

201

**EASY KEGG: UM SISTEMA PARA FILTRAR INFORMAÇÕES DO BANCO DE DADOS KEGG.** *Rejane Apolo Ferreira, Ney Lemke, José Carlos Merino Mombach (orient.)* (Mestrado em Computação Aplicada, Centro de Ciências Exatas e Tecnológicas, UNISINOS).

O metabolismo é um conjunto de reações químicas celulares. Informações sobre o metabolismo de vários organismos podem ser encontradas em diversos bancos de dados como o KEGG (Kyoto Encyclopedia of Genes and Genomes – <http://www.genome.ad.jp/kegg>). Porém, existem muitas inconsistências nesse banco, tais como: ausência de reações químicas, compostos sem código relacionado, reações sem enzimas associadas, mesma informação dita de modos diferentes, erros aleatórios que são detectados somente por especialistas. Com o objetivo de organizar e padronizar as informações de nosso interesse contidas no KEGG, foi desenvolvido um sistema para filtrar e corrigir as informações nele encontradas. Para construir este sistema utilizou-se PERL, uma linguagem capaz de detectar padrões em bancos de dados, principalmente em formato de texto. Este sistema reúne as seguintes funcionalidades: geração de um arquivo que contém todas as reações de todos os organismos que se encontram no KEGG, identificação de todos os compostos principais da reação química, filtragem de informações de um organismo específico e padronização dos dados. O sistema gera dois tipos de arquivos: um que contém todas as reações de todos os organismos e outro que tem reações de determinado organismo somente. Utilizando este sistema, aliado a inspeção manual de especialistas, foi possível reunir um conjunto de dados corretos e correlacionados que podem ser utilizados por um software ou ferramenta de bioinformática para análises. Um exemplo é a construção de grafos do metabolismo que relacionam enzimas e metabólitos.