

Banco de Dados para Bioinformática: Uma Introdução

Emilio Mario Wieczorek
Universidade Federal do Rio Grande
do Sul
Porto Alegre, RS, Brasil
+55 51 3308-3037
emilio@ufrgs.br

ABSTRACT

With the beginning of the Genoma Project, in 1990, some SGBD's (Data Base Management Systems) come being suitable to support the biological data (of genoma), in order to obtain itself to store and to search these information (given) in the possible way most excellent. The majority of the data bases for the bioinformatic (biological) consist of long chains of characters to represent the bases of DNA G (Guanine), A (Adenine), T (Timine) and C (Cytosine). A to be surpassed problem when it speaks in data base for bioinformatic is that data bases have been to a large extent used to manage given enterprise, simple numbers, character or dates. Few data bases had had a native ability to deal with complex data, as multimedia data, space data, or genetic data (succession of genes). For the resolution of such problem, it must be elaborated a standardization of types of data and methods that the data base must support, relating them and describing its functionalities.

Categorias

H.2.4 [Database Manager]: Banco de Dados para Bioinformática – *Tipos de dados, modelo de dados, tipo de banco de dados.*

Termos Gerais

Banco de Dados, Gerenciamento, Biológicos, Bioinformática, Performance, Modelo, Consulta, Construção, Data Warehouse, XML, Gene, Genoma, Genômico, DNA, Fragmento, Consulta.

Keywords

Bioinformatic, data base, consult, search, integration.

1. INTRODUÇÃO

O mapeamento do genoma humano e de outros organismos gera diariamente um elevado volume de informações que são sistematicamente armazenadas em bancos de dados computacionais, sendo estas informações fontes de estudo para a biologia e medicina através da bioinformática.

A bioinformática torna-se essencial para a construção de bases de dados contendo informações sobre os genes e proteínas dos organismos vivos, para a descoberta de novos genes, e de novos medicamentos, pois é através da bioinformática que novas técnicas para o mapeamento e armazenamento das informações extraídas dos genes vem sendo estudadas e estruturadas.

O desafio apresentado pela bioinformática é encontrar a melhor forma de armazenamento e de pesquisa (SQL) para os dados gerados por projetos de pesquisa na área da bioinformática, como

o projeto genoma humano, que possui centenas de gigabytes de dados a espera para serem armazenados e tratados.

Para tanto, surge a necessidade de se possuir formas de armazenamento, acesso e pesquisa sobre tais dados, para que se consiga trazer a informação da melhor maneira desejada possível, devendo existir assim, técnicas diferenciadas para o tratamento destes dados, que são nada mais do que grandes cadeias de DNA (em Banco de Dados, grandes cadeias de caracteres).

Este estudo tem como objetivo o levantamento dos principais esforços computacionais realizados na área de bioinformática, enfocando o uso de banco de dados, servindo para que num futuro breve, consigamos elaborar um padrão a ser utilizado por estes bancos de dados, facilitando assim, a integração de vários institutos de pesquisa que trabalhem com dados genômicos e moleculares, fazendo com que novas descobertas a cerca do genoma, principalmente do genoma humano, sejam realizadas mais rapidamente e com maior eficácia.

O presente artigo está dividido da seguinte forma: a Seção 2 aborda a Revisão de Literatura; a Seção 3 apresenta uma ferramenta para integração de dados e aplicações para dados biológicos; a Seção 4 aborda os trabalhos futuros, a Seção 5 aborda as conclusões e a Seção 6 traz as referências bibliográficas utilizadas na elaboração deste artigo.

2. REVISÃO DE LITERATURA

Há aproximadamente vinte anos atrás, havia uma quantidade de dados genômicos (de genes) relativamente pequena, pois os laboratórios biológicos gastavam meses realizando experiências com pequenos fragmentos de DNA ou de proteínas.

Devido a esta pequena quantidade de dados gerados os próprios biólogos administravam os resultados obtidos nestas pesquisas, geralmente armazenando estes resultados em arquivos de texto, codificados no padrão adotado pelo laboratório onde trabalhavam. Porém, com o Projeto Genoma Humano, a tecnologia avançou e a quantidade de dados gerada pelos laboratórios aumentou drasticamente, e as informações coletadas pelos experimentos utilizando genes do DNA não puderam mais ser armazenadas desta maneira “arcaica” [2].

Para a solução destes problemas, surge a bioinformática, um campo interdisciplinar, que age como interface entre os campos científico e tecnológico. A bioinformática se caracteriza por tentar prover métodos computadorizados para interpretar os dados referentes ao sequenciamento do genoma, que geram grandes volumes de dados espalhados em várias partes do mundo, de forma a trazer novos avanços para a biologia molecular (cura de doenças).

Para [3], a bioinformática representa, hoje, um dos grandes desafios para se tentar decifrar o genoma, pois ao mesmo tempo em que é uma forma de se conseguir informações imediatas para os dados do genoma que vem sendo descobertos, também é a base para um sucesso científico futuro.

Com o grande volume de informação gerado pelos projetos de análise de transcriptomas (fragmentos de DNA), tem se tornado cada vez mais complexo o armazenamento, acesso e a análise dos dados. Para contornar tal dificuldade, devem ser implementados, bancos de dados, que disponibilizem, de modo confiável, os dados e ferramentas de análise. Em muitos casos, esses bancos são abertos, o que aumenta ainda mais a aplicabilidade da pesquisa [6].

Segundo o Departamento Americano de Energia [4], a meta primária dos projetos de genoma públicos e privados é fazer uma série de mapas de diagramas descritivos de cada cromossomo humano a resoluções crescentemente melhores. Isto é feito dividindo os cromossomos em fragmentos menores que podem ser isolados, e ordenando estes fragmentos para corresponder aos locais respectivos dos cromossomos nos fragmentos.

Depois que a ordenação é completada, o próximo passo é determinar a sucessão de bases A (Adenina), T (Timina), C (Citosina) e G (Guanina) em cada fragmento. Então, várias regiões dos cromossomos da sequência serão “marcados” com sua respectiva função. Finalmente podem ser catalogadas diferenças em sucessões entre indivíduos em um cenário global.

Para se tentar solucionar os problemas relacionados com armazenamento, acesso e busca de dados, vários autores e institutos formulam propostas de como estes gigabytes de dados serão armazenados e tratados.

[1] afirma que todos os genes humanos serão achados eventualmente, sendo desenvolvidos diagnósticos para todas as doenças hereditárias, modelos animais para pesquisa de doenças humanas serão mais facilmente desenvolvidos, e curas serão desenvolvidas para muitas doenças.

Muitos destes desenvolvimentos acontecerão, não dentro de laboratórios biológicos, mas sim em plataformas computacionais de alto desempenho, com sistemas gigantesco para armazenar dados do genoma, bancos de dados para procurar pelos dados, semelhanças identificando padrões, como também integração de software para unificar as fatias de conhecimento desenvolvidas em instituições globalmente distribuídas.

Para [1], existem quatro tecnologias poderosas que se mostram promessas para resolver problemas intratáveis em bioinformática: a arquitetura de extensibilidade para armazenar uma sucessão de dados nativamente e executar estruturas de procura no banco de dados; tecnologias de warehousing para dados em padrões genéticos; tecnologias de integração de dados para habilitar questões heterogêneas por fontes biológicas distribuídas, e tecnologias de portal de Internet que permitem publicar informações de pesquisas na área da bioinformática.

Um problema a ser superado quando se fala em banco de dados para bioinformática é que bancos de dados têm, de longe, sido em grande parte usados para administrar dados empresariais, números simples, caráter ou datas. Poucos bancos de dados tiveram uma habilidade nativa para lidar com dados complexos, como dados

multimídia, texto, dados espaciais, ou dados genéticos (sucessão de genes).

A maioria destes dados fica difícil de ser controlados, como questões de achar a semelhança (em grandes cadeias de caracteres), questões sobre sucessões de gene e questões de localização de genes em cadeias de DNA. A figura 1 demonstra as etapas para o armazenamento de segmentos (sucessões de genes) de DNA em um banco de dados.

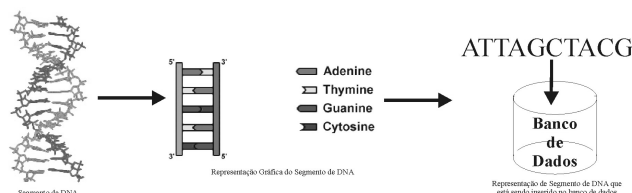


Figura 1. Etapas realizadas para armazenar um segmento (sucessão de genes) de DNA em um banco de dados.

Como pode ser observado na figura 1, os segmentos de DNA são analisados e transformados em uma representação gráfica, sendo depois transcritos para uma sequência textual e então armazenados em uma base de dados.

Segundo [1], para o caso específico de dados genômicos, deveria ser possível procurar por: Propriedades: Quais são as características (propriedades) de um segmento de DNA humano com tamanho igual ou superior a 10Kb e o que está associado a este segmento; Semelhança Estrutural: dado um segmento de genes qualquer (CGTAATGC), que outros segmentos existentes no banco de dados possuirão este mesmo segmento, tanto para este organismo quanto para outros organismos? (A “operação de possuir” deve encontrar somente segmentos que possuem em algum ponto de sua extensão o segmento dado para a procura); Local: dado um fragmento de DNA qualquer (CGTAATGC), qual é a sequência de genes que o antecedem e o precedem.

A menos que bancos de dados possam tratar nativamente de dados complexos, aplicações especializadas têm que ser usadas como intermediárias para executar busca e localização de genes em fragmentos de DNA no banco de dados. A tabela abaixo mostra alguns bancos de dados que permitem trabalhar com sucessões genômicas, sendo a maioria de institutos e universidades, que vêm trabalhando na elaboração de bancos de dados específicos para trabalhar com expressões gênicas.

Tabela 1. Bancos de Dados com capacidade de armazenar e buscar dados genômicos

Banco de Dados	Instituto/Empresa
NIH - Banco de dados de expressão gênica	Molecular Pharmacology of Cancer
SMD - Banco de Dados de Microarrays	Stanford University
YMGV - Visão global sobre Microarray de levedura	http://www.transcriptome.ens.fr/ymgv/

2.1 BANCOS DE DADOS EXTENSÍVEIS

Uma proposta interessante para a solução dos problemas de banco de dados em bioinformática é apresentada por [1] e [5]: devem ser

elaborados bancos de dados que sejam capazes de controlar tipos complexos, de modo a conseguir suprir as necessidades do domínio da aplicação, além de prover apoio a qualquer tipo de dado definido pelo usuário, ou seja, um banco de dados extensível.

Este banco de dados extensível dará apoio às necessidades do sistema para definir tipos de dados novos que sejam capazes de criar entidades de domínio como sucessão genotípica; uso de operadores definidos pelo usuário; indexação de domínio específico, fornecendo apoio para índices específicos de dados genômicos e otimizar a estensibilidade fazendo assim uma ordenação inteligente dos predicados em questão, envolvendo tipos de dados definidos pelo usuário.

Tipicamente, bancos de dados provêm um jogo de operadores pré-definidos para operar em tipos de dados embutidos. Podem ser relacionados os operadores matemáticos (+, -, *, /), de comparação (=, >, <), lógica booleana (NOT, AND, OR), comparação de strings (LIKE) e assim por diante.

Para que se tenham operadores definidos pelo usuário, a Oracle [5] acrescentou a seus bancos de dados (Oracle 8i/9i/10i) a capacidade para definir operadores de domínios específicos, ou seja, se torna possível definir um operador para comparar sucessões genômicas. A implementação do operador é deixado ao usuário, este podendo escolher as funções, os tipos de métodos, pacotes, rotinas de bibliotecas externas e assim por diante. Pode-se ainda, serem invocados os operadores definidos pelo usuário em qualquer lugar, estes podendo ser usados como operadores embutidos, isto é, onde quer que aconteçam nas expressões.

Os operadores definidos pelo usuário podem ainda ser usados em um comando SELECT, na condição de uma cláusula WHERE, na cláusula ORDER BY, e na cláusula GROUP BY. Depois que um usuário define um novo operador, este pode ser usado em comandos SQL juntamente com qualquer outro operador embutido.

Por exemplo: o usuário define um novo operador CONTEM() que possui um FRAGMENTO de DNA decodificado de uma sucessão particular, retornando TRUE se o fragmento contiver a sucessão especificada. Esta pesquisa poderá ser escrita da forma descrita na figura abaixo:

```
SELECT ID FROM TABELADNA
WHERE CONTEM(FRAGMENTO 'GCCATAGACTACA');
```

Figura 2. Exemplo de Consulta SQL para Bancos de Dados Biológicos/Bioinformáticos

Esta habilidade para aumentar a semântica dos operadores de domínio específico é um serviço oferecido pelo banco de dados.

2.2 DATA WAREHOUSE

Uma outra abordagem pode ser através de sistemas envolvendo Data Warehouses (Armazéns de Dados), pois estes são utilizados pela indústria há muitos anos, e como demonstrado pela figura 3, são constituídos tipicamente de 5 camadas: as fontes de dados, que contém os dados a serem integrados (adicionados) ao Data Warehouse através dos Wrapper's (analisadores gramaticais de dados), os mediadores (que traduzem os dados para a representação do Data Warehouse), o próprio Data Warehouse,

que é um grande repositório de dados, geralmente um banco de dados relacional, que apresenta uma visão consistente dos dados provenientes das fontes de dados, e finalmente os usuários, que interagem com o sistema através de uma interface.

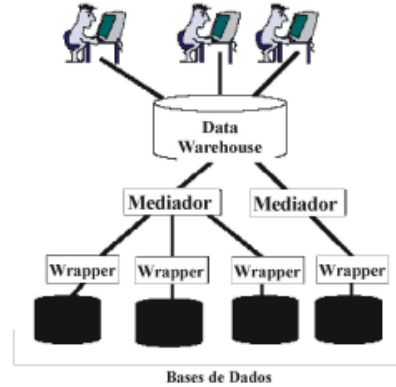


Figura 3. Estrutura de um Data Warehouse

Segundo [2], o desafio para a criação de um Data Warehouse para o ambiente da bioinformática está no fato de que deve-se desenvolver uma infra-estrutura flexível o bastante para controlar a natureza dinâmica do domínio, pois fontes de dados para aplicações científicas são extremamente dinâmicas. Sempre que uma fonte de dados muda seus dados, o Wrapper e o mediador devem ser atualizados para que estas atualizações sejam espelhadas no Data Warehouse. Isto se torna um grande desafio, pois deve-se manter um Data Warehouse extremamente funcional, mesmo integrando várias fontes de dados que sofram mudanças constantemente.

Para o acesso aos dados gerados por projetos de bioinformática (genômicos), o Lawrence Livermore National Laboratory [2] possui um projeto para a criação de um Data Warehouse (chamado de DataFoundry) para o ambiente de bioinformática (dados genômicos). O projeto começou a ser desenvolvido em outubro de 1996 e sua tarefa inicial era desenvolver uma infraestrutura que permitiria criar e manter uma visão consistente de várias fontes de dados autônomas.

A infra-estrutura de meta dados do DataFoundry [2] contém um gerador de mediador, um programa que automaticamente gera um mediador que usa uma coleção de meta dados declarativos, definindo uma biblioteca de classes que podem ser usada pelo wrapper para representar dados obtidos da fonte de dados. Isto simplifica a integração (adição) de novas fontes de dados, pois o administrador somente definirá o conjunto de meta dados apropriados e escreverá um wrapper que usará tais classes resultantes, ao invés de ter de escrever o wrapper e o mediador. Também irá simplificar a manutenção da Data Warehouse, pois é significamente mais fácil atualizar o conjunto de meta dados do que atualizar o mediador.

O DataFoundry proverá acesso para os usuários através de interfaces desenvolvidas basicamente em HTML e Scripts CGI, podendo esta interface ser desenvolvida também em uma linguagem de programação da escolha do laboratório/usuário, como PERL, C/C++, e outras.

Segundo [1], os dados gerados pela bioinformática não serão armazenados em um mesmo banco de dados. É provável que

dados de sucessão pertencentes ao genoma humano sejam colocados em banco de dados públicos ou privados. Também serão armazenados, por centenas de instituições, dados de sucessão para outros tipos de organismos. Novas pesquisas para estes dados farão com que novos dados sejam “criados” continuamente.

2.3 XML

[7] propõe uma solução onde é abordado o desenvolvimento de um sistema de banco de dados XML para o tratamento de dados biológicos. O modelo proposto está baseado em um SGBD XML e aborda a utilização de fontes de dados biológicos diferentes, integrando funcionalidades variadas de bioinformática com um sistema de bancos de dados. Este modelo permite a provisão de funções de bioinformática através da interface do SGBD.

O sistema é projetado como um módulo separado do SGBD XML, possuindo sua própria API que permite que o sistema de banco de dados o utilize através de um plug-in. Esta API provê acesso a uma coleção de dados biológicos através de ferramentas de análise que filtram e procuram por dados durante a execução de uma pesquisa XML.

A figura abaixo representa a interpretação do modelo atual de dados biológicos para análise.

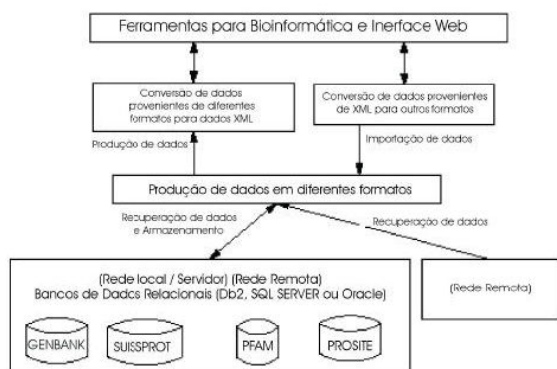


Figura 4. Representação da interpretação do modelo atual de dados biológicos para análise

3. FERRAMENTA DE INTEGRAÇÃO DE FONTES DE DADOS E APLICAÇÕES

As principais ferramentas utilizadas pelos pesquisadores em biologia molecular são associadas a dezenas de arquivos e bancos de dados públicos contendo informações relativas a sub-domínios específicos do conhecimento. Em muitos casos estas informações são produzidas por um único laboratório e suas informações são únicas, isto é, não estão replicadas nos repositórios públicos. Todas essas fontes de dados relevantes para os pesquisadores contêm informações biológicas (ex.: homologias, estrutura e similaridades) e afins (ex.: anotações relevantes e artigos científicos).

Em sua maioria estas fontes de dados diferem na forma de armazenamento de dados e nas informações relevantes a pesquisa. Também estão associadas a aplicativos que diferem nos serviços oferecidos: de visualização dos dados, de busca, de alinhamentos, de comparação de sequências, entre outros. Há implementações

de sistemas que armazenam informações biológicas em arquivos texto (ex.: GenBank), em bancos de dados relacionais (ex.: Swiss-Prot) e em sistemas orientados a objetos persistentes (ex.: AceDB).

Com as limitações das arquiteturas de integração existentes, [11] propõe uma arquitetura de software baseada em um framework orientado a objetos, de forma a prover flexibilidade e extensibilidade.

O framework que foi proposto (Figura 5) se propõe a integrar dados de qualquer fonte de dados de biologia molecular. O projeto inclui *wrapper's* que capturam o esquema e os dados das fontes que participam da integração. Tanto os esquemas como os dados são armazenados em um repositório que utiliza o modelo de dados semi-estruturado. Assim, os esquemas são armazenados em XMLSchema e os dados em XML. Os *wrapper's* têm ainda a função de traduzir os esquemas obtidos nas fontes de dados para XMLSchema e os dados para XML. Este *framework* inova ao tratar da integração de esquemas baseada em um meta-modelo. A integração é feita através de um mediador que captura os esquemas e dados das fontes, faz as conversões necessárias e materializa as informações no repositório.

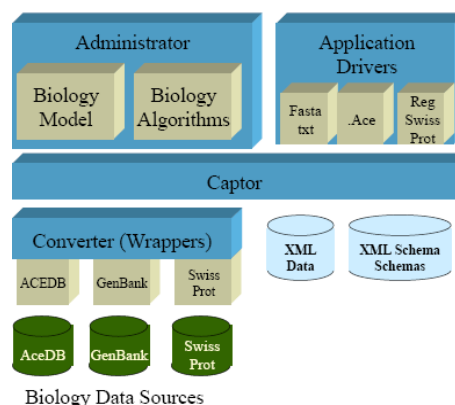


Figura 5. Arquitetura Bio-AXS de integração

A arquitetura do *framework* está subdividida em quatro módulos, cuja interdependência está esquematizada na figura 5. O módulo *Administrador* realiza a interface com os usuários, de forma a prover as seguintes funções: permitir gerenciar o modelo da biologia e esquemas específicos, solicitar a captura de esquemas e/ou de dados, permitir a formulação de consultas, permitir a execução dos programas instanciados no próprio framework e a execução de aplicações externas.

Este módulo contém um repositório, que armazena o modelo global da biologia. O módulo *Capturador* administra o repositório de dados e de esquemas da arquitetura. Os *Wrapper's* implementam o acesso às fontes de dados de biologia, efetuando a tradução dos esquemas das fontes de dados para XML Schema e dos dados para XML.

Os *Driver's* implementam a conversão dos dados do formato interno do framework (XML) para o formato esperado pelas aplicações. Os detalhes da arquitetura de cada módulo, suas classes e métodos, e seus relacionamentos, são descritos em [11].

Fruto de resultados já disponíveis cabe aqui relatar que esta ferramenta, chamada de Bio-AXS, vem sendo utilizada como

infra-estrutura para o sistema de anotação do genoma do *Trypanosoma Cruzi* (FIOCRUZ) e para o sistema de anotação do genoma da *Gluconacetobacter diazotrophicus* (UFRJ).

4. TRABALHOS FUTUROS

Encontrar um banco de dados que suporte tudo o que a é gerado em projetos de pesquisa com genes através da bioinformática é sem sombra de dúvida, complexo, pois o banco de dados deverá se adequar ao domínio da aplicação.

Segundo [8], muitas empresas e institutos vêm pesquisando a área de banco de dados para bioinformática, mas sem conseguir chegar a um padrão a ser adotado para todos os bancos de dados utilizados em bioinformática, pois estas empresas e institutos tentam somente adequar o domínio de suas aplicações aos bancos de dados já existentes no mercado, tentando solucionar suas necessidades imediatas.

Atualmente, a Oracle, uma das grandes empresas do ramo de soluções para banco de dados, iniciou suas pesquisas na área de bioinformática, objetivando atender as necessidades dos institutos e empresas particulares que trabalhem com dados genômicos, sendo assim, uma das primeiras empresas a tentar padronizar o “esquema” de banco de dados para bioinformática, pois até então os esforços para a área eram escassos e individuais, não possibilitando assim a construção de um padrão a ser adotado para o armazenamento e tratamento de dados genômicos.

Segundo [8] e [9] e [10], além da Oracle, outros grandes institutos e centros de pesquisa, tanto de computação quanto de biologia molecular, de Universidades, órgãos do governo de vários países (inclusive o Brasil) e empresas privadas (principalmente farmacêuticas), espalhadas pelo mundo estão tentando entrar em comum acordo para elaborar um padrão que venha a ser adotado pelos bancos de dados adotados na área de bioinformática, a fim de acabar com o problema causado pela falta de padronização, o que dificulta a troca de informações sobre os dados genômicos, além de tentarem desenvolver tecnologias de Data Warehouse e XML para integração de bases biológicas.

5. CONCLUSÕES

Com o grande crescimento que a área de bioinformática vem apresentando, os esforços para se tentar adotar um padrão em bancos de dados voltados para a bioinformática vem aumentando consideravelmente, muitos institutos de pesquisa e empresas de soluções em banco de dados vêm desenvolvendo pesquisas na área para se tentar achar um padrão que seja utilizado em todos os bancos de dados para a bioinformática.

Esta padronização ajudará, não somente o armazenamento, acesso e busca dos dados “criados” por projetos de bioinformática, mas também irá auxiliar em muito a troca de informação sobre estes dados entre os mais diversos institutos de pesquisa espalhados pelo mundo que trabalham com este tipo de dado, facilitando assim a descoberta de curas para doenças e o tratamento preventivo das mesmas.

Soluções como as apresentadas em [5] oferecem um melhor “suporte” para a realização de tais pesquisas, mas tal solução é somente o início de pesquisas que devam surgir nos próximos anos, a fim de fazer com que dados biológicos possuam uma

facilidade de tratamento e busca tal qual existe para dados comuns, como NUMBER, DATE e CHAR.

A utilização de data warehouse é uma solução interessante quando falamos em interligar bases biológicas de várias entidades, mas esta solução não pode ser aplicada separadamente, sem utilizarmos formas de otimização de pesquisas e tratamento dos dados biológicos, pois se somente a integração destes bancos não nos garante que as buscas por informações referentes a dados biológicos vá se dar de uma forma eficaz.

A utilização de tecnologias XML é muito interessante, mas esta tecnologia ainda não está bem formulada para o domínio de dados biológicos, sendo implementada e testada aos poucos, principalmente se apoiando nos conceitos oferecidos pela W3C.

Devido o que foi visto neste artigo, pode-se concluir que o desenvolvimento de soluções para bancos de dados no domínio da bioinformática está em crescente expansão, o que fará com que pesquisas melhores no âmbito da bioinformática venham a acontecer, facilitando assim a descoberta de novos dados biológicos/genômicos que possibilitem a descoberta de novos medicamentos.

6. REFERÊNCIAS

- [1] BANERJEE, S. 2000. A Database Platform for Bioinformatics. Oracle Corporation, Redwood Shores.
- [2] CRITCHLOW, T.; MUSICK, R.; SLEZAK, T. 2000. An Overview of Bioinformatics Research at Lawrence Livermore National Laboratory. Department of Energy by University of California Lawrence. Califórnia U.S.
- [3] LENGAUER, T. 2001. Computational Biology at the Beginning of the Post-genomic Era. University of Bonn, Berlin.
- [4] HUMAN GENOME PROGRAM. 1992. Primer on Molecular Genetics. Department of Energy, Washington D.C.
- [5] ORACLE CORP. 1999. Oracle8i Data Cartridge Developer's Guide: Release 8.1.5 (Part No. A68002-01). Oracle Corporation, Redwood Shores.
- [6] FÉLIX, J. M. 2002. Genoma Funcional. Biotecnologia, Ciência & Desenvolvimento, Nº 24, janeiro a fevereiro.
- [7] SHUI, W. M. 2001. Utilizing Multiple Bioinformatic Information Sources: An XML Database approach 2001 Bioinformatics Honours Thesis. University of New South Wales, Sydney.
- [8] WIECZOREK, E. M. 2002. Caminhos e Tendências do uso de Banco de Dados em Bioinformática. Centro Universitário Lutero de Palmas - CEULP/ULBRA, Palmas-TO.
- [9] WIECZOREK, E. M. 2003. Padrões de Tipos e Métodos para Banco de Dados em Bioinformática. III Congresso Científico do CEULP/ULBRA, Palmas-TO.
- [10] ROCHA, C. P. 2010. Banco de Dados em Bioinformática. DOI=http://artigocientifico.uol.com.br/uploads/artc_1194549401_86.pdf.
- [11] LIFSCHITZ, S. 2006. Algumas Pesquisas em Bancos de Dados e Bioinformática. XXVI Congresso da Sociedade Brasileira da Computação. Campo Grande.