

Preservação da memória com acessibilidade digital: um plugin para descrição de imagens com IA generativa

Edson Gabriel Jacques^{1*}, Carolina Sacramento^{1,2*}, Yndiana Gouveia¹
Wagner Nascimento Silva¹, Ygor Santos Barros¹, Simone Bacellar Leal Ferreira²

¹Casa de Oswaldo Cruz – Fundação Oswaldo Cruz

²Universidade Federal do Estado do Rio de Janeiro (UNIRIO)

{edson.jacques, carolina.sacramento, yndiana.gouveia}@fiocruz.br

{wagner.silva, ygor.barros}@fiocruz.br, simone@uniriotec.br

Abstract. *This article describes the development of a plugin that employs generative Artificial Intelligence (AI) to enhance accessibility in historical archive cataloging systems, particularly for visually impaired users. Developed with generic instructions and a free model from Google, the plugin was tested by regular users, achieving high satisfaction (91.36 on the System Usability Scale). The plugin has been improved based on user feedback and continues to evolve, with plans to refine the instructions and conduct tests with audio description experts and visually impaired individuals.*

Resumo. *Este artigo descreve o desenvolvimento de um plugin que utiliza Inteligência Artificial (IA) generativa para melhorar a acessibilidade em sistemas de catalogação de acervos históricos, especialmente para usuários com deficiência visual. Desenvolvido com instruções genéricas e um modelo gratuito da Google, o plugin foi testado por usuários habituais, atingindo alta satisfação (91.36 no System Usability Scale). O plugin foi aprimorado com o feedback dos usuários e continua em evolução, planejando-se ajustes nas instruções e testes com especialistas em audiodescrição e pessoas com deficiência visual.*

*Ambos os autores contribuíram igualmente para esta proposta.

1. Contexto

Acervo arquivístico ou arquivo é um “conjunto de documentos originais produzidos e acumulados por uma entidade coletiva, pública ou privada, pessoa ou família, no desempenho de suas atividades”[Brasil 2005]. Este conjunto pode ser considerado histórico, caso seja identificado valor para a preservação da memória e da identidade cultural de uma instituição, nação ou cultura. Um arquivo pode incluir uma diversidade de tipos de documentos, como textos, imagens, mapas, registros sonoros e audiovisuais, refletindo a riqueza e complexidade das experiências humanas documentadas [Brasil 2005].

A evolução tecnológica tem revolucionado a preservação de acervos arquivísticos físicos, como cartas e fotografias, por meio da digitalização. Esse processo cria cópias digitais que garantem acesso contínuo aos documentos, protegendo-os do desgaste natural e facilitando sua ampla disseminação.

A Base Arch é o repositório da Fundação Oswaldo Cruz (Fiocruz) que armazena informações sobre seu acervo arquivístico, essencial para pesquisas sobre a história das ciências e da saúde no Brasil, além de preservar a memória institucional e de figuras importantes da saúde pública. A base é administrada pelo Departamento de Arquivo e Documentação da Casa de Oswaldo Cruz (COC)/Fiocruz e utiliza o sistema AtoM (*Access to Memory*), uma plataforma *open source*, web e multilíngue [Artefactual 2025a].

O AtoM, desenvolvido e mantido pela Artefactual - uma empresa canadense que desenvolve *softwares* para acesso e preservação digital, é usado por mais de 220 instituições mundialmente. Permite personalização de temas, criação de *plugins* e contribuições ao seu desenvolvimento devido à sua natureza *open source* [Artefactual 2025b]. A Base Arch utiliza o AtoM com um tema customizado, alinhado à identidade visual da COC/Fiocruz, e recentemente passou por avaliações e ajustes de acessibilidade digital para atender à Lei Brasileira de Inclusão [Brasil 2015] e às políticas de acessibilidade da Fiocruz [Brasil 2019].

Um dos grupos beneficiados pela acessibilidade implementada foi o das pessoas com deficiência visual. Para aquelas com baixa visão, foi essencial oferecer uma interface com contraste de cores adequado e possibilidade de ampliar o zoom sem perda de funcionalidade. Para pessoas com deficiências visuais mais severas, foram feitos ajustes no código fonte para ampliar a compatibilidade com *softwares* leitores de telas. Esses *softwares* estão disponíveis em *smartphones* e podem ser instalados em computadores tradicionais, convertendo texto em voz sintetizada para navegação e acesso autônomo.

Interfaces que seguem os padrões internacionais de acessibilidade [W3C 2024b] viabilizam a interação via leitores de tela, mas nem todo conteúdo é compatível com essa tecnologia. Leitores de tela muitas vezes não conseguem interpretar imagens ou oferecem descrições superficiais. Por isso, é essencial produzir descrições textuais, conhecidas como texto alternativo. Essas descrições devem ser fornecidas pelos produtores de conteúdo [W3C 2024a], como arquivistas e outros profissionais que operam a base.

A criação de textos alternativos pode ser desafiadora para pessoas leigas [Sacramento et al. 2020] que não dominam a audiodescrição, técnica que transforma imagens em palavras, criando descrições de qualidade, em áudio ou texto. A técnica requer a competência de um audiodescriptor roteirista e a consultoria de uma pessoa com deficiência visual treinada na técnica [ABNT 2016]. Devido ao alto custo e complexidade da audiodescrição, torna-se inviável contratar o serviço para descrever as milhares de imagens da Base Arch e futuras adições.

Este artigo contribui com os Grandes Desafios de Sistemas de Informação ao propor uma solução para melhorar a vida das pessoas em seus ambientes [Pereira and Baranauskas 2017]. Para tal, apresenta-se um *plugin* que usa Inteligência Artificial (IA) para descrever imagens em sistemas de catalogação de acervos arquivísticos.

2. Processo adotado

O desenvolvimento do *plugin* foi estruturado em cinco etapas, a saber:

1. **Seleção da ferramenta de IA generativa**, onde um estudo comparativo envolvendo as soluções Chat GPT4 e Gemini avaliou: (a) qualidade das descrições geradas; (b) custo de utilização da API (*Application Programming Interface*) a ser

utilizada na comunicação entre o *plugin* em desenvolvimento e a IA generativa em estudo e (c) facilidade e modularidade da implementação. O estudo resultou na escolha da IA generativa *Gemini* - modelo 1.5 Flash da Google Inc., que apesar de não ter a melhor descrição de imagem (posto obtido pelo Chat GPT4), foi selecionado por ser uma solução gratuita com boa capacidade de compreensão das imagens. Sua versão gratuita permite o processamento de até um milhão de *tokens*¹ por mês, além de oferecer, quinze requisições por minuto e 1.500 requisições/ dia.

2. **Definição do *prompt*² para descrição de imagens:** conforme Tabela 1. O *prompt* definido não seguiu uma normativa específica, mas foi ajustado após testes iniciais para excluir referências a câmeras nas descrições, uma vez que o modelo as incluía, mesmo sem a presença visível de uma câmera (ex.: “a mulher olha para a câmera”). Além disso, a descrição foi limitada a 200 *tokens* e o *prompt* foi formulado em inglês, devido ao desempenho superior observado nesse idioma.
3. **Desenvolvimento de um *plugin* para descrição de imagens**, cujos detalhes serão abordados na próxima seção.
4. **Testes de usabilidade com usuários da base**, realizados por onze profissionais da COC/Fiocruz experientes no uso e alimentação da Base Arch. O teste incluiu duas tarefas: uma na área pública da interface, simulando o uso do recurso de descrição de imagens por usuários finais, e outra na área restrita, de alimentação de conteúdo, onde o *plugin* poderia ser acionado para ajudar no preenchimento do texto alternativo. Os resultados indicaram alta satisfação, com uma pontuação de 91.36 de 100 no SUS (*System Usability Scale*) [Brooke 1996], método que avalia a usabilidade de aplicações através de 10 questões em escala Likert respondidas após o teste. Os valores obtidos nas respostas são submetidos a cálculos propostos no método, para obtenção da pontuação final. Os procedimentos seguiram normas éticas, incluindo a assinatura de um Termo de Consentimento Livre e Esclarecido.
5. **Ajustes na interface**, a partir do *feedback* dado pelos participantes dos testes de usabilidade, com intuito de aprimorar a experiência do usuário.

Tabela 1. Prompt utilizado no plugin (em Inglês)

Describe this image in detail without referencing any camera-related terms such as 'angle', 'focus,' or 'shot'. Focus on the important visual elements, including colors, shapes, sizes, and textures. Provide context if possible, ensuring clarity and accessibility. The response must be written in Portuguese.

3. Solução

O *plugin* foi desenvolvido para ser compatível com o AtoM, que opera sob o *framework* *Symfony* e utiliza a linguagem PHP [Artefactual 2025c]. Foram empregadas as tecnologias Javascript como linguagem de programação, NodeJs (versão 22.8) para execução de *scripts* no lado do servidor e a biblioteca ExpressJs (versão 4.19.2) para a criação da API. O *plugin* e documentação associada está disponível para acesso e download no link: <https://github.com/gabrieljacques-tests/base-arch>.

¹*Token* representa uma fração de texto, podendo ser uma palavra, parte de uma palavra ou um caractere isolado, dependendo do idioma e da estrutura do texto.

²*Prompt* é uma instrução fornecida a um modelo de IA para gerar conteúdo específico.

Quando ativado, o *plugin* é exibido em duas áreas da interface do sistema: na visualização pública de uma descrição arquivística que possui imagem (ou objeto digital) e no cadastro de um objeto digital. Em ambos os contextos, um botão é disponibilizado, permitindo que o usuário solicite a descrição sob demanda. Essa funcionalidade sob demanda foi implementada devido à limitação mensal de *tokens*, mencionada anteriormente.

A Figura 1 mostra (a) o botão exibido quando a imagem não possui descrição e (b) a descrição fornecida pelo modelo de IA utilizado. A estratégia de descrição adotada foi a descrição aberta, que não recorre ao recurso de texto alternativo (atributo “alt” da tag img). Utilizar a descrição no atributo “alt” restringe o acesso à descrição apenas a usuários de leitores de telas. A descrição aberta foi escolhida por ser considerada mais inclusiva e por promover ampla divulgação da iniciativa de descrição de imagens [MWPT 2020]. As descrições criadas pelo modelo de IA generativa começam com a frase: “Gerado por IA”.



Figura 1. Plugin em funcionamento. Crédito: dos autores

Na área administrativa do sistema, o usuário pode gerar e editar descrições ao pressionar o botão dedicado. As descrições aprimoradas são então armazenadas no banco de dados da aplicação. Quando a descrição é fornecida através da interface administrativa, o botão de descrição não é mais visível na interface pública, já que se presume que a descrição foi validada por um humano.

O projeto, ainda em fase inicial, já demonstra um potencial significativo, embora necessite de refinamentos, especialmente no *prompt* utilizado, para melhorar a qualidade das descrições geradas. Nos próximos passos, planeja-se expandir as investigações, explorando outros modelos de IA generativa, inclusive os pagos. Adicionalmente, pretende-se incorporar diretrizes para a descrição de imagens, baseadas em tutoriais especializados e pesquisas acadêmicas, no *prompt*, e aplicar a solução em sistemas de informação de outros domínios. A validação da solução com profissionais de audiodescrição e pessoas com deficiência visual também está prevista.

4. Agradecimentos

Este trabalho foi desenvolvido com o suporte do projeto “Fortalecimento e promoção das ações institucionais da Casa de Oswaldo Cruz”. A revisão textual e a tradução do resumo para o inglês foram realizadas com o apoio do ChatGPT-4.

Referências

- ABNT (2016). Abnt nbr 16452: acessibilidade na comunicação - audiodescrição. Disponível em: <https://www.abntcatalogo.com.br/pnm.aspx?Q=bm51M2JSMWZYa1lLYTRnbU5DZjA2TWlsNjlyVkJVzBHaDhqZThzbERORT0=>. Acesso em: 06 mar. 2025.
- Artefactual (2025a). Atom: Oferecendo acesso à memória desde 2007. Disponível em: <https://www.accesstomemory.org/pt-br/>. Acesso em: 06 mar. 2025.
- Artefactual (2025b). Atom: Users. Disponível em: <https://wiki.accesstomemory.org/wiki/Community/Users>. Acesso em: 06 mar. 2025.
- Artefactual (2025c). What is atom? Disponível em: <https://www.accesstomemory.org/pt-br/docs/2.8/user-manual/overview/intro/>. Acesso em: 06 mar. 2025.
- Brasil (2005). Dicionário brasileiro de terminologia arquivística. Disponível em: https://www.gov.br/conarq/pt-br/centrais-de-conteudo/publicacoes/dicionario_de_terminologia_arquivistica.pdf. Acesso em: 06 mar. 2025.
- Brasil (2015). Lei nº 13.146, de 6 de julho de 2015: Institui a lei brasileira de inclusão da pessoa com deficiência (estatuto da pessoa com deficiência). Disponível em: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/l13146.htm. Acesso em: 06 mar. 2025.
- Brasil (2019). Política da fiocruz para acessibilidade e inclusão das pessoas com deficiência. Disponível em: https://portal.fiocruz.br/sites/portal.fiocruz.br/files/documentos/politica_acessibilidade_inclusao_fiocruz_final.pdf. Acesso em: 06 mar. 2025.
- Brooke, J. (1996). *SUS – a quick and dirty usability scale*, pages 189–194. CRC Press.
- MWPT (2020). Descrição de imagens nas redes sociais: Alt ou hashtags inclusivas no texto? Disponível em: <https://mwpt.com.br/descricaodeimagensnasredessociaisaltou-hashtagsinclusivasnotexto/>. Acesso em: 06 mar. 2025.
- Pereira, R. and Baranauskas, M. C. C. (2017). Systemic and socially aware perspective for information systems. In *I GranDSI-BR: Grand Research Challenges in Information Systems in Brazil 2016-2026*. Disponível em: <https://books-sol.sbc.org.br/index.php/sbc/catalog/view/28/111/261>. Acesso em: 23 mar. 2025.
- Sacramento, C., Nardi, L., Ferreira, S. B. L., and Marques, J. a. M. d. S. (2020). prace-gover: investigating the description of visual content in brazilian online social media. In *Proceedings of the 19th Brazilian Symposium on Human Factors in Computing Systems*, IHC '20, New York, NY, USA. Association for Computing Machinery.
- W3C (2024a). Understanding sc 1.1.1: Non-text content (level a). Disponível em: <https://www.w3.org/WAI/WCAG22/Understanding/non-text-content.html>. Acesso em: 06 mar. 2025.
- W3C (2024b). Web content accessibility guidelines (wcag) 2.2. Disponível em: <https://www.w3.org/TR/WCAG22/>. Acesso em: 06 mar. 2025.