



# Cento Universitário UNA

Sistemas de Informação

Recuperação de Informação

Práticas de Laboratório

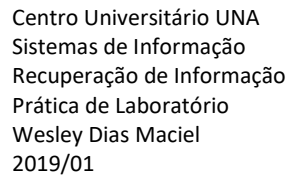
Wesley Dias Maciel

2019/01



Centro Universitário UNA  
Sistemas de Informação  
Recuperação de Informação  
Prática de Laboratório  
Wesley Dias Maciel  
2019/01

# Processador de Consulta



1) Você está recebendo, juntamente com esta prática, uma planilha Excel com um exemplo de cálculo de similaridades no modelo de espaço vetorial e também o projeto da aplicação. Os documentos usados no exemplo são os apresentados nos “slides” empregados na apresentação do conteúdo. A planilha pode ser usada na realização de testes, para validação dos resultados retornados pelo projeto. Ela apresenta os resultados da similaridade seminormalizada e normalizada.

[illegible]

Nessa versão, o projeto apresenta a implementação do processador de consulta. O projeto calcula apenas a similaridade normalizada. Ao iniciar, o projeto insere dados no banco de dados automaticamente. Esses dados da inserção estão no arquivo:

```
src/main/resources/data.sql
```

Para que esses dados sejam carregados automaticamente, foi acrescentado no arquivo `application.properties` a propriedade:

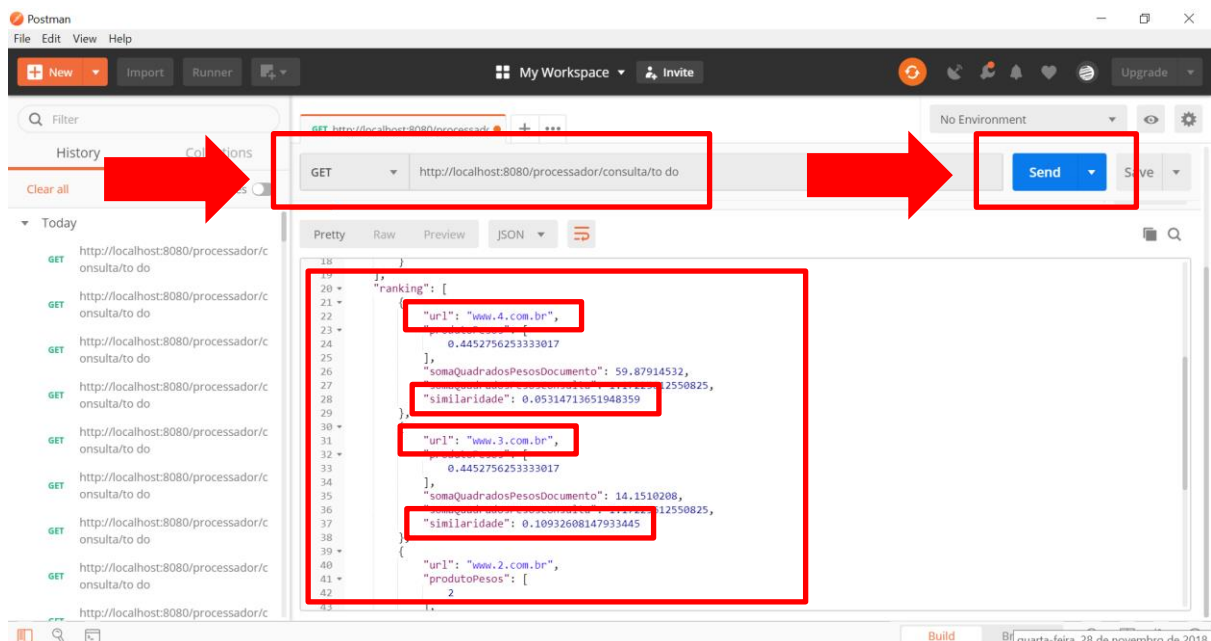
```
#Inicição da base de dados através do conteúdo do arquivo src/main/resources/data.sql.
spring.datasource.initialization-mode = always
```

Esta versão do projeto apresenta algumas modificações em relação ao projeto da prática 16. Por exemplo:



- a) A classe Termo foi renomeada para TermoDocumento.
- b) Foram inseridas as classes: ProcessadorConsulta, Consulta, EntradaRanking, TermoConsulta, IndiceInvertidoRepository, TermoDocumentoRepository, DocumentoService, IndexadorService, ProcessadorConsultaService e TermoDocumentoService.
- c) O atributo “frequenciaNormalizada” foi removido da classe IndiceInvertido, porque o valor do fator  $TF_{i,j}$  foi calculado usando a expressão  $TF_{i,j} = 1 + \log f_{i,j}$  que dispensa o uso da frequência normalizada.

O exemplo abaixo apresenta o ranking retornado para uma consulta:



Analise o código do projeto. Execute o projeto usando o Postman. Em seguida, altere o projeto para que ele:

- 2) Calcule as similaridades com base na visão da consulta. Na criação da visão da consulta é necessário empregar as mesmas operação empregadas na criação da visão dos documentos, por exemplo: remoção de “stop words”, remoção de pontuações, remoção de “tags” HTML, conversão de caracteres para minúsculo, etc...
- 3) Ordene as entradas do “ranking” de resposta em ordem decrescente do valor da similaridade.



- 4) Inclua na API a possibilidade de retornar um “ranking” de resposta com base na similaridade seminormalizada. O administrador do sistema poderá escolher que tipo de similaridade será empregada no processamento da consulta. O usuário comum apenas terá acesso à similaridade normalizada.
- 5) Processe as consultas com base no índice invertido criado para os documentos coletados por seu coletor.
- 6) No projeto, sempre retornar respostas que obedeçam os códigos adequados do protocolo HTTP.
- 7) Analise a API do seu projeto. Sempre que necessário, faça alterações para melhoria da API, adequando-a ao padrão arquitetural REST (Representational State Transfer).

### **Lista de códigos de status HTTP:**

#### **1xx Informativa**

- 100 Continuar
- 101 Mudando protocolos
- 102 Processamento (WebDAV) (RFC 2518)
- 122 Pedido-URI muito longo

#### **2xx Sucesso**

- 200 OK
- 201 Criado
- 202 Aceito
- 203 não-autorizado (desde HTTP/1.1)
- 204 Nenhum conteúdo
- 205 Reset
- 206 Conteúdo parcial
- 207-Status Multi (WebDAV) (RFC 4918)

#### **3xx Redirecionamento**

- 300 Múltipla escolha
- 301 Movido
- 302 Encontrado
- 303 Consulte Outros
- 304 Não modificado
- 305 Use Proxy (desde HTTP/1.1)
- 306 Proxy Switch
- 307 Redirecionamento temporário (desde HTTP/1.1)
- 308 Redirecionamento permanente (RFC 7538[2])

#### **4xx Erro de cliente**

- 400 Requisição inválida
- 401 Não autorizado
- 402 Pagamento necessário
- 403 Proibido
- 404 Não encontrado
- 405 Método não permitido
- 406 Não Aceitável



407 Autenticação de proxy necessária  
408 Tempo de requisição esgotou (Timeout)  
409 Conflito  
410 Gone  
411 comprimento necessário  
412 Pré-condição falhou  
413 Entidade de solicitação muito grande  
414 Pedido-URI Too Long  
415 Tipo de mídia não suportado  
416 Solicitada de Faixa Não Satisfatória  
417 Falha na expectativa  
418 Eu sou um bule de chá  
422 Entidade improcessável (WebDAV) (RFC 4918)  
423 Fechado (WebDAV) (RFC 4918)  
424 Falha de Dependência (WebDAV) (RFC 4918)  
425 coleção não ordenada (RFC 3648)  
426 Upgrade Obrigatório (RFC 2817)  
450 bloqueados pelo Controle de Pais do Windows  
499 cliente fechou Pedido (utilizado em ERPs/VPSS)

**5xx outros erros (erro de servidor)**

500 Erro interno do servidor (Internal Server Error)  
501 Não implementado (Not implemented)  
502 Bad Gateway  
503 Serviço indisponível (Service Unavailable)  
504 Gateway Time-Out  
505 HTTP Version not supported