



**Universidade Federal de Itajubá - UNIFEI - Campus Itabira**  
**Engenharia de Computação**  
**Estrutura de Dados II - ECO029**

**ANDRÉ 'ZANV' VIANA - 25037**  
**LUCAS SAMUEL VIEIRA - 24072**  
**WILLIAM A. PRADO - 24015**

**RELATÓRIO - TRABALHO PRÁTICO I - MÁQUINA DE BUSCA**

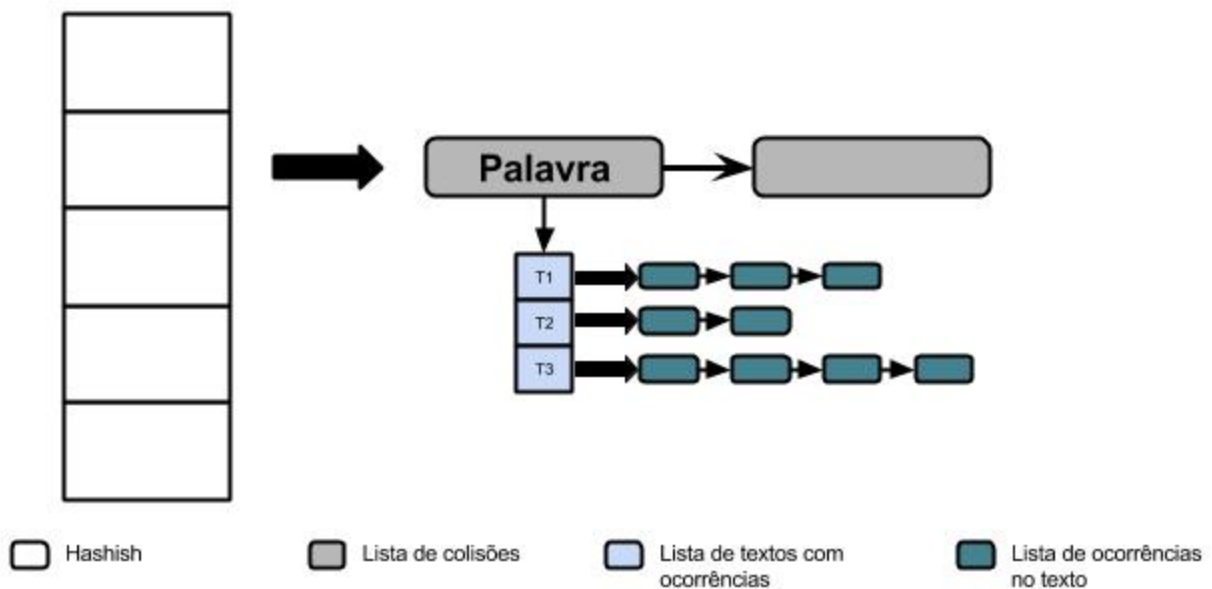
**ITABIRA**  
**2014**

## Descrição do trabalho

O projeto teve como objetivo a criação de um programa capaz de realizar pesquisas por palavras e frases em arquivos de texto pré determinados, usando para o tal o método de Hashing, com tratamento de colisão. As palavras e frases encontradas na busca devem ser impressas juntamente com o nome do arquivo em que se encontram e o contexto.

## Estruturas de Dados Utilizadas

Para a realização das pesquisas, é utilizado o método de Hashing, com listas dinâmicas para o tratamento de colisão. A estrutura completa possui um vetor [Hashish] de listas cujos nós contém sub-listas de nós de arquivo que por sua vez contém mais uma série de sub-listas, cujos nós marcam as posições das ocorrências de determinada palavra em determinado arquivo.



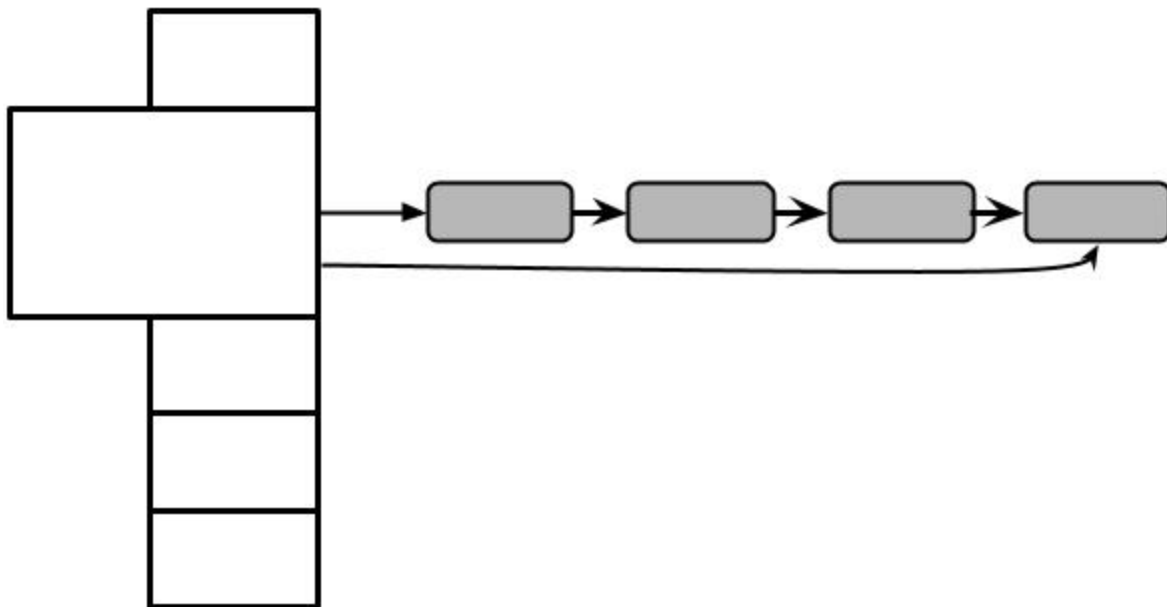
## Vetor Hash (Hashish)

Para o Hashing, é utilizado um vetor de  $10^3$  posições. O índice de cada palavra é definido pela função:

$$\text{Índice} = \left( \sum_{i=1}^n (c_i * 2^i) \right) \bmod 10^3$$

Onde 'n' é o número de caracteres na string, 'i' é a posição do caractere a ser processado incrementada em 1, e 'c', o valor int do caractere a ser processado.

Cada célula do Hashish possui uma lista de colisões.



## Palavra

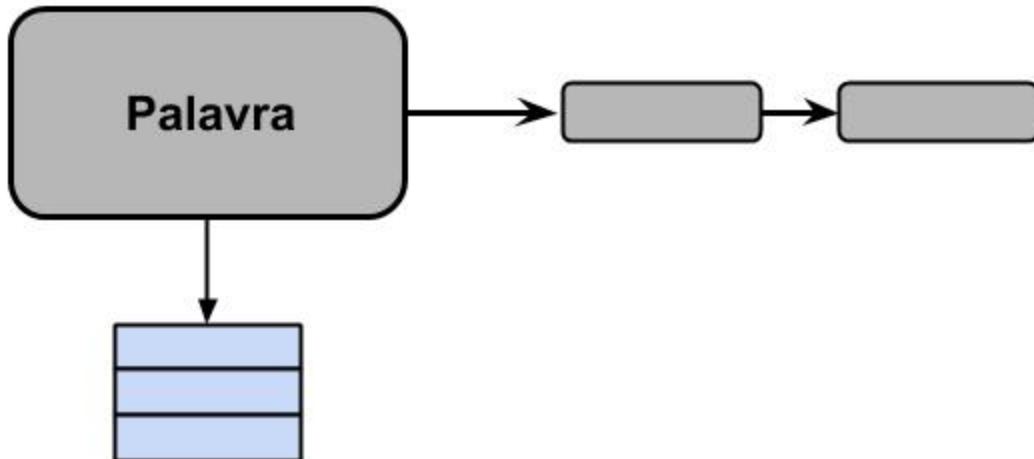
A estrutura Palavra fez-se necessária para a realização das comparações necessárias na busca e adição de ocorrências. Em seu âmago encontra-se uma string contendo uma palavra.

## Listas

O programa usa templates para o tratamento das diversas listas utilizadas. Na estrutura geral de Hash, existem três tipos de lista que diferem no tipo de dado armazenado e propósito.

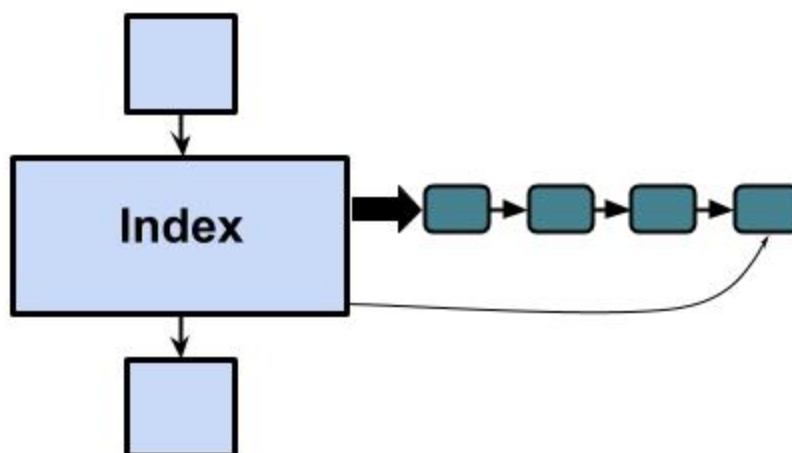
## Lista de Colisões

A lista usada no tratamento de colisões é dinâmica e seus nós usam como chave as palavras cujas ocorrências estão marcadas em suas sub-listas; tal palavra é armazenada com o uso da struct Palavra.



## Sub-lista de Textos

A sub-lista de textos encontra-se sob a lista de colisões, e contém todos os textos nos quais há ao menos uma ocorrência da palavra em questão. Os nós desta lista contém uma referência ao arquivo no qual há ocorrências e uma sub-lista com tais ocorrências.



## Sub-lista de Ocorrências

Por fim, a sub-lista de ocorrências, sob a sub-lista de textos, contém a posição de cada ocorrência da palavra em questão no texto em questão, armazenada em uma variável unsigned long.



## Implementação

A implementação do programa foi dividida em três partes: Text Pool, Hash e Pesquisa.

A classe Text Pool guarda os textos indicados pelo arquivos robots.txt, nos quais ocorrerão as pesquisas. Suas principais funções são usadas para encontrar, processar, guardar e resgatar os textos a serem usados.

Sob Hash, encontram-se as estruturas utilizadas no programa e as funções que preenchem e navegam por tais estruturas. O vetor Hashish e suas listas e sub-listas (definidas por templates) são preenchidos com as palavras e ocorrências dos textos durante a inicialização do programa.

Por fim, sob Pesquisa encontram-se as funções responsáveis pela interface e controle do programa; notavelmente as funções responsáveis por imprimir os resultados da pesquisa.

## Testes

## Busca por Palavra

```
lucas@lucas-xps-l502x: ~/Git/TP01-ECO029-2014/src
lucas@lucas-xps-l502x:~/Git/TP01-ECO029-2014/src$ ./TP01
ATENCAO: Ao utilizar acentos na pesquisa, verifique se o seu terminal ou prompt de comando utiliza codificacao ANSI/WINDOW
S-1252.
Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
boy

Resultado da busca "boy"
1 - tests/braid-epilogue.txt
   "The boy called for the girl to follow him"
   "The boy wanted to protect the girl"
```

```
lucas@lucas-xps-l502x: ~/Git/TP01-ECO029-2014/src
lucas@lucas-xps-l502x:~/Git/TP01-ECO029-2014/src$ ./TP01
ATENCAO: Ao utilizar acentos na pesquisa, verifique se o seu terminal ou prompt de comando utiliza codificacao ANSI/WINDOW
S-1252.
Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
delicia

Resultado da busca "delicia"
ERRO: Palavra nao encontrada.

Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
pai

Resultado da busca "pai"
ERRO: Palavra nao encontrada.
```

```
lucas@lucas-xps-l502x: ~/Git/TP01-ECO029-2014/src
lucas@lucas-xps-l502x:~/Git/TP01-ECO029-2014/src$ ./TP01
ATENCAO: Ao utilizar acentos na pesquisa, verifique se o seu terminal ou prompt de comando utiliza codificacao ANSI/WINDOW
S-1252.
Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
lorem

Resultado da busca "lorem"
1 - tests/lorem_ipsum.txt
   "Lorem ipsum dolor sit amet"
   "Aliquam vitae quam lorem"
   "Lorem ipsum dolor sit amet"
   "Fusce tristique lectus pulvinar lorem consequat"
```

Resultado da busca "a"

- 1 - tests/eggman\_hates\_furries.txt  
"They have been the center of conflict on Mobius since a time no one can recall"  
"A great struggle takes place whenever they are gathered;"
- 2 - tests/samba\_de\_uma\_notas.txt  
"Mas a base 0 uma s0"  
"Como sou a conseq0ncia"
- 3 - tests/pangramas.txt  
"0 noite"
- 4 - tests/lorem\_ipsum.txt  
"Maecenas a volutpat mauris"  
"Morbi ac elit a erat suscipit porta"  
"Proin non enim a ante ultrices pulvinar eget a justo"  
"Proin non enim a ante ultrices pulvinar eget a justo"  
"Aenean a velit vitae felis auctor tempus sed aliquet nisi"
- 5 - tests/tabacaria.txt  
"Para uma rua inacessível a todos os pensamentos"  
"Com a morte a por umidade nas paredes"  
"Com a morte a por umidade nas paredes"  
"Com o Destino a conduzir a carroça de tudo pela estrada de nada"  
"Com o Destino a conduzir a carroça de tudo pela estrada de nada"  
"como se soubesse a verdade"  
"A fileira de carruagens de um comboio"  
"Estou hoje dividido entre a lealdade que devo"  
"A aprendizagem que me deram"
- 6 - tests/braid-epilogue.txt  
"escaping to a life of freedom"  
"or put his arm around her shoulders in a walking embrace"  
"and he picked a path through the jostling crowd"  
"the twisting of metal orbs hanging from a thread"  
"the twisting of metal orbs hanging from a thread"  
"he kneeled behind a bunker in the desert; he held a piece of welder's glass up to his eyes and waited"  
"he kneeled behind a bunker in the desert; he held a piece of welder's glass up to his eyes and waited"  
"Space contracted to a pinpoint"  
"like a sigh"  
"she always walked him on a route that passed in front of the candy store"  
"He can fit them together to create a foundation"  
"a castle"  
""To build a castle of appropriate size"  
"he will need a great many stones"

```

lucas@lucas-xps-l502x: ~/Git/TP01-ECO029-2014/src
lucas@lucas-xps-l502x:~/Git/TP01-ECO029-2014/src$ ./TP01
ATENCAO: Ao utilizar acentos na pesquisa, verifique se o seu terminal ou prompt de comando utiliza codificacao ANSI/WINDOWS-1252.
Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
the

Resultado da busca "the"
1 - tests/eggman_hates_furries.txt
  "The seven chaos emeralds are known to hold the key to manipulation of time and space"
  "The seven chaos emeralds are known to hold the key to manipulation of time and space"
  "They have been the center of conflict on Mobius since a time no one can recall"
  "terror shakes the earth"
  "and destruction rains down on the world"
  "they feed both sides of the battle"
2 - tests/pangramas.txt
  "Quick brown fox jumps over the lazy dog"
3 - tests/braid-epilogue.txt
  ""The boy called for the girl to follow him"
  ""The boy called for the girl to follow him"
  "fighting off the creatures made of smoke and doubt"
  "The boy wanted to protect the girl"
  "The boy wanted to protect the girl"
  "to help her feel supported and close to him amid the impersonal throngs of Manhattan"
  "They turned and made their way toward the Canal St"
  "and he picked a path through the jostling crowd"
  "she said: "You're going the wrong way and you're pulling me with you"
  "He scrutinized the fall of an apple"
  "the twisting of metal orbs hanging from a thread"
  "He was searching for the Princess"
  "implanted tungsten posts into the skulls of water-starved monkeys"
  "" But he would not see her; he only knew how to look at the outside of things"
  ""He scrutinized the fall of an apple"
  "the twisting of metal orbs hanging from a thread"
  "Through these clues he would find the Princess"
  "he kneeled behind a bunker in the desert; he held a piece of welder's glass up to his eyes and waited"
  "It was as though the earth had opened and the skies split"
  "It was as though the earth had opened and the skies split"
  "One felt as though he had been privileged to witness the Birth of the World"
  "One felt as though he had been privileged to witness the Birth of the World"
  "she felt the sadness beneath; she let her breath fall softly"
  "like ashes floating gently on the wind"
  ""She couldn't understand why he chose to flirt so closely with the death of the world"
  ""She couldn't understand why he chose to flirt so closely with the death of the world"
  ""The candy store"
  "Everything he wanted was on the opposite side of that pane of glass"
  "The store was decorated in bright colors"
  "and the scents wafting out drove him crazy"
  "He tried to rush for the door"
  "or just get closer to the glass"
  "or the way he yanked painfully on her braid to make her stop"
  "She followed his gaze toward treats sitting on pillows behind the glass: the chocolate bar and the magnetic monopole"
  "She followed his gaze toward treats sitting on pillows behind the glass: the chocolate bar and the magnetic monopole"
  "She followed his gaze toward treats sitting on pillows behind the glass: the chocolate bar and the magnetic monopole"
  "the It-From-Bit and the Ethical Calculus; and so many other things"
  "the It-From-Bit and the Ethical Calculus; and so many other things"
  "she always walked him on a route that passed in front of the candy store"
  "The moments feel substantial in his mind"
  "reaching down toward the closest one"
  ""He tests the stone's weight; he finds he can lift it"
  "and the others too"

```

```

lucas@lucas-xps-l502x: ~/Git/TP01-ECO029-2014/src
lucas@lucas-xps-l502x:~/Git/TP01-ECO029-2014/src$ ./TP01
ATENCAO: Ao utilizar acentos na pesquisa, verifique se o seu terminal ou prompt de comando utiliza codificacao ANSI/WINDOWS-1252.
Selecione a opcao de pesquisa.
1. Pesquisa por palavra
2. Pesquisa por frase
0. Sair
1
Digite a palavra procurada:
pao

Resultado da busca "pao"
Experimente procurar no ceu,

```



## **Conclusão**

Ao fim do projeto, foi concluído que máquinas de busca simples são menos complexas do que normalmente se espera, porém demandam considerável conhecimento, experiência e hábito para que funcionem de forma correta e estável.