

293

IMPLEMENTAÇÃO DE UM MECANISMO DE COLETA DE DADOS NA WEB. *Oscar Paesi da Silva, Maria Aparecida Martins Souto (orient.) (UFRGS).*

Este trabalho insere-se no contexto do projeto de pesquisa PERXML. O objetivo do PERXML é construir uma ontologia de perfil de usuário que agregue informações cadastrais, competências, qualificações, entre outros, visando a sua utilização em sistemas de recomendação para ambientes de biblioteca digital. Esta bolsa de Iniciação Científica tem como objetivo implementar um protótipo para a extração de informações de usuário na Web. O desenvolvimento do trabalho prevê 2 etapas principais a saber: (i) estudo de técnicas de extração de dados na Web e das tecnologias Java e HTML; e (ii) desenvolvimento de um protótipo que recebe como entrada o endereço da página Web de uma Conferência Científica e captura os nomes dos integrantes do Comitê de Programa através de um parser. Para o desenvolvimento do protótipo foi escolhido o ambiente Eclipse e o servidor Web TomCat. A captura dos nomes envolve o uso de técnicas de busca de informação semi-estruturadas em páginas HTML. Para a extração dos nomes está sendo especificado um conjunto de regras para a construção do parser. O processo de geração de regras é continuamente aprimorado, para extrair, se não todos, o maior percentual possível de nomes válidos. Os resultados deste trabalho servirão como base para a extração de outras informações na Web, as quais deverão compor a ontologia de perfil do usuário no contexto do projeto PERXML. (BIC).