O algoritmo:

A importação da biblioteca que vai permitir a extração em csv, a linha a ser analisada e o conjunto dos símbolos válidos

import csv

linha = 'else 5 > 3 == true : return true'

reservadas = ['while', 'do', 'return', 'if', 'elif', 'else', ':']

operadores = ['<', '=', '+', '-', '==', '!=', '>', '<']

inteiros = ['0', '1', '2', '3', '4', '5', '6', '7', '8', '9']

booleanos = ['true', 'false']

inicia a execução

traduz(linha)

Verifica o se o item pertence a uma dada classificação e retorna o termo classificador

def procura(item):

    if item in reservadas:

        return 'Palavra reservada'

    elif item in operadores:

        return 'Operador'

    elif item in booleanos:

        return 'Booleano'

    elif item in inteiros:

        return 'Constante'

    else:

        return 'Constante'

verifica se há repetição de item

def pesquisa(token, lista):

    for item in lista:

        if token in item:

            return True

    return False

esplita os itens da linha verifica a classificação de cada um, se há repetição de alguns itens, e monta os arquivos CSVs

def traduz(linha):

    itens = linha.replace(';', '').split(' ')

    arq = open('tabelas.csv', 'w', newline='', encoding='utf-8')

    w = csv.writer(arq)

    w.writerow(['TOKEN', 'IDENTIFICACAO', 'TAMANHO', 'POSICAO'])

    listaIdent = []

    listaAtual = []

    for i in range(len(itens)):

        token = itens[i]

        identificacao = procura(itens[i])

        if identificacao == 'Constante':

            if not pesquisa(token, listaIdent):

                listaIdent.append([token, identificacao])

                identificacao = [identificacao, len(listaIdent)]

            else:

                for i, item in enumerate(listaIdent):

                    if item[1] == identificacao and item[0] == token:

                        identificacao = [item[1], i+1]

        tamanho = len(itens[i])

        posicao = 0

        if listaAtual:

            if len(listaAtual) == 1:

                posicao = linha.find(token, len(listaAtual[0])+1)

            else:

                posicao = linha.find(token, listaAtual[-1][1]+1)

        listaAtual.append([token, posicao])

        if token == '=':

            w.writerow(["'"+token, identificacao, tamanho, '(0,{})'.format(posicao)])

        else:

            w.writerow([token, identificacao, tamanho, '(0,{})'.format(posicao)])

    arq = open('simbolos.csv', 'w', newline='', encoding='utf-8')

    w = csv.writer(arq)

    w.writerow(['INDICE', 'SiMBOLO'])

    for item in listaIdent:

        w.writerow([listaIdent.index(item)+1, item[0]])

    arq.close()

Saída em Tabelas

