**1 MENSURAÇÃO Legibilidade (LE)**

**1.1 Literatura Legibilidade (LE)**

**(Moreno & Casasola, 2016)**

O índice FLF será obtido com o auxílio do *software* Microsoft Word 2010, uma vez que este possui uma função de revisão do texto que fornece resultados da FLF. Para a mensuração do índice FLF os arquivos com extensão em *.pdf* serão convertidos para o formato *.doc* do Microsoft Word, sendo as tabelas, cabeçalhos e figuras, excluídas dos arquivos, para não provocar distorções no referido índice (Moreno & Casasola, 2016). A fórmula de facilidade de leitura do índice FLF, considera o comprimento da palavra medido em sílabas e o comprimento da frase medido em número de palavras. O fator palavra mede a dificuldade semântica e a velocidade de reconhecimento, já o fator frase mede a capacidade de memória de curto prazo.

A fórmula do índice FLF é especificada a seguir, conforme a Equação 1.

FLF = 206,835 – (84,6 x CP) – (1,015 x CF) (Equação 1)

Em que:

FLF = é a medida de facilidade de leitura: medido pelo índice FLF;

CP = comprimento da palavra: medido pelo número de sílabas dividido pelo número de palavras; e

CF = comprimento da frase: medido pelo número de palavras dividido pelo número de frases.

Quanto maiores as palavras e as frases pior a legibilidade. O FLF aplicada a fórmula, de 100 em 100 palavras.

Os valores são mensurados numa escala de 0 a 100, classificando o texto, por nível de facilidade de leitura, do nível de legibilidade mais difícil para o mais fácil, conforme a Tabela 1.

Tabela 1

**Níveis de Legibilidade do Índice Flesch**

|  |  |
| --- | --- |
| **Pontuação do FLF** | **Nível de Legibilidade** |
| 00 – 25 | Muito Difícil |
| 25 – 50 | Pouco Difícil |
| 50 – 75 | Fácil |
| 75 – 100 | Muito Fácil |

**Fonte**: Martins *et al.* (1996).

**(Gomes, Ferreira, Martins, 2018)**

As empresas foram analisadas quanto à variação do volume de NEs por meio da diferença do número de páginas, caracteres, palavras, sílabas, frases e parágrafos contidos nas notas explicativas. Para a coleta desses dados, utilizaram-se softwares criados pelos próprios autores combinando funções das seguintes bibliotecas para Python: PyPDF2, Pdf2Txt e regex.

Criou-se um programa para contar as páginas dos arquivos em “.pdf” e exportar os  
resultados para uma tabela com o nome de cada empresa, o ano de referência das notas  
explicativas e o número de páginas de cada um dos arquivos. Enquanto que para a conferência da variação no número de palavras foi necessário extrair o texto de cada arquivo em “.pdf” e salvá-lo em um arquivo “.txt”. Como o tamanho das letras, o espaçamento entre elas ou outro artifício visual podem influenciar a quantidade de páginas, sem necessariamente afetar a quantidade de informações, é importante destacar que não foram encontradas diferenças nos layouts das notas explicativas de uma mesma empresa nos dois anos analisados. Portanto, uma variação positiva (ou negativa) de páginas de fato indica aumento (ou redução) de informação divulgada.

Antes do cálculo do índice de Flesch, utilizou-se a ferramenta Character Count Tool,  
disponível sem custos no sítio eletrônico com o mesmo nome da ferramenta, para coletar os dados quanto ao número de caracteres, sílabas, frases e parágrafos. Ao realizar upload das notas explicativas em formato “.pdf” na ferramenta em questão, ela fornece um relatório completo contendo todos esses dados, entre outros. Essa ferramenta também fornece o tempo estimado de leitura do documento de texto, calculado com base na velocidade média de leitura de um indivíduo que é de aproximadamente 200 palavras por minuto – variável também utilizada neste trabalho.

Este trabalho utiliza como dados o número de páginas, caracteres, palavras, sílabas,  
sentenças e parágrafos e, como indicadores, utiliza-se o tempo estimado de leitura (em  
minutos), a quantidade de palavras sobre o número de parágrafos, a quantidade de palavras sobre o número de frases e a quantidade de sílabas dividida pelo número de palavras. Todos esses dados e indicadores representam o tamanho (rapidez de leitura) das notas explicativas. Já o índice de legibilidade de Flesch é um indicador de facilidade ou qualidade de leitura.

**2 FORMA DE MENSURAÇÃO LEGIBILIDADE DA TESE**

Mensurar pelo Python – fórmulas adaptadas para a língua portuguesa

**1 Teste de facilidade de leitura de Flesch (Flesch reading ease)**

Comprimento da frase: medido pela quantidade de palavras dividido pela quantidade de frases;

Comprimento da palavra: medido pela quantidade de sílabas dividido pela quantidade de palavras.

E após coletar esses dados, inserir na fórmula abaixo de FLF para mensurar a LE.

FLF = 226 – 1,04 x (Quantidade de palavras) – 72 x (Quantidade de sílabas)  
 Quantidade de frases Quantidade de palavras

**2 Índice Gulpease (Indice Gulpease)**

IG = 89 + 300 x (Quantidade de sentenças) – 10 x (Quantidade de letras)  
 Quantidade de palavras

**3 Nível de escolaridade de Flesch-Kincaid (Flesch-Kincaid grade level)**

NG = 0,36 x (Quantidade de palavras) + 10,4 x (Quantidade de sílabas) - 18  
 Quantidade de sentenças Quantidade de palavras

**4 Índice de nebulosidade de Gunning\* (Gunning fog index)**

ING = 0,49 x (Quantidade de palavras) + 19 x (Quantidade de palavras complexas)  
 Quantidade de sentenças Quantidade de palavras

**5 Índice de leiturabilidade automatizado (Automated Readability Index - ARI)**

ILA = 4,6 x (Quantidade de letras) + 0,44 x (Quantidade de palavras) - 20  
 Quantidade de palavras Quantidade de sentenças

**6 Índice de Coleman-Liau (Coleman-Liau index)**

ICL = 5,4 x (Quantidade de letras) - 21 x (Quantidade de sentenças) - 14  
 Quantidade de palavras Quantidade de palavras

**7 Índice de Legibilidade Geral**

ILG = NG + ING + ILA + ICL  
 4

**RESUMO VARIÁVEIS PARA MENSURAÇÃO LEGIBILIDADE NO PYTHON**

São necessárias as informações abaixo da Nota Explicativa de cada empresa e em cada ano de 2010 a 2021.

1 - Quantidade de letras.

2 - Quantidade de sílabas.

3 - Quantidade de palavras.

4 - Quantidade de palavras complexas.

Palavras complexas = palavras com 3 ou mais sílabas.

5 - Quantidade de sentenças.

Sentença = oração (é toda frase que contém um ou mais verbos), normalmente dentro de um parágrafo, a sentença ou oração, é delimitada entre o início da frase (letra maiúscula) até o sinal de pontuação.

Exemplo de uma sentença:

As demonstrações contábeis foram elaboradas seguindo os princípios contábeis.

6 – Quantidade de páginas.

**SCRIPT LEGIBILIDADE NO PYTHON**

PASTAS

Pasta principal

<https://github.com/oziieljuniior/Work_Text/tree/master/python_project/Legibilidade>

Explicações sobre o código

<https://github.com/oziieljuniior/Work_Text/blob/master/python_project/Legibilidade/sobre_legibilidade.txt>

Código final

<https://github.com/oziieljuniior/Work_Text/blob/master/python_project/Legibilidade/legibilidade.py>

SCRIPT

1. Instalar 2 bibliotecas:

1 – nltk - essa biblioteca é de machine learning e é utilizada para realizar a separação de sentenças no texto.

Para instalar digitar no terminal do Visual Code:

python3 -m pip install nltk

2 – pyphen - que realiza a contagem de sílabas no texto.

Para instalar digitar no terminal do Visual Code:

python3 -m pip install pyphen

1. Rodar o código.
2. Ele vai abrir duas janelas para seleção de caminho:

O primeiro caminho é o local onde os relatórios estão salvos.

O segundo caminho é o local onde você quer que a planilha gerada seja salva.

1. A separação dos arquivos para análise de legibilidade deve ser igual ao de violação gramatical.