

Estudo e Desenvolvimento de um Player de Vídeo Acessível Para Libras Explorando a Arquitetura *Speech2Learning*

Resumo

O objetivo principal desta proposta de trabalho de iniciação científica é desenvolver um player de vídeo acessível para a comunidade surda, particularmente aqueles que se comunicam usando a Língua Brasileira de Sinais (Libras). Este player contará com um avatar de Libras integrado à transcrição fornecida por uma instância da arquitetura denominada *Speech2Learning*. A necessidade deste player surge do fato de que a comunidade surda, especialmente aquelas que se comunicam usando línguas de sinais, enfrentam obstáculos significativos para acessar conteúdos audíveis. Esta questão é agravada pela diversidade de línguas de sinais usadas em todo o mundo, o que torna difícil a padronização da interpretação e tradução dessas línguas. No Brasil, estima-se que cerca de 9,7 milhões de pessoas têm deficiência auditiva, representando cerca de 5,1% da população. Destes, muitos usam a Libras para se comunicar. Dada a importância desta língua no país, a proposta de desenvolver um player de vídeo acessível, que integra um avatar de Libras, é particularmente relevante. Além disso, o advento da tecnologia tem apresentado oportunidades significativas para melhorar a acessibilidade no campo da educação. As Tecnologias de Informação e Comunicação (TICs) têm o potencial de tornar as práticas educacionais mais inclusivas, personalizando o processo de ensino e aprendizagem para o contexto dos alunos. Neste contexto, a arquitetura *Speech2Learning*, que se baseia na transcrição de fala para fornecer interpretação visual por meio de um avatar, mostra-se como uma ferramenta poderosa para melhorar a acessibilidade de conteúdos educacionais audíveis para a comunidade surda.

1 Justificativa

Estima-se que 1,57 bilhão de pessoas em todo o mundo tenham algum grau de perda auditiva, o que representa cerca de 20% da população global (FORBES, 2023). Uma parte significativa dessas pessoas se comunica por meio das línguas de sinais. Essas línguas permitem que conversas e informações sejam transmitidas visualmente usando diferentes formas de comunicação sinalizadas. Seus usuários combinam movimentos articulados das mãos, expressões faciais e movimentos corporais para se comunicarem (DUKE, 2009; QUADROS, 2019; HONORA; FRIZANCO, 2017).

No entanto, a *World Federation of the Deaf* (WFD) destaca que não existe uma língua de sinais universal, principalmente devido às singularidades linguísticas em questão, as quais impõem barreiras contundentes para a universalidade das línguas de sinais (ONU, 2023; QUADROS, 2019). Segundo a WFD, há aproximadamente 70 milhões de surdos em todo o mundo, que coletivamente utilizam mais de 300 línguas de sinais diferentes (ONU, 2023).

Dessa forma, a diversidade linguística gera uma falta de consistência na interpretação e tradução das línguas de sinais, evidenciando uma linha de pesquisa relevante (NAPIER; WIT, 2019). De acordo com (DUKE, 2009), em um cenário hipotético em que todas as línguas de sinais fossem unificadas, elas representariam o terceiro idioma mais usado nos Estados Unidos e o quarto no mundo.

No Brasil, a Língua Brasileira de Sinais (Libras) foi oficialmente reconhecida em 2002 (QUADROS, 2019; HONORA; FRIZANCO, 2017). Desde então, várias iniciativas foram promovidas para a inclusão social e cultural dos usuários de Libras. Segundo o último censo do Instituto Brasileiro de Geografia e Estatística (IBGE), há cerca de 9,7 milhões de pessoas com deficiência auditiva no Brasil, representando cerca de 5,1% da população (IBGE, 2010). É importante ressaltar que esses dados não foram coletados especificamente sobre a Libras, mas evidenciam a relevância dessa língua no país.

Por outro lado, o advento da tecnologia tem impactado significativamente o processo de ensino e aprendizagem, com soluções inovadoras, como o uso de dispositivos inteligentes e técnicas de transcrição/tradução simultânea, que podem ajudar no desenvolvimento de aplicações educacionais efetivas e acessíveis para os aprendizes das línguas de sinais (NAPIER; WIT, 2019).

Na prática, as Tecnologias de Informação e Comunicação (TICs) têm alterado as interações pessoais e as práticas de ensino. A globalização das TICs criou um contexto educacional sem precedentes, mais flexível, conectado e inteligente, com tecnologias portáteis e redes de computadores cada vez mais presentes na vida cotidiana (CILLI et al., 2017).

Em um contexto amplo, a educação é o meio pelo qual os hábitos, costumes e valores de um determinado grupo são transferidos de uma geração para a próxima. É um processo contínuo que desenvolve as capacidades físicas, intelectuais e morais do ser humano, permitindo uma integração efetiva dentro da sua comunidade (CILLI et al., 2017; QUADROS, 2019).

Neste sentido, a inclusão social é crucial para garantir o acesso adequado à educação, independentemente da língua dominante do indivíduo, seja ela de sinais ou falada (QUADROS, 2019). A educação é um processo constante e evolutivo que visa a integração das pessoas na sociedade ou grupo ao qual pertencem (CILLI et al., 2017; QUADROS, 2019). Portanto, a inclusão social e digital são aspectos significativos no processo educacional contemporâneo.

As TICs podem desempenhar um papel fundamental neste cenário, contribuindo para tornar as práticas educacionais mais inclusivas e personalizando o processo de ensino e aprendizagem ao contexto dos alunos. Este benefício é especialmente relevante em cenários onde a diversidade linguística é fundamental, como na educação bilíngue para usuários de línguas de sinais. A necessidade de tais tecnologias é realçada quando se considera a diversidade e a singularidade das línguas de sinais e as dificuldades que isso pode trazer para a acessibilidade de conteúdos audíveis.

Diante do cenário apresentado, o trabalho de doutorado de Falvo Jr (FALVOJR; BARBOSA, 2023) se alinha à busca de soluções que empregam as TICs para melhorar a acessibilidade no campo da educação. A proposta de Falvo Jr (FALVOJR; BARBOSA, 2023) é uma arquitetura denominada *Speech2Learning*, fundamentada em um Mapeamento Sistemático (MS) (FALVOJR et al., 2020b; FALVOJR; SCATALON; BARBOSA, 2020; FALVOJR et al., 2020a), que busca favorecer a acessibilidade de conteúdos educacionais audíveis através do reconhecimento de fala. Nesse contexto, o principal objetivo deste projeto é desenvolver um player de vídeo acessível para surdos. Este player contará com um avatar de Libras integrado à transcrição fornecida por uma instância da *Speech2Learning*. Os objetivos e as principais atividades associadas a este projeto serão sumarizados nas próximas seções.

2 Resultados Anteriores

Como comentado anteriormente, a Arquitetura *Speech2Learning* foi proposta após a condução de um MS, cujo objetivo foi identificar o papel da tecnologia no ensino e aprendizagem por meio das línguas de sinais (FALVOJR et al., 2020b; FALVOJR; SCATALON; BARBOSA, 2020; FALVOJR et al., 2020a). O MS identificou 185 estudos primários, proporcionando uma visão geral das principais soluções e delimitando as publicações focadas na Libras.

De modo geral, o MS evidenciou que a tecnologia tem desempenhado um papel significativo no ensino e aprendizagem por meio das línguas de sinais. No entanto, observou-se uma falta de padrões e boas práticas de desenvolvimento que poderiam facilitar o compartilhamento de objetos de aprendizagem. Em resposta a esta lacuna, a Arquitetura *Speech2Learning* foi criada para incentivar a criação de soluções estruturalmente preparadas para a inclusão de pessoas surdas no processo de ensino-aprendizagem por meio da acessibilidade de conteúdos educacionais audíveis.

Tecnicamente, a *Speech2Learning* propõe um arcabouço genérico que vai além das línguas de sinais. No entanto, ele fornece uma abstração que facilita a acessibilidade de objetos de aprendizagem audíveis, permitindo a geração de transcrições e, conseqüentemente, a sinalização desses objetos. Nesse contexto, soluções baseadas em avatares, tais como

*Hand Talk*¹ ou *VLibras*², que se destacaram no MS (FALVOJR et al., 2020b; FALVOJR; SCATALON; BARBOSA, 2020; FALVOJR et al., 2020a), podem ser utilizadas com base nas transcrições, tornando os conteúdos acessíveis para a comunidade das línguas de sinais.

Sendo assim, foi proposta a Arquitetura *Speech2Learning*, uma adaptação da Clean Architecture (MARTIN, 2012) com o objetivo específico de promover a acessibilidade de objetos de aprendizagem por meio do reconhecimento de fala. Por sua vez, a Clean Architecture (MARTIN, 2012) é uma ideia prática que integra algumas das principais referências em Engenharia de Software nas últimas décadas (COCKBURN, 2005; FREEMAN; PRYCE, 2009; PALERMO, 2008; COPLIEN; REENSKAUG, 2012; REENSKAUG, 2009; JACOBSON, 1992).

Essas iniciativas compartilham a ideia central de separar o código em camadas independentes, com o domínio no núcleo da arquitetura. Isso permite a criação de sistemas altamente testáveis, independentes de tecnologia, e adaptáveis às necessidades específicas de um projeto.

A Arquitetura *Speech2Learning*, alinhada a esses princípios, é apresentada na Figura 1. Cada camada desta arquitetura contribui de forma significativa para a inclusão de metadados nos objetos de aprendizagem, os quais são cruciais para melhorar a acessibilidade.

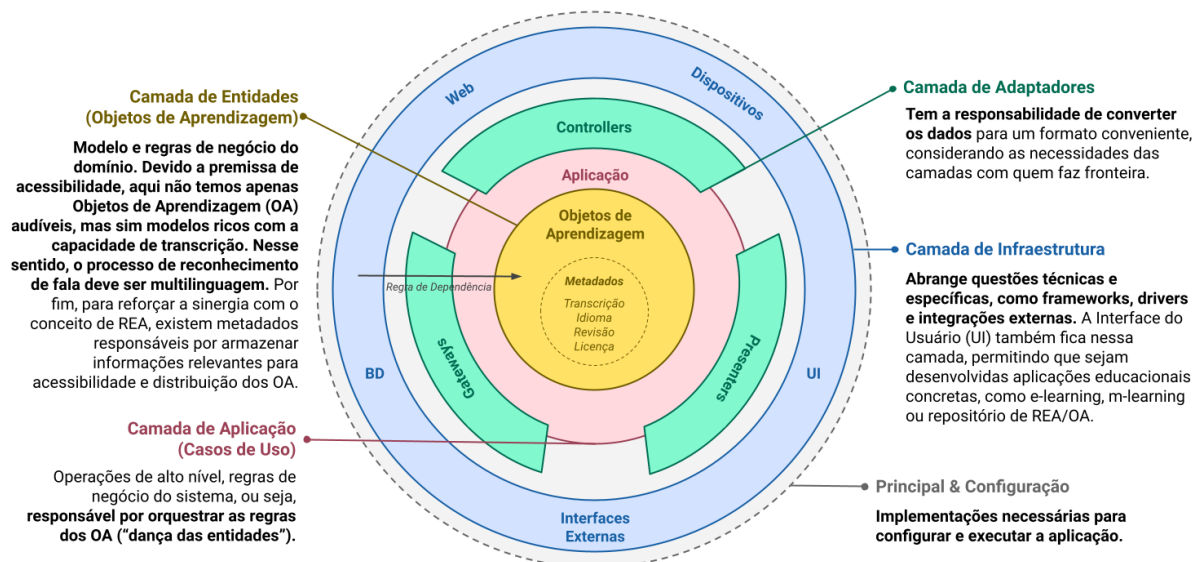


Figura 1 – Diagrama Arquitetural *Speech2Learning*

Esta representação visual da Arquitetura *Speech2Learning* é essencial para entender o funcionamento de suas instâncias. No contexto do player de vídeo, por exemplo, tem-se uma instância construída como uma API REST (estilo arquitetural baseado no protocolo HTTP). A Figura 2 demonstra esta interação em um diagrama de sequência.

¹ <https://www.handtalk.me>

² <https://www.gov.br/governodigital/pt-br/vlibras>

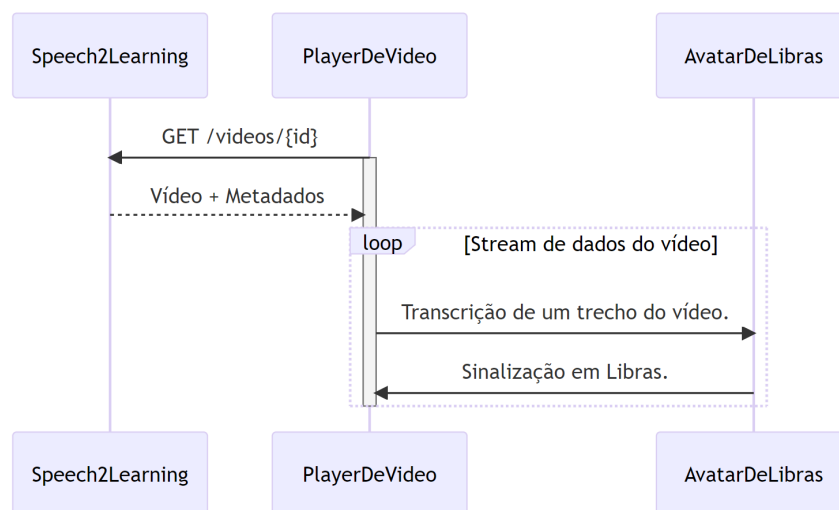


Figura 2 – Diagrama de sequência com a integração entre o Player de Vídeo e uma Instância da *Speech2Learning* (API REST)

No diagrama apresentado, uma instância *Speech2Learning* é vista desempenhando seu papel como intermediária no fornecimento de metadados relevantes para o player de vídeo. Esses metadados, que podem incluir transcrições ou legendas, podem promover a acessibilidade do conteúdo do vídeo.

Esta interação entre a instância *Speech2Learning* e o player de vídeo destaca a flexibilidade e eficácia da Arquitetura *Speech2Learning* no apoio à criação de soluções mais inclusivas de aprendizado. Ao adotar essa arquitetura, os desenvolvedores têm a capacidade de criar sistemas que promovem a acessibilidade, são altamente testáveis, independentes de tecnologia e adaptáveis às necessidades específicas de seus projetos.

3 Objetivos

Tendo em vista o contexto apresentado, este projeto tem como principal objetivo desenvolver, avaliar e evoluir um player de vídeo acessível e integrado a uma instância da arquitetura *Speech2Learning*, visando a educação inclusiva para surdos por meio da Libras. Em linhas gerais, será desenvolvido um componente de software genérico para a reprodução de vídeos e interpretação de transcrições/legendas por meio de um avatar de Libras.

4 Métodos

Na busca de um modelo eficiente para o desenvolvimento de software, este trabalho adota uma abordagem híbrida, que combina elementos de várias metodologias ágeis para se ajustar de forma mais adequada às especificidades do projeto. Essa estratégia tem como objetivo minimizar riscos por meio do desenvolvimento de software em ciclos curtos

e contínuos, conhecidos como iterações, geralmente variando de uma a quatro semanas (BECK et al., 2001).

Cada iteração é tratada como um mini-projeto de software, envolvendo todas as tarefas necessárias para entregar um pequeno incremento de funcionalidade: planejamento, análise de requisitos, design, codificação, teste e documentação (KNIBERG; SKARIN, 2014). Em vez de focar apenas em adicionar um conjunto substancial de novas funcionalidades a cada iteração, o projeto busca a capacidade de entregar uma nova versão do software no final de cada ciclo (BECK et al., 2001). Esta versão é então usada para reavaliar as prioridades do projeto.

O método híbrido a ser adotado neste trabalho une aspectos do Scrum e do Kanban, criando um ambiente propício para a implementação de uma variedade de processos e técnicas (AHMAD; MARKKULA; OIVO, 2013). Ambas as estruturas são voltadas para tornar visível a eficácia das práticas de desenvolvimento, possibilitando seu aprimoramento, ao mesmo tempo em que proporcionam um ambiente em que produtos complexos podem ser desenvolvidos (SCHWABER; BEEDLE, 2002).

Após a conclusão do desenvolvimento, estudos empíricos devem ser realizados para avaliar a aplicabilidade do software no contexto real. Os testes de usabilidade podem ser realizados com os usuários para avaliar aspectos da interação usuário-computador, complementados por avaliações heurísticas realizadas por especialistas na área. Adicionalmente, os testes em campo com uma amostra de usuários finais do software também devem ser realizados para comprovar a aplicabilidade em seu contexto real, bem como para identificar possíveis melhorias (NIELSEN, 1993).

O uso combinado do Scrum e Kanban neste projeto híbrido permite a implementação de um processo de desenvolvimento flexível, eficiente e adaptável. Esta metodologia híbrida visa proporcionar uma visão mais clara da eficácia relativa das práticas de desenvolvimento, facilitando sua melhoria e, ao mesmo tempo, fornecer um ambiente eficaz para o desenvolvimento de produtos complexos (AHMAD; MARKKULA; OIVO, 2013).

5 Atividades a Serem Desenvolvidas

As principais atividades a serem conduzidas para o cumprimento dos objetivos do trabalho proposto são apresentadas a seguir:

1. **Estudo dos conceitos de desenvolvimento de software:** entendimento dos principais conceitos referentes ao desenvolvimento de software, incluindo a metodologia Agile, Scrum e Kanban.
2. **Estudo de tecnologias e soluções para players de vídeo:** compreensão dos principais conceitos e tecnologias envolvidas na reprodução de vídeo, incluindo codecs,

padrões de transmissão e protocolos. Além disso, envolve a pesquisa e análise de soluções open-source existentes para players de vídeo, incluindo a análise do código, avaliação da comunidade de desenvolvimento e revisão da documentação.

3. **Estudo de tecnologias e soluções de avatar para Libras:** percepção das diferentes tecnologias e soluções de avatares para sinalização em Libras que poderão ser integradas ao player de vídeo.
4. **Benchmark de players de vídeo comerciais:** análise e comparação dos principais players de vídeo comerciais do mercado, focando na avaliação de funcionalidades, facilidade de uso e recursos de acessibilidade.
5. **Definição das funcionalidades do Player de Vídeo:** definição de quais funcionalidades devem estar presentes na nova versão do player de vídeo, incluindo funcionalidades de acessibilidade como legendas e sinalização em Libras.
6. **Desenvolvimento do Player de Vídeo:** implementação do player de vídeo com base nos resultados das atividades anteriores, com especial atenção para a integração com a API do *Speech2Learning* e a incorporação de avatares para Libras.
7. **Avaliação do Player de Vídeo:** condução de estudos empíricos para avaliar o player de vídeo desenvolvido, considerando a perspectiva de diferentes tipos de usuários, incluindo aqueles com necessidades especiais.
8. **Evolução do Player de Vídeo:** evolução do player de vídeo com base nos resultados obtidos na Atividade 7. Identificadas as melhorias necessárias, o player de vídeo passará por uma fase de aperfeiçoamento e evolução, com foco na acessibilidade e na integração com avatares para Libras.
9. **Elaboração de artigos e relatórios:** formalizar as etapas conduzidas durante o desenvolvimento do trabalho e os resultados obtidos a partir dessa experiência. Os artigos desenvolvidos devem ser submetidos a congressos de iniciação científica e outros congressos nas áreas de interesse.

6 Resultados Esperados e Indicadores de Avaliação

O objetivo deste projeto de iniciação científica é contribuir para o desenvolvimento e aprimoramento das tecnologias de reprodutores de vídeo acessíveis. De forma mais específica, os resultados esperados são:

- Desenvolver um player de vídeo agnóstico, capaz de se integrar a qualquer sistema que suporte transcrição ou legenda, com funcionalidade adicional para ativação de avatares para sinalização em Libras.

- Ampliar as pesquisas iniciadas por Falvo Jr (FALVOJR; BARBOSA, 2023) sobre acessibilidade em conteúdos educacionais audíveis, com potencial de aplicação em edtechs de grande escala.
- Fornecer insights valiosos para a comunidade de desenvolvimento de software, destacando a importância da acessibilidade e proporcionando uma solução prática para a implementação de recursos de acessibilidade em reprodutores de vídeo.

Nesse contexto, os indicadores de avaliação do projeto, qualitativos e quantitativos, permitirão mensurar o sucesso da solução desenvolvida. Os indicadores mais relevantes são:

- Feedback dos usuários (qualitativa): A opinião dos usuários será coletada por meio de pesquisas de satisfação e usabilidade, fornecendo insights sobre a facilidade de uso e a eficácia do reprodutor de vídeo acessível, particularmente para usuários surdos.
- Facilidade de integração com sistemas existentes (quantitativa): A facilidade de integração do player de vídeo será avaliada, comparando-a com a de seus principais concorrentes, em consonância com o benchmarking planejado nas atividades a serem desenvolvidas.

7 Cronograma de Execução

O cronograma de execução das atividades anteriormente definidas é apresentado na Tabela 1.

Tabela 1 – Cronograma de atividades.

Atividades:	2023					2024						
	Ago	Set	Out	Nov	Dez	Jan	Fev	Mar	Abr	Mai	Jun	Jul
1. Conceitos - Metodologia Ágil												
2. Conceitos - Players de Vídeo												
3. Conceitos - Avatares de Libras												
4. Benchmark de Mercado												
5. Definição das Funcionalidades												
6. Desenvolvimento do Player												
7. Avaliação do Player												
8. Evolução do Player												
9. Elaboração de Artigos												

Referências

- AHMAD, M. O.; MARKKULA, J.; OIVO, M. Kanban in software development: A systematic literature review. In: IEEE. *2013 39th Euromicro Conference on Software Engineering and Advanced Applications*. [S.l.], 2013. p. 9–16. Citado na página 6.
- BECK, K. et al. *Manifesto for Agile Software Development*. [S.l.]: Agile Alliance, 2001. Citado na página 6.
- CILLI, T. et al. *A Tecnologia da Informação e Comunicação nas Práticas Educacionais*. São Paulo-SP: Edição Independente, 2017. Citado na página 2.
- COCKBURN, A. *Hexagonal Architecture*. 2005. Disponível em: <<https://bit.ly/3NiKuOu>>. Citado na página 4.
- COPLIEN, J. O.; REENSKAUG, T. *Agile Software Architecture: Chapter 2. The DCI Paradigm*. [S.l.]: Morgan Kaufmann, 2012. Citado na página 4.
- DUKE, I. *The Everything Sign Language Book: American Sign Language Made Easy*. Avon, MA: Adams Media Corporation, 2009. Citado 2 vezes nas páginas 1 e 2.
- FALVOJR, V.; BARBOSA, E. F. *Speech2Learning: Uma Arquitetura de Apoio à Acessibilidade de Objetos de Aprendizagem Audíveis*. 2023. Projeto de doutorado em andamento. Instituto de Ciências Matemáticas e de Computação – ICMC/USP, São Carlos-SP. Citado 2 vezes nas páginas 3 e 8.
- FALVOJR, V. et al. Tecnologias Aplicadas ao Ensino e Aprendizagem com Línguas de Sinais: Um Mapeamento Sistemático Sob as Perspectivas Nacional e Internacional. *Revista Novas Tecnologias na Educação (RENOTE)*, p. 1–13, 2020. Citado 2 vezes nas páginas 3 e 4.
- FALVOJR, V. et al. Tecnologias Aplicadas ao Ensino e Aprendizagem de LIBRAS: Um Mapeamento Sistemático. In: *Anais do Simpósio Brasileiro de Informática na Educação (SBIE 2020)*. Natal, RS, Brasil: [s.n.], 2020. p. 1–10. Citado 2 vezes nas páginas 3 e 4.
- FALVOJR, V.; SCATALON, L. P.; BARBOSA, E. F. The Role of Technology to Teaching and Learning Sign Languages: A Systematic Mapping. In: *Proceedings of the 50th Annual Frontiers in Education Conference (FIE 2020)*. Uppsala, Sweden: [s.n.], 2020. p. 1–10. Citado 2 vezes nas páginas 3 e 4.
- FORBES. *Deafness Statistics: 2023*. Forbes, 2023. Disponível em: <<https://bit.ly/3P0w8ni>>. Citado na página 1.
- FREEMAN, S.; PRYCE, N. *Growing Object-Oriented Software, Guided by Tests*. [S.l.]: Addison-Wesley, 2009. Citado na página 4.
- HONORA, M.; FRIZANCO, M. L. E. *Livro Ilustrado de Língua Brasileira de Sinais – Desvendando a comunicação usada pelas pessoas com surdez*. [S.l.]: Ciranda Cultural, 2017. Citado 2 vezes nas páginas 1 e 2.
- IBGE. *Conheça o Brasil - População: Pessoas com Deficiência*. 2010. Disponível em: <<https://bit.ly/2Zf79lV>>. Citado na página 2.
- JACOBSON, I. *Object-Oriented Software Engineering: A Use Case Driven Approach*. [S.l.]: ACM Press, 1992. Citado na página 4.

KNIBERG, H.; SKARIN, M. *Scrum and XP from the trenches*. [S.l.]: C4Media, 2014. Citado na página 6.

MARTIN, R. C. *Clean Architecture*. [S.l.]: Prentice Hall, 2012. Citado na página 4.

NAPIER, J.; WIT, M. D. *WFD Position Paper on Accessibility: Sign Language Interpreting and translation and technological developments*. 2019. Disponível em: <<https://bit.ly/36rwhYP>>. Acesso em: 02/12/2020. Citado na página 2.

NIELSEN, J. *Usability engineering*. [S.l.]: Elsevier, 1993. Citado na página 6.

ONU. *United Nations: International Day of Sign Languages*. 2023. Disponível em: <<https://bit.ly/3CBOCTX>>. Citado na página 2.

PALERMO, J. *Onion Architecture*. 2008. Disponível em: <<https://bit.ly/43v9rvN>>. Citado na página 4.

QUADROS, R. M. de. *LIBRAS – Língua para o Ensino Superior*. [S.l.]: Parábola Editorial, 2019. Citado 2 vezes nas páginas 1 e 2.

REENSKAUG, T. *The DCI Architecture: A New Vision of Object-Oriented Programming*. [S.l.], 2009. Disponível em: <<https://bit.ly/42rEAyZ>>. Citado na página 4.

SCHWABER, K.; BEEDLE, M. *Agile software development with Scrum*. [S.l.]: Prentice Hall, 2002. Citado na página 6.