



Trabalho 2 – DigestCalculator

Construir um programa Java que (i) use a JCA; (ii) não use interface gráfica; e (iii) seja executado em uma linha de comando com argumentos, da seguinte forma:

DigestCalculator <SP> *Tipo_Digest* <SP> *Caminho_da_Pasta_dos_Arquivos* <SP> *Caminho_ArqListaDigest*

onde,

Tipo_Digest – Tipo do digest a ser calculado (MD5/SHA1/SHA256/SHA512)

Caminho_da_Pasta_dos_Arquivos - Informa a localização (caminho) da pasta que contém os arquivos que devem ser processados.

Caminho_ArqListaDigest - Informa a localização do arquivo que contém uma lista de digests de arquivos conhecidos.

<SP> - Caractere espaço em branco.

O arquivo com a lista de digests utiliza o formato XML e é formado por zero ou mais linhas formatadas da seguinte maneira:

```
<CATALOG>
  <FILE_ENTRY>
    <FILE_NAME>Nome_Arq</FILE_NAME>
    <DIGEST_ENTRY>
      <DIGEST_TYPE>Tipo_Digest</DIGEST_TYPE>
      <DIGEST_HEX>Digest_Hex</DIGEST_HEX>
    </DIGEST_ENTRY>
  </FILE_ENTRY>
</CATALOG>
```

onde,

Nome_Arq - Nome de um arquivo qualquer, sem informar o caminho.

TipoDigest - Indica o digest em seguida (MD5/SHA1/SHA256/SHA512).

Digest_Hex - Digest em hexadecimal referente ao tipo de digest especificado anteriormente.

<CATALOG> - Tag raiz do XML.

<FILE_ENTRY> - Tag que marca o início das entradas de um arquivo.

<FILE_NAME> - Tag que fornece o nome do arquivo.

<DIGEST_ENTRY> - Tag que marca o início da entrada de um digest.

<DIGEST_TYPE> - Tag que fornece o TipoDigest do digest.

<DIGEST_HEX> - Tag que fornece o Digest_hex do digest.

Observações:

- Um <CATALOG> é formado por zero ou mais <FILE_ENTRY>;
- Um <FILE_ENTRY> possui um <FILE_NAME> e um ou mais <DIGEST_ENTRY>.
- Um <DIGEST_ENTRY> possui um <DIGEST_TYPE> e um <DIGEST_HEX>.

Exemplo:

```
<CATALOG>
  <FILE_ENTRY>
    <FILE_NAME>Arquivo1.dat</FILE_NAME>
    <DIGEST_ENTRY>
      <DIGEST_TYPE>SHA1</DIGEST_TYPE>
      <DIGEST_HEX> 8d901bb3a2840ac030f7dbdd7cb823808858cb2f</DIGEST_HEX>
    </DIGEST_ENTRY>
    <DIGEST_ENTRY>
      <DIGEST_TYPE>MD5</DIGEST_TYPE>
      <DIGEST_HEX>42b83991bd1b47b373074111c34fb428</DIGEST_HEX>
    </DIGEST_ENTRY>
  </FILE_ENTRY>
  <FILE_ENTRY>
    <FILE_NAME>Arquivo2.dat</FILE_NAME>
    <DIGEST_ENTRY>
      <DIGEST_TYPE>SHA256</DIGEST_TYPE>
      <DIGEST_HEX>c8db093d264aa744d178470ad97aa64e67e84ab96e3b3310fb6f0eda429e6622</DIGEST_HEX>
    </DIGEST_ENTRY>
  </FILE_ENTRY>
</CATALOG>
```

Neste exemplo, o Arquivo1.dat tem dois digests na sua lista (SHA1 e MD5) enquanto o Arquivo2.dat tem apenas um digest na sua lista. Porém, ambos poderiam ter até 4 digests nas suas respectivas listas, um de cada tipo, em qualquer ordem.

O programa deve executar o seguinte procedimento:

- 1 - Calcular o digest solicitado do conteúdo de todos os arquivos presentes na pasta fornecida;
- 2 - Comparar os digests calculados com os respectivos digests registrados para cada arquivo no arquivo ArqListaDigest, se existirem, e com os digests dos arquivos existentes na pasta;
- 3 – Imprimir, na saída padrão, uma lista com o seguinte formato:

```
Nome_Arq1<SP>Tipo_Digest<SP>Digest_Hex_Arq1<SP>(STATUS)
Nome_Arq2<SP>Tipo_Digest<SP>Digest_Hex_Arq2<SP>(STATUS)
.....
Nome_ArqN<SP>Tipo_Digest<SP>Digest_Hex_ArqN<SP>(STATUS)
```

onde:

<SP> - Caracter espaço em branco.

Nome_Arq1 .. Nome_ArqN - Correspondem aos nomes dos arquivos encontrados na pasta fornecida para o cálculo dos digests (sem a informação do caminho da pasta).

Tipo_Digest - Tipo do digest calculado (MD5/SHA1/SHA256/SHA512)

Digest_Hex_ArqN – Digest formatado em hexadecimal calculado para o arquivo N.

STATUS - Corresponde a um dos status definidos abaixo:

OK = Status do arquivo cujo digest calculado é igual ao digest fornecido no arquivo ArqListaDigest e não colide com o digest de outro arquivo na pasta.

NOT OK = Status do arquivo cujo digest não é igual ao digest fornecido no arquivo ArqListaDigest e não colide com o digest de outro arquivo na pasta.

NOT FOUND = Status do arquivo cujo digest não foi encontrado no arquivo ArqListaDigest e não colide com o digest de outro arquivo na pasta.

COLISION = Status do arquivo cujo digest calculado colide com o digest de outro arquivo de nome diferente encontrado no arquivo ArqListaDigest ou com o digest de um dos arquivos presentes na pasta.

4 - Os digests calculados para os arquivos com status NOT FOUND devem ser acrescentados no registro de um nome de arquivo existente ou como uma nova entrada de um novo arquivo no final do arquivo de lista de digests, mantendo seu formato padrão. Os digests calculados para os arquivos com status COLISION não devem ser acrescentados no arquivo de lista de digests.

Observação 1: O nome do programa executável deve ser DigestCalculator.

Observação 2: O código fonte deve ser compilado com o Sun JDK 1.8.

Observação 3: Estude os métodos *update* da classe *MessageDigest* e selecione o método adequado para este trabalho.

Observação 4: Os digests devem ser calculados para o *conteúdo dos arquivos* presentes na pasta fornecida na linha de comando e NÃO para o *nome dos arquivos* que estão na pasta.

Observação 5: Se os argumentos da linha de comando forem omitidos ou insuficientes para a execução do programa, deve-se imprimir uma mensagem com a orientação de execução e, em seguida, o programa deve ser encerrado.

O programa fonte deve ser submetido na seção de tarefa deste trabalho, no site de EAD da PUC-Rio. **Prazo de submissão: 17/4/2023 – 13:00h. Prazo máximo para submissão: 17/4/2023 – 23:59h.**