



Ciências Exatas e da Terra

EXTRAÇÃO AUTOMÁTICA DE METADADOS

Oswaldo Luiz de Souza Ferreira Fabio Duncan de Souza

Oswaldo Luiz de Souza Ferreira “As bibliotecas digitais vem sendo empregadas de forma crescente nas mais diversas áreas. Em consequência deste crescimento, cresce também a demanda por ferramentas capazes de automatizar os processos que envolvem a submissão dos documentos para tais bibliotecas. Um exemplo de tarefa a ser realizada quando da submissão de um conteúdo para uma biblioteca digital é o cadastro dos metadados deste conteúdo. Isto se deve ao fato de que tais bibliotecas armazenam o documento digital e precisam de informações sobre este para que consultas possam ser executadas e apresentadas aos usuários. Assim, torna-se fundamental a extração de metadados específicos do documento armazenado para que as informações sejam geradas de maneira a satisfazer os critérios estabelecidos (MANICA, Edimar ; CERVI, C. R. ; GALANTE, R. 2008). Durante o processo de criação de conteúdos digitais é muito comum que os seus autores não possuam conhecimento ou não tenham a preocupação de preencher os metadados disponibilizados nos formatos de arquivos utilizados. Porém, para realizar o cadastro destes metadados de forma manual, em uma biblioteca digital, gera uma grande quantidade de esforço e tempo. O objetivo deste projeto é criar uma ferramenta que possa extrair automaticamente tais metadados, buscando-os no corpo do texto de documentos como Trabalhos de Conclusão de Curso (TCC), Artigos de Eventos e Periódicos. Cabe ressaltar que o desenvolvimento desta ferramenta faz parte de uma projeto maior que é a Biblioteca Digital da EPCT, desenvolvida pelo Núcleo de Pesquisa em Sistemas de Informações do Instituto Federal Fluminense (NSI/IFF).

Palavras-chave: metadata, extrator, automático

Instituição de fomento: IFF