



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Visualização de Vias Metabólicas com Banco de Dados em Grafo

Gabriella de Oliveira Esteves

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Orientador

Prof. Dr.^a Maria Emília Machado Telles Walter

Brasília
2016

Universidade de Brasília — UnB
Instituto de Ciências Exatas
Departamento de Ciência da Computação
Bacharelado em Ciência da Computação

Coordenador: Prof. Dr. Rodrigo Bonifácio de Almeida

Banca examinadora composta por:

Prof. Dr.^a Maria Emília Machado Telles Walter (Orientador) — CIC/UnB
Prof. Dr. Professor I — CIC/UnB
Prof. Dr. Professor II — CIC/UnB

CIP — Catalogação Internacional na Publicação

Esteves, Gabriella de Oliveira.

Visualização de Vias Metabólicas com Banco de Dados em Grafo /
Gabriella de Oliveira Esteves. Brasília : UnB, 2016.

53 p. : il. ; 29,5 cm.

Monografia (Graduação) — Universidade de Brasília, Brasília, 2016.

1. Bioinformática, 2. Redes Metabólicas, 3. Banco de Dados Não
Relacional, 4. Grafo, 5. OrientDB

CDU 004.4

Endereço: Universidade de Brasília
Campus Universitário Darcy Ribeiro — Asa Norte
CEP 70910-900
Brasília-DF — Brasil



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Visualização de Vias Metabólicas com Banco de Dados em Grafo

Gabriella de Oliveira Esteves

Monografia apresentada como requisito parcial
para conclusão do Bacharelado em Ciência da Computação

Prof. Dr.^a Maria Emília Machado Telles Walter (Orientador)
CIC/UnB

Prof. Dr. Professor I Prof. Dr. Professor II
CIC/UnB CIC/UnB

Prof. Dr. Rodrigo Bonifácio de Almeida
Coordenador do Bacharelado em Ciência da Computação

Brasília, 08 de Julho de 2016

Dedicatória

Dedicatória

Agradecimentos

Agradecimento

Resumo

Resumo em português

Palavras-chave: Bioinformática, Redes Metabólicas, Banco de Dados Não Relacional, Grafo, OrientDB

Abstract

Abstract in english

Keywords: Bioinformatics, Metabolic Networks, Non-Relational Database, Graph, OrientDB

Sumário

1	Introdução	1
1.1	Justificativa	1
1.2	Problema	1
1.3	Objetivo	1
1.4	Descrição dos Capítulos	1
2	Redes Metabólicas	3
2.1	Conceitos Básicos de Biologia Molecular	3
2.1.1	Ácidos Nucléicos	3
2.1.2	Síntese de Proteína	5
2.2	Conceitos Básicos de Metabolismo	7
2.2.1	Metabolismo Primário	7
2.2.2	Metabolismo Secundário	7
2.3	Banco de Dados de Redes Metabólicas	7
2.4	Banco de Dados NoSQL	8
2.5	Propriedades	8
2.6	OrientDB	8
2.6.1	KEGG	9
2.6.2	BioCyc	9
2.6.3	Reactome	9
3	Ferramentas de visualização de Redes Metabólicas	10
3.1	KEGG	10
3.2	MetaCyc	10
3.3	Reactome Browser	10
3.4	Cytoscape	10
4	IHC	11
5	2Path: Aplicação Web	12
5.1	Implementação	12
5.2	Visualização das redes metabólicas	13
5.3	Desafios	13
6	Método e Resultados	14

7	Conclusão e Trabalhos Futuros	15
7.1	Conclusão	15
7.2	Trabalhos Futuros	15
8	Cronograma	16
	Referências	17

Lista de Figuras

2.1	imagem de um nucleotídeo e das bases nitrogenadas. Mostrar backbone da pentose 1'...5'. Adaptado de : [14].	3
2.2	Adaptado de : [14]	5
2.3	Adaptado de : [2]	6

Lista de Tabelas

8.1	Cronograma	16
-----	----------------------	----

Capítulo 1

Introdução

1.1 Justificativa

Atualmente, a quantidade de dados ômicos estudados pelos pesquisadores é extensa e complexa. Uma maneira de amenizar o esforço feito para analisá-los e compreendê-los é oferecer uma ferramenta que aproxime o usuário (pesquisador) e os dados e a maneira mais natural é representar tais dados em forma de grafo (redes metabólicas). Esta ferramenta deverá permitir que o usuário visualize e interaja com os dados dinamicamente, além de disponibilizar mecanismos de busca em grafos, úteis para sua pesquisa.

1.2 Problema

Construir uma visualização interativa de redes metabólicas armazenadas em banco de dados de grafos que permita ao pesquisador explorar os aspectos biológicos do organismo estudado.

1.3 Objetivo

Construir um sistema que acesse redes metabólicas armazenadas em bancos de dados em grafo e gere uma visualização interativa

- Implementar uma busca das vias metabólicas de interesse a a partir de parâmetros informados pelo pesquisador no sistema
- Recuperar a informação desejada e exibi-la para o pesquisador de forma ergonômica
- Implementar algoritmos de busca em grafos para recuperar a informação solicitada e/ou sugerir informação relevante

1.4 Descrição dos Capítulos

No Capítulo 1 serão descritos os conceitos básicos de biologia molecular, metabolismo primário e secundário e os bancos de dados mais utilizados para armazenar as informações referentes às redes metabólicas. No Capítulo 2 serão apresentados detalhadamente quatro

ferramentas de visualização de redes metabólicas: *KEGG Pathway*, *MetaCyc*, *Reactome Browser* e *Cytoscape*.

Os Capítulos 3 e 4 já são relacionados ao sistema 2Path desenvolvido neste projeto. Enquanto o Capítulo 3 aborda o tema de interação humano-computador para a concepção de uma interface auto-explicativa e consistente, o Capítulo 4 descreve as linguagens e ambientes utilizados para a construção de tal. O Capítulo 5 especifica como são feitas as buscas por vias metabólicas no sistema pelo usuário e apresenta os resultados obtidos.

O Capítulo 6 finaliza o trabalho com a conclusão e sugestão de trabalhos futuros.

Capítulo 2

Redes Metabólicas

Neste capítulo serão descritos os conceitos básicos da biologia molecular, metabolismo e bancos de dados específicos para redes metabólicas. A primeira seção detalha a origem das principais estruturas que promovem o metabolismo tais como DNA e enzima, enquanto que a segunda seção descreve de fato como ocorre o processo. Por fim, a última seção apresenta os principais bancos de dados em grafo voltados para redes metabólicas: *KEGG*, *MetaCyc* e *Reactome*.

2.1 Conceitos Básicos de Biologia Molecular

2.1.1 Ácidos Nucléicos

Os ácidos nucleicos são biomoléculas responsáveis pelo armazenamento, transmissão e tradução das informações genéticas dos seres vivos. Isto é possível devido ao processo de síntese de proteínas que permite, assim, a base da herança biológica. Os ácidos nucleicos são polímeros, macromoléculas formadas por estruturas menores chamadas monômeros, que nesse caso são nucleotídeos. Nucleotídeos são compostos de três elementos: um radical fosfato (HPO_4), uma pentose, ou seja, um monossacarídeo formado por cinco átomos de carbono, e uma base nitrogenada. Existem cinco tipos de bases nitrogenadas que podem compor um nucleotídeo: Adenina(A), Timina(T), Citosina(C), Guanina(G) e Uracila(U).

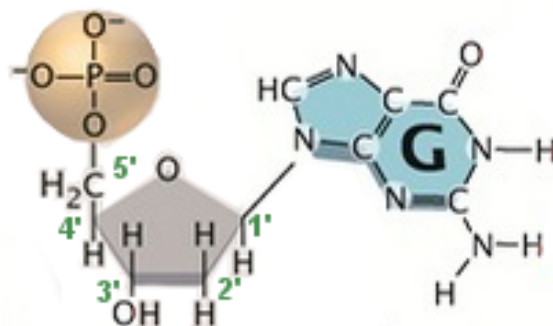


Figura 2.1: imagem de um nucleotídeo e das bases nitrogenadas. Mostrar backbone da pentose 1'...5'. Adaptado de : [14].

Na figura 2.1, observa-se que no *backbone* do nucleotídeo existe uma numeração de 1' à 5', que representam os carbonos presentes na pentose. Para a criação de uma fita de ácido nucleico, no processo de polimerização formar-se uma ligação fosfodiéster entre o carbono da posição 5' do *backbone* de um nucleotídeo e o carbono de posição 3' do *backbone* de outro [15]. Por definição o sentido da leitura de uma fita de ácido nucleico é $5' \rightarrow 3'$, o que é deve ser levado em consideração ao se fazer interpretação de dados do material genético.

Dois tipos de ácidos nucleicos são encontrados nos seres vivos: ácido desoxirribonucleico (DNA ou ADN) e ácido ribonucleico (RNA ou ARN). Eles diferenciam-se tanto na estrutura do *backbone* e nas bases nitrogenadas, quanto em suas funções. Os DNAs são as biomoléculas que armazenam as informações referentes ao funcionamento de todas as células dos seres vivos de maneira específica: sequências de pares de bases nitrogenadas. Nesse sentido, além de haver a ligação fosfodiéster entre os nucleotídeos, cada um também se liga a partir de suas bases nitrogenadas, formando assim um eixo helicoidal tridimensional chamada de dupla hélice [15]. Esta estrutura foi descoberta em 1953, pelo biólogo James Watson e pelo físico Francis Crick [14], porém os ácidos nucleicos já eram estudado desde 1869 na Suíça pelo químico-fisiológico Friedrich Miescher.

Em relação à estrutura dos monômeros do DNA, o *backbone* dos nucleotídeos é uma desoxirribose, indicada na figura 2.2. Para a formação da dupla hélice, os pares são feitos com uma base nitrogenada do grupo de purinas, composto orgânico que possui um anel duplo de carbono, e outra base do grupo de pirimidinas, composto orgânico que possui um anel simples de carbono. No caso do DNA, somente quatro das cinco bases são empregadas: as purinas Adenina(A) e Guanina(G), que se ligam com as pirimidinas Timina(T) e Citosina(C) respectivamente. Desta forma, A e T são bases complementares, assim como G e C. Uma fita de DNA pode conter centenas de milhões de nucleotídeos.

A representação do DNA, seja nos livros ou computacionalmente, é dada por um par em paralelo de strings de letras A, T, G e C. Como explicado no início dessa seção, o sentido padrão da leitura de uma fita é de $5' \rightarrow 3'$, mas no caso do DNA, as hélices são dispostas de maneira antiparalela, ou seja, uma é lida de $5' \rightarrow 3'$ e a outra, de $3' \rightarrow 5'$. Observa-se que a partir de uma hélice, pode-se inferir a sequência de sua hélice complementar. Seja, por exemplo, uma hélice H1 igual a AGTAAGC; então H2 em seu sentido oposto é H2' igual a TCATTCG, e no sentido regular, igual a GCTTACT. A figura 2.2 apresenta a estrutura do DNA como explicada nesta seção.

Os RNAs são biomoléculas semelhantes ao DNA, porém contam com três diferenças básicas. A primeira é a estrutura do *backbone* dos nucleotídeos, que é composta por uma ribose ao invés de um desoxirribose. A segunda diferença é em relação às bases nitrogenadas, onde a pirimidina Uracila(U) substitui a Timina(T). Por fim, o RNA é formado por apenas uma hélice tridimensional.

Existem três tipos de RNAs presentes no citoplasma - espaço entre a membrana plasmática e o núcleo da célula. Cada um possui funções específicas que serão detalhadas na seção 2.1.2. Em suma, O RNA mensageiro (mRNA) é responsável pela transferência de informação do DNA para o RNA ribossômico (rRNA), que por sua vez irá desanexar a proteína do RNA transportador (tRNA) combinando-o com o rRNA, executando assim, a síntese de proteína.

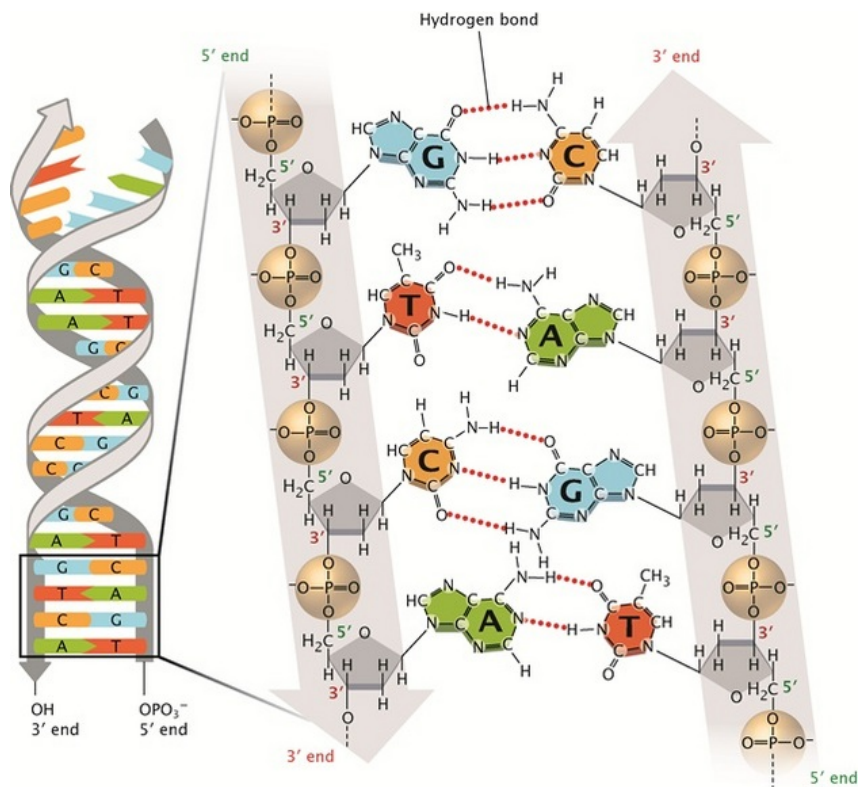


Figura 2.2: Adaptado de : [14]

2.1.2 Síntese de Proteína

As proteínas são biomoléculas com diversas responsabilidades no corpo dos seres vivos. Se fizerem parte do no grupo de proteínas fibrosas, como o colágeno, irão compor a estrutura do corpo e para isso precisam ser resistentes e insolúveis em água. Caso estejam no grupo de proteínas globulares, como a hemoglobina, realizarão processos dinâmico pelo corpo tais como transportações e cataliações [1]. Cada tarefa é realizada por um proteína com uma estrutura específica e otimizada pra tal.

Assim como os ácidos nucleicos, as proteínas são polímeros, macromoléculas cujos monômeros são aminoácidos. Aminoácidos são moléculas que possuem cinco componentes: amina (NH_2), carbono (C), hidrogênio (H), ácido carboxílico (COOH) e uma cadeia lateral que funciona como identificador de cada um dos 20 tipos de aminoácidos presentes nos seres vivos. A maneira como eles são criados será explicada com mais detalhes mais à frente, pois envolve um processo complexo de síntese de proteína executado pelo ribossomo. A ligação, ou polimerização, de dois aminoácidos é feita unindo a amida de um com o ácido carboxílico do outro, liberando uma molécula de água (H_2O) e formando uma cadeia chamada de dipeptídeo. Como houve liberação de água na ligação, o dipeptídeo não é formado por aminoácidos, mas sim resíduos dos mesmos. Nesse sentido, cadeias peptídicas de 100 à 5000 diferentes resíduos aminoácidos, ou cadeia polipeptídicas, constituem a proteína.

Existem quatro estruturas para caracterização de uma proteína [15]. A mais simples é chamada de estrutura primária e é composta por uma sequência linear de resíduos aminoácidos. A estrutura secundária é tridimensional e estabiliza-se por meio de ligações

de hidrogênio na cadeia principal, chamada de *backbone*. Dependendo da disposição dos resíduos de aminoácidos, esta cadeia pode se dar forma de hélice (α -Helix) ou em forma de folha (β -Helix). A estrutura terciária é dada pela união de várias estruturas secundárias e, por fim, a estrutura quaternária é composta de múltiplas estruturas terciárias [2]. A figura 2.3 ilustra os quatro tipos de proteínas descritos.

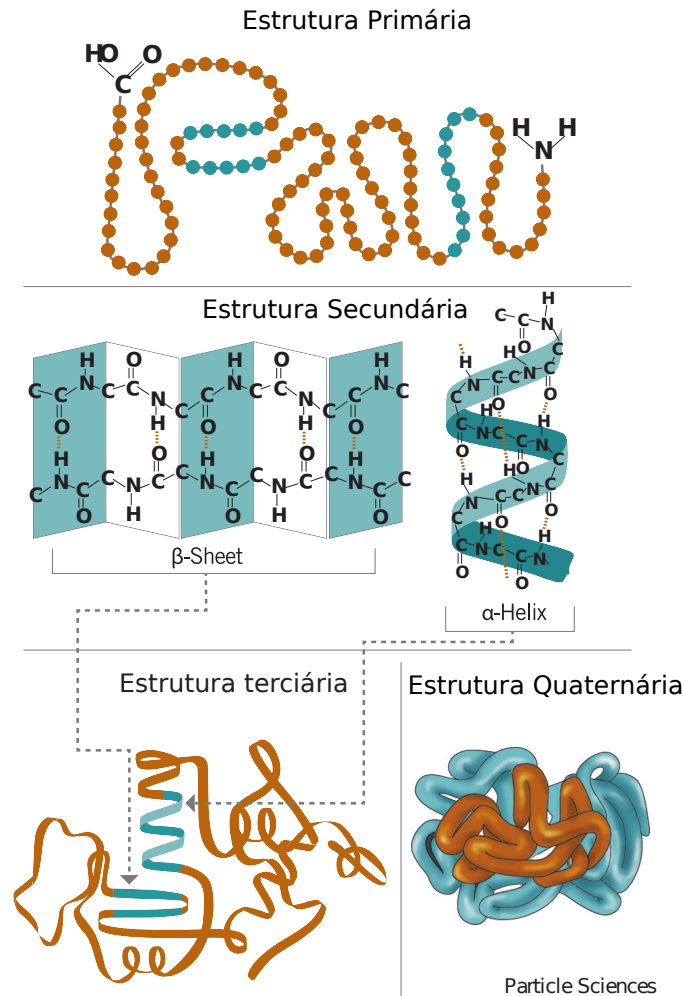


Figura 2.3: Adaptado de : [2]

A **transcrição** é o processo de produção de mRNA a partir do DNA e ele ocorre da seguinte forma: O início de cada gene possui um identificador em uma das fitas para indicar o local da codificação e, a partir dali, uma cópia inversa (A, T, C, G são traduzidos para U, A, G, C respectivamente) do mesmo é feita sob forma de molécula de mRNA que, por consequência, obterá a mesma sequência que a cadeia codificadora (a qual não possui o identificador), porém trocando os U's por Ts.

O mRNA deixa, então, o núcleo celular e inicia a **tradução** no citoplasma. O processo ocorre no interior de uma organela celular chamada de ribossomo, constituído de proteínas e rRNA e cuja função é construir a molécula de proteína a partir de duas entradas, o mRNA e tRNA. A estrutura do tRNA é tal que de um lado se encaixa exatamente um códon (sequência de três nucleotídeos) e no oposto, seu aminoácido correspondente.

O processo de tradução se dá da seguinte forma: a medida em que o mRNA passa pelo interior do ribossomo, este atrai quaisquer tRNAs das proximidades cujos códons sejam correspondentes ao da subsequência corrente do mRNA. No momento em que o códon do tRNA se conecta com um dos códons do mRNA, a molécula de proteína em desenvolvimento é liberada e, com o auxílio da catálise de uma enzima, agregada no aminoácido que estava fixado naquele tRNA. Esta fase é finalmente completa quando o mRNA apresenta um códon de parada, pois nenhum tRNA possui correspondência para tal [15]. Uma proteína simples é, então, formada.

2.2 Conceitos Básicos de Metabolismo

As reações bioquímicas são alterações químicas que fornecem um ou mais produtos a partir de um ou mais substratos. Uma via metabólica é uma sequência de reações bioquímicas, cujo produto e substrato são denominados de metabólitos, que podem ser catalisados por enzimas, as quais muitas vezes necessitam de compostos químicos chamados de co-fatores para realizarem suas atividades na célula. O conjunto de vias metabólicas de um organismo é chamado de rede metabólica.

Enzimas são proteínas responsáveis por auxiliar a realização de biossíntese (construção) e biodegradação de moléculas no metabolismo com o propósito de catalisar (acelerar) reações bioquímicas. As enzimas possuem um local pré-determinado em formato côncavo chamado de sítio ativo, que comporta um ou mais substratos. Se a enzima comporta apenas um substrato, a estrutura que se forma com o preenchimento do sítio ativo é um complexo enzima-substrato, porém se ela comporta mais de um substrato, a estrutura é chamada de complexo ternário intermediário. Quando a atividade catalítica não pode ser realizada apenas pela enzima, co-fatores auxiliam o processo. Eles podem ser coenzimas, associadas momentaneamente às enzimas, ou grupos prostéticos, associados firmemente às enzimas. Quando duas enzimas possuem a mesma atividade enzimática porém estruturas físicas diferentes são chamadas isoenzimas.

No caso em que o metabolismo exerce uma função fundamental no organismo, ele é classificado como metabolismo primário. Mitose e meiose são exemplos de metabolismos primários. Já quando o metabolismo não está relacionado a reprodução, desenvolvimento ou crescimento, ele não é essencial no organismo e, portanto, secundário. Os metabólitos secundários, apesar da aparente insignificância, podem ser antibióticos, por exemplo, e deste modo são bastante aplicados na medicina e na indústria [?].

Biosíntese/Degradação. Atividade enzimática (relação metabólito/enzima). Co-fatores

2.2.1 Metabolismo Primário

2.2.2 Metabolismo Secundário

2.3 Banco de Dados de Redes Metabólicas

No início dos anos 50, uma química britânica chamada Rosalind Frankling usou a técnica de difração de raios-X para determinação da estrutura da biomolécula do DNA e concluiu que sua forma era helicoidal. Seu trabalho foi empregado nos experimentos de dois pesquisadores, Francis Crick e James Watson, em um laboratório em Cambridge, na

Inglaterra. No mesmo ano, a dupla decifrou a estrutura do DNA: duas longas fitas enroscadas uma na outra em espiral para a direita, ligadas por pares de bases complementares, formando o que chamaram de dupla-hélice. Apesar da grande descoberta, isto não era o suficiente para entender como eram produzidas as proteínas, portanto os cientistas mudaram o foco das pesquisas para o RNA, uma vez que sabiam o quanto sua concentração aumentava sempre que as células começavam a produzir proteínas [?]. Em 1958, Crick e Watson anunciaram mais uma descoberta: A partir do DNA, o processo de *transcrição* fornece uma fita de RNA, que por sua vez, a partir do processo de *tradução*, fornecem a proteína. Esta sequência de processos ficou conhecida como Dogma Central da biologia molecular.

Conforme o número de sequências de proteínas cresciam, aumentava também a necessidade de criar-se um banco de dados para indexá-las. A físico-química norte-americana Margaret Dayhoff, com colaboração de alguns membros do *National Biomedical Research Foundation* em Washington, foi a primeira a construir um banco de dados com este propósito em um tipo de atlas de proteínas na década de 60. Somente em 1984 esta coleção foi intitulada de *Protein Information Resource* [10]. Os dados eram organizados de acordo com o grau de similaridade das sequências, onde o agrupamento das mesmas era dado em forma de árvore filogenética representando famílias e superfamílias de proteínas. Caso a semelhança seja alta, é provável que tenham as mesmas funções bioquímicas e estrutura tridimensional. A partir da árvore gerada, foi possível calcular as mutações que ocorreram nos aminoácidos durante a evolução genética e, então, produzir uma tabela utilizada até hoje, chamada PAM (*percent accept mutation*), que apresenta tais dados. Outro banco de dados de grande porte e bastante utilizado nos dias de hoje é o GenBank, estabelecido em 1982 por Walter Goad e demais colaboradores e, agora, com o patrocínio do *National Center for Biotechnology Information*. Os dois bancos são públicos e continuam crescendo exponencialmente [10].

2.4 Banco de Dados NoSQL

NoSQL

[6]

Comparação com SQL

2.5 Propriedades

ACID, BASE

Consistency, availability and tolerance of network partition (consistência, disponibilidade e tolerância de partição de redes)

2.6 OrientDB

sobre ACID

Modelo CAP

JAVA API

2.6.1 KEGG

2.6.2 BioCyc

2.6.3 Reactome

Capítulo 3

Ferramentas de visualização de Redes Metabólicas

3.1 KEGG

3.2 MetaCyc

3.3 Reactome Browser

3.4 Cytoscape

Capítulo 4

IHC

Capítulo 5

2Path: Aplicação Web

O sistema desenvolvido para este projeto é uma aplicação web chamada *2Path*. O usuário deve se cadastrar no *website* para ter acesso às redes metabólicas do banco de dados do sistema, bem como pesquisar por palavras chaves no mesmo. Neste capítulo serão apresentadas as linguagens e ferramentas utilizadas no desenvolvimento do *website*, as características, funcionalidades e limites do sistema e, por fim, as dificuldades enfrentadas na implementação do projeto.

5.1 Implementação

O sistema foi desenvolvido no ambiente de desenvolvimento integrado *open source* Eclipse Java EE - *Java Platform, Enterprise Edition*, versão Mars 4.5.2. A plataforma Eclipse foi projetada com o objetivo de agilizar o desenvolvimento de recursos integrados baseando-se em um modelo de *plug-in*. Na *workbench* no Eclipse, cada *plug-in* é responsável por pequenas tarefas, tais como compilar, testar ou debugar [5].

Para simplificar a obtenção das dependências do projeto, ou seja, pacotes de arquivos java (extensão .jar), foi utilizada o Apache Maven, *software* de gerenciamento de projeto e ferramenta de compreensão de programa. Este *software* opera sobre o arquivo *pom.xml*, onde POM significa *Project Object Model* e contém as especificações de cada projeto que se tornará dependência do sistema em desenvolvimento, além de outros aspectos do código. No exemplo abaixo, o fragmento do *pom.xml* indica o *groupId* - código único entre a organização ou projeto, *artifactId* - nome do projeto, *version* - versão do projeto que será baixada e *scope* - escopo em que o projeto será necessário no sistema (compilação, execução ou teste).

```
<dependencies>
(...)
    <!-- PrimeFaces (biblioteca de componentes) -->
    <dependency>
        <groupId>org.primefaces</groupId>
        <artifactId>primefaces</artifactId>
        <version>3.5</version>
        <scope>compile</scope>
    </dependency>
(...)
```

</dependencies>

O servidor selecionado para «<» o sistema na rede, *localhost* porta 8080, foi o Apache TomCat versão 7.0. Este software é uma implementação *open source* das quatro tecnologias [12] a seguir:

- *Java Servlet:*
- *JavaServer Pages:*
- *Java Expression Language:*
- *Java WebSocket:*

COLOCAR IMAGEM DA ARQUITETURA MVC : An MVC EBookShop with Servlets, JSPs, and JavaBeans Deployed in Tomcat [7]

As quatro páginas da aplicação foram desenvolvidas na linguagem de marcação XHTML, *Extensible Hypertext Markup Language*, e a estilização em CSS, *Cascading Style Sheets*. Com a primeira é possível criar objetos na página *web* através de componentes nativos e não nativos da linguagem chamados *tags*. As principais *tags* são apresentadas na Tabela ???. Já com CSS é possível customizar cada objeto da página *web*, alterando seu tamanho, posição, cor, fonte, e várias outras características. Para tal, o objeto por ser alterado individualmente através de seu ID; em conjunto, com objetos da mesma classe ou *tag*.

JSF
PRIMEFACES

5.2 Visualização das redes metabólicas

JAVASCRIPT
ANGULARJS
ORIENTBD

5.3 Desafios

O que foi o trabalho. Dever de todo o ambiente usado Neste capítulo serão apresentados os primeiros resultados experimentais obtidos.

Capítulo 6

Método e Resultados

Capítulo 7

Conclusão e Trabalhos Futuros

7.1 Conclusão

Neste capítulo serão apresentadas as considerações finais do trabalho, assim como as limitações e dificuldades encontradas.

7.2 Trabalhos Futuros

A partir deste trabalho, foi possível identificar os seguintes pontos a serem melhorados:

- x

Capítulo 8

Cronograma

O cronograma está apresentado na Tabela a seguir, mostrando o início das atividades em Janeiro de 2016 com a revisão literária e com término previsto para Junho de 2016, juntamente com a defesa do Trabalho de Conclusão de Curso.

Tabela 8.1: Cronograma

Atividades	2016					
	Jul	Ago	Set	Out	Nov	Dez
Revisão bibliográfica	X	X				
Familiaridade com ambiente de desenvolvimento		X	X			
Implementação da aplicação		X	X	X		
Interpretação dos resultado				X	X	X
Defesa						X

Referências

- [1] Proteínas. <http://www.professoraangela.net/documents/proteinas.html>, visitado em 2016-01-02. 5
- [2] Protein structure, 2009. <http://www.particlesciences.com/news/technical-briefs/2009/protein-structure.html>, visitado em 2016-01-02. vii, 6
- [3] Bonnie Berger, Jian Peng, and Mona Singh. Computational solutions for omics data. *Nature reviews. Genetics*, 14(5):333–346, May 2013.
- [4] Laura Bonetta. Protein-protein interactions: Interactome under construction. *Nature*, 468(7325):851–854, Dec 2010.
- [5] Eclipse Foundation, Inc., 102 Centrepointhe Drive, Ottawa, Ontario,. *Eclipse documentation - Current Release*, 4.6 edition, 2016. http://help.eclipse.org/neon/index.jsp?topic=%2Forg.eclipse.platform.doc.isv%2Fguide%2Fint_eclipse.htm, visitado em 2016-07-01. 12
- [6] Jing Han, E. Haihong, Guan Le, and Jian Du. Survey on NoSQL database. In *Pervasive Computing and Applications (ICPCA), 2011 6th International Conference on*, pages 363–366. IEEE, October 2011. 8
- [7] Chua Hock-Chuan. Java web database applications, 2011. <https://www.ntu.edu.sg/home/ehchua/programming/java/JavaWebDBApp.html>, visitado em 2016-08-19. 13
- [8] Vincent Lacroix, Ludovic Cottret, Patricia Thébault, and Marie-France Sagot. An introduction to metabolic networks and their structural analysis. *IEEE/ACM Trans. Comput. Biology Bioinform.*, 5(4):594–617, 2008.
- [9] Elaine R Mardis. Next-generation dna sequencing methods. *Annu Rev Genomics Hum Genet*, 9:387–402, 2008.
- [10] David W. Mount. *Bioinformatics : sequence and genome analysis*. Cold Spring Harbor, N.Y. Cold Spring Harbor Laboratory Press, 2001. 8
- [11] R. Nussbaum. *Genética Médica*. Elsevier Editora Ltda., 2008.
- [12] Oracle. *Java Platform, Enterprise Edition The Java EE Tutorial, Release 7*, 2014. <https://docs.oracle.com/javaee/7/tutorial>, visitado em 2016-08-19. 13

- [13] Mihaela Pertea and Steven L. Salzberg. Between a chicken and a grape: estimating the number of human genes. *Genome Biology*, 11(5):1–7, 2010.
- [14] Leslie A. Pray. Discovery of dna structure and function: Watson and crick, 2008. <http://www.nature.com/scitable/topicpage/discovery-of-dna-structure-and-function-watson-397>, visitado em 2016-01-15. vii, 3, 4, 5
- [15] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. PWS Publishing Company, 1997. 4, 5, 7