

Glossário Online em Libras para Informática como uma estratégia para aprendizagem de termos técnicos

Fábio A. P. Paiva¹, Álvaro H. Silva¹, Diego S. Pereira¹, Valério G. Medeiros Jr.¹
Vitor D. L. Santos¹, Rickson A. C. Soares¹

¹Instituto Federal do RN
Parnamirim – RN – Brasil

{fabio.procopio,alvaro.silva,diego.silva,valerio.medeiros}@ifrn.edu.br

{vitor036daniel,rickson280806}@gmail.com

Abstract. *Despite the fact that the deaf community of students in Brazil has a significant percentage, several technical school subjects are not inclusive based. This reality also affects deaf students at Computing courses. This article proposes a collaborative glossary online for Libras in Computing course classes. It aims at reducing the difficulties and barriers of inclusive education for deaf students. Most of these students are not literate in Portuguese language. The contributions of the work are: 1) capacity of self-construction of a prototype through the extraction of new words and expressions available on online monolingual glossaries; 2) video productions in Libras to define Computing technical expressions; and 3) the sharing of glossary expressions through an API REST for different types of applications.*

Resumo. *Apesar da comunidade surda ter um percentual significativo no Brasil, diversos conteúdos técnicos não são inclusivos. Isso se estende também a estudantes surdos de informática. Este artigo propõe um glossário colaborativo online em Libras para informática cujo objetivo é reduzir a barreira de inclusão de estudantes surdos, uma vez que muitos deles não são alfabetizados na língua portuguesa. As contribuições do artigo são: 1) capacidade de autoconstrução do protótipo através da extração de novos termos disponíveis em glossários online monolíngue; 2) produção de vídeos em Libras para definir termos técnicos de informática; e 3) compartilhamento de termos do glossário através de uma API REST para uso de outros tipos de aplicação.*

1. Introdução

Segundo os dados do Instituto Brasileiro de Geografia e Estatística (IBGE), aproximadamente 5% da população do país possui algum tipo de deficiência auditiva [IBGE 2022]. Assim, como forma de proporcionar que os cidadãos com alguma deficiência auditiva (parcial ou total) sejam indivíduos ativos na sociedade, é fundamental a criação de mecanismos que proporcionem acessibilidade.

Em geral, a população surda do Brasil é pouco alfabetizada na língua portuguesa e, por isso, muitas vezes apresenta dificuldades no acesso a textos escritos. Na área de Tecnologia da Informação (TI), essa dificuldade tende a aumentar ainda mais. Muitas

soluções, como o VLibras¹, não dão suporte a termos comuns da TI e, assim, causam um distanciamento dos usuários surdos e de promissores profissionais.

Nos últimos anos, diversos estudos vêm sendo conduzidos [Faqueti et al. 2005] [Barwaldt et al. 2017] [Calé et al. 2017] [da Cruz et al. 2017] [Fabris et al. 2018] [Granada et al. 2022] [Silva Junior et al. 2023] a fim de apresentar soluções de tecnologias assistivas para fomentar o progresso educacional desses alunos.

Considerando a necessidade dos surdos em conhecerem conceitos do contexto de informática, os diversos recursos educacionais criados (vídeos, blogs, textos e outros) não são suficientes para um aprendizado adequado. Todavia, algumas instituições de ensino que ofertam cursos de formação em TI, como o Instituto Federal do Rio Grande do Sul (IFRS) e a Universidade Federal do Rio de Janeiro (UFRJ), disponibilizam diversas aulas na plataforma Youtube com palavras de temáticas variadas que são direcionados às pessoas surdas.

Este trabalho é uma nova versão do protótipo funcional de um glossário *online* bilíngue (Português e Língua Brasileira de Sinais, Libras) cujo objetivo é disponibilizar um conjunto de termos da área de informática com acessibilidade para pessoas surdas. Os termos e suas definições, em língua portuguesa, são associados a vídeos que foram produzidos por intérpretes em Libras. Cada par termo/definição pode ser incluído, colaborativamente, por usuários ou de forma automática. A inclusão automática é realizada por meio de um módulo *web scraping* que é programado para visitar glossários monolíngues, extrair termos adequados, associá-los às suas definições e armazená-los no glossário para que uma pessoa surda, não alfabetizada em língua portuguesa, possa compreender o sinal e a sua definição. Quando comparada à versão anterior, a novidade desta versão consiste em três novas *features*, a saber: 1) realização de buscas semânticas no glossário, 2) implementação de uma API (*application programming interface*) REST (*representational state transfer*) para compartilhar o glossário com aplicações clientes e 3) categorização automática de termos dentro do glossário.

2. Trabalhos relacionados

Voltada a docentes que ministram conteúdos a pessoas surdas, a pesquisa de [Faqueti et al. 2005] propôs uma solução que foca no ensino de Libras, utilizando termos técnicos de informática. Por outro lado, o trabalho de [Barwaldt et al. 2017], que utiliza uma lista composta por palavras reservadas da linguagem LOGO, buscou melhorar a convivência de estudantes surdos na matemática fundamental.

Os trabalhos de [Calé et al. 2017] e [da Cruz et al. 2017] propuseram uma aplicação móvel que correlaciona termos de informática a sinais em Libras. Essa aplicação disponibiliza um acervo de vídeos que auxilia no processo ensino-aprendizagem por meio de termos de informática interpretados em Libras.

A publicação de [Fabris et al. 2018] construiu um glossário colaborativo de símbolos em Libras que aborda termos técnicos utilizados na engenharia de software. Outro trabalho semelhante é o de [Lima and Araújo 2018] que disponibiliza uma plataforma *online* denominada Dicionário de Informática em Libras (DINF).

¹<https://www.gov.br/governodigital/pt-br/vlibras>

No trabalho de [Granada et al. 2022] foi apresentado o iTecDeaf, um glossário técnico para surdos que se baseia na teoria de Vygotsky. A proposta utiliza vídeos, animações em Libras e datilologia na forma escrita. Os autores utilizaram a avaliação preditiva a fim de identificar ajustes no design.

Por fim, a proposta de [Silva Junior et al. 2023] é mapear os sinais em Libras que correspondem às terminologias específicas da área de engenharia de software. Os autores apontaram a carência de recursos para a tradução de termos específicos dessa área.

Para identificar características entre os trabalhos relacionados apresentados nesta seção, foram definidos os seguintes critérios de avaliação e, posteriormente, comparados na Tabela 1 (👍 sinaliza que o trabalho possui a característica indicada, 👎 sinaliza que o trabalho não possui a característica indicada e ? os autores não conseguiram identificar essa característica no texto):

- (C1) – constrói novos sinais em Libras para termos técnicos;
- (C2) – apresenta glossário em língua portuguesa contendo termos técnicos;
- (C3) – apresenta glossário em Libras contendo termos técnicos;
- (C4) – atualiza o glossário de forma automática;
- (C5) – categoriza os termos do glossário;
- (C6) – disponibiliza o glossário por meio de API REST; e
- (C7) – plataforma em que a aplicação é executada.

Tabela 1. Comparação das características entre os trabalhos relacionados.

Trabalho	(C1)	(C2)	(C3)	(C4)	(C5)	(C6)	(C7)
[Faqueti et al. 2005]	👍	👎	👎	👎	👎	👎	Web
[Barwaldt et al. 2017]	👍	👎	👎	👎	👎	👎	?
[Calé et al. 2017]	👍	👎	👎	👎	👍	👎	Mobile
[da Cruz et al. 2017]	👍	👎	👎	👎	👍	👎	Mobile
[Fabris et al. 2018]	👍	👎	👎	👎	👎	👎	?
[Lima and Araújo 2018]	👍	?	?	👎	👎	👎	Web
[Granada et al. 2022]	👎	👍	👍	👎	👍	👎	Web
[Silva Junior et al. 2023]	👎	👎	👎	👎	👎	👎	?
Esta nova versão	👎	👍	👍	👍	👍	👍	Web

Os critérios escolhidos foram definidos considerando a relevância para o desenvolvimento de uma aplicação voltada para a disseminação de termos técnicos em Libras. Eles foram escolhidos por proporcionar uma comparação abrangente dos trabalhos existentes, identificando lacunas que necessitam de melhorias em relação a este projeto.

3. Metodologia

A abordagem metodológica, utilizada no desenvolvimento do protótipo apresentado, é amplamente utilizada no contexto de engenharia de software por proporcionar uma estrutura adequada para desenvolvedores. Portanto a seguir, são apresentadas e detalhadas as etapas realizadas, a saber: 1) levantamento de requisitos, 2) especificação, 3) implementação e 4) testes.

3.1. Levantamento de requisitos

Para entender algumas das várias necessidades dos estudantes surdos, foram realizadas reuniões com intérpretes de Libras e uma psicopedagoga, com especialização em Libras. Todas elas são servidoras do IFRN.

Durante os quatro anos letivos em que o curso técnico integrado em informática é ministrado no IFRN, duas intérpretes em Libras desempenharam um papel crucial acompanhando um aluno surdo da instituição. Rotineiramente, elas se revezavam para fornecer interpretação simultânea durante as aulas frequentadas pelo aluno. Como resultado desse acompanhamento, elas contribuíram para a compreensão dos requisitos funcionais, enumerados abaixo:

1. **glossário técnico** – disponibilizar um conjunto de termos técnicos e suas respectivas definições para estudantes surdos (e não-surdos) de informática;
2. **colaborativismo** – permitir que usuários surdos e intérpretes de Libras produzam vídeos explicando a definição dos termos presentes no glossário;
3. **autoconstrução** – permitir que o glossário seja atualizado automaticamente com novos termos sem que, necessariamente, o usuário final precise incluir;
4. **compartilhamento** – disponibilizar o glossário para aplicações clientes através de uma API REST.

Vale ressaltar também que seria importante a participação do aluno surdo na elucidação dos requisitos, porém por ter concluído o curso técnico e se desligado da instituição, ele não conseguiu participar das reuniões. De todo modo, as intérpretes conseguiram identificar as principais necessidades dele, enquanto aluno do IFRN.

3.2. Especificação

Construiu-se um diagrama de caso de uso (DCU) para compreender os requisitos e identificar os atores e suas funcionalidades. Por causa da limitação do número de páginas, optou-se por não apresentar o DCU, porém os principais casos de uso são listados abaixo:

1. Atualizar glossário automaticamente através do *web scraping*;
2. Pesquisar textual de termos do glossário;
3. Gerenciar atualizações no glossário;
4. Sugerir a inclusão de novos vídeos;
5. Validar sugestões de vídeos e de termos;
6. Compartilhar glossário através de uma API REST;
7. Realizar pesquisa semântica no glossário; e
8. Categorizar termos de forma automática.

Os casos de uso (6), (7) e (8) são as novidades da nova versão do protótipo apresentado. As novas *features* incluem: a) a implementação de uma API REST para disponibilizar o conjunto de termos do glossário às aplicações clientes; b) a pesquisa de termos do glossário através de buscas semânticas; e c) a classificação automática de novos termos, de acordo com as categorias pré-estabelecidas no glossário.

3.3. Implementação

Nesta etapa, construíram-se *mockups* com base nos principais casos de uso. Além disso, também foram discutidos os tipos de testes e as tecnologias a serem utilizados. A escolha das tecnologias foi fundamentada em suas capacidades técnicas e na adequação às necessidades específicas do projeto.

Para desenvolver o protótipo, optou-se por utilizar as tecnologias Figma, Node.js, TypeScript, MySQL, React, Next.js e Docker. Além delas, foram ainda usadas:

- Python – biblioteca *scrapy* para implementar o *web scraping*;
- Kysely – abstração da linguagem SQL para construir consultas no TypeScript;
- *Embeddings* da API OpenAI – medem o grau de similaridade entre strings;
- Tailwind CSS – *framework* CSS para construir interfaces do usuário;
- InstantSearch.js – biblioteca para construir interfaces de pesquisa no *front-end*; e
- CapRover – automação de *deploys* e gerenciamento de aplicações em servidores.

3.4. Testes

Para avaliar um protótipo, podem ser aplicados diferentes tipos de testes. Os aplicados para avaliar esta proposta foram os seguintes:

- de navegação – para verificar a facilidade de o usuário navegar pela interface;
- de compatibilidade – para garantir o funcionamento correto do protótipo em diferentes dispositivos; e
- de informação – para verificar se os pares termo/definição e sinal/interpretação são compreensíveis.

O protótipo está em desenvolvimento constante e, assim, ainda não foram realizadas validações com usuários finais. Todavia uma das estratégias pensadas para validação é por meio de *feedbacks* fornecidos pelos usuários surdos e não-surdos. O objetivo é garantir que esses usuários obtenham o máximo de entendimento sobre os termos do glossário.

4. GOLI: Glossário Online em Libras para Informática

Esta proposta é mais uma extensão do trabalho de conclusão de curso (TCC) de um aluno surdo, egresso do curso técnico integrado em informática (*campus* Parnamirim, IFRN). A concepção do TCC surgiu da dificuldade do estudante em entender termos de informática como “laço”, “chave estrangeira”, “objeto” e outros que conheceu durante o curso.

Esta proposta utiliza um *web scraping* para acessar glossários monolíngues e extrair o par termo/definição. O *web scraping* é uma abordagem de coleta de dados automatizada que permite, constantemente, a autoconstrução do glossário garantindo que novos termos e definições sejam incluídos no glossário. Além disso, outra contribuição deste trabalho é a disponibilização dos termos do glossário – por meio da implementação de uma API REST – para que os termos possam ser acessados por outras aplicações.

Na literatura pesquisada, inclusive em plataformas populares de vídeos como YouTube e TikTok, foram encontrados poucos trabalhos que explicam o significado de termos técnicos em Libras. É comum encontrar vídeos que apresentam sinais referentes a termos, contudo, uma pessoa surda que deseja conhecer conceitos da área de informática, apenas a apresentação do sinal não é suficiente para o seu entendimento.

A Figura 1 demonstra 08 (oito) interações representadas pelas letras A, B, C, D, E, F, G e H. Na figura, as interações têm o objetivo de mostrar como os fluxos de dados se comportam dentro da arquitetura. O fluxo A é disparado quando o módulo *web scraping* visita glossários monolíngues (em Português) e, como resposta através do fluxo B, obtém-se uma lista de novos termos extraídos. Dessa forma, o glossário é constantemente

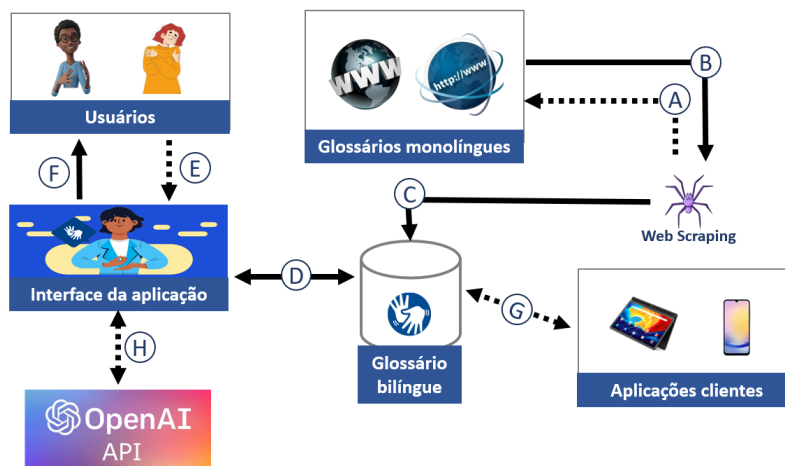


Figura 1. Arquitetura do protótipo proposto.

enriquecido com novos termos e definições em língua portuguesa, conforme fluxo C. A execução do módulo *web scraping* é agendada para disparar, nos domingos, às 3h.

Já o fluxo E representa um usuário que pesquisa termos do glossário. Esse fluxo também pode representar um usuário colaborando com o glossário como, por exemplo, sugerindo um vídeo que interpreta um termo. O fluxo D são as chamadas que a aplicação faz ao glossário e o F representa a exibição do resultado de uma pesquisa.

Nesta versão do protótipo, foi desenvolvida uma nova *feature* por meio da implementação de uma API REST, representada no fluxo G. Ela permite que desenvolvedores acessem o glossário facilmente e, assim, garantem que outras aplicações obtenham essas informações de forma transparente.

O protótipo permite a busca no glossário de diferentes maneiras, a saber: 1) digitando o termo de pesquisa; 2) explorando o termo clicando na primeira letra correspondente ao seu nome; 3) clicando na categoria a qual o termo está associado; ou 4) digitando uma expressão para realizar uma pesquisa semântica.

A pesquisa mais simples é pelo nome do termo e, assim, os termos que atenderem à condição de busca serão listados. Outra alternativa é a pesquisa pela letra inicial do termo, que corresponde aos índices do glossário. Ao clicar em um índice, é criada uma lista contendo os termos iniciados pela letra escolhida.

Além da pesquisa de termos pelo nome ou pela letra inicial, o protótipo também permite que sejam realizadas buscas categorizadas. Para isso, foram criadas 12 (doze) categorias, dentro das quais os autores acreditam que os termos do glossário podem ser agrupados. Termos como “*Query*”, “*Psb*” e “*Odbc*” estão classificados na categoria “Banco de Dados” e, portanto, ao selecionar essa categoria os referidos termos (e outros que atendem a esse critério de busca) serão listados. Caso o usuário deseje conhecer mais sobre o conceito, ele pode clicar no termo para exibir sua definição e para assistir ao vídeo correspondente em Libras como se observa na Figura 2. Na figura, também observam-se a descrição e a fonte (isto é, o glossário monolíngue) de onde foi extraída a definição.

A pesquisa semântica é uma nova *feature* desenvolvida nesta versão. Ela é implementada integrando-se à API OpenAI através de *embeddings*, como se vê no fluxo H. Ao

submeter uma consulta como “golpe online”, conforme Figura 3, o módulo realiza uma pesquisa considerando os termos e as definições que apresentam um determinado grau de similaridade com a consulta informada. O resultado da pesquisa considera *strings* exatas e *strings* que fazem parte do contexto da pesquisa. O modelo de *embedding* utilizado foi o *text-embedding-3-large*, a métrica de avaliação é a similaridade de cosseno e um *threshold* de similaridade igual a 80%. O motivo da escolha foi o fato de essa combinação ter apresentado os melhores resultados.

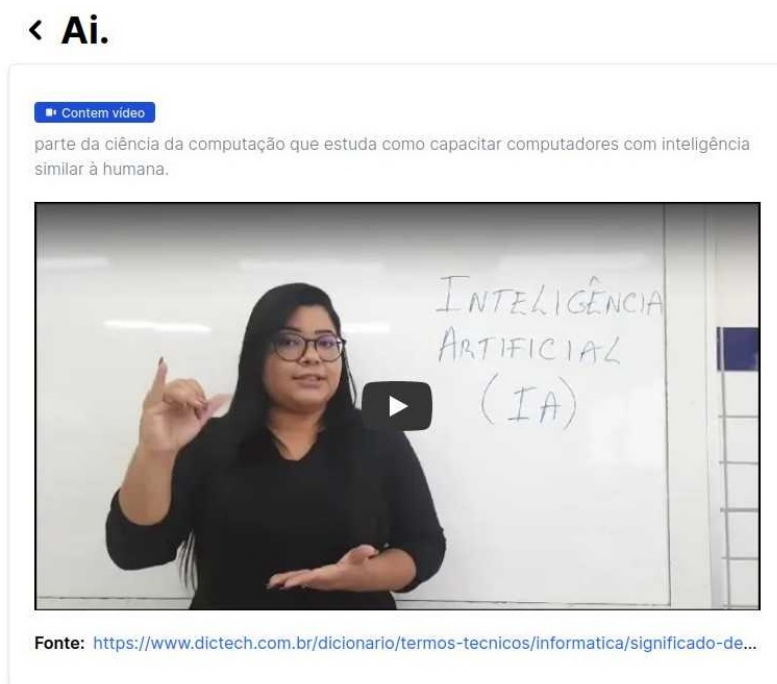


Figura 2. Interpretação de “Inteligência Artificial” e não apenas o sinal em Libras.

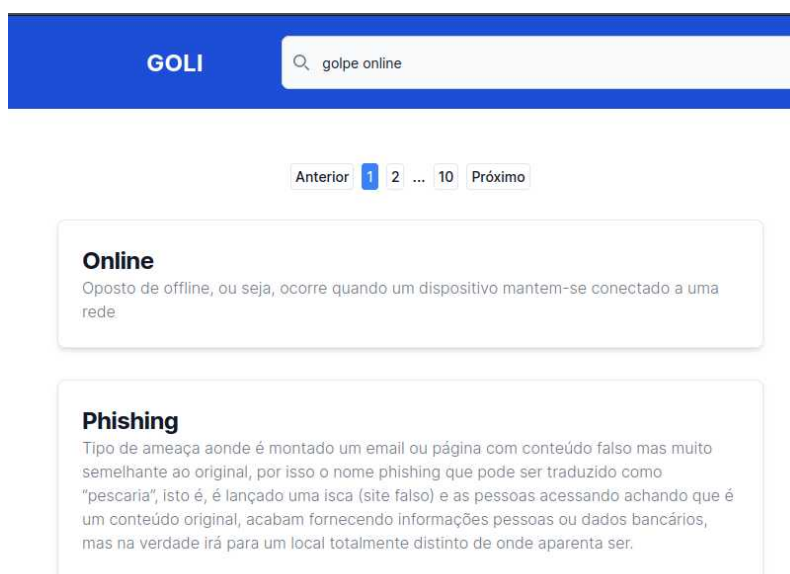


Figura 3. Pesquisa semântica usando a expressão “golpe online”.

5. Considerações finais

Este trabalho propõe uma alternativa para estudantes surdos da área de informática. O estudo desenvolveu um protótipo que oferece um glossário *online* em Libras, abrangendo termos técnicos para indivíduos surdos, independentemente de possuírem conhecimento prévio na área. O glossário é colaborativo e pode ser elaborado por usuários finais ou administradores ou, ainda, por meio de um técnicas de *web scraping*.

O protótipo apresentado pode gerar impactos na comunidade surda e na de profissionais da educação. Para os estudantes surdos, o GOLI promove o acesso ao conhecimento técnico com menores barreiras linguísticas. Isso pode contribuir para a redução das desigualdades educacionais e para a inclusão desses indivíduos nos cursos de informática. Para os educadores, o glossário oferece uma ferramenta integrada às práticas pedagógicas, promovendo aprendizagem mais inclusiva. Além disso, a contribuição colaborativa enriquece o conteúdo disponível garantindo a atualização contínua do glossário.

Como trabalhos futuros, pretende-se implementar estratégias de SEO (*search engine optimization*) visando melhorar a utilidade do glossário para a comunidade surda, bem como para outros usuários interessados na área de informática e, além disso, disponibilizar o serviço para usuários externos ao IFRN para maiores *feedbacks*.

Referências

- Barwaldt, R., Cesário, V., Domingues, D., and Granada, R. (2017). A linguagem logo como alternativa para o aprendizado de alunos surdos.
- Calé, F. R., Paim, L. N., Volpasso, L. H. S., Benassi, P., Miranda, R., Meneghini, T. L., and Cruz, S. (2017). Librasti: Uma aplicação móvel para levar o vocabulário da tecnologia da informação ao público surdo. *XXII ENAPET-Brasília*, 6pp.
- da Cruz, S. M. S., Calé, F., Paim, L., Benassi, P., and Miranda, R. (2017). Uma ferramenta para auxiliar o ensino da tecnologia da informação para surdos. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, volume 6, page 244.
- Fabris, J. A., Prietch, S. S., and Ricardi, K. (2018). Construção colaborativa de signos específicos da língua brasileira de sinais para termos da subárea de engenharia de software. In *Anais do XXVI Workshop sobre Educação em Computação*. SBC.
- Faqueti, C. G., Grandi, G., Fantini, L. S., and Lorenzetti, M. L. (2005). Infolibras—o uso da web para o aprendizado da língua de sinais com termos da informática. In *Anais do Workshop de Informática na Escola*, volume 1.
- Granada, R. P., Barwaldt, R., Gentil, N., and Lussanrriaga, M. D. (2022). Itecdeaf-glossário técnico em libras com interface dinâmica para surdos. *Educação em Foco*, 27(1):27005–27005.
- IBGE (2022). Pessoas com deficiência e as desigualdades sociais no brasil.
- Lima, G. and Araújo, R. (2018). Dinf: Uma plataforma online para termos da área de informática em libras. In *Anais do XXIV Workshop de Informática na Escola*, pages 652–659. SBC.
- Silva Junior, Nildo Pereira e Melo, S. M., Rech, G. C., and Prates, J. M. (2023). Mapeamento de sinais em libras para o apoio ao ensino de engenharia de software. *Revista Novas Tecnologias na Educação*, 21(1):180–189.