Aula\_3\_e\_4\_Python\_para\_PLN

{

"nbformat": 4,

"nbformat\_minor": 0,

"metadata": {

"colab": {

"name": "Aula 3 e 4 - Python para PLN",

"provenance": []

},

"kernelspec": {

"name": "python3",

"display\_name": "Python 3"

},

"language\_info": {

"name": "python"

}

},

"cells": [

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "kLKAdrEZVAFP",

"outputId": "35113fa5-f5c3-4ff1-edc7-aeaa85adb267"

},

"source": [

"infile = open(\"/content/drive/MyDrive/arquivos\_txt/corpus\_teste.txt\", 'r')\n",

"texto = infile.read()\n",

"print(len(texto.split()))\n",

"infile.close()"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"758\n"

]

}

]

},

{

"cell\_type": "code",

"metadata": {

"id": "djwRd3xFWb-C"

},

"source": [

"outfile = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'w')\n",

"outfile.write(\"Números de 1 a 10\"+'\\n')\n",

"for i in range(1,11):\n",

" outfile.write(str(i)+\"\\n\")\n",

"outfile.close()"

],

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "h5oGFf5lW2cR",

"outputId": "1b6e52d1-34b3-424c-8f2c-6270141b31f2"

},

"source": [

"arquivo = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'r')\n",

"print(arquivo.read())"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"Números de 1 a 10\n",

"1\n",

"2\n",

"3\n",

"4\n",

"5\n",

"6\n",

"7\n",

"8\n",

"9\n",

"10\n",

"\n"

]

}

]

},

{

"cell\_type": "code",

"metadata": {

"id": "mhYhRlHaXnJd"

},

"source": [

"outfile = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'w')\n",

"outfile.write(\"Adicionando números de 11 a 20\"+'\\n')\n",

"for i in range(11,21):\n",

" outfile.write(str(i)+\"\\n\")\n",

"outfile.close()"

],

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "IQ6S7VHoXyHZ",

"outputId": "64c2768d-b571-4836-f5e3-6831d3e83a2b"

},

"source": [

"arquivo = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'r')\n",

"print(arquivo.read())"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"Adicionando números de 11 a 20\n",

"11\n",

"12\n",

"13\n",

"14\n",

"15\n",

"16\n",

"17\n",

"18\n",

"19\n",

"20\n",

"\n"

]

}

]

},

{

"cell\_type": "code",

"metadata": {

"id": "kpRkpTK4X1qu"

},

"source": [

"outfile = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'a')\n",

"outfile.write(\"Adicionando números de 21 a 30\"+'\\n')\n",

"for i in range(21,31):\n",

" outfile.write(str(i)+\"\\n\")\n",

"outfile.close()"

],

"execution\_count": null,

"outputs": []

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "hJGFwqmsX7qb",

"outputId": "68709696-80ba-47d8-965b-6e7dfe86fdaa"

},

"source": [

"arquivo = open(\"/content/drive/MyDrive/arquivos\_txt/result.txt\", 'r')\n",

"print(arquivo.read())"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"Adicionando números de 11 a 20\n",

"11\n",

"12\n",

"13\n",

"14\n",

"15\n",

"16\n",

"17\n",

"18\n",

"19\n",

"20\n",

"Adicionando números de 21 a 30\n",

"21\n",

"22\n",

"23\n",

"24\n",

"25\n",

"26\n",

"27\n",

"28\n",

"29\n",

"30\n",

"\n"

]

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "NQx2zuskg4BT",

"outputId": "e7174208-bf01-49c4-d5c3-1b19c0a0fb5d"

},

"source": [

"'aspas simples'"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'aspas simples'"

]

},

"metadata": {},

"execution\_count": 3

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "\_g68HlGfg9uc",

"outputId": "e8d94697-2b24-4720-8b5a-c768d821e714"

},

"source": [

"\"aspas duplas\""

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'aspas duplas'"

]

},

"metadata": {},

"execution\_count": 4

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "G-JpkSr7hAsE",

"outputId": "f6320935-ce88-4883-a5ae-52c630d2b481"

},

"source": [

"texto = \"aspas simples\"\n",

"type(texto)"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"str"

]

},

"metadata": {},

"execution\_count": 6

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "wRVBHLw5hMtY",

"outputId": "7f9cd817-92c3-475f-ebda-f595d6978917"

},

"source": [

"\"\"\"aspas triplas tem uma propriedade que ignoram\n",

" a quebra de \n",

"linha\"\"\""

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'aspas triplas tem uma propriedade que ignoram\\n a quebra de \\nlinha'"

]

},

"metadata": {},

"execution\_count": 8

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "Lfb-0tfXixw6",

"outputId": "c25a33d5-a64c-49b7-8f00-0618ca44650e"

},

"source": [

"variavel = 1 + 1\n",

"variavel\n",

"variavel + 3\n",

"print(variavel)"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"2"

]

},

"metadata": {},

"execution\_count": 9

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "DZgCLBDZi3RB",

"outputId": "72fb6e2e-9006-4070-8125-f1c21f70df23"

},

"source": [

"r = 'roney'\n",

"l = ' lira'\n",

"r = r + ' lira'\n",

"r + l"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'roney lira'"

]

},

"metadata": {},

"execution\_count": 13

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "myCNagGajkHL",

"outputId": "ed8ef12d-c556-49dd-f512-6d9a0f6abd4c"

},

"source": [

"pi = 3.14\n",

"info = 'o valor de pi é = ' + str(pi)\n",

"print(info)"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"o valor de pi é = 3.14\n"

]

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "rmG3Ig5mkos1",

"outputId": "5743c20f-2ea1-47b4-b331-5fc9b91452b2"

},

"source": [

"string = 'instituto de ciências matemáticas e de computação'\n",

"string[1:10]"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'nstituto '"

]

},

"metadata": {},

"execution\_count": 17

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "JkIdSOAXmhvU",

"outputId": "e571e4aa-2381-4e2b-c791-c4aac3857303"

},

"source": [

"string.upper()"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO'"

]

},

"metadata": {},

"execution\_count": 18

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "unmAPhSenWfY",

"outputId": "3698aae1-96f2-4c09-de5a-eb8e41e92379"

},

"source": [

"string.replace('A', 'X')"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'instituto de ciências matemáticas e de computação'"

]

},

"metadata": {},

"execution\_count": 22

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "\_OPOyRVvpdMy",

"outputId": "08634234-6efc-4ef4-9063-89444bb99dd5"

},

"source": [

"len(string.split())"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"7"

]

},

"metadata": {},

"execution\_count": 23

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "o9xUEc6RqqcT",

"outputId": "26e4e6ff-21a9-4d56-c60d-5bd333108a92"

},

"source": [

"lista\_split = string.split()\n",

"lista\_split.append(\"ICMC\")\n",

"lista\_split"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"text/plain": [

"['instituto', 'de', 'ciências', 'matemáticas', 'e', 'de', 'computação', 'ICMC']"

]

},

"metadata": {},

"execution\_count": 32

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/",

"height": 36

},

"id": "MqkQYQoxq0DR",

"outputId": "932ef9d3-427f-4f1f-e705-47cf2f7c75b4"

},

"source": [

"juntar = \" \".join(lista\_split)\n",

"juntar"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "execute\_result",

"data": {

"application/vnd.google.colaboratory.intrinsic+json": {

"type": "string"

},

"text/plain": [

"'instituto de ciências matemáticas e de computação ICMC'"

]

},

"metadata": {},

"execution\_count": 33

}

]

},

{

"cell\_type": "code",

"metadata": {

"colab": {

"base\_uri": "https://localhost:8080/"

},

"id": "iKcHbXGTsJUO",

"outputId": "50aae8f0-f128-4689-bb0e-58ae6b64d765"

},

"source": [

"string = 'instituto de \\\\ ciências matemáticas \\\\ e de computação'\n",

"print(string)"

],

"execution\_count": null,

"outputs": [

{

"output\_type": "stream",

"name": "stdout",

"text": [

"instituto de \\ ciências matemáticas \\ e de computação\n"

]

}

]

}

]

}