

Relatório Final de Atividades

Integração de dados na biologia sistêmica: um estudo de caso em *Arabidopsis Thaliana*

vinculado ao projeto

Bioinformática e Reconhecimento de Padrões

Gabriel Rubino

Bolsista CNPq

Engenharia de Computação

Data de ingresso no programa: 08/2015

Prof. Dr. Fabrício Martins Lopes

Área do Conhecimento: 1-03.00.00-7 Ciência da Computação

CAMPUS CORNÉLIO PROCÓPIO, 2016

GABRIEL RUBINO

FABRÍCIO MARTINS LOPES

**INTEGRAÇÃO DE DADOS NA BIOLOGIA SISTÊMICA: UM ESTUDO DE
CASO EM ARABIDOPSIS THALIANA**

Relatório de pesquisa do Programa de
Iniciação Científica da Universidade
Tecnológica Federal do Paraná.

CORNÉLIO PROCÓPIO, 2016

SUMÁRIO

INTRODUÇÃO	4
MATERIAIS E MÉTODOS	7
RESULTADOS E DISCUSSÕES	14
CONCLUSÕES	16
REFERÊNCIAS	17

INTRODUÇÃO

Com o aumento da disponibilidade dos dados proporcionado pelo desenvolvimento de tecnologias cada vez mais avançadas em diversas áreas, como por exemplo imagens, áudio, astronomia, biologia, entre outras, dentre as quais muitas delas são disponibilizadas de forma online, houve o surgimento de plataformas online para organizar essas informações, sendo essas mantidas por universidades, empresas e governos. Esses dados muitas vezes são confiáveis e são vastamente utilizados para pesquisa e desenvolvimento tecnológico.

Neste contexto apresentado, os bancos de dados biológicos se caracterizam como um exemplo dessa prática. Essas entidades recebem vários dados, preenchidos por pesquisadores espalhados por todo o mundo. O acesso se dá através de plataformas online como por exemplo o TAIR, utilizado para guardar dados de genes de *Arabidopsis thaliana*.

Unir os dados de genes de um mesmo organismo a fim de inferir uma rede gênica é um desafio. Esse problema acontece devido ao elevado número de genes e a baixa quantidade de experimentos em proporção, esse é um fenômeno conhecido como maldição da dimensionalidade.

Uma alternativa para contribuir neste cenário é usar a integração de dados biológicos a partir de bancos de dados biológicos públicos na internet. Para abordar o problema da integração dos dados muitas vezes são usados organismos modelos como é o caso *Arabidopsis thaliana*, pois dessa forma validar as relações inferidas entre os genes torna-se mais adequado devido à disponibilidade de dados sobre esse organismo.

Depois dos dados estarem integrados em forma de um grafo, muito usados para mostrar dependências, sua visualização se torna possível. Para isso o grafo deve representar os seus nós, genes e características, por círculos e suas relações como arestas.

Problema

Como descrito na seção anterior existem muitos dados biológicos disponíveis e a integração deles pode ser feita para o estudo do organismo em questão. Assim a resolução do problema de integração dos dados e sua visualização em forma de grafo foi dividida em três etapas. As etapas são interdependentes e devem ser resolvidas na ordem que serão apresentadas a seguir.

Uma das preocupações quando se integra dados é escolher quais fontes de dados serão mais relevantes para conseguir completar os objetivos propostos. Visto que cada base de dado possui perfis diferentes, portanto os dados são focados em diferentes áreas. Como resumidamente descrito na Etapa 1 da Figura 1.

Depois da obtenção dos dados das fontes de dados é necessário processar esses arquivos de modo a organizá-los para a futura criação da rede. Por isso a definição das informações relevantes deve ser explorada. Como resumidamente descrito na Etapa 2 da Figura 1.

Os grafos são uma boa solução para a visualização de redes, mas seu armazenamento e gerenciamento muitas vezes são complexos devido ao grande número de nós incluídos em sua estrutura. Por isso métodos de visualização foram criados. Como resumidamente descrito na Etapa 3 da Figura 1.

O armazenamento dos grafos pode ser feito em banco de dados relacionais estruturados em tabelas que guardam as entidades e suas relações. Esse tipo de persistência dos dados é eficaz em alguns casos, mas muitas vezes a extração das informações do grafo não é fácil, pois é preciso criar instruções de requisição de dados complexos.

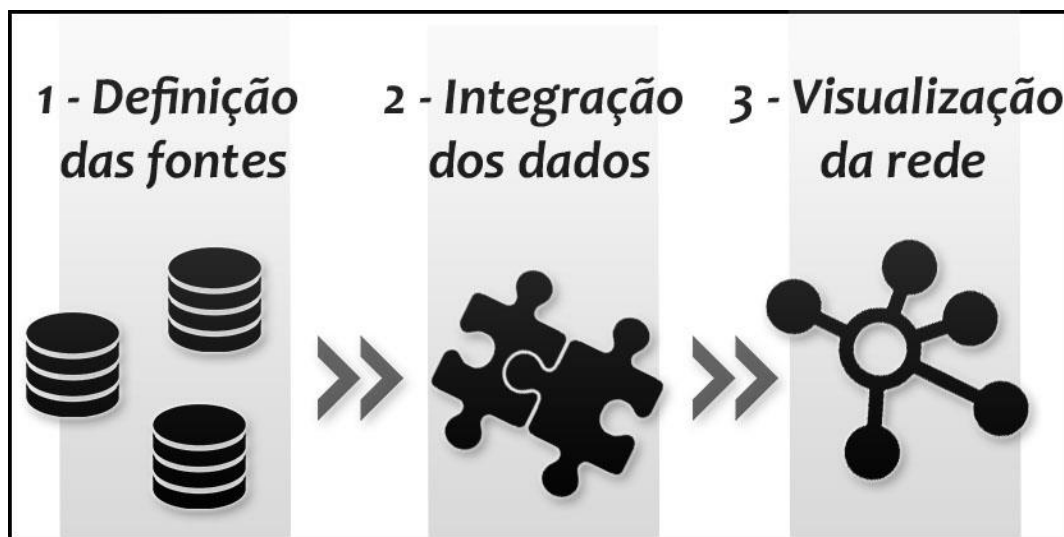


Figura 1. Etapas principais do projeto.

Este trabalho é proposto com a finalidade de facilitar o estudo dos relacionamentos entre os genes de um organismo, pois irá automatizar muitas etapas da integração de dados disponíveis em bancos de dados biológicos. Além disso esse projeto tem como objetivo disponibilizar uma ferramenta de visualização da rede em forma de grafo possibilitando a extração de características do mesmo.

Uma rede de genes auxilia na visualização de relações entre várias entidades direta e indiretamente. Isso é possível, pois medidas de grafos podem ser aplicadas como por exemplo: mensurar os nós mais conectados e as distâncias entre as relações. Com um modelo de grafo simulações podem ser feitas a fim de validar algumas hipóteses sem a necessidade de muitos testes *in vitro*.

Objetivos

Desenvolvimento de sistema para integração dos dados e visualização de redes gênicas.

Implementação de rotinas de leitura, pré-processamento e visualização de grafos genéricos.

Desenvolvimento e implementação de metodologias para a integração de informações biológicas nas redes gênicas.

Desenvolvimento e implementação de metodologias para a visualização de redes gênicas.

Desenvolvimento e implementação de metodologias para extração de informações de redes gênicas.

MATERIAIS E MÉTODOS

Este capítulo apresenta resumidamente as principais etapas adotadas para o desenvolvimento do projeto. Primeiramente é relatada as escolhas das tecnologias e ferramentas para implementar o trabalho. Posteriormente o foco será dado aos métodos usados para a solução do problema. Depois são apresentados os resultados. Para comprovar a eficácia dos métodos, meios para a validação dos resultados foram adotados.

Tecnologias

Para alcançar alguns dos objetivos descritos, foi realizada uma pesquisa com o banco de dados Neo4j que é considerado um dos mais utilizados para armazenar e gerenciar grafos. Essa pesquisa visou identificar o funcionamento desse banco de dados em grafos e como é possível interagir com sua linguagem, gerar consultas e resultados.

A linguagem usada para fazer consultas no Neo4J é a Cypher, que tem sua arquitetura voltada para grafos. Uma maneira de interagir com esse banco fazendo consultas em Cypher é através de programas feitos em Java, pois o Neo4J disponibiliza várias bibliotecas com esse objetivo.

Um modo de visualizar os grafos resultantes é através da biblioteca d3.js feita para JavaScript com o objetivo de criar elementos visuais interativos.

Banco de dados públicos

Os dados biológicos estão espalhados em vários bancos de dados públicos e cada um contém tipos diferentes de informações. Esses dados possuem intersecções onde é possível unir vários bancos em forma de grafo. A seguir serão descritos os bancos de dados usados nesse projeto e suas características principais e como os dados são expostos.

NATIONAL CENTER FOR BIOTECHNOLOGY INFORMATION

O *National Center for Biotechnology Information*, NCBI, é uma instituição dos Estados Unidos e tem como objetivo centralizar vários tipos de informação sobre biologia. (NCBI, 2015)

Uma de suas missões consiste em desenvolver tecnologias da informação que visam entender o funcionamento genético a fim de prevenir e combater doenças. Para isso foram criados sistemas automáticos que ajudam a comunidade médica e de biotecnologia a obter tais dados.

Os seus dados podem ser acessados por sua plataforma online como exemplificado na Figura 2. Dentre esses dados estão sua identificação, tipo de gene, sinônimos do seu nome, nome do RNA dentre outros.

DG1 pentatricopeptide repeat-containing protein delayed greening 1 [*Arabidopsis thaliana* (thale cress)]

Gene ID: 836893, updated on 8-May-2016

Summary	
Gene symbol	DG1
Gene description	pentatricopeptide repeat-containing protein delayed greening 1
Primary source	TAIR:AT5G67570
Locus tag	AT5G67570
Gene type	protein coding
RNA name	pentatricopeptide repeat-containing protein delayed greening 1
RefSeq status	REVIEWED
Organism	Arabidopsis thaliana (ecotype: Columbia)
Lineage	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; Gunneridae; Pentapetales; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis
Also known as	DELAYED GREENING 1; DG1; EMB1408; EMB246; embryo defective 1408; EMBRYO DEFECTIVE 246; K9I9.14; K9I9_14

Figura 2. Exemplo de uma tela com as informações de um gene no NCBI

GENE ONTOLOGY

O Gene Ontology, GO, mantém e desenvolve toda a linguagem e vocabulário usado para representar os genes e seus produtos e disponibiliza ferramentas que ajudam a manipular todos os dados armazenados por ele caracterizando assim uma ontologia.

De maneira geral o termo ontologia é a representação de algum conhecimento com todos os elementos que os compõem e como estão relacionados. Por isso a maioria das coisas que pode ser observada tem sua ontologia. Esse termo é bastante usado no campo da bioinformática (GO, 2015)

Os seus dados podem ser acessados por sua plataforma online como exemplificado na Figura 3. Dentre esses dados estão sua identificação, tipo de característica, sinônimos do seu nome, definições sobre sua atuação na planta dentre outros.

Outro ponto importante a se destacar sobre os dados do GO é que eles são estruturados de forma hierárquica, como pode ser analisado um exemplo na Figura 4 no qual a ontologia GO:0003700, fator de transcrição, faz parte de vários outros grupos. Dessa forma existem ontologias mais específicas e outras mais genéricas.

transcription factor activity, sequence-specific DNA binding

Term Information	
Accession	GO:0003700
Ontology	Molecular Function
Synonyms	alt_id: GO:0000130 exact: sequence-specific DNA binding transcription factor activity broad: transcription factor activity
Definition	Interacting selectively and non-covalently with a specific DNA sequence in order to modulate transcription. The transcription factor may or may not also interact selectively with a protein or macromolecular complex. <i>Source:</i> GOC:curators, GOC:txnOH
Comment	None
Subset	Plant GO slim Prokaryotic GO subset
Community	Add usage comments for this term on the GONUTS wiki.

[Back to top](#)

Figura 3. Exemplo de uma tela com as informações de uma ontologia no GO

```
GO:0003674 molecular_function [989119 gene products]
  GO:0001071 nucleic acid binding transcription factor activity [38198 gene products]
    GO:0003700 transcription factor activity, sequence-specific DNA binding [38188 gene products]
```

Figura 4. Exemplo de uma árvore hierárquica de uma ontologia do GO

THE ARABIDOPSIS INFORMATION RESOURCE

O *The Arabidopsis Information Resource*, TAIR (TAIR, 2016), é um banco de dados que mantém informações sobre biologia molecular da *Arabidopsis thaliana*. Os dados disponíveis no TAIR incluem a sequência genômica completa junto com sua estrutura, produto e expressão de genes, mapas genômicos. Os dados das funções dos genes são atualizados semanalmente, com as últimas pesquisas realizadas, através das submissões feitas pela comunidade.

O centro de recursos biológicos da *Arabidopsis thaliana* na universidade de Ohio que coleta, reproduz e preserva sementes e DNA da *Arabidopsis thaliana* está totalmente integrado com a plataforma do TAIR, tornado-a sempre atualizada e confiável.

Os seus dados podem ser acessados por sua plataforma online como exemplificado na Figura 5. Dentre esses dados estão sua identificação, tipo de gene,

sinônimos do seu nome, definições sobre sua atuação na planta e informações sobre suas características.

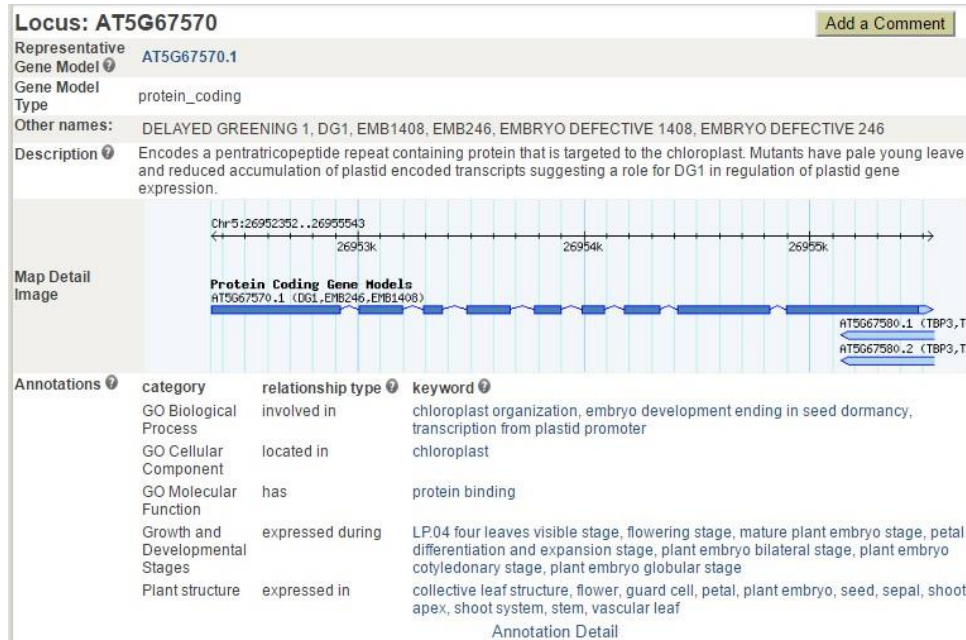


Figura 5. Exemplo de uma tela com as informações de um gene no TAIR

PLANT ONTOLOGY

O *Plant Ontology*, PO (PO, 2016), controla o vocabulário que descreve a anatomia, morfologia e estágios do desenvolvimento de todas as plantas. O objetivo do PO é criar uma *framework* para consultas entre diversas espécies considerando a expressão genética e fenótipo. Desde o início de janeiro de 2011 o PO foi unido em uma única base que antes eram sobre a anatomia e outra sobre o desenvolvimento estrutural das plantas.

Os seus dados podem ser acessados por sua plataforma online como exemplificado na Figura 6. Dentre esses dados estão sua identificação, tipo de característica, sinônimos do seu nome em várias línguas, definições sobre sua atuação na planta e comentários adicionais.

Outro ponto importante a se destacar sobre os dados do PO é que eles são estruturados de forma hierárquica, como pode ser analisado um exemplo na Figura 7 no qual a raiz (*root*) faz parte de vários outros grupos.

root

Term information ↓ Term lineage ↓ External references ↓ Term annotations →

Term Information

Accession	PO:0009005
Aspect	plant anatomy
Synonyms	narrow: aerial root narrow: climbing root exact: raíz (Spanish) exact: radices exact: radix exact: 根 (Japanese) alt_id: PO:0003006
Definition	A plant axis (PO:0025004) that lacks shoot axis nodes (PO:0005004), grows indeterminately, and is usually positively geotropic. [source: ISBN:978-0879015329 , ISBN:9780964022157 , POC:curators , POC:Laurel_Cooper]
Comment	Roots function in the absorption of water and inorganic nutrients, anchoring the plant body to the ground, and supporting it, storage of food and nutrients, and vegetative reproduction. The roots of most vascular plant species enter into symbiosis with soil-borne microorganisms. Roots are usually found underground, although there are many exceptions, such as the aerial roots of orchids. Roots often form secondary thickening from the root lateral meristem (PO:0006308). Commonly thought of as one of the three basic parts of the plant body, along with the shoot axis (PO:0025029) and leaves (PO:0025034).

Back to top

Figura 6: Exemplo de uma tela com as informações de uma ontologia no PO

- ⊕ all : all [150687]
 - ⊕ ⓘ PO:0025131 : plant anatomical entity [150663]
 - ⊕ ⓘ PO:0009011 : plant structure [150663]
 - ⊕ ⓘ PO:0025497 : collective plant structure [138737]
 - ⊕ ⓘ PO:0025007 : collective plant organ structure [138737]
 - ⊕ ⓘ PO:0025025 : root system [48853]
 - ⊕ ⓘ PO:0009005 : root [48388]

Figura 7: Exemplo de uma árvore hierárquica de uma ontologia do PO

Métodos

O projeto desenvolvido é apresentado em três partes principais como exibidas na Figura 8. A primeira etapa é referente a captura dos dados das bases online (TAIR, GO e PO), a segunda parte mostrará como foi feito o processamento dos dados para armazená-lo em forma de grafo e a terceira parte terá como foco a visualização e consulta dos grafos persistidos no banco.

Todos esses métodos em conjunto formam o programa Visual Ontogrator, que é um sistema web no qual é possível fazer consultas sobre os genes de *Arabidopsis thaliana* junto com suas relações e características.

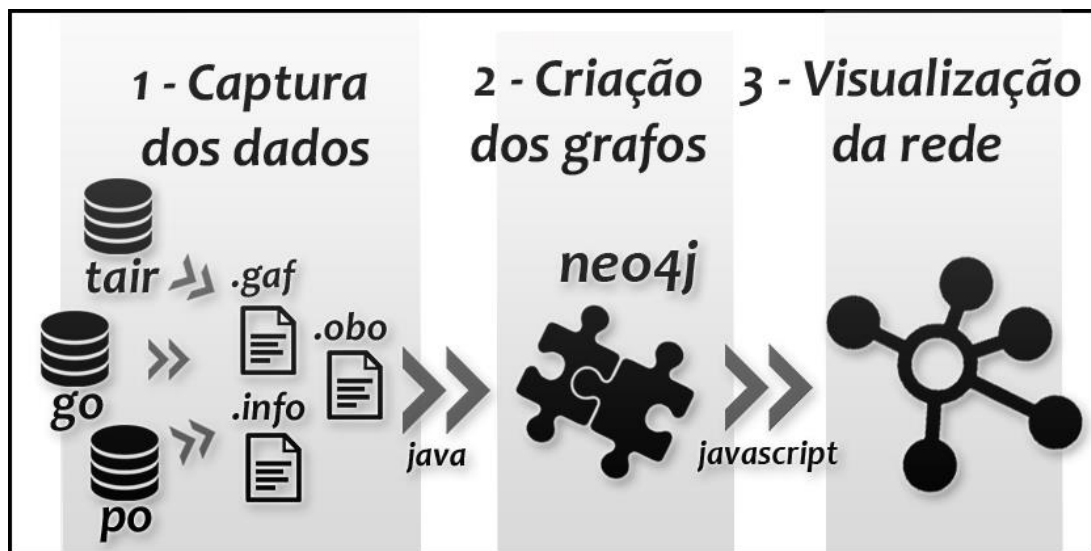


Figura 8. Etapas principais do projeto.

Captura dos dados e pré-processamento - Primeira Etapa

Os bancos usados nesse projeto são o TAIR, GO e o PO que podem ser acessados através de plataformas online. Seus dados podem ser baixados em forma de arquivos. Esses arquivos possuem três formatos sendo eles, GAF que armazena as relações entre os genes, OBO que armazena as ontologias e o INFO que tem o foco na descrição dos genes.

Depois dos arquivos estarem armazenados eles foram carregados em um programa em Java que os processou de modo a organizá-los criando as relações em um banco de dados MySQL. Os dados foram armazenados no banco de dados MySQL para remover possíveis redundâncias presentes nos arquivos e para facilitar futuras consultas visto que os arquivos estavam estruturados em forma de tabelas.

O banco de dados em MySQL resultou em 3 tabelas. A primeira é sobre os genes, ela contém colunas com a identificação usada no TAIR e no NCBI, o tipo de gene e em qual cromossomo o gene está inserido. A segunda tabela contém as ontologias, nelas se encontram as informações dos bancos GO e PO e suas colunas armazenam sua identificação, categoria, definição, palavras chave referentes a ontologia dentre outros dados. A última tabela é responsável por criar as associações entre os genes e as ontologias.

Criação dos grafos - Segunda Etapa

A próxima etapa foi criar um algoritmo para a gerar a estrutura do grafo, esse algoritmo utiliza as informações contidas no banco de dados MySQL, preenchidos com

os arquivos GAF, OBO e INFO, e os estruturava em forma de grafo utilizando o banco de dados Neo4J e a linguagem Cypher.

O grafo criado possui dois tipos de nós, um deles é o tipo Gene que armazena a descrição do gene, sua identificação de locus em qual cromossomo ele está presente dentre outras. O segundo tipo de nó é referente às características dos genes, esses nós possuem dados sobre a categoria, palavras chaves e as identificações nos bancos de dados GO e PO.

Todas as relações entre os nós são feitas entre um nó do tipo gene e um nó do tipo característica, por isso não é possível um nó do mesmo tipo estar relacionado diretamente um com o outro.

Visualização e consulta - Terceira Etapa

Após todo o processamento descrito anteriormente foi criado um grafo resultante com 33583 genes. Para acessar esses dados foi criada uma interface na qual o usuário pode interagir com a rede gerada mudando a posição e informações dos nós, a Figura 9 mostra um exemplo de grafo gerado a partir da base criada.

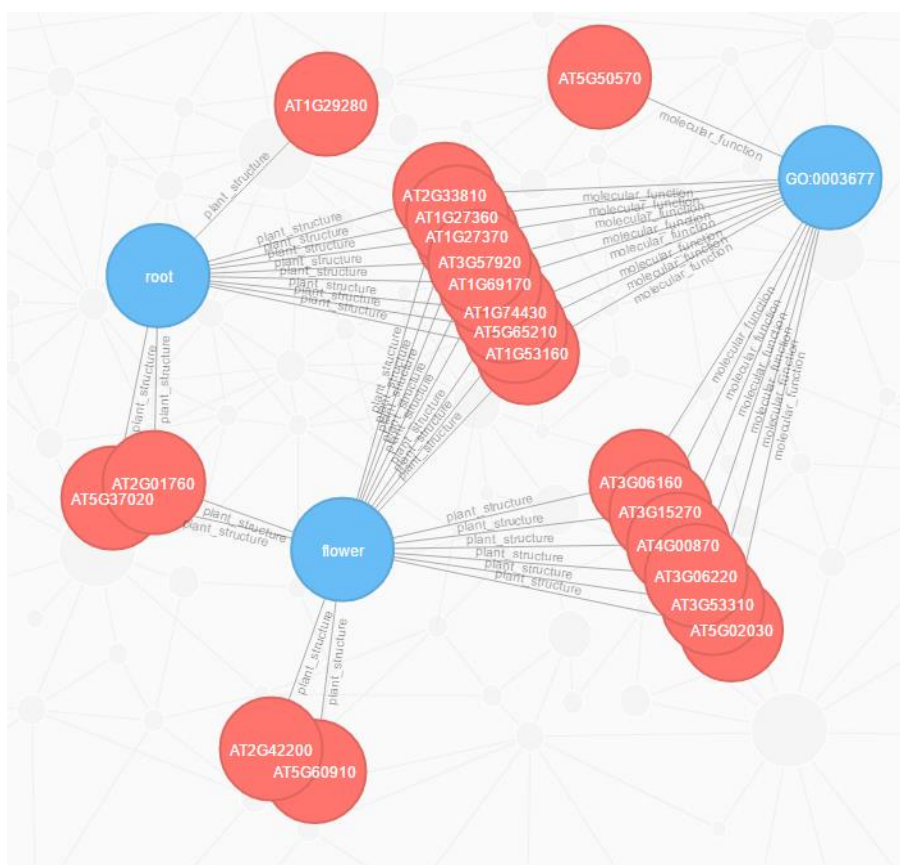


Figura 9. Exemplo de grafo gerado pelo Visual Ontogrator.

A interface foi feita usando a linguagem Javascript e a biblioteca d3.js especializada em visualização de grafos. Dessa forma o processamento gráfico da rede consultada será feita no computador do cliente.

RESULTADOS E DISCUSSÕES

Este capítulo apresenta os resultados desse projeto, com alguns exemplos e descrições do sistema final, juntamente com o *front-end* desenvolvido para a visualização das redes de genes.

Programas

Como resultado do projeto foram criados 2 programas, o primeiro deles é usado no processamento dos arquivos das bases de dados online e o segundo na consulta e visualização do grafo.

Processamento de arquivos

Um dos resultados desse trabalho foi a criação de um programa responsável pelo processamento dos arquivos das bases de dados online (TAIR, GO e PO) e seu armazenamento no banco de dados MySQL. Esse programa pode ser baixado no link <https://github.com/gabrielrubinobr/VisualOntogrator>.

Visual Ontogrator

A maior contribuição desse projeto foi a criação do programa Visual Ontogrator. Esse programa é um sistema web que tem como objetivo integrar dados de genes de *Arabidopsis thaliana* e os exibir em forma de grafo.

Na sua página inicial, como mostrado na Figura 10, o usuário pode colocar 3 listas principais. A primeira é a lista “Gene” onde podem ser colocados os genes, a segunda a lista “Relação” contendo as relações e a terceira, lista “Característica”, serão os nós referentes as características dos genes.

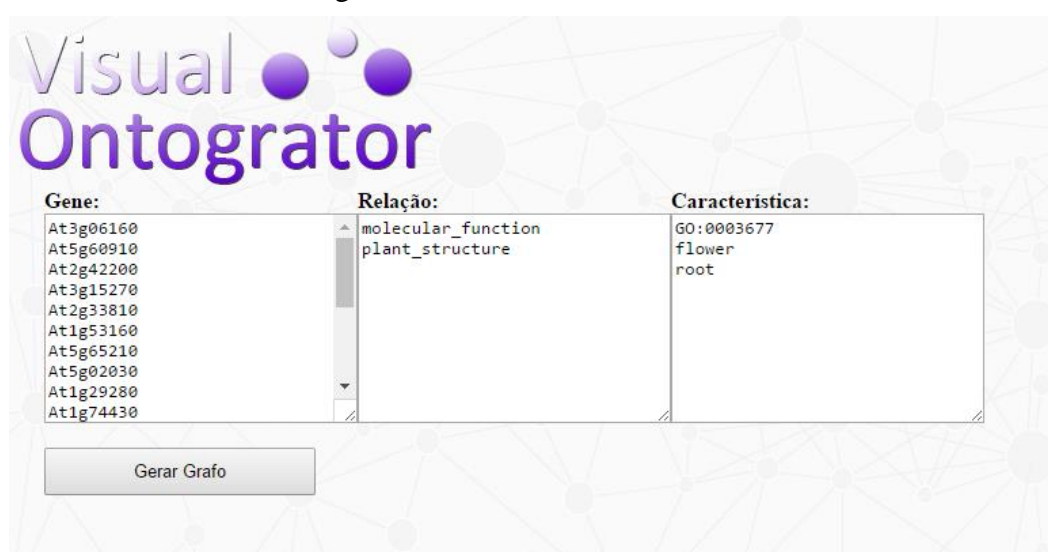


Figura 10. Exemplo de tela principal do Visual Ontogrator.

Depois de clicar no botão “Gerar Grafo”, apresentado na Figura 10, o usuário é redirecionado para a página onde será mostrado o grafo resultante.

Na página de visualização existe a opção para remover ou mostrar os nomes presentes nas relações, para isso basta clicar no botão "Ocultar/Mostrar Relações", o resultado desta ação pode ser vista na Figura 11.

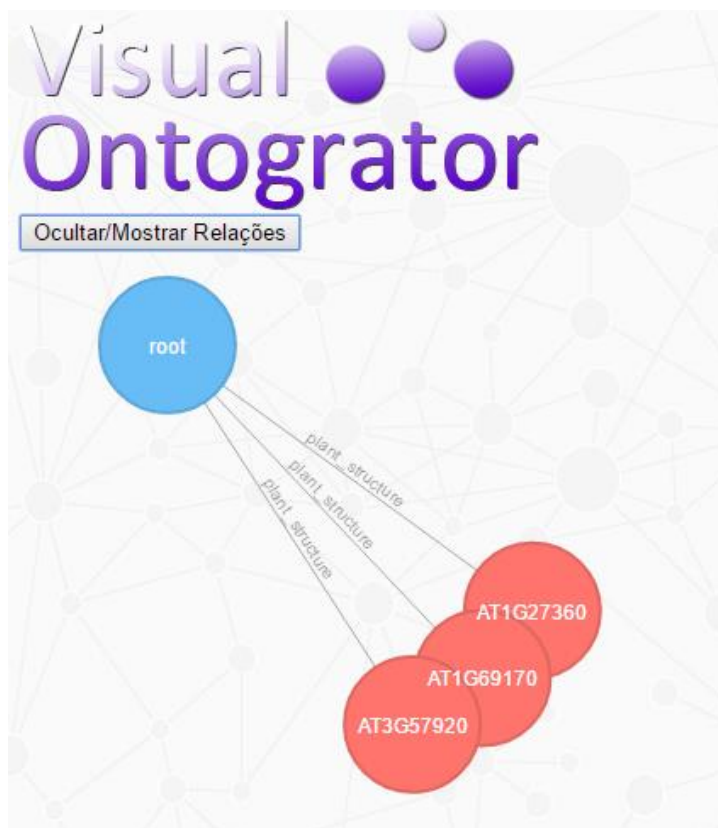


Figura 11. Exemplo de tela de grafos do Visual Ontogrator.

O programa Visual Ontogrator, juntamente com sua base de dados para o Neo4J podem ser baixados no link <https://github.com/gabrielrubinobr/VisualOntogrator>.

Desse modo o programa traz informações biológicas filtrando o grande volume de dados de dos bancos de genes e exibindo-as em forma de grafo. A Figura 9 exemplifica quais os genes estão relacionados com a raiz e a flor da *Arabidopsis thaliana* esse tipo de consulta ajuda o pesquisador, por exemplo, a checar informações ou comprovar teorias.

CONCLUSÕES

Neste capítulo são apresentadas quais as limitações que o projeto apresenta para que o escopo do problema fique bem definido permitindo o maior entendimento das tarefas realizadas.

Extração de informações de grafos

Depois da estruturação e armazenamento da rede em forma de grafo é possível medir e analisar suas informações. Por isso métodos de medida de redes complexas poderão ser adicionados em trabalhos futuros a fim de se obter novas informações sobre a rede tais como: clusterização e seus coeficientes, distância média, entropia e grau de distribuição, diâmetro e caminho mais curto.

Outro modo de extrair mais informações do grafo é mudando sua estrutura. Essa mudança poderá ser feita pela substituição dos nós de características por arestas de relação. Dessa forma a rede resultante teria somente nós de genes o que facilitará os cálculos de métricas desses grafos.

Limitações

Este trabalho se dispõe a integrar dados biológicos e gerar sua visualização através do desenvolvimento de metodologias adotadas. Observa-se que organismos possuem muitas peculiaridades tornando extremamente complexa a generalização de métodos eficazes para todos os casos. Nesse sentido o presente trabalho é focado em métodos de integração para um único organismo, a *Arabidopsis thaliana*. Dessa forma a validação dos métodos foi mais eficiente. Assim trabalhos futuros podem tentar generalizar essas metodologias a fim de abranger mais organismos.

Outro ponto a se destacar são os dados que são integrados as redes. Esses dados serão restritos às características dos genes. Futuramente outros tipos dados, como o de expressão de genes, poderão ser acoplados ao programa, o que pode reduzir a dimensionalidade do sistema e pode aumentar a precisão na inferência das redes.

REFERÊNCIAS

- [1] VAN EKENSTEIN, G.O.R.A.; DEURING, H.; TEN BRINKE, G.; ELLIS, T.S. Blends of Caprolactam/Caprolactone Copolymers and Chlorinated Polymers. *Polymer*, v. 38, p.3025-3034, 1997.
- [2] ALON, U. An introduction to systems biology: design principles of biological circuits. CRC press, 2006.
- [3] AUER. Dbpedia: A nucleus for a web of open data. [S.l.]: Springer, 2007. BISHOP, C. M. Neural networks for pattern recognition. [S.l.]: Oxford university press, 1995.
- [4] BOCCALETTI. Complex networks: Structure and dynamics. Physics reports, Elsevier, v. 424,n. 4, p. 175–308, 2006.
- [5] CSMBIO. MicroArray. 2015. Disponível em: <<http://pt.dbpedia.org/>>. Acesso em: 10 de setembro de 2015.
- [6] D3.JS. Data-Driven Documents. 2016. Disponível em: <<https://d3js.org/>>. Acesso em: 24 de maio de 2016.
- [7] DBPEDIA. DBpedia. 2015. Disponível em: <<http://pt.dbpedia.org/>>. Acesso em: 10 de setembro de 2015.
- [8] DEY, P. M.; HARBORNE, J. B. Plant biochemistry. [S.l.]: Academic Press, 1997.
- [9] DUFVA, M. Introduction to microarray technology. In: DNA Microarrays for Biomedical Research. Springer, 2009. p. 1–22.
- [10] GO. Gene Ontology. 2015. Disponível em: <<http://geneontology.org/>>. Acesso em: 10 de setembro de 2015.
- [11] INDYK, P.; MOTWANI, R. Approximate nearest neighbors: towards removing the curse of dimensionality. In: ACM. Proceedings of the thirtieth annual ACM symposium on Theory of computing. [S.l.], 1998. p. 604–613.
- [12] INITIATIVE, A. G. Analysis of the genome sequence of the flowering plant arabidopsis thaliana. *nature*, v. 408, n. 6814, p. 796, 2000.
- [13] KELEMEN. Computational intelligence in bioinformatics. [S.l.]: Springer, 2008.
- [14] KLUG. Conceitos de genética. [S.l.]: Artmed, 2010.
- [15] LEARN, S. Affinity Propagation. 2015. Disponível em: <<http://scikitlearn.org/stable/modules/clustering.html>>. Acesso em: 12 de setembro de 2015.
- [16] LOPES, F. M. Redes complexas de expressão gênica: síntese, identificação, análise e aplicações. Tese (Doutorado)— Universidade de São Paulo, 2011.

- [17] MARX, V. Biology: The big challenges of big data. Nature, Nature Publishing Group, v. 498, n. 7453, p. 255–260, 2013. 41
- [18] NATURE. Arabidopsis Thaliana. 2015. Disponível em: <<http://www.nature.com/>>. Acesso em: 10 de setembro de 2015.
- [19] NCBI. National Center for Biotechnology Information. 2015. Disponível em: <<http://www.ncbi.nlm.nih.gov/>>. Acesso em: 10 de setembro de 2015.
- [20] NEO4J. Cypher, SQL-inspired language. 2016. Disponível em: <<http://neo4j.com/developer/cypher-query-language/>>. Acesso em: 24 de maio de 2016.
- [21] PAGE. The pagerank citation ranking: bringing order to the web. Stanford InfoLab, 1999.
- [22] PAREJA, P. bio4j. 2015. Disponível em: <<http://bio4j.com/>>. Acesso em: 14 de setembro de 2015.
- [23] PEARL, J. Probabilistic reasoning in intelligent systems: networks of plausible inference. Morgan Kaufmann, 2014.
- [24] PENTEADO. Um estudo sobre bancos de dados em grafos nativos. 2014.
- [25] PO. Plant Ontology. 2016. Disponível em: <<http://www.plantontology.org/>>. Acesso em: 24 de maio de 2016.
- [26] ROBINSON. Graph databases. [S.l.]: "O'Reilly Media, Inc.", 2013.
- [27] SMITH, C. DBpedia. 2011. Disponível em: <<http://cs.smith.edu/>>. Acesso em: 12 de setembro de 2015.
- [28] TAIR. The Arabidopsis Information Resource. 2016. Disponível em: <<https://www.arabidopsis.org/>>. Acesso em: 24 de maio de 2016.
- [29] THORNALLEY, P. J. The glyoxalase system: new developments towards functional characterization of a metabolic pathway fundamental to biological life. Biochemical Journal, Portland Press Ltd, v. 269, n. 1, p. 1, 1990.
- [30] WANG. Function annotation of an sbp-box gene in arabidopsis based on analysis of coexpression networks and promoters. International journal of molecular sciences, Molecular
- [31] Diversity Preservation International, v. 10, n. 1, p. 116–132, 2009.
- [32] WEST. Introduction to graph theory. Prentice hall Upper Saddle River, 2001.
- [33] YOSHIDA. Estimating time-dependent gene networks from time series microarray data by dynamic linear models with markov switching. In: IEEE. Computational

Systems Bioinformatics Conference, 2005. Proceedings. 2005 IEEE. [S.l.], 2005. p. 289–298.