



Universidade de Brasília  
Departamento de Ciência da Computação

# **DODFCorpus - contratos e licitações**

## **Anotação e validação do dataset de contratos e licitações do DODF**

Artur Hugo Cunha Pereira

21 de setembro de 2022

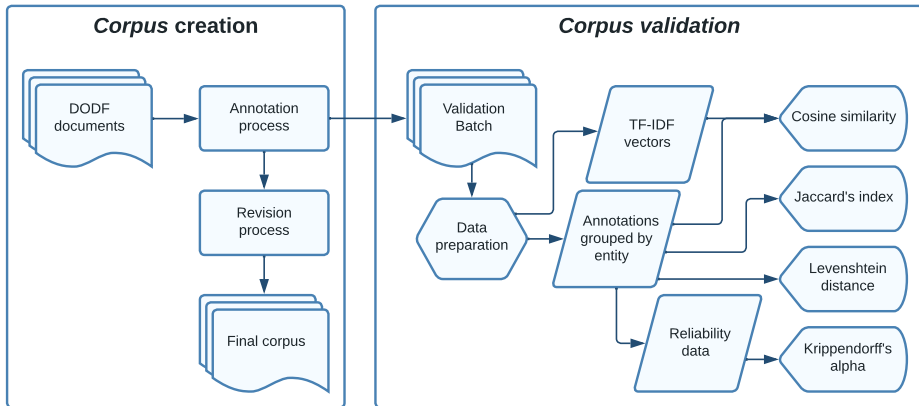
Introdução  
Motivação

Metodologia  
Criação do corpus  
Validação do corpus  
Preparação de dados

Resultados

Conclusão

- ▶ Auxiliar profissionais
- ▶ Modelos precisam de dados rotulados
- ▶ Verificar qualidade da anotação
  - ▶ Concordância de rótulo
  - ▶ Concordância estrutural



- ▶ Tipos de ato:
  - ▶ EXTRATO\_CONTRATO
  - ▶ EXTRATO\_ADITAMENTO\_CONTRATUAL
  - ▶ EXTRATO\_CONVENIO
  - ▶ AVISO\_LICITACAO
  - ▶ AVISO\_SUSPENSAO\_LICITACAO
  - ▶ AVISO\_ANUL\_REV\_LICITACAO
- ▶ +40 tipos de entidade

AVISO DE LICITAÇÃO PREGÃO ELETRÔNICO Nº 58/2021 A presente licitação tem por objeto a aquisição de  
material de consumo - Registro de Preços para aquisição, por demanda, de Brita Graduada Simples - BGS, tudo  
conforme especificado no Edital e em seus anexos. Processo 00113-00007505/2021-70. Data e horário para  
recebimento das propostas: até 09h00min do dia 01 de julho de 2021, com valor estimado de R\$ 9.938.200,00. O  
respectivo Edital poderá ser retirado exclusivamente nos endereços eletrônicos www.der.df.gov.br e  
www.licitacoes-e.com.br. Demais informações no próprio Edital. Brasília/DF, 09 de junho de 2021 ANA HILDA DO  
CARMO SILVA Diretora

Figura: Exemplo de um ato anotado

- ▶ Concordância de rótulo: os anotadores concordaram no rótulo da entidade?
  - ▶ Kappa de Cohen
  - ▶ Alpha de Krippendorff
  - ▶ AC de Gwet

- ▶ Concordância de rótulo: os anotadores concordaram no rótulo da entidade?
  - ▶ Kappa de Cohen
  - ▶ **Alfa de Krippendorff**
  - ▶ AC de Gwet



- ▶ Concordância estrutural: os anotadores concordam em onde começa onde termina a entidade?
  - ▶ Edit distance
  - ▶ Índice de Jaccard
  - ▶ Similaridade cosseno

- ▶ Agrupamento de anotações
  - ▶ Critério de offset
  - ▶ Critério de tamanho
- ▶ Vetores TF-IDF para similaridade cosseno
- ▶ Matriz de confiabilidade

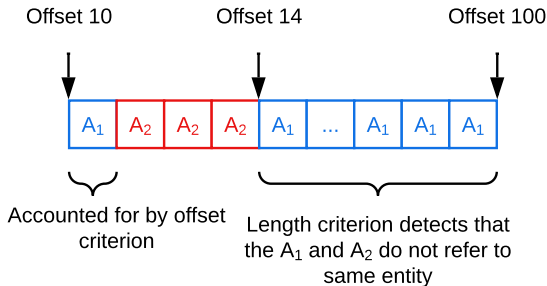


Figura: Exemplo dos critérios de agrupamento

Tabela: Exemplo de matriz de confiabilidade

|           | <i>E1</i> | <i>E2</i> | <i>E3</i> | <i>E4</i> |
|-----------|-----------|-----------|-----------|-----------|
| <i>A1</i> |           | <i>L1</i> | <i>L2</i> |           |
| <i>A2</i> |           |           | <i>L1</i> | <i>L3</i> |
| <i>A3</i> | <i>L1</i> | <i>L1</i> |           | <i>L3</i> |
| <i>A4</i> | <i>L2</i> | <i>L1</i> |           | <i>L3</i> |

- ▶ Alfa de Krippendorff:

$$\alpha_k = 1 - \frac{D_o}{D_e} \quad (1)$$

- ▶  $D_o$ : discrepância observada
- ▶  $D_e$ : discrepância esperada ao acaso
- ▶ Resultado:  $\alpha_k = 0,97 \Rightarrow D_o > D_e$

| Tipo de entidade              | Edit dist.                  | Jaccard                         | Cossim                          | Anotações |
|-------------------------------|-----------------------------|---------------------------------|---------------------------------|-----------|
| numero_contrato               | $2 \pm 3$                   | $0.8 \pm 0.3$                   | $0.8 \pm 0.3$                   | 1724      |
| orgao_contratante             | $8 \pm 17$                  | $0.7 \pm 0.4$                   | $0.7 \pm 0.4$                   | 1972      |
| codigo_siggo                  | <b><math>1 \pm 2</math></b> | $0.8 \pm 0.4$                   | $0.8 \pm 0.4$                   | 141       |
| EXTRATO_CONTRATO              | $148 \pm 486$               | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 1083      |
| data_assinatura_contrato      | $1 \pm 3$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 726       |
| entidade_contratada           | $6 \pm 13$                  | $0.8 \pm 0.4$                   | $0.8 \pm 0.3$                   | 1084      |
| cnpj_entidade_contratada      | $1 \pm 4$                   | $0.9 \pm 0.2$                   | $0.9 \pm 0.2$                   | 225       |
| objeto_contrato               | $38 \pm 125$                | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 1039      |
| valor_contrato                | $1 \pm 3$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 800       |
| nota_empenho                  | $1 \pm 4$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 680       |
| vigencia_contrato             | $17 \pm 44$                 | $0.8 \pm 0.4$                   | $0.8 \pm 0.4$                   | 938       |
| cnpj_orgao_contratante        | $2 \pm 7$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 71        |
| EXTRATO_ADITAMENTO_CONTRATUAL | $65 \pm 293$                | $0.9 \pm 0.2$                   | $0.9 \pm 0.2$                   | 607       |
| numero_termo_aditivo          | $1 \pm 5$                   | <b><math>0.9 \pm 0.2</math></b> | <b><math>0.9 \pm 0.2</math></b> | 628       |
| objeto_aditamento_contratual  | $97 \pm 266$                | $0.8 \pm 0.4$                   | $0.8 \pm 0.4$                   | 606       |

| Tipo de entidade          | Edit dist.                  | Jaccard                         | Cossim                          | Anotações |
|---------------------------|-----------------------------|---------------------------------|---------------------------------|-----------|
| EXTRATO_CONVENIO          | 647 $\pm$ 591               | 0.3 $\pm$ 0.5                   | 0.3 $\pm$ 0.5                   | 9         |
| entidade_conveniente      | 21 $\pm$ 18                 | 0.2 $\pm$ 0.4                   | 0.2 $\pm$ 0.4                   | 10        |
| cnpj_entidade_conveniente | 14 $\pm$ 8                  | 0.2 $\pm$ 0.4                   | 0.3 $\pm$ 0.4                   | 4         |
| orgao_concedente          | 34 $\pm$ 11                 | 0.0 $\pm$ 0.0                   | 0.0 $\pm$ 0.0                   | 7         |
| cnpj_orgao_concedente     | 18 $\pm$ 0                  | 0.0 $\pm$ 0.0                   | 0.0 $\pm$ 0.0                   | 1         |
| objeto_convenio           | 111 $\pm$ 179               | 0.5 $\pm$ 0.5                   | 0.5 $\pm$ 0.5                   | 11        |
| vigencia_convenio         | 25 $\pm$ 23                 | 0.4 $\pm$ 0.5                   | 0.4 $\pm$ 0.5                   | 9         |
| valor_convenio            | 4 $\pm$ 6                   | <b>0.7 <math>\pm</math> 0.5</b> | <b>0.7 <math>\pm</math> 0.5</b> | 3         |
| data_assinatura_convenio  | 10 $\pm$ 11                 | 0.5 $\pm$ 0.5                   | 0.5 $\pm$ 0.5                   | 8         |
| numero_convenio           | <b>7 <math>\pm</math> 5</b> | 0.2 $\pm$ 0.4                   | 0.2 $\pm$ 0.4                   | 9         |

| <b>Tipo de entidade</b>              | <b>Edit dist.</b>           | <b>Jaccard</b>                  | <b>Cossim</b>                   | <b>Anotações</b> |
|--------------------------------------|-----------------------------|---------------------------------|---------------------------------|------------------|
| modalidade_licitacao                 | $2 \pm 4$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 830              |
| numero_licitacao                     | $3 \pm 51$                  | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 1104             |
| orgao_licitante                      | $6 \pm 17$                  | $0.7 \pm 0.4$                   | $0.7 \pm 0.4$                   | 774              |
| AVISO_LICITACAO                      | $29 \pm 169$                | <b><math>1.0 \pm 0.1</math></b> | <b><math>1.0 \pm 0.1</math></b> | 873              |
| tipo_objeto                          | $4 \pm 6$                   | $0.7 \pm 0.5$                   | $0.7 \pm 0.5$                   | 873              |
| valor_estimado_contratacao           | $0 \pm 2$                   | $1.0 \pm 0.2$                   | $1.0 \pm 0.2$                   | 560              |
| data_abertura_licitacao              | $1 \pm 3$                   | $0.9 \pm 0.2$                   | $0.9 \pm 0.2$                   | 758              |
| sistema_compras                      | $2 \pm 7$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 900              |
| codigo_licitacao_<br>sistema_compras | <b><math>0 \pm 2</math></b> | $0.9 \pm 0.2$                   | $0.9 \pm 0.2$                   | 611              |
| AVISO_SUSPENSAO_<br>LICITACAO        | $27 \pm 204$                | $1.0 \pm 0.2$                   | $1.0 \pm 0.2$                   | 84               |
| decisao_tcdf                         | $23 \pm 37$                 | $0.2 \pm 0.3$                   | $0.4 \pm 0.4$                   | 19               |
| AVISO_ANUL_<br>REV_LICITACAO         | $34 \pm 165$                | $1.0 \pm 0.2$                   | $1.0 \pm 0.2$                   | 30               |
| identificacao_ocorrencia             | $2 \pm 5$                   | $0.8 \pm 0.3$                   | $0.8 \pm 0.3$                   | 33               |

| Tipo de entidade         | Edit dist.                  | Jaccard                         | Cossim                          | Anot. | Tipos de ato             |
|--------------------------|-----------------------------|---------------------------------|---------------------------------|-------|--------------------------|
| processo_gdf             | $1 \pm 5$                   | $0.9 \pm 0.2$                   | <b><math>0.9 \pm 0.2</math></b> | 2570  | Todos                    |
| nome_<br>responsavel     | $1 \pm 5$                   | <b><math>1.0 \pm 0.2</math></b> | $0.9 \pm 0.2$                   | 1183  | Todos                    |
| unidade_<br>orcamentaria | $1 \pm 3$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 558   | Contratos e convênios    |
| programa_<br>trabalho    | $3 \pm 6$                   | $0.8 \pm 0.3$                   | $0.9 \pm 0.3$                   | 712   | Contratos e convênios    |
| fonte_recurso            | $4 \pm 10$                  | $0.7 \pm 0.4$                   | $0.7 \pm 0.4$                   | 801   | Contratos e convênios    |
| natureza_<br>despeza     | <b><math>1 \pm 2</math></b> | $0.8 \pm 0.4$                   | $0.8 \pm 0.4$                   | 564   | Contratos e convênios    |
| objeto_licitacao         | $31 \pm 90$                 | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 951   | Licitações e suspensões  |
| data_escrito             | $1 \pm 4$                   | $0.9 \pm 0.3$                   | $0.9 \pm 0.3$                   | 560   | Aditamentos e suspensões |



- ▶ Altos desvios padrão
- ▶ Extratos de convênio
- ▶ Resultados inesperados: entidades numéricas
- ▶ Trabalhos futuros:
  - ▶ Validação extrínseca: problemas de classificação e segmentação
  - ▶ Avaliar o efeito do processo de revisão