



Evento	Salão UFRGS 2022: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
Ano	2022
Local	Campus Centro - UFRGS
Título	Fundamentos linguísticos para a acessibilidade da informação científica para leitores de escolaridade limitada: simplificação textual, gramatical, lexical e terminological em ciências da saúde
Autores	GUILLERMO SILVA VILLAR FRANCINE FACCHIN ESTEVES
Orientador	MARIA JOSE BOCORNY FINATTO

RESUMO

TÍTULO DO PROJETO: FUNDAMENTOS LINGUISTICOS PARA A ACESSIBILIDADE DA INFORMACAO CIENTÍFICA PARA LEITORES ADULTOS DE ESCOLARIDADE LIMITADA: SIMPLIFICACAO TEXTUAL, GRAMATICAL, LEXICAL E TERMINOLOGICA EM CIENCIAS DA SAUDE

Aluno: Guillermo Silva Villar

Orientador: Maria José Bocorny Finatto

RESUMO DAS ATIVIDADES DESENVOLVIDAS PELO BOLSISTA

O trabalho do bolsista trabalho visou ajudar a ampliar as temáticas cobertas pela Ferramenta MedSimples (veja em: www.ufrgs.br/textecc/acessibilidade/page/cartilha/), um sistema *on-line* gratuito de apoio à produção de textos escritos sobre temas médicos. MedSimples visa dar suporte à produção de informações escritas sobre temas de Saúde para pessoas adultas de escolaridade limitada, com apresentação de sugestões de reformulação que podem ser incorporadas ao texto do usuário. Essa ferramenta tem duas funcionalidades de reconhecimento de palavras que partem de duas principais bases de dados interconectadas: a) base de palavras difíceis com indicações de sinônimos mais fáceis; b) base de terminologias especializadas com definições simplificadas.

Na fase atual, o bolsista está ajudando a construir novas bases de dados para utilização com textos das áreas de Oncologia e Cuidados Paliativos: será destacado aqui somente o trabalho no módulo de Oncologia. A partir de um conjunto documental de referência voltado para a divulgação sobre Câncer para o público leigo, que contém um universo de 42.145 palavras (*tokens*), são extraídas informações para abastecer as funcionalidades da ferramenta. A extração é feita com o *software* AntConc (ANTHONY, L. (2021), que produz diferentes tipos de listas de palavras e expressões a partir de arquivos de texto e aponta seus *scores* de frequência. A partir do reconhecimento do vocabulário dos documentos de referência, foram identificadas as palavras potencialmente difíceis e as terminologias mais frequentemente empregadas nesses textos. O grau de potencial dificuldade das palavras que não são termos técnicos ou científicos é validado pela base textual Corpop (www.ufrgs.br/textecc/porlexbras/corpop/index.php) construída por Pasqualini (2018), que serve como referência de um vocabulário mais simples ou acessível. As terminologias e suas definições são verificadas por meio do uso de dicionários de tipo geral, dicionários especializados e de fontes documentais especializadas na divulgação de assuntos de Medicina.

No momento, após o processamento dos documentos de referência de Oncologia, já identificamos: 85 palavras e expressões potencialmente difíceis para o público-alvo que tem escolaridade limitada ao Ensino Fundamental, tais como MONITORAMENTO DE DADOS e ADJUVANTE; e 493 termos técnicos tais como APOPTOSE e ANGIOGÊNESE TUMORAL, que carecem de explicações ou definições facilitadas. O bolsista, junto com a orientadora e uma mestrandia, está construindo propostas de definições simplificadas para essas terminologias, as quais serão conceitualmente validadas por profissionais da Saúde.

Um exemplo de enunciado definitório simplificado em construção, ainda a ser revisado pelo médico parceiro, é: [BIÓPSIA= [<Quando se corta um pedacinho de uma parte do corpo para ser examinado em um laboratório>]].

Como resultado do trabalho em andamento, as informações levantadas dos textos, processadas sob a forma de listas, abastecem duas novas bases de dados que serão incorporadas à MedSimples, no módulo de Oncologia. Antes disso, cada palavra das listas geradas recebe um enquadramento automático, denominado *parsing*. Nesse enquadramento, cada item é etiquetado por sua categoria e função gramatical, assim, uma palavra como PINTAS é transformada em PINTA, identificada como palavra difícil ou terminologia. Se for palavra difícil, é então categorizada: [**substantivo** (uma marca na pele)]; ou [**verbo** (faz pintura)] ou [**adjetivo** (padrão, aparência)] e esses significados são mostrados para o redator-usuário. Caso um item seja identificado como uma terminologia, como BIÓPSIA, ele será conectado a sua respectiva definição simplificada, que será apresentada como sugestão na saída do processamento. A definição simplificada, por sua vez, deve ser formulada utilizando-se, preferentemente, um repertório de palavras mais fáceis, como vê-se no exemplo acima para BIÓPSIA.

Esse processo de produzir novas definições facilitadas em Oncologia vai inclusive servir como ponto de melhoria para outras definições em diferentes módulos já ativos da Ferramenta MedSimples, onde já há, por exemplo um mesmo termo definido de modo diferente. Esse é o caso do termo [BIÓPSIA], no módulo de COVID, que está assim apresentado: [<Procedimento no qual se colhe uma amostra de tecidos ou células do nosso corpo para ser estudada em laboratório; exame muito usado para verificar câncer>]. Essa definição poderá ser substituída pela nova, mais simples. Assim sendo, quando estiver concluído o processo de identificação do vocabulário do novo módulo Oncologia e estiver concluída a geração das suas duas novas bases de dados, o usuário do sistema poderá inserir um texto de seu interesse, iniciar o processamento e, como resultado, visualizar um texto marcado com sugestões de melhorias. Receberá itens sublinhados em azul, conjuntamente com seus respectivos sinônimos facilitados e alertas, e as terminologias sublinhadas em verde, acompanhadas de definições ou explicações mais facilitadas, podendo assim incorporar ou não esse conjunto de sugestões e melhorias ao seu texto.

Referências:

ANTHONY, L. (2022). AntConc (Versão 4.0.0) [Software de computador]. Universidade de Waseda. Tokyo, Japão. Disponível em: <https://www.laurenceanthony.net/software>
PASQUALINI, Bianca Franco. **CorPop: um corpus de referência do português popular escrito do Brasil**. Orientadora: Maria José Bocorny Finatto. Tese (Doutorado) - Universidade Federal do Rio Grande do Sul, Instituto de Letras, Programa de Pós-Graduação em Letras, Porto Alegre, BR-RS, 2018.