeyword().
OPERATOR	== != <= >= \+ - * /= > <	Matches arithmetic, relational, and assignment operators.
PUNCTUATION	;\, \(\) \{ \}	Matches punctuation symbols (semicolon, comma, parentheses,

braces).

ERROR / UNKNOWN

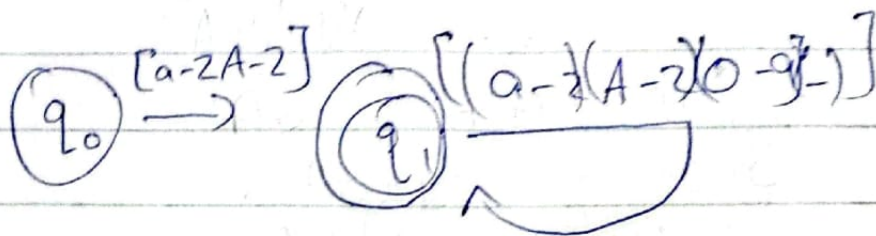
Matches any single
unrecognized character.

Keyword List (Matched Separately with isKeyword())

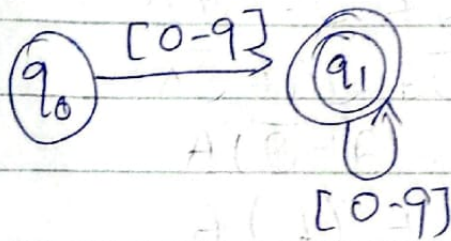
agar, warna, jabtak, karo, har, wapas, integer, decimal, lafz, sahi, ghalat, shuru, khatam, amal,
dikhao

5. FA Diagram.

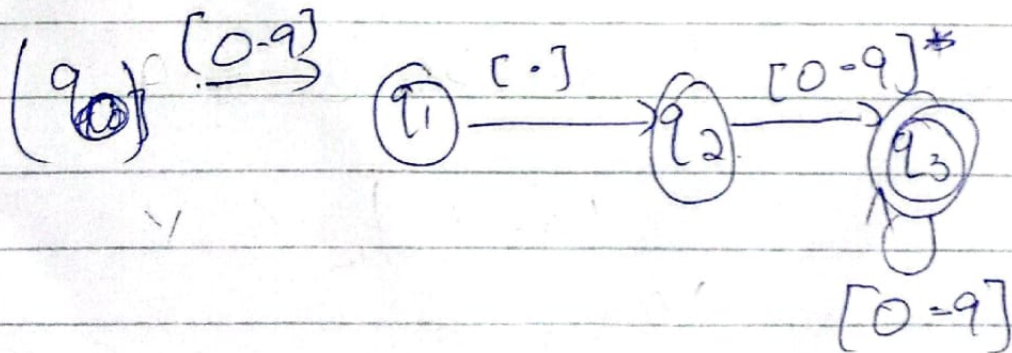
A. identifier



B. integer.



C. float.



6.

Explanation

Why these keywords?

The keywords were chosen based on commonly used Urdu words that connect with programming concepts. Words like agar link with if & jabtak link with "while".

Unique How?

~~C++~~

- Cultural Relevance
- Learning & beginner friendly
- C++ like structure
- Lightweight focused on learning & basics.

