

NOME: TÚLIO ROBERTO DE LIMA LARA

TÍTULO: Extração e Integração de Dados Web em Fontes Heterogêneas

AUTORES: TÚLIO ROBERTO DE LIMA LARA

ORIENTADOR: THIAGO MAGELA RODRIGUES DIAS

AGÊNCIA FINANCIADORA (se houver): PAPq

PALAVRA CHAVE: Dados Web, Integracao de Dados, Extracao de Dados

RESUMO

A quantidade de dados produzidos na Web aumenta a cada dia. Novos serviços como blogs, redes de relacionamentos, sítios dinâmicos e feeds RSS, são alguns exemplos de como a chamada Web 2.0 tem contribuído consideravelmente na quantidade de dados gerados, permitindo que usuários não apenas acessem conteúdo disponível, mas também produzam dados de natureza diversa a partir de sua interação com a Web. A quantidade enorme de dados gerados exige o uso de técnicas automatizadas como Descoberta de Conhecimento em Banco de Dados e Mineração de Dados para extração de informações úteis. Com os dados integrados, técnicas de Mineração de Dados como agrupamento, classificação e associação, entre outras, podem ser aplicadas para a extração de conhecimento útil a cerca destes dados. Este trabalho tem como objetivo o desenvolvimento de um Mashup framework para a extração e integração de dados em fontes heterogêneas para mineração.

A Web 2.0 se caracteriza muito mais por uma mudança de postura por parte dos usuários do que por uma infra-estrutura tecnológica inovadora. O que vem acontecendo é que o público está deixando de ser apenas consumidor de informação, mas também para se tornar produtor [1].

Para Musser et al. [1] a Web 2.0 pode ser definida como um conjunto de "tendências econômicas, sociais e tecnológicas que coletivamente fundam a próxima geração da Internet – uma mídia mais madura e distintiva, caracterizada pela participação dos usuários, abertura, e efeitos de rede."