

Mineração de texto em pedidos de Lei de Acesso à informação - LAI

Packages for this routine

BASE DE DADOS

Importação dos dados

Caminho do projeto

```
PATH = "../proj_eSIC_v10/textmining_pt/DATA/"
```

- Pedidos e-SIC

```
FILE = "DATA/relatorio_pedidos.ods"
db_raw = readODS::read.ods(file = paste0(PATH,FILE), sheet = 1); # dim(db_raw)
dbnames = db_raw[1,]; db_raw = db_raw[-1,];
colnames(db_raw) = c("ProtocoloPedido", "DataRegistro", "DATA_PRAZOATEND", "DESCRI_PEDIDO",
                    "RESUMO_PEDIDO", "DATA_RESPOSTA")
#View(head(db_raw))
LAI = db_raw
```

- Respostas e-SIC

```
FILE1 = "DATA/relatorio_respostas.xlsx"
db1_raw = readxl::read_excel(paste0(PATH,FILE1), sheet = "DADOS", col_names = TRUE);
# dim(db1_raw); names(db1_raw)
colnames(db1_raw) = c("ID", "DATA", "SOLICITACAO", "DIRETORIA", "DATA_RESPOSTA")
#View(head(db1_raw))
LAI1 = db1_raw
```

- Stopwords

```
FILE2 = "DATA/stopwords_PT_FINAL.csv"
stopwords_pt = read.csv(paste0(PATH,FILE2), sep = ';', header = F, encoding = "UTF-8")
stopwords_pt = stopwords_pt[,-2];
cat(paste0("O nosso vetor de stopwords contém ",length(stopwords_pt), " palavras únicas"))
```

```
## O nosso vetor de stopwords contém 605 palavras únicas
```

```
## dim(stopwords_pt); class(stopwords_pt)
stopwords_pt = as.character(stopwords_pt)
stopwords_pt[1:14]
```

```
## [1] ", "      "a"      "à"      "acerca" "adeus"  "agora"  "aí"
## [8] "ainda"  "alem"   "além"   "alguas" "algo"   "algumas" "alguns"
```

- Dicionário > BASE DE DADOS - REAL PRO TEXTO DO TCC

Dicionário de variáveis - PEDIDOS

```
dicionario = "DATA/Dicionario-Dados-Exportacao.txt"
dic_pedidos = read.delim(dicionario, sep = "-", skip = 3, header = FALSE, nrow = 21) %>%
  select(-V1)
```

```
colnames(dic_pedidos) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_pedidos); View(dic_pedidos)
```

Dicionário de variáveis - RECURSOS

```
dic_recursos = read.delim(dicionario, sep = "-", skip = 30, header = FALSE, nrows = 17) %>%
  select(-V1)
colnames(dic_recursos) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_recursos); View(dic_recursos)
```

Dicionário de variáveis - SOLICITANTES

```
dicionario = "DATA/Dicionario-Dados-Exportacao.txt"
dic_solicitantes = read.delim(file = dicionario, sep = "-", skip = 53, header = FALSE, nrows = 10) %>%
  select(-V1)
colnames(dic_solicitantes) = c("Nome das variáveis", "Tipo e descrição da variável")
#dimnames(dic_solicitantes); View(dic_solicitantes)
```

- Anexo 02: Tabela - Dicionário de variáveis da tabela de pedidos

```
dic_pedidos %>%
kable("latex", caption = "Dicionário de variáveis da tabela de pedidos", booktabs = T) %>%
  kable_styling(latex_options = c("striped", "hold_position"), full_width = F) %>%
  #column_spec(1:1, width = "3cm") %>%
  column_spec(2:2, width = "15cm") %>%
landscape()
```

Table 1: Dicionário de variáveis da tabela de pedidos

Nome das variáveis	Tipo e descrição da variável
IdPedido	inteiro: identificador único do pedido (não mostrado no sistema);
ProtocoloPedido	texto(17): número do protocolo do pedido;
OrgaoSuperiorAssociadoaoDestinatario	texto(250): Quando o órgão for vinculado, este campo traz o nome do seu órgão superior;
OrgaoDestinatario	texto(250): nome do órgão destinatário do pedido;
Situacao	texto(200): descrição da situação do pedido;
DataRegistro	Data DD/MM/AAAA HH:MM:SS : data de abertura do pedido;
ResumoSolicitacao	texto(255): resumo do pedido;
DetalhamentoSolicitacao	texto(2048): detalhamento do pedido;
PrazoAtendimento	Data DD/MM/AAAA HH:MM:ss : data limite para atendimento ao pedido;
FoiProrrogado	texto(3) Sim ou Não : informa se houve prorrogação do prazo do pedido;
FoiReencaminhado	texto(3) Sim ou Não: informa se o pedido foi reencaminhado;
FormaResposta	texto(200): tipo de resposta escolhida pelo solicitante na abertura do pedido;
OrigemSolicitacao	texto(50): informa se o pedido foi aberto em um Balcão SIC ou pela Internet;
IdSolicitante	inteiro: identificador único do solicitante (não mostrado no sistema);
CategoriaPedido	texto(200) : categoria do pedido atribuída pelo SIC de acordo com o VCGE (Vocabulário COntrolado do GOverno Eletrônico);
SubCategoriaPedido	texto(200) : subcategoria do pedido atribuída pelo SIC de acordo com o VCGE (Vocabulário COntrolado do GOverno Eletrônico);
NumeroPerguntas	inteiro : número de perguntas feitas no pedido;
DataResposta	Data DD/MM/AAAA HH:MM:SS : data da resposta ao pedido (campo em branco para pedidos que ainda estejam na situação Em Tramitação);
Resposta	texto(8000): resposta ao pedido;
TipoResposta	texto(100) : tipo resposta dada ao pedido (campo em branco para pedidos que ainda estejam na situação Em Tramitação);
ClassificacaoTipoResposta	texto(200): subtipo da resposta dada ao pedido (campo em branco para pedidos que ainda estejam na situação Em Tramitação);

- Anexo 03: Tabela - Dicionário de variáveis da tabela de recursos

```
dic_recursos %>%  
kable("latex", caption = "Dicionário de variáveis da tabela de recursos", booktabs = T) %>%  
  kable_styling(latex_options = c("striped", "hold_position"), full_width = F) %>%  
  #column_spec(1:1, width = "3cm") %>%  
  column_spec(2:2, width = "15cm") %>%  
landscape()
```

Table 2: Dicionário de variáveis da tabela de recursos

Nome das variáveis	Tipo e descrição da variável
IdRecurso	inteiro: identificador único do recurso (não mostrado no sistema);
IdRecursoPrecedente	inteiro: identificador único do recurso que precedeu este (não mostrado no sistema e em branco no caso de Recursos de 1ª Instância e Reclamações);
DescRecurso	texto(8000): descrição do recurso;
IdPedido	inteiro: identificador único do pedido ao qual o recurso pertence (não mostrado no sistema);
IdSolicitante	inteiro: identificador único do solicitante (não mostrado no sistema);
ProtocoloPedido	texto(17): número do protocolo do pedido ao qual o recurso pertence;
OrgaoSuperiorAssociadoaoDestinatario	texto(250): Quando o órgão for vinculado, este campo traz o nome do seu órgão superior;
OrgaoDestinatario	texto(250): nome do órgão destinatário do recurso;
Instancia	texto(80): descrição da instância do recurso;
Situacao	texto(80): descrição da situação do recurso;
DataRegistro	Data DD/MM/AAAA HH:MM:SS : data de abertura do recurso;
PrazoAtendimento	Data DD/MM/AAAA HH:MM:SS : data limite para atendimento ao recurso;
OrigemSolicitacao	texto(50): informa se o recurso foi aberto em um Balcão SIC ou pela Internet;
TipoRecurso	texto(80): motivo de abertura do recurso;
DataResposta	Data DD/MM/AAAA HH:MM:SS : data da resposta ao recurso (campo em branco para recursos que ainda estejam na situação Em Tramitação);
RespostaRecurso	texto(8000): resposta ao recurso;
TipoResposta	texto(80): tipo resposta dada ao recurso (campo em branco para recursos que ainda estejam na situação Em Tramitação);

- Anexo 04: Tabela - Dicionário de variáveis da tabela de solicitantes

```
dic_solicitantes %>%  
kable("latex", caption = "Dicionário de variáveis da tabela de solicitantes", booktabs = T) %>%  
  kable_styling(latex_options = c("striped", "hold_position"), full_width = F) %>%  
  #column_spec(1:1, width = "3cm") %>%  
  column_spec(2:2, width = "15cm") %>%  
landscape()
```

Table 3: Dicionário de variáveis da tabela de solicitantes

Nome das variáveis	Tipo e descrição da variável
IdSolicitante	inteiro: identificador único do solicitante (não mostrado no sistema);
TipoDemandante	texto(15): Pessoa Física ou Pessoa Jurídica;
DataNascimento	Data DD/MM/AAAA : data de nascimento do solicitante;
Sexo	texto(13) : Masculino ou Feminino (em branco para pessoa jurídica);
Escolaridade	texto(200): Escolaridade do solicitante (em branco para pessoa jurídica);
Profissao	texto(200): Profissão do solicitante (em branco para pessoa jurídica);
TipoPessoaJuridica	texto(200): tipo de Pessoa Jurídica do solicitante (em branco para pessoa física)
Pais	texto(200): país de residência do solicitante;
UF	texto(2): UF de residência do solicitante;
Municipio	texto(200): Município de residência do solicitante;