017 **IMPLEMENTAÇÃO DE UMA FERRAMENTA DE TRADUÇÃO DE ENTIDADES HTML EM CARACTERES ACENTUADOS.** *Andrea Raymundo Balle, Renata de Matos Galante (orient.)* (UFRGS).

Entidades HTML são representações, em forma de código, de caracteres especiais ou acentuados. Têm por característica começarem pelo símbolo "&" e terminarem com ";". Quando lidas por um navegador de internet, são interpretadas e é mostrado na tela o símbolo correspondente. Algumas aplicações, no entanto, não identificam caracteres como o "&", mas lêem perfeitamente caracteres especiais e acentuados. O objetivo desse trabalho é desenvolver uma ferramenta que leia um arquivo XML, identifique as entidades HTML que estão contidas em cada um de seus nodos e as substitua por seus caracteres correspondentes. Anteriormente ao início do desenvolvimento do código, foi feito o estudo das entidades HTML e seus padrões, o que resultou numa maior otimização da aplicação. Tendo em vista que o objetivo primordial é a eliminação das entidades no arquivo XML da DBLP, um site de referências bibliográficas em Ciência da Computação que reúne milhares de artigos, sabe-se os que os documentos a serem tratados são grandes. Por esse motivo, a implementação não poderia montar toda a estrutura do XML na memória. Isso determinou as ferramentas a serem utilizadas na implementação: a linguagem de programação Java e o parser SAX, uma api do Java 1.4 que faz o parsing de um XML por eventos. Esta ferramenta está inserida no contexto de um projeto de mestrado que especifica um mecanismo para representar e detectar diferentes versões de um mesmo documento XML em bibliotecas digitais. A ferramenta desenvolvida será um componente para o mecanismo de detecção de versões, sendo posteriormente expandida com novas funcionalidades. Cabe ressaltar, entretanto, que a ferramenta desenvolvida pode ser utilizada em outros projetos que envolvam XML e que necessitem de um documento sem entidades HTML. (CNPq).