

Proposta de Trabalho

Jéssica Yumi Kojima

Supervisora
Profa. Dra. Kelly Rosa Braghetto

Departamento de Ciência da Computação
Instituto de Matemática e Estatística
Universidade de São Paulo

Tema da Monografia

Bancos de dados de imagens: uma aplicação em Neurociência.

Resumo

Experimentos em neurociência com frequência envolvem a aquisição de imagens como as de ressonância magnética e ressonância magnética funcional. Experimentos desse tipo são chamados pelos cientistas de experimentos de neuroimagem.

Atualmente não existem soluções padronizadas para o armazenamento dessas imagens, nem dos dados relacionados ao protocolo experimental sob o qual elas foram coletadas.

Além disso, muitas vezes um experimento de neurociência acaba utilizando documentos médicos, os quais não são relatados com detalhes suficientes para o nível de uma pesquisa. E existem casos nos quais pesquisadores, ao tentarem reproduzir seus trabalhos e os de outrem com precisão, sentem falta de algumas informações.

E ainda, o tamanho e o volume de imagens gerado a cada experimento pode ser gigantesco, dificultando seu armazenamento.

Logo, visando o armazenamento adequado dos dados e metadados apropriados relacionados aos experimentos de neurocientistas, este projeto irá propor uma solução para o gerenciamento de dados desses experimentos que envolvam imagens. Essa solução deverá comportar a quantidade de imagens produzidas a partir desses experimentos e todas as informações associadas a eles.

Contexto Científico e Justificativa do Trabalho

O grande aumento da capacidade de armazenamento de dados em equipamentos eletrônicos nos últimos anos, unido ao extremo barateamento do seu custo, tem possibilitado que cientistas coletem volumes cada vez maiores de dados provenientes de experimentos científicos. A área de neurociência exemplifica bem esse fato, com os seus dados de experimentos de eletrofisiologia e neuroimagem sendo relatados com maior frequência e detalhes em sistemas de arquivamento e bancos de dados, tanto nos públicos como nos de acesso restrito.

No entanto, existe uma grande variabilidade com a qual são descritos experimentos de neurociência. Não há um padrão único para os protocolos desses estudos, o que dificulta a reprodução fiel deles por outros pesquisadores. Além disso, tais experimentos podem apresentar um extenso volume de dados a serem armazenados, dentre quais também podem ser imagens.

Alguns estudos de neurociência, por exemplo, envolvem a obtenção de uma série de imagens da estrutura ou funcionamento parcial do cérebro. Os voluntários do pesquisa precisam ficar completamente parados enquanto a imagem estrutural é adquirida. Já enquanto a funcional é produzida, eles fazem atividades específicas, que dependem do estudo sendo realizado, como mexer uma parte do corpo repetidamente ou imaginar que estão se movimentando, sem estarem realmente.

Dessa forma, neurocientistas conseguem monitorar as atividades cerebrais, e verificar quais seções são encarregadas dessas funções. Essas partes do cérebro que apresentam maior atividade, em experimentos envolvendo ressonância magnética, por exemplo, ficam luminosas enquanto as outras se mantêm escuras na neuroimagem. Segundo o site "Telemed Providers" [2.1], esses experimentos que utilizam ressonância magnética geralmente produzem arquivos de imagens de 5 a 6MB, com 100 a mais de 1000 imagens contidas em 4 a 19 séries, sendo assim um desafio para o armazenamento em um banco de dados.

Já existem, nessa área de pesquisa em neurociência, algumas soluções para o armazenamento de neuroimagens e seus metadados, assim como artigos contendo as informações que precisam ser relatadas nesse tipo de experimentos. Dentre eles está o projeto Open Access Series of Imaging Studies (OASIS) e o artigo "Guidelines for reporting an fMRI study".

A área médica apresenta outras soluções, como, por exemplo, o PACS (Picture Archiving and Communication System). O PACS é um sistema de arquivamento e comunicação de acesso remoto voltado para o diagnóstico por imagem, como pode ser visto no artigo "PACS: Sistemas de Arquivamento e Distribuição de Imagens" [5.2]. Conforme seus autores, há também o padrão DICOM (Digital Imaging and Communications in Medicine), usado para transferir as imagens radiológicas e outras informações médicas. Outro exemplo de padrão utilizado pelos médicos é o HL7. Esse foi criado pela organização Health Level 7, a qual se envolve com o desenvolvimento de padrões de interoperabilidade internacionais de informática relacionada à saúde.

E ainda, existem soluções de propósito geral que envolvem o gerenciamento de imagens, como sistemas de bancos de dados e softwares de gerenciamento de arquivos que possam ser usados no armazenamento de imagens. Por exemplo, algumas extensões do PostgreSQL são utilizadas para tanto, como a extensão IE (Image-handling Extension), que implementa um novo tipo de dado chamado *PGImage*. Outro exemplo é o sistema Hadoop Distributed File System (HDFS), que também possibilita o armazenamento de imagens, desde que essas sejam convertidas em SequenceFiles – arquivos consistindo de chaves-valores binários.

Além disso, existem também bancos de dados baseados em NoSQL que gerenciam imagens, como o MongoDB. Ele possui uma especificação chamada GridFS, feita para o armazenamento e recuperação de arquivos que ultrapassam o tamanho limite de um documento BSON de 16MB. Ao invés de armazenar um arquivo em um único documento, GridFS o divide em partes, e armazena cada uma delas como um documento separado.

Analisando tais ferramentas e as soluções apresentadas nas áreas biológicas e biomédicas, este trabalho visa a proposição de uma arquitetura completa que suporta a quantidade de neuroimagens exigida pelos estudos de neurociência. Essa arquitetura utilizará ferramentas de código aberto, e conterá os dados necessários para possibilitar a reprodução com precisão de experimentos de neurocientistas.

Objetivos

1. Identificar os tipos, as restrições e os formatos dos dados e metadados envolvidos em um experimento de neuroimagem;
2. Modelar um banco de dados que represente esse domínio;
3. Propor uma arquitetura baseada em software livre para o armazenamento e gerenciamento das imagens coletadas em experimentos de neuroimagem. Nessa arquitetura, as imagens serão mantidas associadas aos dados do protocolo dos experimentos que as originaram;

Atividades já realizadas

- Estudo preliminar do tema;
- Delimitação do escopo do trabalho;

Atividades previstas para 2015

1. Estudo e a análise de requisitos de dados de experimentos de neurociência envolvendo neuroimagens, como, por exemplo, as informações a serem relatadas propostas no artigo "Guidelines for reporting an fMRI study" [3.1];
2. Estudos sobre ferramentas que podem ser usadas no gerenciamento de imagens, como, por exemplo, a extensão IE do PostgreSQL;
3. Usando como base padrões de representação de dados de domínios biológicos e biomédicos já existentes, juntamente com a análise realizada na primeira atividade, criar um modelo de banco de dados para protocolos de experimentos de neuroimagem;
4. Projetar uma arquitetura para o gerenciamento de imagens de experimentos de neurociência, usando software livre e o modelo de banco de dados desenvolvido na atividade anterior;
5. Redação da monografia;

Cronograma de Atividades

Atividade	Maio	Junho	Julho	Agosto	Setembro	Outubro	Novembro
1	X	X	X				
2		X	X	X			
3			X	X	X		
4			X	X	X	X	
5					X	X	X

Estrutura Esperada da Monografia

1. Introdução
 1. Motivação
 2. Objetivos
2. Conceitos
 1. Experimentos de Neurociência
 1. Protocolo geral
 2. Exemplos
 2. Neuroimagens
 1. Ressonância magnética (MRI, do inglês *Magnetic Resonance Imaging*)
 2. Ressonância magnética funcional (fMRI, do inglês *functional Magnetic Resonance Imaging*)
 3. Outras (Raio-X, Ultrassonografia e Cintilografia)
 3. Armazenamento de Imagens
3. Soluções médicas já existentes
 1. PACS
 2. DICOM
 3. HL7
4. Sistemas para o gerenciamento de imagens
 1. Sistemas gerenciadores de banco de dados com apoio a imagens
 2. Sistemas de arquivos distribuídos
 3. Sistemas especialistas
5. Uma proposta de solução para o gerenciamento de dados de experimentos de neuroimagem
 1. Banco de dados para protocolos de experimentos de neuroimagem
 1. Modelo conceitual
 2. Modelo lógico
 2. Arquitetura para o armazenamento de neuroimagens
 3. Protótipo do sistema
6. Conclusão
7. Bibliografia

Referências

1. Informações sobre neuroimagens:
 1. <https://en.wikipedia.org/wiki/Neuroimaging>
2. Informações sobre ressonância magnética:
 1. "Magnetic Resonance Imaging (MRI)": <http://www.telemedproviders.com/telemedicine-articles/91-magnetic-resonance-imaging-mri.html>
 2. "Magnetic resonance imaging": http://en.wikipedia.org/wiki/Magnetic_resonance_imaging
 3. "Ressonância magnética funcional mapeia regiões do cérebro": <http://globo.com/rede-globo/globo-ciencia/v/ressonancia-magnetica-funcional-mapeia-regioes-do-cerebro/2659584/>
3. Requisitos para um relatório de ressonância magnética funcional:
 1. R. A. Poldrack, P. C. Fletcher, R. N. Henson, K. J. Worsley, M. Brett, T. E. Nichols - "Guidelines for reporting an fMRI study": <http://www.sciencedirect.com/science/article/pii/S1053811907011020>
4. Armazenando imagens em banco de dados já existentes:
 1. PostgreSQL:
 1. <http://www.postgresql.org/docs/7.4/static/jdbc-binary-data.html>
 2. Guliato D, de Melo EV, Rangayyan RM, Soares RC - "POSTGRES-IE: an image-handling extension for PostgreSQL": <http://www.ncbi.nlm.nih.gov/pubmed/18214614>
 3. Guliato D, de Melo EV, Rangayyan RM, Soares RC - "POSTGRES-IE: an image-handling extension for PostgreSQL" – Seção 3 – PostgreSQL with Image-handling Extension: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3043679/#Sec3>
 2. Hadoop HDFS:
 1. <http://stackoverflow.com/questions/16546040/store-images-videos-into-hadoop-hdfs>
 3. NoSQL:
 1. <http://www.mongodb.org/>
 2. <http://docs.mongodb.org/manual/core/gridfs/>
5. Soluções empregadas para o armazenamento de neuroimagens:
 1. Picture Archiving and Communication System (PACS):
 1. http://en.wikipedia.org/wiki/Picture_archiving_and_communication_system
 2. Paulo Mazzoncini de Azevedo-Marques, Samuel Covas Salomão - "PACS: Sistemas de Arquivamento e Distribuição de Imagens": http://www.abfm.org.br/rbfm/publicado/rbfm_v3n1_131-9.pdf
 2. Digital Imaging and Communications in Medicine (DICOM):
 1. <http://en.wikipedia.org/wiki/DICOM>
 2. <http://dicom.nema.org/>
 3. Health Level 7 (HL7):
 1. http://en.wikipedia.org/wiki/Health_Level_7
 2. <https://www.hl7.org/>
 3. <http://www.interfaceware.com/hl7.html>
6. Bancos de dados abertos de Neurociência:

1. http://en.wikipedia.org/wiki/List_of_neuroscience_databases
7. Bases públicas para MRI e fMRI:
 1. OpenfMRI: <https://openfmri.org/>
 2. Open Access Series of Imaging Studies (OASIS): <http://www.oasis-brains.org/>