

## 10 CLASSIFICAÇÃO DAS TAGS DOS PEDIDOS DE ACESSO À INFORMAÇÃO POR MEIO DO CHATGPT

*Classification of Tags for access to information requests through ChatGPT*

**Zenóbio Santos Júnior**

Universidade FUMEC

<https://orcid.org/0000-0003-1380-7391>

zenojr19@gmail.com

**Frederico Giffoni de Carvalho Dutra**

Universidade FUMEC

<https://orcid.org/0000-0002-8666-0354>

fgcdutra@gmail.com

**Eduardo José da Silva Luz**

Universidade Federal de Ouro Preto

<https://orcid.org/0000-0001-5249-1559>

eduluz@ufop.edu.br

## RESUMO

**INTRODUÇÃO:** A Lei de Acesso à Informação garante o direito de acesso à informação por meio do Fala.BR, que contempla os tipos de manifestação de ouvidoria (denúncia, elogio, reclamação, solicitação e sugestão) e acesso à informação aos órgãos do poder público (Brasil, 2011). Ao realizar uma manifestação, o usuário se depara com 21 campos de cadastramento, dentre eles: Órgão Destinatário, Prazo de Atendimento, Forma de Resposta, Assunto Pedido, Subassunto do Pedido e *Tag*, este o foco do estudo. Pesquisa realizada em 2021 apontou a ausência de preenchimento das *Tags* em 76,81% pelos usuários que utilizaram o Fala.BR naquele ano (Santos Júnior; Corrêa; De Faria, 2023). Diante do exposto, surge a seguinte questão: Quais as contribuições o ChatGPT pode trazer para a classificação das *Tags* dos pedidos de acesso à informação das Universidades Federais no Fala.BR? O objetivo é propor ferramenta de inteligência artificial, por meio do ChatGPT, no auxílio da classificação das *Tags* dos pedidos de acesso à informação das Universidades Federais do Brasil no Fala.BR. **FUNDAMENTAÇÃO TEÓRICA:** Para facilitar o acesso às informações de forma clara, ágil e confiável, faz-se necessário uma boa classificação. A classificação tem importância na padronização dos termos, palavras-chave ou descritores resultantes do processo de indexação, cumprindo a função de criar rótulos que, junto com o número de classificação dos documentos, representam pontos de acesso para a recuperação dos documentos demandados pelos usuários (Sousa; Araújo Júnior, 2017). O ChatGPT é um sistema de Processamento de Linguagem Natural (PNL) que pode gerar conversas semelhantes às humanas (Deng; Lin, 2022) e poderá reduzir o tempo de classificação e apresentar resultados impressionantes para o Fala.BR. **METODOLOGIA:** A pesquisa possui caráter exploratório, aplicado, de abordagem qualitativa-quantitativa, e utilizou pesquisas bibliográficas e experimentais como procedimentos técnicos. Os registros dos pedidos de acesso às

informações das Universidades Federais de 2023 foram realizados por meio do *Download* de Dados do Fala.BR. Utilizou-se o Código de Classificação de Documentos de Arquivo das atividades-fim das Instituições Federais de Ensino Superior (Brasil, 2013) e o ChatGPT para a classificação das *Tags*. Selecionou-se 50 textos de acesso à informação para análise de 10 Arquivistas, a fim de avaliar o grau de eficiência da ferramenta. Após a análise, somaram-se os resultados do ChatGPT e para avaliação das *Tags* foi utilizado o modelo de linguagem BERTimbau (Souza; Nogueira; Lotufo; 2020), calculando a distância de cosseno para determinar a similaridade via *Word Embedding*,<sup>2</sup> sendo que, quanto mais próximos de 1 serão considerados altos e baixos quando forem inferiores a 0,6. O grau de efetividade dos resultados entre o homem x máquina foi avaliado por meio da similaridade por cosseno, que realiza a medida entre dois vetores num espaço vetorial compreendido entre eles. Essa medida é usada em várias áreas onde a magnitude dos vetores não é (tão) importante quanto sua direção, em especial em busca e recuperação de informação, para medir a semelhança entre uma consulta e um documento (Ribeiro-Neto; Baeza-Yates, 2011). **RESULTADOS:** Verificou-se que em 2023, 81,51% das *Tags* dos pedidos de informação das Universidades encontravam-se sem preenchimento. Entretanto, “todas” as *Tags* dos 50 textos foram preenchidas tanto pelo ChatGPT quanto por arquivistas e ao menos uma palavra foi sugerida ao usuário, chegando até cinco opções. A média da similaridade de cosseno entre o ChatGPT e Arquivistas para as *Tags* é de aproximadamente 0,7702 o que sugere uma concordância relativamente alta entre as *Tags* classificadas pelo especialista humano e pelo ChatGPT. **CONSIDERAÇÕES FINAIS:** A pesquisa demonstrou que o uso do ChatGPT, poderá trazer benefícios na classificação das *Tags* dos pedidos de acesso à informação para o Fala.BR. Uma delas, a inversão de valores, pois o que era negativo (ausência de preenchimento em

81%) passou a ser positivo com uso do ChatGPT (77% das *Tags* preenchidas) comparado com a análise dos Arquivistas.

**Palavras-chave:** classificação; *tags*; acesso à informação; Fala.BR; ChatGPT.

**Data de submissão:** 20/09/2024

**Data de aprovação:** 03/10/2024

## REFERÊNCIAS

- BRASIL. Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**, 18 nov. 2011.
- BRASIL. Ministério da Gestão e da Inovação em Serviços Públicos. Conselho Nacional de Arquivos. Portaria MEC Nº 1.224, de 18 de dezembro de 2013. **Código de Classificação de Documentos de Arquivo Relativos às Atividades-fim das Instituições Federais de Ensino Superior e na Tabela de Temporalidade e Destinação de Documentos de Arquivo Relativos às Atividades-Fim das Instituições Federais de Ensino Superior**. Rio de Janeiro: Arquivo Nacional, 2013.
- DENG, Jianyang; LIN, Yijia. The benefits and challenges of ChatGPT: An overview. **Frontiers in Computing and Intelligent Systems**, v. 2, n. 2, p. 81-83, 2022.
- RIBEIRO-NETO, Berthier; BAEZA-YATES, R. **Modern information retrieval: the concepts and technology behind search**. 2011.
- SANTOS JÚNIOR, Zenóbio; CORRÊA, Fábio; DE FARIA, Vinícius F. Classificação de documentos de arquivo: principais temas dos pedidos de acesso à informação nas universidades federais mais demandadas pela LAI. In: PINHA, Mariana Lousada; VITORIANO, Marcia Cristina de Carvalho Pazin; SANTOS, Paulo Roberto Elianos. **Arquivos, democracia e justiça social [livro eletrônico]**. São Paulo: ARQ-SP, 2023.
- SOUSA, Renato Tarciso Barbosa; ARAÚJO JUNIOR, Rogerio Henrique de. A classificação e o vocabulário controlado como instrumentos efetivos para a recuperação da informação arquivística. In: VAQUINHAS, Nelson; CAIXAS, Marisa; VINAGRE, Helena. (org.).

2 É uma técnica na qual palavras individuais são representadas como vetores de valor real em um espaço de menor dimensão e captura semântica Inter palavra. Turing, Disponível em: <https://www.turing.com/kb/guide-on-word-embeddings-in-nlp> Acesso em: 30 jun. 2024.

**Da produção à preservação informacional:** desafios e oportunidades. [S. l.: s. n.], 2017.

SOUZA, F., Nogueira, R., & LOTUFO, R. (2020). BERTimbau: pre-trained BERT models for Brazilian Portuguese. *In INTELLIGENT SYSTEMS: 9th BRAZILIAN CONFERENCE, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I 9* (pp. 403–417). Springer International Publishing.

#### Notas:

##### Conflito de interesse

Informa se há conflitos de interesse financeiros ou de outra natureza por parte dos autores.

##### Contribuição dos autores

Os autores dessa publicação contribuíram com a mesma sob os seguintes aspectos principais:

**1** – Zenóbio dos Santos Júnior desenvolveu sua Tese, no qual deu base para a pesquisa e elaboração do pôster. Elaborou o planejamento da pesquisa e levantou a coleta de dados no Fala.BR, bem como nas análises dos dados, assim como na discussão dos resultados e redação do pôster.

**2** – Frederico Giffoni de Carvalho Dutra, orientador do trabalho, apoiou no planejamento do estudo, coleta de dados a campo, análise dos dados, discussão dos resultados e revisão do pôster.

**3** – Eduardo José da Silva Luz participou das análises dos dados obtidos, apoiou na discussão dos resultados e revisão do pôster.

##### Informar se a publicação é oriunda de uma dissertação ou tese

**Aprovação Ética:** Pesquisa aprovada no Comitê de Ética em Pesquisa da Universidade FUMEC sob o Parecer nº 6.715.968 em 21 de março de 2024 com base na Resolução nº 510, de 7 de abril de 2016, para as pesquisas que realizam recrutamento de participantes e coleta de dados envolvendo seres humanos em pesquisas.

A pesquisa é parte da análise dos dados e resultados da Tese de Doutorado de Zenóbio dos Santos Júnior, intitulada “O uso do ChatGPT no Fala.BR: Proposta para indexação dos pedidos de Acesso à Informação Das Universidades Federais do Brasil”.

#### Agradecimentos

À minha família e a toda comunidade acadêmica da Universidade FUMEC.