

DESEMPENHO DE COERÊNCIA DO INDEXADOR AUTOMÁTICO DA NLM – METAMAP APLICADO EM BASE DE ARTIGOS DA SCIELO¹

Gabrieli Fonseca², Prof. Dr. Divino Ignácio Ribeiro Junior³

¹ Vinculado ao projeto “Indexação automática: desenvolvimento de uma plataforma de indexação de acervos textuais digitais para repositórios institucionais”

² Acadêmica do Curso de Biblioteconomia – FAED – Bolsista PROBITI

³ Orientador, Departamento de Biblioteconomia e Gestão da Informação – FAED –
divinoirj@gmail.com

Uma área de estudo de grande importância no campo da Ciência da Informação é a Indexação, por sua natureza multidisciplinar e por ser aplicada por todas as áreas de conhecimento. A Indexação consiste em técnicas de análise de conteúdo que possibilitam sintetizar as informações importantes de um documento por meio da delegação de termos que o representem, com o intuito de tornar os documentos recuperáveis. Existem três tipos de indexação: manual, automática e semiautomática. Nosso foco é no aprofundamento da indexação automática e suas ferramentas, das quais o MetaMap¹ – indexador automático da NLM – National Library of Medicine – é o objeto de estudo deste trabalho. Nosso objetivo principal foi analisar a medida de coerência entre os termos obtidos com a indexação automática do software da NLM para com os termos de assunto atribuídos pelo indexador humano (bibliotecários da FURB). Os dados coletados foram organizados em planilhas para cálculo das medidas de coerência e obtenção dos resultados.

Foi realizado uma pesquisa quali-quantitativa, de natureza básica e de caráter bibliográfico e, possui como objetivos ser uma pesquisa exploratória. Esse trabalho justifica-se por ser pioneiro nesse assunto e dessa forma, agregar novos conhecimentos para a área biblioteconômica.

Os dados utilizados na pesquisa estão publicados em <http://www.labtecgc.udesc.br/alessa-dados-abertos/> e foram produzidos durante a pesquisa de mestrado da acadêmica Alessa Fabíola dos Santos, acadêmica do PPGInfo/UDESC em 2017. Nessa pesquisa foram gerados dados a partir de 52 artigos da área da saúde que passaram por processos de indexação manual e indexação automática, que foram consolidados e publicados no portal do LabTecGC. Para cada monografia foram registrados os termos propostos através da indexação manual pelos bibliotecários, e, os descritores sugeridos pelo indexador automático Metamap da NLM que foram acolhidos pelos bibliotecários indexadores.

Nesta etapa da pesquisa do projeto os dados desses 52 documentos foram analisados para produzir os indicadores de coerência mencionados no objetivo principal. Na síntese dos dados foram analisados um total de 351 descritores gerados pelo indexador automático da NLM – MetaMap, e que desses 351 descritores, 152 são equivalentes semanticamente com os descritores propostos pelo indexador humano (bibliotecários da FURB). Desse modo, o sistema foi 43% coerente com o indexador humano. Ou seja, 43% dos termos atribuídos pelo indexador automático foram equivalentes aos termos atribuídos pelos bibliotecários.

Então, considerando o escopo da coleta dos dados verificou-se a capacidade do sistema computacional em agregar valor ao trabalho do bibliotecário, visto que, sua atuação se mostrou

¹ https://www.nlm.nih.gov/research/umls/implementation_resources/metamap.html

eficaz enquanto indexador automático, pois o mesmo sugeriu termos cujos conceitos são os mesmos que aqueles obtidos pela indexação manual. Dessa maneira, o indexador automático pode ser utilizado em uma lógica de cooperação com o trabalho do bibliotecário e não de substituição do mesmo por uma ferramenta tecnológica.

Por fim, espera-se que nas etapas futuras do projeto de pesquisa outras áreas do conhecimento possam ser investigadas e propostas novas tecnologias para a indexação automática, consolidando um conjunto de pesquisas nessa área, agregar novos conhecimentos e ferramentas para a Biblioteconomia e a Ciência da Informação.

Palavras-chave: Indexação. Indexação automática. Análise de Coerência.