

Resumo de Poster

Autores:

Arthur Alberto Corrêa Treuherz

Ana Cristina Espindola Campos

Heitor Barbieri

Olga Pedroza Ribeiro

E-mail: arthur.treuherz@bireme.org (apenas o do autor principal)

Título: Adequação do vocabulário DeCS à nova ortografia da língua portuguesa

Área Temática: Informação para a decisão em saúde

País: Brasil

Introdução:

O trabalho visa atender determinação do Decreto 6583/08 do governo brasileiro para atualização da ortografia da língua portuguesa até 31/12/2012, conforme Acordo Ortográfico aprovado em 1990 pelos países cujo idioma oficial é o português, para sua unificação. Assim, a versão do DeCS 2013 em português compreenderá a nova ortografia. Em virtude do grande número de termos, o objetivo do trabalho foi permitir o máximo de automatização das alterações necessárias, com o mínimo de erro.

Métodos:

- Coleta e avaliação da documentação oficial sobre a nova ortografia;
- Identificação dos campos DeCS contendo palavras em português;
- Identificação primeiramente de termos contendo acentos e/ou hífen e geração de listas com estas palavras para cada campo;
 - Palavras com hifens e número, são quebradas nos hifens e a parte contendo número é descartada; nestes casos, palavras começadas/terminadas com hífen, perdem o hífen.
- Uso de API do corretor ortográfico LibreOffice, permitindo acesso ao dicionário por programação, e separação das palavras em três listas de trabalho;
- Análise e enxugamento manual destas listas para alteração automática do vocabulário:
 - (a) **Corretas**: confirmação se palavras identificadas como corretas realmente estavam corretas, após o que sai do fluxo;
 - (b) **incorretas com sugestões pelo corretor**: avaliação das sugestões do corretor para as palavras incorretas, eliminando erros do corretor;
 - (c) **incorretas**: inclusão de sugestão pelo setor e identificação e eliminação manual de palavras corretas não existentes no dicionário do corretor.

- Separação das listas enxutas (b) e (c) para descritores e sinônimos em duas: só diferem na acentuação e diferem na grafia. Para a primeira, alteração automática e para a segunda, manual.
- Após tratamento das palavras contendo acento e/ou hífen, as demais passaram pelo mesmo método acima.

Resultado e discussão

Das 23755 palavras em português com hífen e/ou acento, o corretor identificou 76% como corretas, sugeriu correção para 5%, restando 19% para avaliação visual e correção automática ou manual. Dessas, aproximadamente 90% estavam corretas, não sendo reconhecidas pelo corretor por serem técnicas.

Conclusões e comentários:

Os dados até o momento indicam que o DeCS 2013 terá 6,9% das palavras modificadas devido ao acordo ortográfico e a erros encontrados. Por ser o DeCS um vocabulário técnico, muitas das palavras não existem no Acordo, no Vocabulário Ortográfico da Língua Portuguesa (VOLP) ou em Dicionários como Houaiss e Aurélio, o que faz do processo manual de revisão uma tarefa árdua. A versão do DeCS 2013 abrangerá a nova ortografia, porém não se descarta a possibilidade de pequenos ajustes pontuais.

Confirmação de recebimento do Resumo:

Recebemos o resumo de sua experiência para a Sessão de Pôsteres do **CRICS9 - 9o Congresso Regional de Informação em Ciências da Saúde**. A divulgação dos Pôsteres selecionados será realizada por email no dia **10 de Septiembre de 2012**. Obrigado por enviar seu resumo! Comitê Organizador - CRICS9