

Desenvolvimento e validação de uma ferramenta distribuída para coleta temática de páginas da Web baseada em gênero.

MARCOS VINICIUS OLIVEIRA SOUZA (Autor), Guilherme Tavares de Assis (Orientador), Vítor Mangaravite (Co-Orientador)

Coletores temáticos apresentam o propósito maior de coletar páginas da Web que sejam relevantes a um tópico ou interesse específico do usuário, sendo importantes para uma grande variedade de aplicações. Em geral, eles funcionam tentando localizar e coletar páginas que estejam relacionadas a um determinado tópico de interesse. Nesse contexto, foi proposta e desenvolvida uma abordagem para coleta temática onde o tópico de interesse pode ser expresso por termos que descrevem o conteúdo e o gênero das páginas da Web desejadas, possibilitando a construção de coletores temáticos que realizam processos de coleta eficazes e eficientes, conforme já demonstrado experimentalmente. Com o objetivo de melhorar a escalabilidade de tal abordagem para coleta temática baseada em gênero, esse projeto de iniciação científica propõe uma arquitetura nova de funcionamento para a mesma, onde etapas relativas a processos de coleta temática podem ser realizadas de forma distribuída. Experimentos comprovaram a melhoria da escalabilidade da abordagem em relação à sua forma original de funcionamento: de uma forma geral, ao utilizar 8 computadores, o ganho foi, na média, de 83,50% considerando o tempo de execução total de processos de coleta temática relativos a dois tópicos de interesse distintos.

Instituição de Ensino: Universidade Federal de Ouro Preto