065 **ANOTAÇÃO AUTOMÁTICA DE KEYWORDS PARA PROTEÍNAS RELACIONADAS A MYCOPLASMATACEAE USANDO TÉCNICAS DE APRENDIZADO DE MÁQUINA.** *Abner N. Pitinga, Ana L. C. Bazzan* (Instituto de Informática, Departamento de Informática Teórica, UFRGS).

Uma das grandes necessidades nos projetos genoma de hoje é a de analisar e anotar uma grande quantidade de seqüências de proteínas de forma rápida e eficaz. Entretanto, os pesquisadores envolvidos nessas pesquisas não possuem tempo suficiente para fazer a anotação completa dessas seqüências, o que torna óbvia a necessidade de se criar ferramentas para automatizar esse processo. A proposta deste trabalho é a utilização de técnicas de aprendizado de máquina que, a partir de um conjunto de dados já anotados, gerem regras para anotação de keywords para novas seqüências de proteínas. O conjunto de dados usados para geração dessas regras são provenientes do banco de dados de seqüências de proteínas Swiss-Prot. Todas as seqüências usadas estão relacionadas à família Mycoplasmataceae, a qual pertence o Mycoplasma hyopneumonia - bactéria que afeta os suínos. Essa escolha se deve ao fato de que esse trabalho visa ser usado no Projeto Pigs que está fazendo o sequênciamento dessa bactéria. Esse trabalho automatiza as seguintes etapas: obter os dados do Swiss-Prot, formatá-los e gerar as regras de anotação aplicando o pacote WEKA sobre esses dados (este pacote implementa o algoritmo C4.5); viabilizando, assim, a utilização de um grande volume de dados.(CNPq).