

## Problema 2: Health Search

### Tema

Arquivos, ordenação, estruturas heterogêneas de dados.

### Cronograma

Data	Atividade
25/05/22	Apresentação do Problema 3
01/06/22	Sessão Tutorial - Problema 3
08/06/22	Sessão Tutorial - Problema 3
15/06/22	Sessão Tutorial - Problema 3
22/06/22	Sessão Tutorial - Problema 3
03/07/22	Entrega do código-fonte do Problema 3
05/07/22	Entrega do relatório do Problema 3

### Problema

Um grande hospital de Feira de Santana está enfrentando um problema de organização de seus documentos. Há uma grande quantidade de prontuários de pacientes, laudos de exames, protocolos e manuais de procedimentos que ficam armazenados em computadores. Estes documentos, no entanto, estão dispersos em vários lugares, além de ser difícil localizar estes documentos pelo seu conteúdo. Os prontuários ficam armazenados nos computadores das unidades de enfermagem, e são acessados e complementados à medida que ocorre a evolução do paciente. Os protocolos e manuais ficam no computador da diretoria, com várias versões conforme novos regulamentos são aprovados. Os laudos de exames ficam armazenados no computador do técnico de radiologia e são editados pelo médico responsável pelo exame. Para localizar um documento de interesse, é preciso solicitar a um funcionário para localizá-lo, o que muitas vezes é feito abrindo e verificando vários arquivos no computador.

Pensando em ampliar as possibilidades de organização e busca de documentos, os representantes da diretoria do hospital, em uma reunião geral, solicitaram ao colegiado de engenharia da computação, que seus alunos do 1º semestre promovessem novas alternativas de busca para melhorar a consulta de documentos, permitindo buscas a partir de palavras contidas no texto do documento. Um dos professores já indicou que a busca deve ser rápida e eficiente, e um dos veteranos sugeriu o uso de um índice invertido (*inverted index*), uma estrutura de dados muito utilizada em Recuperação da Informação (*Information Retrieval*).

O sistema de busca deve efetuar a leitura de todos os arquivos (formato \*.txt) presentes em um diretório, e o usuário pode sempre indicar um novo diretório para

incrementar a lista de documentos. Caso o usuário indique um diretório previamente indexado, o sistema deve atualizar as informações de documentos já indexados. O usuário também pode solicitar a remoção de documentos indexados, indicando um diretório ou um arquivo específico. O sistema também deve efetuar a busca de documentos a partir de uma palavra (termo), indicando todos os documentos que contêm aquela palavra e quantas vezes a palavra ocorre no documento, além de informar a quantidade de documentos encontrados. O resultado desta busca deve estar ordenado de maneira decrescente pela frequência do termo no documento. A operação do sistema deve ser unicamente por argumentos de linha de comando, que podem fazer a operação de atualização do índice pela indicação de um diretório ou arquivo, visualização do índice ou a operação de busca no índice por palavra-chave. Caso os argumentos não sejam adequados, o sistema deve fornecer ajuda ao usuário.

### Produto

Você deve fazer a implementação do sistema de busca e recuperação de documentos, com modularização adequada e estruturas de dados bem planejadas, usando a linguagem de programação Python. Só podem ser utilizadas bibliotecas básicas do Python, como 'os', mas não é permitido o uso de bibliotecas específicas para qualquer fim.

Você também deve entregar um relatório conforme modelo e instruções disponibilizados por seu tutor.

O código-fonte do *software* deve ser enviado ao seu respectivo tutor, via Google Classroom, até às 23h59m do dia **03/07/22**. O relatório deve ser enviado ao seu respectivo tutor, via Google Classroom, até às 23h59m do dia **05/07/22**. Haverá penalidade de **2 pontos** por um dia de atraso na entrega do código-fonte. Após um dia de atraso, o trabalho não será mais aceito. O relatório será aceito somente com o aceite do código-fonte. Haverá penalidade de **2 pontos** por um dia de atraso na entrega do relatório. Após um dia de atraso, o relatório não será mais aceito. Tanto o código-fonte quanto o relatório devem ser desenvolvidos **individualmente**. Por fim, deve constar no código-fonte declaração de ausência de plágio, disponibilizada no Google Classroom.

### Recursos para Aprendizagem

Vide Plano de Ensino no Sagres.