

Carlos Eduardo Benevides Bezerra

Disciplina: CMP112 - Sistemas de Informação Distribuídos

Prof.: José Palazzo Moreira de Oliveira

Autoanálise do trabalho desenvolvido

Introdução

O trabalho que está sendo desenvolvido na disciplina tem como objetivo a criação de um sistema de busca e agregação de informações na área de saúde em um repositório acessível através do sistema. Devido ao tempo disponível para realização deste projeto, o escopo do trabalho foi reduzido para informações a respeito do mal de Alzheimer. Pretende-se agrupar estas informações, obtidas de repositórios distribuídos e disponíveis online – além de normalizá-las e classificá-las de acordo com critérios de qualidade a serem determinados.

Foram definidas quatro etapas: padronização, localização, extração/normalização e classificação de acordo com a qualidade de cada informação. Foram criados quatro grupos, cada um com três integrantes. A cada grupo foi associada uma das etapas a serem executadas pelo sistema, sendo que o grupo do qual faço parte é o que está tratando da etapa de **localização**.

Nas seções seguintes, será descrito os atuais resultados alcançados no projeto, assim como a expectativa de como os conceitos aprendidos em aula poderão ser utilizados para ajudar no êxito do trabalho. Na última seção, estão destacadas as contribuições pessoais e a nota auto-atribuída.

Desenvolvimento do trabalho

Para atingir os objetivos traçados no trabalho, nós – eu e meus colegas de disciplina, com orientações do professor – definimos as etapas que o sistema deveria seguir e, baseado nisso cada grupo fragmentou sua etapa em partes menores. Por exemplo, o grupo de localização dividiu seu trabalho em três partes principais:

- definição dos motores de busca para fazer a pesquisa de informações de saúde na Internet;
- definição dos critérios a serem utilizados nas buscas e
- indexação dos resultados encontrados com as buscas em uma base de dados.

Após nós termos feito essa divisão, a cada membro foi designada uma das partes do trabalho de localização, sendo que a minha parte foi referente à **definição dos motores de busca**. Alguns motores de busca que encontramos foram:

- **Google Health**: através do *Google Health*, que requer *login* de usuário, é possível criar um perfil pessoal com informação a respeito do estado de saúde, assim como

- buscar informações a respeito de doenças e buscar médicos de áreas específicas;
- **Google Scholar**: permite busca de artigos científicos e deverá ser usado para encontrar informações mais técnicas, voltadas para um público com bases acadêmica e médica maiores;
 - **CurbSide.MD**: também voltado para a busca de artigos científicos, porém é focado apenas na área médica. Seus resultados já vêm classificados de acordo com o tipo de informação que trazem. Por exemplo, para a busca de informações sobre uma doença, ele já traz os resultados separados por: diagnóstico, tratamento, epidemiologia, relatos etc.;
 - **MedStory**: faz uma busca na web em geral, mas os classifica de acordo com diversas palavras-chave, indicando quais palavras-chave aparecem com maior frequência nos resultados;
 - **PubMed**: uma ferramenta relativamente simples de busca por artigos científicos da área médica;
 - **SearchMedica**: busca por artigos científicos na área médica, com resultados classificados de acordo com o tipo de artigo (como por exemplo artigos teóricos, artigos práticos com resultados de testes, informativos para pacientes e outros). A pesquisa também pode ser feita de acordo com a área médica (neurológica, cardiovascular etc.);
 - além dos mais conhecidos, como **Google (Web)**, **Yahoo!** e **Windows Live Search**.

Além dos motores de busca de páginas web, existe a possibilidade também de fazer buscar utilizando redes P2P, como foi proposto em (Balakrishnan et al., 2003) e em (Yu et al., 2004).

Motores escolhidos

Deseja-se obter, basicamente, três tipos de informações: textos em linguagem apropriada para o público em geral, textos técnicos/científicos com informações mais detalhadas e dados estatísticos e informações a respeito de médicos que possam atender às necessidades dos pacientes.

Yahoo! e Windows Live Search não serão utilizados, pois seus resultados são bastante semelhantes aos do Google Web. O motor PubSide não será utilizado, pois seus resultados, apesar de estarem estruturados, são bastante semelhantes aos do Google Scholar, sendo que seu escopo de busca é menor e seus resultados não são classificados. Portanto, dividindo os motores com base nessas características, temos:

- Textos com linguagem acessível: Google (Web), Google Health e MedStory;
- Textos técnicos/científicos: Google Scholar, CurbSide e SearchMedica;
- Dados de médicos e serviços: Google Health.

As redes P2P para busca de informações, apesar de serem escaláveis, descentralizadas e eficientes, não serão utilizadas por duas razões: primeiro, não foram

encontradas implementações de fato desse tipo de ferramenta de busca para dados na web sobre a área de saúde e, segundo, a implementação deste tipo de ferramenta não caberia no tempo disponível para a disciplina.

Conceitos aprendidos e expectativas para o projeto

Os conceitos de localização, padronização etc. foram aprendidos e servirão para encontrar e extrair informações de maneira a disponibilizar em um único repositório. O conceito de qualidade e, principalmente, de métricas de qualidade, terá um peso significativo para a avaliações das informações encontradas, especialmente no que diz respeito à fidedignidade das suas fontes, que também serão avaliadas segundo critérios a definir. Tal questão será fundamental, considerando que autenticidade de informações é um ponto crítico quando se trata da área de saúde.

Contribuições e tarefas realizadas pelo aluno

Para que o estado do trabalho chegasse aonde está, eu – assim como outros colegas – participei e continuo participando do seu andamento de diferentes formas: pesquisa de artigos científicos relacionados, participação ativa durante as aulas e discussão com outros alunos, dentro e fora do *moodle*. Na determinação dos grupos, contribuí para a definição do escopo deste trabalho, através da exposição de opiniões na sala de aula e na redação da tarefa atribuída ao grupo – e a cada integrante. Neste último caso, o texto foi enviado pelo aluno **Luiz Svoboda**.

Para ser mais claro quanto aos resultados obtidos por mim, eu ajudei na definição do escopo do trabalho – em especial durante a aula –, consegui artigos e ferramentas relacionadas e classifiquei os motores de busca de acordo com o tipo de informação que retornam, eliminando alguns que eram ou redundantes, ou cujos resultados não estavam agrupados de acordo com algum critério.

Nota auto-atribuída: 9, devido a uma ou duas faltas à aula.

Referências

Balakrishnan, H., Kaashoek, M. F., Karger, D., Morris, R., Stoica, Ion. Looking up data in P2P systems. In Communications of the ACM, 2003.

Yu, S., Liu, J., Le, J. DHT Facilitated Web Service Discovery Incorporating Semantic Annotation. In Lecture Notes in Computer Science, Springer, 2004.