018 **UMA FERRAMENTA PARA DETECÇÃO DE VERSÕES DE DOCUMENTOS XML ATRAVÉS DA IMPLEMENTAÇÃO DE ALGORITMOS DE SIMILARIDADE.** *Rodrigo Otávio Silva Santos, Renata de Matos Galante (orient.)* (UFRGS).

O problema da detecção de versões de arquivos é importante em vários cenários, como identificação de clones de software, ranking de páginas Web, detecção de plágio, e busca em sistemas P2P. Existem pesquisas em detecção de diferenças que podem ser utilizadas como base para a identificação de versões. Estas abordagens utilizam algoritmos para detectar similaridades/diferenças entre arquivos. Por exemplo, se dois arquivos possuem um conjunto pequeno de diferenças, então eles são considerados duas versões; caso contrário, dois documentos diferentes. Estes algoritmos geralmente levam em consideração o conjunto de operações atômicas que deveria ser aplicado para obter um arquivo alvo a partir de um arquivo original. No entanto, a saída destes algoritmos é geralmente um documento que representa as diferenças entre os dois documentos de entrada. Semanticamente, este conjunto de diferenças não traz muita informação em relação ao grau de similaridade entre os arquivos analisados. Este trabalho visa à implementação de funções de similaridade para arquivos XML, com o objetivo de detectar versões de um mesmo documento. Para isso, são implementados dois mecanismos: um para análise de similaridade entre arquivos que possuem somente diferenças de conteúdo, e outro para arquivos que possuem diferenças de conteúdo e estrutura. As funções implementadas consideram as seguintes características: (i) o percentual de elementos que são comuns, diferentes, foram adicionados e/ou removidos entre os dois arquivos, com base num algoritmo diff;(ii) a relevância da diferença entre certos elementos (por exemplo, uma diferença na data de nascimento de uma pessoa é mais significativa que uma diferença no endereço desta pessoa). Como ponto positivo do trabalho, as funções implementadas consideram várias características, não são restritas a um domínio específico e podem ser adaptadas para considerar outros cenários.