Teste de Text mining: análise humana e de máquina do Relatório P-36

Viviane Schneider

Última versão: 26/01/2021

* [1 Contexto do estudo](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#contexto-do-estudo)
  + [1.1 Objetivo do estudo](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#objetivo-do-estudo)
  + [1.2 Pergunta de pesquisa](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#pergunta-de-pesquisa)
  + [1.3 Métodos, técnicas e materiais](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#m%C3%A9todos-t%C3%A9cnicas-e-materiais)
* [2 Análise do documento P-36](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#an%C3%A1lise-do-documento-p-36)
  + [2.1 Análise do contexto feita por humano](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#an%C3%A1lise-do-contexto-feita-por-humano)
    - [2.1.1 Local do acidente](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#local-do-acidente)
    - [2.1.2 Agente investigador](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#agente-investigador)
    - [2.1.3 Objetivo da investigação](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#objetivo-da-investiga%C3%A7%C3%A3o)
    - [2.1.4 Causas mais prováveis](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#causas-mais-prov%C3%A1veis)
    - [2.1.5 Possíveis não coonformidades](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#poss%C3%ADveis-n%C3%A3o-coonformidades)
    - [2.1.6 Possíveis causas a partir de fatores humanos](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#poss%C3%ADveis-causas-a-partir-de-fatores-humanos)
    - [2.1.7 Vitmas fatais](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#vitmas-fatais)
    - [2.1.8 Eventos de falhas](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#eventos-de-falhas)
  + [2.2 Identificação de tipo de acidentes em relação com eventos causais e fatores humanos](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-tipo-de-acidentes-em-rela%C3%A7%C3%A3o-com-eventos-causais-e-fatores-humanos)
    - [2.2.1 Explosões](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#explos%C3%B5es)
    - [2.2.2 Naufrágio](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#naufr%C3%A1gio)
    - [2.2.3 Recomendações a partir das lições aprendidas](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#recomenda%C3%A7%C3%B5es-a-partir-das-li%C3%A7%C3%B5es-aprendidas)
    - [2.2.4 Conclusões gerais sobre análise feita por humano](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#conclus%C3%B5es-gerais-sobre-an%C3%A1lise-feita-por-humano)
  + [2.3 Analise do contexto feita por máquina](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#analise-do-contexto-feita-por-m%C3%A1quina)
    - [2.3.1 Conclusões sobre análise humana do contexto](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#conclus%C3%B5es-sobre-an%C3%A1lise-humana-do-contexto)
  + [2.4 Identificação de frases com fatores humanos feita por máquina](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-frases-com-fatores-humanos-feita-por-m%C3%A1quina)
    - [2.4.1 Identificação de frases com termos “conformidade” e “conformidades”](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-frases-com-termos-conformidade-e-conformidades)
    - [2.4.2 Identificação de frases com termos “procedimentos”, “procedimento” e “regras”](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-frases-com-termos-procedimentos-procedimento-e-regras)
    - [2.4.3 Identificação de frases com termos “gestão”, “operacional”](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-frases-com-termos-gest%C3%A3o-operacional)
    - [2.4.4 Identificação de frases com termos “risco”](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#identifica%C3%A7%C3%A3o-de-frases-com-termos-risco)
* [3 Conclusões gerais](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#conclus%C3%B5es-gerais)
* [4 Referências](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#refer%C3%AAncias)
* [5 Apêndice A - Análise exploratória do contexto por máquina](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#ap%C3%AAndice-a---an%C3%A1lise-explorat%C3%B3ria-do-contexto-por-m%C3%A1quina)
  + [5.0.1 Palavras mais frequentes](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#palavras-mais-frequentes)
  + [5.0.2 Rede de palavras](https://rstudio-pubs-static.s3.amazonaws.com/717154_f647a0bff6fc4c90803514784b565d32.html#rede-de-palavras)

1 Contexto do estudo

Este é um estudo exploratório de técnicas de Text Mining. O intuito é descobrir que técnicas que podem ser utilizadas para identificar fatores humanos em relatórios de acidentes fornecidos pela ANP.

1.1 Objetivo do estudo

Formação de uma base de dados que possa ser utilizada para estimar a probabilidade de ocorrência de um acidente a partir de fatores humanos identificados. Para realizar este estudo foram feitas análises com algoritmos computacionais e análises feitas por humanos. Desta forma busca-se avaliar como um humano identifica determinadas palavras e frases em um texto, para então reproduzir o método em um algoritmo. Como resultado final, espera-se a criação de um método que possa identificar e estruturar em uma base de dados os fatores humanos descritos no modelo HF2.

1.2 Pergunta de pesquisa

Como identificar fatores humanos relacionados ao modelo HF2 nos relatórios de investigação de acidentes da ANP?

1.3 Métodos, técnicas e materiais

**Descrever método ao finalizar as analises**

2 Análise do documento P-36

2.1 Análise do contexto feita por humano

Quando lemos o contexto do relatório temos as seguintes informações:

2.1.1 Local do acidente

Acidente ocorrido com a plataforma P-36, instalada no campo de Roncador, na Bacia de Campos.

2.1.2 Agente investigador

Agência Nacional do Petróleo (ANP) e a Diretoria de Portos e Costas (DPC) do Comando da Marinha do Brasil constituíram uma comissão conjunta de investigação

2.1.3 Objetivo da investigação

O objetivo era examinar as causas do acidente e, em conseqüência, adquirir subsídios para implementar medidas corretivas e regulamentações visando à melhoria das práticas e procedimentos operacionais na execução de atividades marítimas de exploração e produção de petróleo e gás natural.

2.1.4 Causas mais prováveis

A análise das causas mais prováveis do acidente permitiu identificar o evento crítico como sendo a operação de esgotamento de água do tanque de drenagem de emergência da coluna de popa bombordo, iniciada na noite do dia 14 de março de 2001.

A água contaminada com resíduos oleosos presente no tanque seria bombeada para o manifolde (header) de produção da plataforma que recebe o fluxo de petróleo e gás natural proveniente dos poços produtores. Daí, juntamente com produção de hidrocarbonetos, escoaria para a planta de processo.

Entretanto, dificuldades operacionais para a partida da bomba de esgotamento desse tanque permitiram que houvesse fluxo reverso de óleo e gás pelas linhas de escoamento dos tanques e sua entrada no outro tanque (popa boreste) através de válvula presumivelmente danificada ou parcialmente aberta. A partida da bomba após 54 minutos fez diminuir o fluxo reverso de hidrocarbonetos e a água bombeada passou a entrar no tanque de popa boreste.

A pressurização contínua deste tanque levou a seu rompimento mecânico cerca de duas horas após o início da operação de esgotamento do outro tanque, caracterizando o evento relatado como sendo a primeira explosão, ocorrido às 0 h 22 min do dia 15 de março de 2001.

Os fluidos do tanque rompido e de linhas e demais equipamentos também danificados passaram a ocupar o compartimento do quarto nível da coluna. Houve escapamento de gás para os conveses superiores através de aberturas nesse compartimento e por linhas de suspiro e ventilação rompidas. Cerca de 20 minutos após o rompimento do tanque, houve a explosão do gás que atingira a área do convés do tank top e segundo convés junto à coluna.

Estava, assim, caracterizado o evento relatado como a ocorrência da segunda explosão, quando foram vitimadas onze pessoas da brigada de incêndio da plataforma.

2.1.5 Possíveis não conformidades

A análise realizada permitiu que fossem identificadas várias não conformidades relativas a procedimentos regulamentares de operação, manutenção e projeto, destacando-se aquelas referentes à movimentação freqüente de água nos tanques de drenagem de emergência, à operação de esgotamento do tanque de popa bombordo e à classificação da área de risco em torno desses tanques.

O colapso mecânico do tanque de drenagem de emergência de popa boreste, seguido imediatamente pelo rompimento da linha de recalque de água salgada que passava pelo quarto nível, iniciou o alagamento da coluna. A migração de água para a parte inferior da coluna se deu quando a água no compartimento do quarto nível atingiu os dampers do sistema de ventilação que deveriam fechar automaticamente; porém, devido a falha no funcionamento de seus atuadores, permitiram a passagem de fluidos.

A quantidade de líquido no interior da coluna e em parte do flutuador provocou o adernamento da plataforma que foi intensificado com a progressão da água para o tanque de lastro da coluna de popa boreste e para a caixa de estabilidade contígua. Esses espaços foram inundados porque as elipses de acesso aos mesmos haviam sido deixadas abertas desde o dia anterior ao acidente para possibilitar a inspeção do reparo de trinca verificada na caixa de estabilidade.

Para compensar a inclinação da plataforma ocasionada pela entrada de água na coluna avariada, passou-se a admitir água no tanque de lastro situado no lado diametralmente oposto.

Essa medida acelerou o aumento indesejável do calado da plataforma. A submersão contínua foi intensificada pelo alagamento da coluna avariada, da inundação do tanque de lastro de popa boreste, da caixa de estabilidade contígua e da admissão deliberada de água de lastro nos tanques de proa bombordo.

A evacuação de 138 pessoas consideradas não essenciais às operações de emergência foi iniciada à 1 h 44 min de 15 de março e durou cerca de 2 h 30 min, tendo sido utilizados guindaste e cesta de transferência para tal finalidade. Às 6 h 03 min do mesmo dia, após esgotadas as possibilidades de manter a plataforma nivelada, a equipe que permanecera a bordo para operações de emergência abandonou a unidade. A análise dos eventos marcantes relativos ao alagamento da plataforma conduziu à identificação de várias não conformidades quanto a procedimentos regulamentares de operação e manutenção, destacando-se a não observância de estanqueidade e compartimentagem em áreas críticas para a preservação da estabilidade da unidade.

Além disso, cabe destacar a ineficácia das ações para conter o alagamento ou efetuar o desalagamento antes da plataforma ser inteiramente abandonada.Após o abandono da plataforma, foram efetuadas diversas tentativas de salvamento da unidade, particularmente a injeção de nitrogênio e ar comprimido nos compartimentos alagados para expulsão da água.

Contudo, não houve êxito em manter a unidade estabilizada e sua submersão lenta e progressiva teve continuidade até às 11 h e 40 min do dia 20 de março quando foi a pique.

2.1.6 Possíveis causas a partir de fatores humanos

A análise efetuada pela Comissão de Investigação ANP/DPC permitiu constatar deficiências no sistema de gestão operacional das atividades marítimas de petróleo e gás natural da Petrobras, na condução das atividades específicas à plataforma P-36, caracterizadas por não conformidades identificadas quanto a procedimentos regulamentares de operação e manutenção.

Além disso, os planos de contingência para acidentes de grande proporção e os esquemas de resposta a emergência de grande risco necessitam ser imediatamente aprimorados, bem como a revisão de critérios de projetos de engenharia em unidades flutuantes de produção para assegurar maior proteção intrínseca.

O acidente foi causado por uma série de fatores que, isoladamente, não seriam suficientes para determiná-lo. O exame desses fatores levou à classificação de alguns deles como críticos e determinantes, sendo o critério para essa classificação o de que a ausência de qualquer um deles interromperia a seqüência dos eventos que resultaram no desfecho verificado.

2.1.7 Vitimas fatais

Acidente com 11 vitimas fatais

2.1.8 Eventos de falhas

O Encadeamento de eventos ocasionou o acidente, sendo que se um desses eventos tivesse sido barrado o acidente provavelmente não ocorreria com tanta gravidade. Foram relatados os seguintes eventos:

2.2 Identificação de tipo de acidentes em relação com eventos causais e fatores humanos

Nesta identificação foram escolhidos os termos que aparecem no sumário executivo e que apareceram também na análise feita por máquina.

2.2.1 Explosões

2.2.1.1 Evento 1 - Movimentação freqüente de água nos tanques de drenagem de emergência

* **deficiência de projeto** foi constatada: • vulnerabilidade do esquema de ligação dos tanques de drenagem de emergência com o manifolde de produção por não apresentar nenhuma proteção adicional ou redundância no caso de falha simples de uma das válvulas de admissão dos tanques. (PLANEJAMENTO)
* não conformidade quanto a **procedimentos de projeto**: • inadequação da classificação da área em torno do tanque de drenagem de emergência. (PLANEJAMENTO)
* não conformidade quanto a procedimentos regulamentares de **resposta a emergências**, particularmente no que tange a:• sistema de **comunicação** e **coordenação deficientes entre a equipe de resposta à emergência e o comando da plataforma**. (COMUNICAÇÃO) (RESPOSTA)
* A principal não conformidade identificada diz respeito ao armazenamento de grande quantidade de água contaminada nos tanques de drenagem de emergência durante parte considerável do período em que a plataforma esteve em produção, **contrariando o Manual de Operações da Planta de Processo da Plataforma** (Operating Manual – Process – ET 3010.38-1200-941-AMK-924 e DE-3010.38-5336-943-AMK-033) (PROCEDIMENTOS) linha 6 condicoes de trabalho
* não conformidades relativas a **procedimentos regulamentares de operação e manutenção**: • erros sistemáticos na sondagem volumétrica manual e inoperância dos indicadores de nível dos tanques de drenagem de emergência; • entupimento do vaso de dreno aberto, que recebe a água que escoa pelas bandejas dos equipamentos da plataforma.(PROCEDIMENTOS)

2.2.1.2 Manutenção no tanque de drenagem de emergência de popa boreste

* não conformidade relativa a **procedimentos regulamentares de operação** pode ser caracterizada: • Isolamento da linha de suspiro do tanque sem que sua linha de admissão fosse também isolada. (PROCEDIMENTOS)

2.2.1.3 Operação de esgotamento do tanque de drenagem de emergência de popa bombordo

* não conformidades quanto a **procedimentos regulamentares de operação** e **manutenção**: • operação realizada **sem a supervisão do Coordenador da Plataforma ou do Supervisor de Produção** (não foi possível constatar se a válvula na linha do manifolde de produção, que, conforme depoimentos, exigia senha para sua abertura controlada, fora aberta com autorização do Coordenador da Plataforma); • falha mecânica ou fechamento incompleto da válvula de admissão do tanque de boreste. (PROCEDIMENTOS)

2.2.2 Naufrágio

2.2.2.1 Alagamento da coluna e flutuador

* não conformidades relativas a **procedimentos de manutenção e operação**, conforme relatado a seguir: • falha no funcionamento dos dampers do sistema de ventilação da coluna, decorrente de defeito em seus atuadores eletro-hidráulicos;

• elipses de acesso ao tanque de lastro de popa boreste e à caixa de estabilidade contígua mantidas abertas por tempo além do necessário à realização de inspeção e reparo, alterando a compartimentagem da plataforma considerada nos estudos de estabilidade intacta e em avaria, **contrariando os procedimentos do Manual de Operação** – Item Casco e Estrutura – Estanqueidade à água; Embora não possa ser caracterizada como uma não conformidade, a utilização do sistema fail set para a válvula da caixa de mar impossibilitou qualquer atuação do operador no sentido de modificar seu estado após a falha, uma vez que o **sistema não dispunha de alternativas que permitissem contornar a restrição imposta.** (SISTEMA, MANUTENÇÃO E OPERAÇÃO, PROCEDIMENTOS)

2.2.2.2 Admissão de água de lastro em proa bombordo

* não conformidades No que tange ao controle de estabilidade de uma unidade flutuante em condições de avaria, as seguintes não conformidades puderam ser identificadas: • **ineficácia das ações para conter** o alagamento na coluna avariada, efetuar o seu desalagamento, ou promover a movimentação de água de lastro entre colunas não avariadas; (RESPOSTA),

• **coordenação e treinamento deficientes de pessoal nas ações de controle de estabilidade em emergência**. A seguinte **não conformidade quanto a procedimentos de manutenção** também foi identificada: • existência de duas bombas de recalque de água salgada fora de operação por **problemas de manutenção**. (MANUTENÇÃO, COORDENAÇÃO, TREINAMENTO)

2.2.2.3 Submersão contínua da plataforma

* O fato do tanque de lastro e caixa de estabilidade mencionados terem sido mantidos abertos caracterizou uma **não conformidade crítica relativa a procedimentos regulamentares de operação em áreas que exigem estanqueidade e compartimentagem**, em desacordo com o Manual de Operação (MA-3010.38-1320-915-NBD-909-01) – Item de Estabilidade em Avaria, constituindo causa **determinante para o naufrágio**. (PROCEDIMENTOS DE OPERAÇÃO)
* **a ineficácia das ações para conter** o alagamento ou efetuar o desalagamento pode ser identificado como uma não conformidade relativa a **procedimentos operacionais** de controle de estabilidade em avaria. (RESPOSTA,PROCEDIMENTOS DA OPERAÇÃO)

2.2.3 Recomendações a partir das lições aprendidas

* **Aprimoramento do sistema de gestão operacional**: Revisão e aplicação do sistema de gestão de modo a assegurar estrita observância de procedimentos regulamentares , inclusive efetuando a revisão da definição de responsabilidades relativas à manutenção, operação e segurança.(GESTÃO OPERACIONAL)
* **Revisão de critérios de projeto**: Efetuar a revisão de regras e procedimentos aplicáveis a projetos de forma a assegurar proteção intrínseca de sistemas e componentes críticos de unidades marítimas. (REVISÃO, REGRAS)
* **Classificação de áreas de risco**: Estabelecer critérios adicionais para aplicação simultânea de normas relativas à classificação de Zonas de Risco. (GESTÃO DE RISCO)
* **Ações simultâneas de comissionamento, manutenção e operação**: Estabelecer critérios para identificar limites para atividades de comissionamento concomitantes com a operação e manutenção de unidades marítimas. (MANUTENÇÃO)
* **Dimensionamento e capacitação de pessoal** Reavaliar o dimensionamento e a qualificação das equipes de operação e manutenção de unidades marítimas, bem como daquelas responsáveis pela resposta a emergências de grande risco.(CAPACITAÇÃO)
* **Gerenciamento de projetos de conversão de unidades**. Implementar regras e procedimentos para compatibilizar sistemas originais e as alterações de projeto de modo a garantir a segurança operacional e a proteção ambiental. (PROCEDIMENTOS, REGRAS)
* **Código Internacional de Gerenciamento de Segurança**: Avaliar a conveniência de antecipar a adoção do Código Internacional de Gerenciamento de Segurança, aprovado pela Resolução A 741(18) da Organização Marítima Internacional (IMO), para unidades marítimas.(REGRAS E NORMAS)

**Resposta a emergências de grande risco**: Elaboração de plano de emergência e implementação de esquema de resposta a situações que envolvam grande risco à salvaguarda de vidas humanas, segurança da navegação, proteção ambiental e propriedade. (RESPOSTA)

2.2.4 Conclusões gerais sobre análise feita por humano

É possível identificar a correlação das falhas com fatores humanos apontados nos relatórios. A partir deste relatório foi possível identificar correlações que poderão ser utilizadas na formação de vetores de busca para fatores humanos, além da formação de uma base de dados que possa correlacionar eventos tipos de acidentes, eventos causais, fatores humanos e ações de resiliência (a partir das recomendações dos relatórios)

2.3 Analise do contexto feita por máquina

Esta análise feita por máquina busca identificar se é possível chegar a alguma conclusão a partir dos resultados trazidos pela máquina. Todas as análises e códigos fontes, e gráficos estão detalhados no apêndice A. Alguns códigos fonte foram omitidos no documento para facilitar a leitura.

2.3.1 Conclusões sobre análise humana do contexto

2.3.1.1 Conclusões sobre análisa partir da rede de palavras

Há diversos clusteres de palavras, dos quais destaca-se:

1. Gás Natural, petróleo, liberado.
2. Tanque devido ao acesso.
3. lastro avaria seguida consequencia contigua estabilidade em decorrencia torno proa. Bombordo popa boreste anexo. Aguá, salgada, contaminada, nos tanques, os eventos marcantes compartimentos alagados.
4. Evento relatado sendo bem como por seguinte. Não conformidades quanto conformidade criticas. Foi plataforma foram.
5. regulamentares, procedimentos operacionais.
6. Alagamento ou efetuar
7. Efetuada análise realizada pela entrada comissão.
8. Gestão operacional.
9. Risco grande
10. mesmo dia
11. tank top
12. dreno aberto
13. Manutenção operação
14. Segundo convés principal
15. Linha na coluna avariada
16. continua submersão
17. ar comprimido
18. Dreno aberto 19: Linha na coluna avariada
19. grande risco

2.3.1.2 Conclusão sobre palavras mais frequentes

O acidente ocorrido em uma plataforma parece ser proveniente de uma emergência que ocorreu envolvendo equipamentos como boreste, popa, caixa de agúa, em uma coluna, talvez com a ocorrência de gas, água e tanque. Parece ter ocorrido uma alagamento e explosão, palavras que são bastante mensionadas. Quanto aos fatores humanos, parece haver ocorrido problemas de manutenção e não conformidade, problemas de procedimentos.

2.4 Identificação de frases com fatores humanos feita por máquina

Nesta etapa buscou-se identificar as frases que contém os fatores humanos identificados, os quais mantém relação com o modelo HF2.

Para realizar esta análise buscou-se resgatar as frases que contém os seguintes termos identificados no passo exploratório:

2.4.1 Identificação de frases com termos “conformidade” e “conformidades”

Frases identificadas:

# Tidying text

Texto <- arquivoPdf %>%

read\_pdf()

Texto$text[str\_detect(

Texto$text,

pattern = "conformidades|conformidade"

)]

## [1] "A análise realizada permitiu que fossem identificadas várias não conformidades relativas a"

## [2] "identificação de várias não conformidades quanto a procedimentos regulamentares de"

## [3] "conformidades identificadas quanto a procedimentos regulamentares de operação e"

## [4] "a identificação de não conformidades com regras e procedimentos regulamentares de"

## [5] "a caracterização das não conformidades constatadas."

## [6] "Identificação de não conformidades"

## [7] "A principal não conformidade identificada diz respeito ao armazenamento de grande"

## [8] "Além disso, foram identificadas as seguintes não conformidades relativas a procedimentos"

## [9] "Identificação de não conformidades"

## [10] "A seguinte não conformidade relativa a procedimentos regulamentares de operação pode ser"

## [11] "Identificação de não conformidades"

## [12] "A operação acima relatada caracterizou-se como uma não conformidade crítica relativa a"

## [13] "Foram também identificadas as seguintes não conformidades quanto a procedimentos"

## [14] "Embora não possa ser caracterizada como uma não conformidade, a seguinte deficiência de"

## [15] "Identificação de não conformidades"

## [16] "A análise efetuada conduziu à identificação da seguinte não conformidade quanto a"

## [17] "Identificação de não conformidades"

## [18] "A análise efetuada conduziu à identificação de não conformidade quanto a procedimentos"

## [19] "como a caracterização das não conformidades constatadas. 11"

## [20] "Identificação de não conformidades"

## [21] "Foram constatadas não conformidades relativas a procedimentos de manutenção e operação,"

## [22] "Embora não possa ser caracterizada como uma não conformidade, a utilização do sistema fail"

## [23] "Identificação de não conformidades"

## [24] "seguintes não conformidades puderam ser identificadas:"

## [25] "A seguinte não conformidade quanto a procedimentos de manutenção também foi identificada:"

## [26] "Identificação de não conformidades"

## [27] "caracterizou uma não conformidade crítica relativa a procedimentos regulamentares de"

## [28] "ser identificado como uma não conformidade relativa a procedimentos operacionais de controle"

## [29] "não conformidades quanto a procedimentos regulamentares de operação, manutenção e"

## [30] "sistema de ventilação da coluna de popa boreste caracterizou-se como uma não conformidade"

Texto$text[str\_detect(

Texto$text,

pattern = "relativa a|relativas|regulamentares de|tange a:"

)]

## [1] "A análise realizada permitiu que fossem identificadas várias não conformidades relativas a"

## [2] "procedimentos regulamentares de operação, manutenção e projeto, destacando-se aquelas"

## [3] "identificação de várias não conformidades quanto a procedimentos regulamentares de"

## [4] "conformidades identificadas quanto a procedimentos regulamentares de operação e"

## [5] "a identificação de não conformidades com regras e procedimentos regulamentares de"

## [6] "Além disso, foram identificadas as seguintes não conformidades relativas a procedimentos"

## [7] "regulamentares de operação e manutenção:"

## [8] "A seguinte não conformidade relativa a procedimentos regulamentares de operação pode ser"

## [9] "A operação acima relatada caracterizou-se como uma não conformidade crítica relativa a"

## [10] "procedimentos regulamentares de operação e processo, constituindo-se em causa"

## [11] "regulamentares de operação e manutenção:"

## [12] "regulamentares de resposta a emergências, particularmente no que tange a:"

## [13] "Foram constatadas não conformidades relativas a procedimentos de manutenção e operação,"

## [14] "caracterizou uma não conformidade crítica relativa a procedimentos regulamentares de"

## [15] "ser identificado como uma não conformidade relativa a procedimentos operacionais de controle"

## [16] "não conformidades quanto a procedimentos regulamentares de operação, manutenção e"

## [17] "responsabilidades relativas à manutenção, operação e segurança."

## [18] "Estabelecer critérios adicionais para aplicação simultânea de normas relativas à classificação"

2.4.2 Identificação de frases com termos “procedimentos”, “procedimento” e “regras”

Frases identificadas:

# Tidying text

Texto <- arquivoPdf %>%

read\_pdf()

Texto$text[str\_detect(

Texto$text,

pattern = "procedimentos|procedimento|regras"

)]

## [1] "procedimentos operacionais na execução de atividades marítimas de exploração e produção"

## [2] "procedimentos regulamentares de operação, manutenção e projeto, destacando-se aquelas"

## [3] "identificação de várias não conformidades quanto a procedimentos regulamentares de"

## [4] "conformidades identificadas quanto a procedimentos regulamentares de operação e"

## [5] "das práticas e procedimentos operacionais referentes a atividades marítimas de exploração e"

## [6] "a identificação de não conformidades com regras e procedimentos regulamentares de"

## [7] "Além disso, foram identificadas as seguintes não conformidades relativas a procedimentos"

## [8] "ser indevidamente pressurizado. Tal procedimento somente poderia ter sido efetuado após"

## [9] "que tal procedimento foi determinante para permitir a pressurização posterior do tanque e sua"

## [10] "A seguinte não conformidade relativa a procedimentos regulamentares de operação pode ser"

## [11] "procedimentos regulamentares de operação e processo, constituindo-se em causa"

## [12] "do manifolde de produção, o procedimento regulamentar de esgotamento de água prescrevia"

## [13] "Foram também identificadas as seguintes não conformidades quanto a procedimentos"

## [14] "procedimentos de projeto:"

## [15] "A análise efetuada conduziu à identificação de não conformidade quanto a procedimentos"

## [16] "Foram constatadas não conformidades relativas a procedimentos de manutenção e operação,"

## [17] "estabilidade intacta e em avaria, contrariando os procedimentos do Manual de"

## [18] "A seguinte não conformidade quanto a procedimentos de manutenção também foi identificada:"

## [19] "caracterizou uma não conformidade crítica relativa a procedimentos regulamentares de"

## [20] "ser identificado como uma não conformidade relativa a procedimentos operacionais de controle"

## [21] "não conformidades quanto a procedimentos regulamentares de operação, manutenção e"

## [22] "No que diz respeito a procedimentos de manutenção, a falha no fechamento dos dampers do"

## [23] "procedimentos regulamentares , inclusive efetuando a revisão da definição de"

## [24] "Efetuar a revisão de regras e procedimentos aplicáveis a projetos de forma a assegurar"

## [25] "Implementar regras e procedimentos para compatibilizar sistemas originais e as alterações de"

2.4.3 Identificação de frases com termos “gestão”, “operacional”

Texto$text[str\_detect(

Texto$text,

pattern = "Gestão|operacional"

)]

## [1] "sistema de gestão operacional das atividades marítimas de petróleo e gás natural da 5"

## [2] "de mantê-la nivelada devido a perda total do sistema de controle operacional da unidade. A"

## [3] "falhas no esquema operacional de controle de estabilidade de unidade flutuante, em condições"

## [4] "O quadro delineado evidencia deficiências no sistema de gestão operacional das atividades"

## [5] "a) Aprimoramento do sistema de gestão operacional"

## [6] "projeto de modo a garantir a segurança operacional e a proteção ambiental."

2.4.4 Identificação de frases com termos “risco”

Texto$text[str\_detect(

Texto$text,

pattern = "risco"

)]

## [1] "operação de esgotamento do tanque de popa bombordo e à classificação da área de risco em"

## [2] "esquemas de resposta a emergência de grande risco necessitam ser imediatamente"

## [3] "Como as áreas do terceiro e quarto nível não foram classificadas como zona de risco,"

## [4] "inadequação da classificação da área de risco em torno dos tanques de drenagem de"

## [5] "c) Classificação de áreas de risco"

## [6] "risco."

## [7] "h) Resposta a emergências de grande risco"

## [8] "que envolvam grande risco à salvaguarda de vidas humanas, segurança da navegação,"

3 Conclusões gerais

Quando realizada em somente um documento, a análise do contexto do relatório torna-se mais rápida, pois consistem somente em ler o sumário executivo do documento. Contudo, essa tarefa não pode ser replicada e automatizada. Já a análise por máquina demora mais em um primeiro documento, até o estabelecimento do algoritmo, contudo a análise poderá ser replicada para uma quantidade ilimitada de documentos.

4 Referências

National Commission on the BP Deepwater Horizon Oil Spill and Offshore Drilling (NCDWHSOD). Deep Water: The Gulf Oil Disaster and the Future of Offshore Drilling. Report to the President. January 2011 Cover Photo: © Steadfast TV. ISBN: 978-0-16-087371-3. <https://www.govinfo.gov/content/pkg/GPO-OILCOMMISSION/pdf/GPO-OILCOMMISSION.pdf>

<http://data7.blog/grafo-de-palavras-anitta-twitter/>

Analise de palavras. Disponivel em: <https://www.ufrgs.br/wiki-r/index.php?title=Frequ%C3%AAncia_das_palavras_e_nuvem_de_palavras> Esta página foi modificada pela última vez em 12 de dezembro de 2018, às 19h30min Conteúdo disponível sob Creative Commons - Atribuição - Compartilha nos Mesmos Termos, salvo indicação em contrário.

<https://p4husp.github.io/material/tutorial11/>

Principal: <https://www.tidytextmining.com/ngrams.html>

Corpus and Machine Learning: <https://rstudio-pubs-static.s3.amazonaws.com/265713_cbef910aee7642dc8b62996e38d2825d.html>

Machine learning: <https://kenbenoit.net/pdfs/text_analysis_in_R.pdf>

Mineração de texto: <https://www.rpubs.com/LaionBoaventura/mineracaodetexto>

MANIPULAÇÃO DE STRINGS E TEXT MININGhttps://gomesfellipe.github.io/post/2017-12-17-string/string/

@article{JSSv025i05, author = {Ingo Feinerer and Kurt Hornik and David Meyer}, title = {Text Mining Infrastructure in R}, journal = {Journal of Statistical Software, Articles}, volume = {25}, number = {5}, year = {2008}, keywords = {}, abstract = {During the last decade text mining has become a widely used discipline utilizing statistical and machine learning methods. We present the tm package which provides a framework for text mining applications within R. We give a survey on text mining facilities in R and explain how typical application tasks can be carried out using our framework. We present techniques for count-based analysis methods, text clustering, text classification and string kernels.}, issn = {1548-7660}, pages = {1–54}, doi = {10.18637/jss.v025.i05}, url = {<https://www.jstatsoft.org/v025/i05>} }

5 Apêndice A - Análise exploratória do contexto por máquina

5.0.1 Palavras mais frequentes

Aqui é possível identificar as palavras mais frequentes que aparecem, para então tentar trazer contextos do que ocorreu, antes de uma análise humana.

# Counting single words\*\*

frequenciaPalavras <- CleanW %>%

count(Palavra, sort = TRUE) %>%

filter(Palavra != "") %>%

arrange(desc(n))

# Visualiza frequencia de palavras

DT::datatable(frequenciaPalavras)

Show  entries

Search:

|  | **Palavra** | **n** |
| --- | --- | --- |
| 1 | plataforma | 105 |
| 2 | tanque | 69 |
| 3 | agua | 67 |
| 4 | analise | 58 |
| 5 | coluna | 58 |
| 6 | gas | 52 |
| 7 | acidente | 48 |
| 8 | emergencia | 47 |
| 9 | boreste | 46 |
| 10 | popa | 46 |

Showing 1 to 10 of 1,229 entries

Previous12345…123Next

head(frequenciaPalavras, n=20) %>%

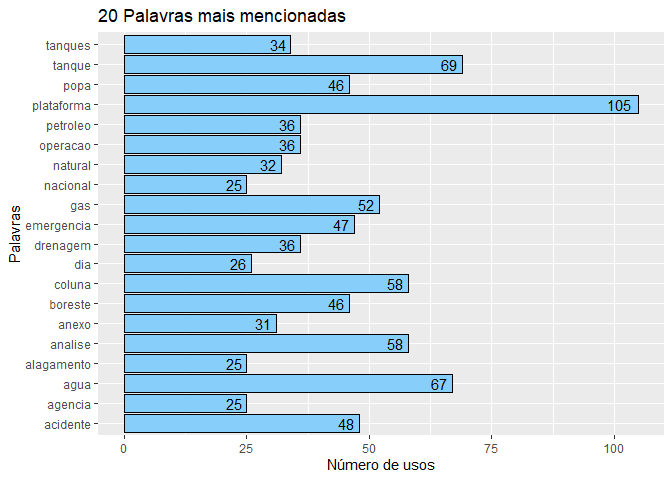
ggplot(aes(Palavra, n)) +

geom\_bar(stat = "identity", color = "black", fill = "#87CEFA") +

geom\_text(aes(hjust = 1.3, label = n)) +

coord\_flip() +

labs(title = "20 Palavras mais mencionadas", x = "Palavras", y = "Número de usos")



# Cria nuvem de palavras

wordcloud(

words = frequenciaPalavras$Palavra,

freq = frequenciaPalavras$n,

min.freq = 2,

max.words = 300,

random.order = FALSE,

rot.per = 0.35,

colors = brewer.pal(8, "Dark2")

)



5.0.2 Rede de palavras

A rede de palavras também pode ser um mapa para compreender mais rápido o que um documento revela. A rede apresentada aqui foi feita a partir de bigrams, que são estruturas que verificam a frequencia de pares de palavras de um documento.

par\_Palavras <- TidyT %>%

unnest\_tokens(Palavra, text, token = "ngrams", n = 2) %>%

mutate(Palavra = na.omit(NormalizaParaTextMining(Palavra))) %>%

anti\_join(palavrasRemover) %>%

anti\_join(my\_stopwords) %>%

separate(Palavra, c("word1", "word2"), sep = " ") %>%

filter(!word1 %in% stop\_words$word) %>%

filter(!word2 %in% stop\_words$word) %>%

filter(!word1 %in% my\_stopwords$Palavra) %>%

filter(!word2 %in% my\_stopwords$Palavra) %>%

count(word1, word2, sort = TRUE)

## Joining, by = "Palavra"

## Joining, by = "Palavra"

# remove NA

par\_Palavras <- na.omit(par\_Palavras)

DT::datatable(par\_Palavras)

Show  entries

Search:

|  | **word1** | **word2** | **n** |
| --- | --- | --- | --- |
| 1 | Popa | boreste | 37 |
| 2 | Gas | natural | 32 |
| 3 | Agencia | nacional | 25 |
| 4 | Petróleo | gas | 23 |
| 5 | Não | conformidades | 19 |
| 6 | Quarto | nivel | 16 |
| 7 | Procedimentos | regulamentares | 12 |
| 8 | Não | conformidade | 11 |
| 9 | Agua | salgada | 10 |
| 10 | Na | coluna | 10 |

Showing 1 to 10 of 1,875 entries

Esses pares são agrupados e acabam formando uma rede, de acordo com a frequência em que existem no documento. As ligações dos nós representam essa frequência. Assim, quanto mais larga for a linha de conecta as palavras, maior é a frequência que ela aparece no texto.

## Function to show word network

.Par\_net = function(z) {

par\_Palavras %>%

filter(n >= z) %>%

graph\_from\_data\_frame() %>%

ggraph(layout = "fr") +

geom\_edge\_link(aes(edge\_alpha = n, edge\_width = n)) +

geom\_node\_point(color = "darkslategray4", size = 4) +

geom\_node\_text(color = "red", aes(label = name), vjust = 1.8, size=3) +

labs(title= "Word graph - The Gulf Oil Disaster Report",

subtitle = paste("Pairwise analysis words >= ", z, " frequency",

x = "", y = ""))

}

.Par\_net(3)

