Lexikálny analyzátor

Lexikálny analyzátor je implementovaný ako konečný automat, ktorý využíva pomocné stavy(napr. pri načítavaní čísla). Tento konečný automat rozdeľuje zdrojový program na lexémy(napr. identifikátory, čísla, operátory a podobne), ktoré sú reprezentované tokeny. Pričom token je reprezentovaný abstraktnou datovou štruktúrou, ktorá obsahuje jeho typ a riadok, na ktorom sa nachádza. Pokiaľ daný lexém je identifikátor, číslo alebo reťazec, tak lexikálny analyzátor vracia aj názov/hodnotu. Po načítaní identifikátoru prebehne kontrola jeho názvu, či nie je totožný s niektorým z kľúčových slóv, pretože pre kľúčové slová sú vytvorené špeciálne typy tokenov.

Komentáre a biele znaky sú lexikálnym analyzátorom ignorované.

V prípade chybnej štruktúry aktuálneho lexému program končí s odpovedajúcou návratovou hodnotou.