Pré-processamento de textos em R

Micael Filipe

13/02/2020

Linguagem de programação R

A linguagem de programação R pode ser utilizada para desempenhar inumeras tarefas, desde cálculos básicos a, até a realização de análises estatísticas complexas. Além disso, o R também pussui uma gama de recursos para plotagem de gráficos, como personalização de: cor, tipo e tamanho de letra, símbolos, títulos e subtítulos, pontos, linhas, legendas, planos de fundo e entre outros. Os gráficos podem ser usados para obter informações visuais significativas durante a análise de dados ou podem ser exportados em um relatório para apresentações.

Mais que um software que realiza análises estatísticas, R é um ambiente e uma linguagem de programação orientada a objeto. Nele, números, vetores, matrizes, arrays, data frames e listas podem ser armazenados em objetos. A linguagem R pode ser utilizada em diversas áreas e para diversos fins, tais como: pesquisa científica, business analytics, desenvolvimento de software, relatórios statísticos, econometria e análise financeira, ciência sociais, e big data analytics.

Mineração de textos em R

Além das tarefas já citadas anteriormente, a linguagem R pode ser utilizada para mineração de textos. Mineração de textos pode ser definida como uma metodologia para extração de informações a partir de dados textuais, que tem como base o uso de técnicas de processamento de linguagem natural, recuperação de informação e aprendizado de maquina para tratar os dados a fim de filtar apenas os dados que possam ser úteis para o usuário final.

O processo de mineração de textos pode ser resumido em quatro etapas fundamentais:

- **Coleta de documentos**

  Essa etapa tem como ojetivo a coleta de documentos que irão compor o banco de textos a ser analisado.

- **Pré-processamento**

  Nesa etapa os documentos coletados são estruturados de forma padronizada para que os algoritimos que serão utilizados prosteriormente sejma capazes de fazer a manipulação de todos os documentos. Também são definidos os termos e caracteres especiais que não possuem significado relevante (preposições, artigos, pontuação, cabeçalhos, etc.) que serão removidos do conjunto de textos.

- **Extração de conhecimento**

  Essa etapa consiste na extração do conhecimento por meio da aplicação de algoritimos de extração automática de conhecimento. Os termos são agrupados conforme sua similaridade e interação entre si no conjunto de textos.

- **Avaliação e interpretação dos resultados**

  Por fim, os reultados obtidos são avaliados a fim de determinar se os algoritimos reveleram informações relevantes sobre o conjunto de textos.

Pré-processamento

Os pacotes UDpipe e TM foram utilizados para o pré processamento dos textos.

Importação dos pacotes a seren utilizados:

```
options("encoding" = "UTF-8")
options(scipen = 999)
library(stringr)
library(tm)
library(SnowballC)
library(lexiconPT)
library(tidytext)
library(tidyverse)
library(magrittr)
library(stm)
library(ggridges)
library(formattable)
```

Leitura de todos os dados do diretorio para um dataframe:

```
dados<-"/home/micael/R_envs/text-mining/dados/sbrt_txts/dossies"
txtdf<-readtext::readtext(dados, encoding = "latin1")
```

Tentativa de extração dos títulos dos documentos:

```
titulo<-NULL
for (i in 1:length(txtdf$text)){
tx<-strsplit(x=txtdf[i,2], "\n")[[1]][2]
titulo[i]<-tx
}
txtdf$titulo<-titulo
```

Limpeza dos arquivos de texto:

```
txtdf$text<-sub('.*\nConteúdo',"",txtdf$text)
txtdf$text<-sub('.*\nCONTEÚDO',"",txtdf$text)
txtdf$text<-sub('.*\nTítulo',"",txtdf$text)
txtdf$text<-gsub('[1-9][0-9]* Copyright © Serviço Brasileiro de Respostas Técnicas -
SBRT - http://www.sbrt.ibict.br',"",txtdf$text)
txtdf$text<-gsub('[1-9][0-9]* Copyright © Serviço Brasileiro de Respostas Técnicas -
SBRT - http://www.respostatecnica.org.br',"", txtdf$text)
txtdf$text<-gsub('[1-9][0-9]*\nCopyright © Serviço Brasileiro de Respostas Técnicas
 - SBRT - http://www.respostatecnica.org.br',"", txtdf$text)
txtdf$text<-gsub('\nCopyright © Serviço Brasileiro de Respostas Técnicas - SBRT - ht
tp://www.sbrt.ibict.br\n\n[1-9][0-9]*',"", txtdf$text)
txtdf$text<-gsub('Disponível em: ',"",txtdf$text)
txtdf$text<-str_replace_all(txtdf$text, "[^[:alnum:].:,?!;]", " ")
txtdf$text<-gsub("\\s+", " ", str_trim(txtdf$text))
txtdf$text<-gsub('Copyright Serviço Brasileiro de Respostas Técnicas SBRT http: www.
respostatecnica.org.br [1-9][0-9]*',"",txtdf$text)
txtdf$text<-gsub('INTRODUÇÃO',"",txtdf$text)
txtdf$text<-gsub('Introdução',"",txtdf$text)
txtdf$text<-gsub('www.*.br',"",txtdf$text)
txtdf$text<- iconv(txtdf$text, from = "UTF-8", to = "ASCII//TRANSLIT")
```

Palavras que serão removidas dos arquivos de texto:

```
sw_pt_tm <- tm::stopwords("pt") %>% iconv(from = "UTF-8", to = "ASCII//TRANSLIT")
sbrt_sw <- c("http", "senai", "deve","acesso", "brasil", "devem","www.sbrt.ibict.br"
,
             "serviço", "brasileiro", "respostas", "técnicas", "técnico",
             "www.respostatecnica.org.br", "pode", "ser","norma","iso", "kg",
             "fig", "fonte", "sbrt", "abnt", "nbr", "tecnica")
sw_pt_tm <- c(sbrt_sw, sw_pt_tm)
```

Apresenta as 50 palavras mais fequentes de todo o o conjunto de textos:

```
df_palavra <- txtdf %>%
  unnest_tokens(palavra, text) %>%
  filter(!palavra %in% sw_pt_tm)

df_palavra %>%
  count(palavra) %>%
  arrange(desc(n)) %>%
  head(50) %>%
  formattable()
```

palavra	n
agua	4968
producao	4065
processo	3822
1	3237
figura	3146
produtos	3117
produto	2976
forma	2900
2	2626
podem	2497
3	2360
uso	2267
sistema	2198
qualidade	2179
solo	2073
maior	2041
sendo	2032
plantas	2018
temperatura	1977
sobre	1946
tipo	1875

palavra	n
residuos	1831
apos	1784
outros	1684
material	1682
controle	1662
5	1644
durante	1582
4	1581
alem	1538
20	1504
cada	1490
parte	1464
atraves	1460
2007	1434
10	1433
grande	1429
c	1428
bem	1415
al	1401
et	1398
meio	1393
condicoes	1380
area	1341
ambiente	1330
materiais	1326
desenvolvimento	1315
onde	1304
ar	1296
tempo	1293

Faz o processamento do texto:

```
proc <- stm::textProcessor(txtdf$text, metadata = txtdf, language = "portuguese",
                           customstopwords = sw_pt_tm)
```

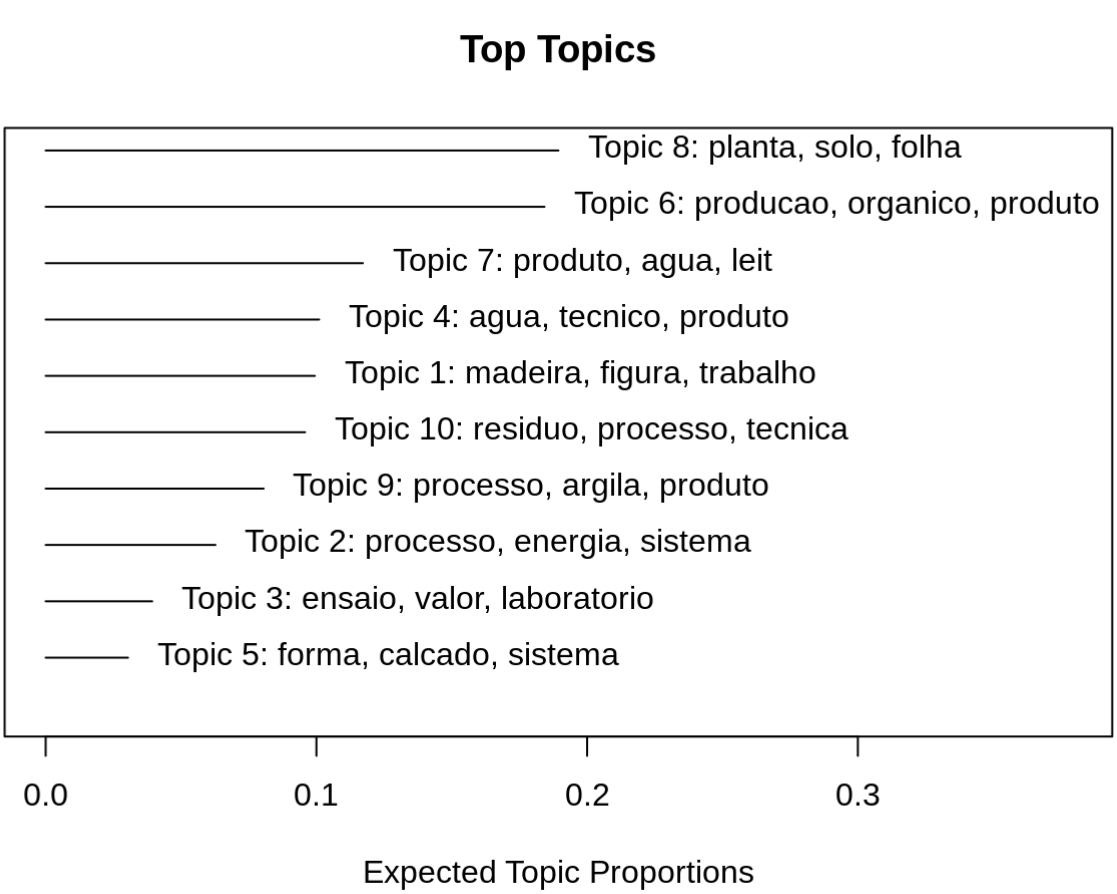
```
## Building corpus...
## Converting to Lower Case...
## Removing punctuation...
## Removing stopwords...
## Remove Custom Stopwords...
## Removing numbers...
## Stemming...
## Creating Output...
```

```
out <- stm::prepDocuments(proc$documents, proc$vocab, proc$meta,
                          lower.thresh = 10)
```

```
## Removing 34905 of 41020 terms (78416 of 361013 tokens) due to frequency
## Your corpus now has 464 documents, 6115 terms and 282597 tokens.
```

Faz a modelagem de tópicos (10 tópicos):

```
fit <- stm(
  documents = out$documents, vocab = out$vocab, data = out$meta,  K = 10,
  max.em.its = 75, init.type = "Spectral", verbose = FALSE
)
plot(fit, "summary")
```



Apresenta as palavras que mais representa cada tópico (FREX)

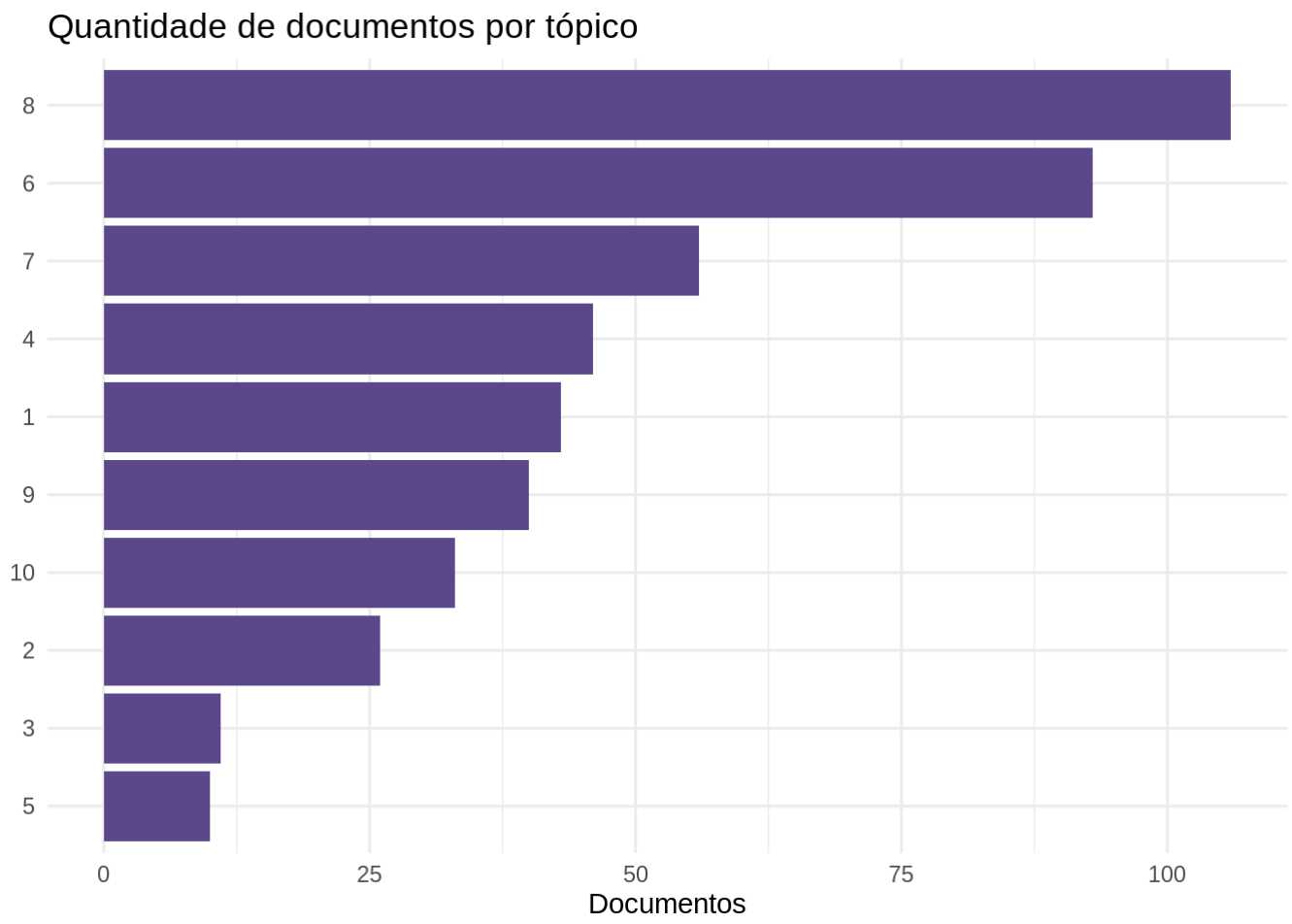
```
stm::labelTopics(fit)
```

```
## Topic 1 Top Words:
##       Highest Prob: madeira, figura, trabalho, cort, peca, produto, maquina
##       FREX: dossi, movei, madeira, cimento, cort, trabalhador, lixa
##       Lift: dossi, mdf, usinagem, macica, lixadeira, layout, confinado
##       Score: dossi, peca, usinagem, madeira, mobiliario, mdf, pastilha
## Topic 2 Top Words:
##       Highest Prob: processo, energia, sistema, agua, temperatura, gas, forma
##       FREX: gas, extrusao, injecao, secador, combustao, solar, etanol
##       Lift: inducao, aquecedor, combustao, ventilador, cabecot, sensor, valvula
##       Score: inducao, extrusao, aquecedor, extrusora, moldagem, film, etanol
## Topic 3 Top Words:
##       Highest Prob: ensaio, valor, laboratorio, amostra, resultado, analis, test
##       FREX: ensaio, laboratorio, retencao, amostra, erro, determinacao, bloco
##       Lift: retencao, chemic, technolog, ensaio, desvio, participant, rubber
##       Score: retencao, ensaio, rubber, participant, medicao, polimero, ceramico
## Topic 4 Top Words:
##       Highest Prob: agua, tecnico, produto, substancia, alimento, podem, uso
##       FREX: tecnico, rdc, desinfeccao, desinfetant, cloro, substancia, anvisa
##       Lift: gram, tecnico, veterinario, febr, rdc, cronica, oral
##       Score: tecnico, rdc, agua, microrganismo, tensoativo, regulamento, anvisa
## Topic 5 Top Words:
##       Highest Prob: forma, calcado, sistema, couro, tratamento, modelo, figura
##       FREX: calcado, lodo, dejeto, bolsa, couro, salto, lagoa
##       Lift: numeracao, dejeto, salto, moda, curtimento, lagoa, calcado
##       Score: numeracao, calcado, lodo, dejeto, couro, suino, efluent
## Topic 6 Top Words:
##       Highest Prob: producao, organico, produto, servico, animai, sistema, organi
ca
##       FREX: servico, animai, cana, organico, agricultura, cogumelo, abelha
##       Lift: pasto, servico, bonotto, degradada, biodiversidad, avicultura, raco
##       Score: servico, cana, raca, certificadora, abelha, cogumelo, mel
## Topic 7 Top Words:
##       Highest Prob: produto, agua, leit, processo, temperatura, acucar, massa
##       FREX: sal, batata, fermentacao, fruta, leit, acucar, fermento
##       Lift: glucos, lactico, vigor, licor, sensori, desnatado, fermentaco
##       Score: vigor, leit, acucar, fermentacao, fruta, carn, queijo
## Topic 8 Top Words:
##       Highest Prob: planta, solo, folha, fruto, sement, plantio, especi
##       FREX: plantio, muda, cultivar, flore, planta, cova, fruto
##       Lift: crisantemo, lagarta, amarelecimento, antracnos, apodrecimento, areno,
besouro
##       Score: fruto, plantio, sement, colheita, cultivar, adubacao, praga
## Topic 9 Top Words:
##       Highest Prob: processo, argila, produto, tipo, temperatura, peca, adesivo
##       FREX: argila, ceramica, adesivo, termoplastico, corant, solda, pigmento
##       Lift: polimerica, ondulado, termoplastico, inchamento, latex, caulim, silic
ato
##       Score: termoplastico, elastomero, argila, solda, ondulado, adesivo, ceramic
o
## Topic 10 Top Words:
##       Highest Prob: residuo, processo, tecnica, produto, ambient, producao, empre
sa
##       FREX: tecnica, residuo, reciclagem, limpa, implementacao, emisso, gestao
##       Lift: sucata, tecnica, gerenci, emisso, conama, minimizacao, josean
##       Score: tecnica, efluent, emisso, reciclagem, implementacao, residuo, conama
```

Apresenta gráfico de documento por tópico:

```
nomes_topicos <- c("1", "2", "3",
                   "4", "5", "6", "7",
                   "8", "9", "10")
maior_prob <- apply(fit$theta, 1, max)
topico_doc <- nomes_topicos[apply(fit$theta, 1, which.max)]

df_topico <- txtdf %>%
  mutate(maior_prob = maior_prob,
         topico = topico_doc)
df_topico %>%
  count(topico) %>%
  mutate(topico = forcats::fct_reorder(topico, n)) %>%
  ggplot(aes(x = topico, y = n)) +
  geom_col(fill = "mediumpurple4") +
  theme_minimal() +
  labs(x = NULL, y = "Documentos",
       title = "Quantidade de documentos por tópico") +
  coord_flip()
```



Cria dataframe com nome do arquivo, titulo e tópico:

```
topico_doc<-df_topico %>%
  group_by(topico)%>%
  arrange(desc(maior_prob))%>%
  select(titulo,doc_id, topico, maior_prob)
```

Apresenta dos 10 primeiros titulos relacionados ao tópico 1:

```
topico_doc%>%
  filter(topico==1)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Sistema de Proteção Coletiva contra Queda de Altura na Indústria da Construção Civil Aledson Damasceno Costa	157.txt	1	0.9569258
Ergonomia	158.txt	1	0.9453324
PROCESSOS DE FABRICAÇÃO DE PROTÓTIPOS DE MÓVEIS	242.txt	1	0.9448476
As condições da falta de segurança dos andaimes como fonte potencial de risco de quedas na construção Civil Aledson Damasceno Costa Rede de Tecnologia da Bahia RETEC/BA	80.txt	1	0.9232797
Pastilhas alisadoras Ivan Leandro Debiasi	48.txt	1	0.9005532
TIPOS DE LAYOUT E SUA APLICAÇÃO NA INDÚSTRIA MOVELEIRA	240.txt	1	0.8792224
Esquadrias em madeira para portas e janelas Cecilia Chicoski da Silva	187.txt	1	0.7966342
Montagem e Instalação de móveis	159.txt	1	0.7670434
Fabricação de shapes de skate Elizabeth Martines	6112.txt	1	0.7334199
Segurança do trabalho em espaços confinados Mônica Belo Nunes	5665.txt	1	0.7323084

Apresenta dos 10 primeiros titulos relacionados ao tópico 2:

```
topico_doc%>%
  filter(topico==2)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Conversão de equipamentos residenciais a gás	53.txt	2	0.9408900
Sistema de aquecimento solar para uso em instalações hidráulicas residenciais	272.txt	2	0.9354074
Gás natural veicular (GNV) Sonia Maria Marques de Oliveira	65.txt	2	0.9212214
Energia térmica na indústria: usos, eficiências e fontes Lothar Hoppe	5716.txt	2	0.8907280
Eficiência Energética em Processos de Combustão ganhos ambientais e econômicos	164.txt	2	0.8646584
Equipamentos e processos de secagem Marina Fernanda Stocco Zempulski Ladislau Nelson Zempulski Instituto de Tecnologia do Paraná	179.txt	2	0.7992745
Sistemas de Exaustão e Ventilação Industrial Renata Martins	141.txt	2	0.7923538
Instalação e utilização adequada de uma câmara escura para processamento de filmes radiográficos Otávio Souza Rocha Liz	264.txt	2	0.7520430
Processo de transformação de plásticos por sopro	5656.txt	2	0.7074919

titulo	doc_id	topico	maior_prob
Energia Eólica	5707.txt	2	0.6996507

Apresenta dos 10 primeiros titulos relacionados ao tópico 3:

```
topico_doc%>%
  filter(topico==3)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Ensaios de Proficiência (ensaios interlaboratoriais) em borracha	176.txt	3	0.9857795
Ensaios de proficiência para laboratórios de controle da qualidade-componentes para	47.txt	3	0.9805683
Materiais de Referência Elastoméricos	271.txt	3	0.9632809
	6.txt	3	0.8310904
Análise das Propriedades Dinâmico-Mecânicas em Compostos de Borracha	268.txt	3	0.7446280
Instrumentalização de um Laboratório para Avaliação do Desempenho, Segurança e Eficiência Energética de Aparelhos Domésticos de Cocção a Gás	270.txt	3	0.6081691
Análises Instrumentais Aplicadas a Materiais Poliméricos	269.txt	3	0.5524763
Produção de Embalagem de Papel	200.txt	3	0.5417856
Reologia escoamento e deformação da matéria	27657.txt	3	0.5051225
Toxicidade em efluentes industriais	5648.txt	3	0.4123845

Apresenta dos 10 primeiros titulos relacionados ao tópico 4:

```
topico_doc%>%
  filter(topico==4)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Água mineral Maria Paula Assis Yamada	223.txt	4	0.8481740
Desinfetante doméstico Allan George A. Jaigobind	265.txt	4	0.8234960
Água potável	9057.txt	4	0.7651539
Fabricação de Cosméticos ea	32.txt	4	0.7604360
Qualidade microbiológica do ar de ambientes condicionados	189.txt	4	0.7553373
Qualidade da água de hemodiálise Ivete Keiko Shimada Coimbra Carmen Etsuko Higaskino Eder José dos Santos Maria Paula Assis Yamada Quelcy Barreiros Correa	216.txt	4	0.6972216

titulo	doc_id	topico	maior_prob
Gestão de Resíduos Sólidos de Saúde Elisabeth Flávia Roberta Oliveira da Motta	61.txt	4	0.6921906
Alimentos para atletas Janaína Szwaidak Marcelino Marlene Szwaidak Marcelino Instituto de Tecnologia do Paraná	27615.txt	4	0.6897966
Segurança no trabalho em laboratórios	5700.txt	4	0.6759809
Tratamento e Controle de Água de Piscina Nelson Alves Góes	114.txt	4	0.6509216

Apresenta dos 10 primeiros titulos relacionados ao tópico 5:

```
topico_doc%>%
  filter(topico==5)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Modelagem de bolsas	15.txt	5	0.9858904
Biossistemas integrados na suinocultura Rogério Moreira de Oliveira	312.txt	5	0.9821754
Modelagem de bolsas em tecido	5649.txt	5	0.9719976
Modelagem Técnica do Calçado Mauri Rubem Schmidt	162.txt	5	0.9431793
Confecção e moda surfwear, beachweare streetwear	322.txt	5	0.8326793
Fôrmas e sistemas de medidas para calçados Elenilton Gerson Berwanger	301.txt	5	0.8001477
	7.txt	5	0.5929000
Processo de fabricação do calçado	169.txt	5	0.5476081
Identificação da microfauna presentes em sistema de tratamento tipo Lodo Ativado em curtume	5677.txt	5	0.5315400
Sumário	27544.txt	5	0.4782103

Apresenta dos 10 primeiros titulos relacionados ao tópico 6:

```
topico_doc%>%
  filter(topico==6)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
Cunicultura	5694.txt	6	0.9563981
Bovinocultura	6105.txt	6	0.9282116
Caprinocultura Orgânica	5662.txt	6	0.9185144
CUNICULTURA ORGÂNICA Rosa Maria Beraldo	5718.txt	6	0.9172908

titulo	doc_id	topico	maior_prob
Colheita e processamento de cogumelos comestíveis e medicinais para comercialização	148.txt	6	0.9051355
Frango Orgânico Lucas José Campanha Ricardo Augusto Bonotto Barboza Universidade Estadual Paulista	5436.txt	6	0.8965738
Estrutiocultura (criação de avestruz) Lucimar Tunes Leite Instituto de Tecnologia do Paraná	311.txt	6	0.8740076
Wilton Neves Brandão Rede de Tecnologia da Bahia RETEC/BA	122.txt	6	0.8724550
Criação de Codornas Eduardo Henrique da Silva F. Matos Centro de Apoio ao Desenvolvimento Tecnológico da Universidade de Brasília CDT/UnB	192.txt	6	0.8659494
Indicações Geográficas aplicadas ao setor de gemas e joias	27694.txt	6	0.8590737

Apresenta dos 10 primeiros titulos relacionados ao tópico 7:

```
topico_doc%>%
  filter(topico==7)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
ABRIL/2007	102.txt	7	0.9949206
Produtos de Soja Lilian Guerreiro REDETEC Rede de Tecnologia do Rio de Janeiro Dezembro / 2006	28.txt	7	0.9839339
Doce em Pasta e em Calda Renata Martins	234.txt	7	0.9766675
Produção de Polpa de Fruta Congelada e Suco de Frutas	117.txt	7	0.9748032
Fabricação de iogurtes Noely Forlin Robert	320.txt	7	0.9561191
Processamento de Frutas Cristalizadas Eduardo Henrique da Silva Figueiredo Matos Centro de Apoio ao Desenvolvimento Tecnológico	109.txt	7	0.9484876
PANIFICAÇÃO Lilian Guerreiro REDETEC Rede de Tecnologia do Rio de janeiro	27.txt	7	0.9309640
Processamento de conservas e temperos	55.txt	7	0.8992215
MASSAS ALIMENTÍCIAS Lilian Guerreiro	26.txt	7	0.8702365
T ecnologia do pescado Ingrid Vieira Machado de Moraes Rede de Tecnologia do Rio de Janeiro	58.txt	7	0.8630480

Apresenta dos 10 primeiros titulos relacionados ao tópico 8:

```
topico_doc%>%
  filter(topico==8)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob

titulo	doc_id	topico	maior_prob
Cultivo de Lírio de corte e de vaso Glecimar Fabrin Pozza	326.txt	8	0.9930252
Cultivo de Violeta-africana José dos Anjos Soares Júnior	325.txt	8	0.9743225
Cultivo de Violeta-africana José dos Anjos Soares Júnior	327.txt	8	0.9743225
Cultivo do Morango Nilva Chaves	152.txt	8	0.9727960
Cultivo Comercial de Anturium (Anthurium Andreanum Linl)	323.txt	8	0.9726164
ADUBAÇÃO VERDE Ivo Pessoa Neves	108.txt	8	0.9695336
Cultivo comercial de Palma de Santa Rita	83.txt	8	0.9576792
CULTIVO E PROCESSAMENTO DE PIMENTA	124.txt	8	0.9523446
CULTIVO DE ARROZ IVO PESSOA NEVES Rede de Tecnologia da Bahia-RETEC/BA	256.txt	8	0.9467304
CULTIVO DE CENOURA IVO PESSOA NEVES	254.txt	8	0.9451062

Apresenta dos 10 primeiros titulos relacionados ao tópico 9:

```
topico_doc%>%
  filter(topico==9)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob
ADESIVOS	113.txt	9	0.9540590
Argila Propriedades e utilizações Lucia Helena de Araújo Jorge	5661.txt	9	0.9328864
Elastômeros Termoplásticos Fabiane Trombetta SENAI-RS	156.txt	9	0.9292553
Argila Propriedades e utilizações Lucia Helena de Araújo Jorge	5687.txt	9	0.9179109
Fosfatização Ladislau Nelson Zempulski Marina Fernanda Stocco Zempulski Instituto de Tecnologia do Paraná	166.txt	9	0.9124704
Fabricação de peças em fibra de vidro (compósitos)	118.txt	9	0.9084903
Fabricação de Artefatos de Látex	174.txt	9	0.8908625
Cerâmica Ivo Mezzadri Filho Instituto de Tecnologia do Paraná	107.txt	9	0.8781456
Serigrafia Marcelo Shiniti Uchimura Instituto de Tecnologia do Paraná	167.txt	9	0.8470397
Fabricação de Tintas Cristine Canaud	138.txt	9	0.8350176

Apresenta dos 10 primeiros titulos relacionados ao tópico 10:

```
topico_doc%>%
  filter(topico==10)%>%
  head(10)%>%
  formattable()
```

titulo	doc_id	topico	maior_prob

titulo	doc_id	topico	maior_prob
Produção mais Limpa no setor gráfico Joseane Machado de Oliveira SENAI-RS	49.txt	10	0.9895548
Minimização de efluentes e resíduos na indústria galvânica	274.txt	10	0.9450772
Produção mais Limpa no Setor de Peças Brutas	245.txt	10	0.9423215
Produção mais Limpa no setor de panificação Joseane Machado de Oliveira SENAI-RS	50.txt	10	0.9148629
Produção mais Limpa no Setor Madeira e Mobiliário	243.txt	10	0.8940143
!"#$ % &’ ( )**+	275.txt	10	0.8684979
Avaliação de aspectos e impactos ambientais, legislação ambiental e gerenciamento de resíduos	161.txt	10	0.8328841
Produção mais Limpa no Setor Plástico	246.txt	10	0.8319060
Gestão de Resíduos Sólidos em Canteiros de Obras Raquel Naves Blumenschein	43.txt	10	0.8112102
GERENCIAMENTO DE RESÍDUOS EM OFICINAS AUTOMOTIVAS	248.txt	10	0.7750688