



ESTRUTURA DE DADOS (T390/T923)	
Nome do Projeto: Índice Remissivo	
TERMO DE ABERTURA DO PROJETO	
Professor: Gilson Pereira do Carmo Filho	Semestre: 2020.2
Avaliação: AV3	Pontos: 7

## 1. Justificativa do projeto

Várias aplicações necessitam de um relatório de referências cruzadas. Por exemplo, a maioria dos livros apresentam um índice remissivo, que corresponde a uma lista alfabética de palavras-chave ou palavras relevantes do texto com a indicação dos locais no texto onde cada palavra-chave ocorre.

Como exemplo, suponha um arquivo contendo um texto constituído por:

**Good programming is not learned from  
generalities, but by seeing how significant  
programs can be made clean, easy to  
read, easy to maintain and modify,  
human-engineered, efficient, and reliable,  
by the application of common sense and  
by the use of good programming practices.**

Assumindo que o índice remissivo seja constituído das palavras-chave:

**programming, programs, easy, by, human-engineered, and, be, to,**

o programa para criação do índice deve produzir a seguinte saída:

and	4	5	6
be	3		
by	2	6	7
easy	3	4	
human-engineered	5		
programming	1	7	
programs	3		
to	3	4	

Note que a lista de palavras-chave está em ordem alfabética. Adjacente a cada palavra está uma lista de números de linhas, um para cada vez que a palavra ocorre no texto. Uma estrutura de dados desse tipo é conhecida como **arquivo invertido**. O arquivo invertido é um mecanismo muito utilizado em arquivos constituídos de texto, como as **máquinas de busca na Web**.

## 2. Descrição do produto do projeto

Projete um programa para produzir um índice remissivo. O programa deverá ler um número arbitrário de palavras-chave que deverão constituir o índice remissivo, seguido da leitura de um texto de tamanho arbitrário, que deverá ser esquadrinhado à procura de palavras que pertençam ao índice remissivo. Cabe ressaltar que uma palavra é definida como uma sequência de letras e dígitos, começando com uma letra.

Utilize um método eficiente para verificar se uma palavra lida do texto pertence ao índice. Para resolver esse problema, você deve utilizar duas estruturas de dados distintas: **árvore de busca** e **tabela de dispersão** (*hashing*).

Observe que, apesar de o *hashing* ser mais eficiente do que árvore de busca, existe uma desvantagem na sua utilização: após atualizado todo o índice remissivo, é necessário imprimir suas palavras em ordem alfabética. Isso é imediato em árvores de busca, mas, quando se usa *hashing*, isso é problemático, sendo necessário ordenar a tabela de dispersão que contém o índice remissivo.

### 3. Principais entregas do projeto

As principais entregas do projeto são:

GRUPO	REQUISITOS
Entrada de Dados	<ul style="list-style-type: none"><li>• Ler um arquivo do tipo TXT (texto) contendo um número arbitrário de palavras-chave que deverão constituir o índice remissivo;</li><li>• Ler um arquivo do tipo TXT contendo o texto a ser esquadriado à procura de palavras que pertençam ao índice remissivo.</li></ul>
Geração do Índice Remissivo	<ul style="list-style-type: none"><li>• Incluir as palavras-chave no índice;</li><li>• Utilizar um método eficiente para verificar se uma palavra lida do texto pertence ao índice;</li><li>• Utilizar duas estruturas de dados distintas: <b>árvore de busca</b> e <b>tabela de dispersão</b> (<i>hashing</i>).</li></ul>
Saída de Dados	<ul style="list-style-type: none"><li>• Gerar um arquivo do tipo TXT contendo o índice remissivo.</li></ul>

### 4. Estratégia de condução do projeto

- Dúvidas acerca dos requisitos do projeto deverão ser esclarecidas com o professor;
- O código fonte do projeto deverá ser enviado por meio do **AVA** (Moodle) até o prazo estipulado para a entrega. Não serão aceitas entregas após o prazo;
- Na data de apresentação dos projetos, o código fonte deverá estar disponível no AVA para o professor, que então executará e testará o programa na presença dos alunos. Estes, por sua vez, serão **arguidos pelo professor** sobre o código fonte;
- O código fonte do projeto será submetido a uma ferramenta de verificação de plágio. Qualquer tentativa de cópia do projeto de outro aluno ou da Internet, ou qualquer outra tentativa de fraudar o projeto, incluindo cópia de trechos do código fonte, resultará em aplicação de nota **ZERO**.

## 5. Critério de avaliação do projeto

Na avaliação do projeto serão consideradas a execução correta das funcionalidades do programa e a conformidade do código fonte ao conteúdo abordado na disciplina.

A tabela a seguir mostra a pontuação das entregas do projeto.

GRUPO	PONTOS
Entrada de Dados	1,0
Geração do Índice Remissivo	5,0
Saída de Dados	1,0
<b>TOTAL</b>	<b>7,0</b>

## 6. Premissas e restrições para o projeto

PREMISSAS	RESTRIÇÕES
<ul style="list-style-type: none"><li>Os alunos deverão ter lido e realizado as atividades do livro <b>Estrutura de Dados: algoritmos, análise da complexidade e implementações em Java e C/C++</b>.</li></ul>	<ul style="list-style-type: none"><li>O projeto tem um prazo total de 2 (duas) semanas;</li><li>O projeto deverá ser desenvolvido na linguagem <b>Java</b>;</li><li>O projeto deverá ser realizado preferencialmente em dupla ou individualmente.</li></ul>