

066

PROCESSAMENTO DE LINGUAGEM NATURAL APLICADA AOS BANCOS BIOLÓGICOS DA NCBI. *Rafael Teodósio Pereira, Giovani Rubert Librelotto (orient.) (UNIFRA).*

O objetivo da Bioinformática é tornar possível a descoberta de novas introspecções biológicas, bem como criar uma perspectiva global na qual os princípios unificados da biologia possam ser discernidos. No começo da “revolução genômica”, um interesse da bioinformática foi a criação e manutenção de um banco de dados para armazenar informações biológicas como seqüências de nucleotídeos e aminoácidos. Com esse objetivo em mente, o NCBI (National Center for Biotechnology Information) criou uma série de bancos para armazenar os dados referentes a cada um dos tipos de informações biológicas: PubMed (para artigos científicos), GenBank (seqüências genéticas), Entrez Protein (para seqüências de proteínas) e Cancer Chromosomes (alterações cromossômicas em câncer), entre outros. Entretanto, o método de busca utilizado por tais bancos é baseado apenas em palavras-chave, podendo-se usar alguns limitadores para restringir o resultado das buscas. De forma a apresentar uma nova abordagem de busca nestes bancos mantidos pela NCBI, este projeto de pesquisa tem como objetivo permitir a busca nestes bancos através de linguagem natural. Desta forma, o usuário pode indicar frases completas (em inglês) e o sistema proposto fará um processamento da referida sentença. Este pré-processamento será realizado pelo Genia Tagger, o qual gera uma tabela que caracteriza cada elemento da sentença. A partir deste resultado, um parser mapeia esta tabela para um banco de dados MySQL, onde os elementos são analisados e dão origem a uma série de ligações Web para ocorrências nos bancos da NCBI citados previamente. Assim, a partir de uma sentença em inglês, o usuário terá acesso aos registros nestes bancos que estejam relacionados com a mesma.