

Como criar uma biblioteca em python baseada no ALT?

A principal vantagem seria utilizar a ferramenta em massa, aplicando a vários documentos utilizando programação

Na amostra de texto:

“No [Sistema Único de Saúde - SUS](#), a Assistência Farmacêutica e Insumos Estratégicos é a área responsável pela promoção, proteção e recuperação da saúde, tanto individual como coletiva, tendo o medicamento como insumo essencial, visando ao seu acesso e ao seu uso racional e garantindo à população o acesso a medicamentos considerados essenciais e promover o uso racional dos mesmos. Estes medicamentos devem ser seguros, eficazes e de qualidade.

O SUS seleciona um elenco de medicamentos com o objetivo de atender as demandas prioritárias da população. Esta seleção é feita utilizando a metodologia de Saúde Baseada em Evidências, adotando, de forma hierárquica, os critérios de eficácia, segurança, comodidade e custo, comparando-se as alternativas existentes.

A gestão das Políticas Nacionais de Assistência Farmacêutica, de Medicamentos de Plantas e Fitoterápicos é de responsabilidade do Departamento de Assistência Farmacêutica e Insumos Estratégicos - DAF, que tem como principal objetivo a execução de ações que garantam e ampliem o acesso da população a medicamentos promovendo o seu uso racional, sob orientação técnica e em estreita consonância com a Política Nacional de Saúde e de Assistência Farmacêutica.

Para que a população possa usufruir de todas essas ações, é feito um esforço coletivo entre municípios, estados e o Ministério da Saúde.”

O ALT e Python retornam as seguintes métricas:

Métrica	ALT	Python
Letras	1124	1124
Sílabas	518	521
Palavras	205	203
Sentenças	6	6
Letras/palavra	5.5	5,54
Sílabas/palavra	2.5	2,57
Palavras/sentença	34.2	33,83

Um dos principais desafios está sendo a reprodutibilidade dessas métricas, já que as formas de legibilidade se baseiam nelas. Embora esse exemplo tenha dado uma pequena variação, já testei exemplos que davam diferenças substanciais nas métricas.

Um outro ponto, seria a inclusão do corpus de texto para as 5000 palavras mais complexas, de forma que se calculasse a métrica de palavras complexas.

Por fim, a biblioteca poderia ser disponibilizada através do Python Package Index, de forma a ser utilizado via pip install; além disso, podem ser adotadas medidas de colaboração ou não do código. Acredito que, uma vez finalizado, seja um produto passível de registro.