

Construção de um Léxico Computacional de Peixes Ornamentais Marinhos do Brasil com Aplicações em PLN e Funções Hash

André Akira Muraoka, DC/UEL, andre.akira.muraokaa@uel.br,
<https://orcid.org/0009-0007-9745-8286>

Bruno Henrique S. Viola, DC/UEL, bruno.henrique.silva.viola@uel.br,
<https://orcid.org/0009-0005-3914-9085>

Bruno Massaru Futata Kato, DC/UEL, brunoww.futata@uel.br,
<https://orcid.org/0009-0003-2368-5938>

Larissa Maria dos Santos Fonseca, DC/UEL, larissa.fonseca@uel.br,
<https://orcid.org/0009-0006-2754-0869>

Cinthyana Renata Sachs C. de Barbosa, PPGCC/UEL, cinthyana@uel.br,
<https://orcid.org/0000-0001-6248-6630>

Lucas Pereira Cassanho, PPGCC/UEL, lucas.cassanho@uel.br,
<https://orcid.org/0009-0004-6232-0461>

Resumo: A construção de léxicos estruturados é fundamental para o Processamento de Linguagem Natural em domínios técnicos. Este trabalho apresenta um léxico computacional voltado a peixes ornamentais marinhos, com 14 atributos definidos a partir de fontes especializadas. Foram coletadas e padronizadas informações de 150 espécies, acessíveis por meio de uma interface gráfica desenvolvida em Python. Foram consideradas diversas funções hash quanto à dispersão, colisões e tempo de acesso. A função One at a Time apresentou maior velocidade, enquanto a Universal obteve melhor dispersão. O léxico proposto possibilita consultas rápidas e pode ser utilizado como recurso educacional voltado ao estudo da ictiologia e à conscientização ambiental.

Palavras-chave: léxico, tabela hash, processamento de linguagem natural, ictiologia computacional, educação ambiental.

Computational Lexicon Construction of Brazilian Marine Ornamental Fish with Applications in NLP and Hash Functions

Abstract: Structured lexicon construction is fundamental for Natural Language Processing (NLP) in specialized domains. This study presents a computational lexicon focused for Brazilian marine ornamental fish, featuring 14 attributes derived from specialized sources. Data from 150 species were collected, standardized, and made accessible through a Python-developed graphical interface. Multiple hash functions were evaluated for dispersion, collisions handling, and access time. The One-at-a-Time hash function demonstrated the highest speed, while Universal hash function achieved superior dispersion. The proposed lexicon enables fast queries and can serve as an educational resource for ichthyology studies and environmental awareness initiatives.

Keywords: lexicon, hash table, natural language processing, computational ichthyology, environmental education.

1. Introdução

O comércio de Organismos Aquáticos Ornamentais, principalmente de peixes, caracteriza-se como uma atividade econômica com grande importância, sendo esse mercado responsável por movimentar bilhões de dólares por ano ao redor do globo (Sampaio e Nottingham, 2008), (Rezende e Fujimoto, 2021), (Carvalho *et al.*, 2023). Segundo o IBAMA (Brasil, 2022), por meio de uma portaria publicada em 2022 pelo órgão, o termo ‘ornamental’ se refere a espécies mantidas com o objetivo decorativo e ilustrativo, sendo a prática de criação e comércio desses animais conhecida como ‘aquariofilia’.

Carvalho *et al.* (2023) ressalta que muitos aquaristas (nomenclatura atribuída a quem cria essas espécies ornamentais por lazer ou comércio) são atraídos por características chamativas dos animais, como: cores vivas e variadas, formatos únicos, comportamentos em cardumes, entre outros atributos. Carvalho *et al.* (2023) e Sampaio e Nottingham (2008) ressaltam o potencial do Brasil por acomodar uma diversidade ampla de espécies que atendem a esses requisitos. No entanto, o extrativismo pode acarretar sérias consequências ao meio ambiente brasileiro.

Nesse contexto, a ictiologia, ramo da biologia dedicado ao estudo dos peixes, oferece subsídios relevantes para atividades relacionadas à criação, identificação e manejo de espécies ornamentais. A partir dessa disciplina, é possível acessar informações detalhadas sobre aspectos morfológicos, comportamentais e ecológicos das espécies, contribuindo para uma compreensão mais acurada da diversidade presente em ambientes naturais e artificiais. No contexto da aquariofilia e da fiscalização ambiental, esse conhecimento auxilia na diferenciação de espécies com características fenotípicas semelhantes, na identificação de exigências específicas de manutenção em cativeiro e na adoção de práticas mais compatíveis com a conservação e o uso sustentável da fauna aquática.

Por essa via, Sampaio e Nottingham (2008) lançaram o ‘*Guia Para Identificação de Peixes Ornamentais Brasileiro*’ como uma ferramenta auxiliar à fiscalização, permitindo a identificação de espécies permitidas ou não para o comércio. Ressalta-se a importância de um guia prático e visual como esse devido à ausência de um documento similar no período publicado, lacuna que, inclusive, permanece até o presente momento. Esses autores ainda evidenciam possíveis usos desse documento por estudantes e pesquisadores, pois esse conta com informações detalhadas e ricas sobre cada espécime abordada.

Ademais, Lisbôa e Barbosa (2013) apontam para a relevância do processo de identificação, extração e organizações de termos técnicos e conhecimentos em áreas como a biologia. Esse procedimento permite a construção do que é entendido como “dicionários” pela área de Processamento de Linguagem Natural, os quais, segundo Lisbôa e Barbosa (2013) e Faria e Barbosa (2020), oportunizam que atividades como a transferência e o acesso a conhecimento específicos sejam facilitadas. Faria e Barbosa (2020) ainda demonstram a possibilidade de construção de ferramentas de apoio ao ensino, as quais dependem da base estabelecida pelo dicionário.

Nesse sentido, o presente trabalho propõe a criação de um Léxico Computacional de Peixes Ornamentais Marinhos do Brasil com base nos dados disponíveis no guia de identificação, oferecendo uma ferramenta alinhada às demandas de fiscalização, pesquisa e ensino. Ademais, o armazenamento desses dados é realizado com o auxílio de funções hash, visando aproveitar os benefícios associados à sua eficiência na recuperação de

informações. Outrossim, este artigo apresenta uma interface gráfica que propicia o acesso aos dados extraídos e organizados.

O artigo está estruturado da seguinte forma: a Seção 2 apresenta o Guia para Identificação de Peixes Ornamentais Brasileiros, fonte dos dados do léxico; a Seção 3 discute as aplicações pedagógicas do léxico em contextos formais e informais; a Seção 4 diferencia léxicos computacionais de dicionários tradicionais, ressaltando sua relevância para o Processamento de Linguagem Natural; a Seção 5 detalha as funções hash utilizadas para otimizar o armazenamento e o acesso aos dados; a Seção 6 apresenta trabalhos correlatos que fundamentam esta pesquisa; a Seção 7 descreve a metodologia adotada para coleta e organização dos dados; a Seção 8 aborda a implementação da interface gráfica para consulta ao léxico; a Seção 9 expõe os resultados dos testes comparativos das funções hash; e, por fim, a Seção 10 traz as considerações finais e perspectivas para trabalhos futuros.

2. Guia para Identificação de Peixes Ornamentais Brasileiros

O *Guia para Identificação de Peixes Ornamentais Brasileiros* (Sampaio e Nottingham, 2008) foi desenvolvido diante da ausência de publicações semelhantes e da demanda dos órgãos fiscalizadores por uma ferramenta que auxilie no controle das atividades de coleta, transporte e comercialização dessas espécies. Além de apresentar a distribuição geográfica dos peixes e suas características morfológicas e comportamentais – fundamentais para a identificação – o documento também aborda aspectos como comportamentos típicos, exigências para manutenção em cativeiro e detalhes relacionados à comercialização.

As descrições morfológicas das espécies utilizam termos técnicos também presentes na definição dos atributos estruturais e padrões de coloração no dicionário. A Figura 1 ilustra os termos referentes às características estruturais e a Figura 2 aqueles relacionados à coloração.

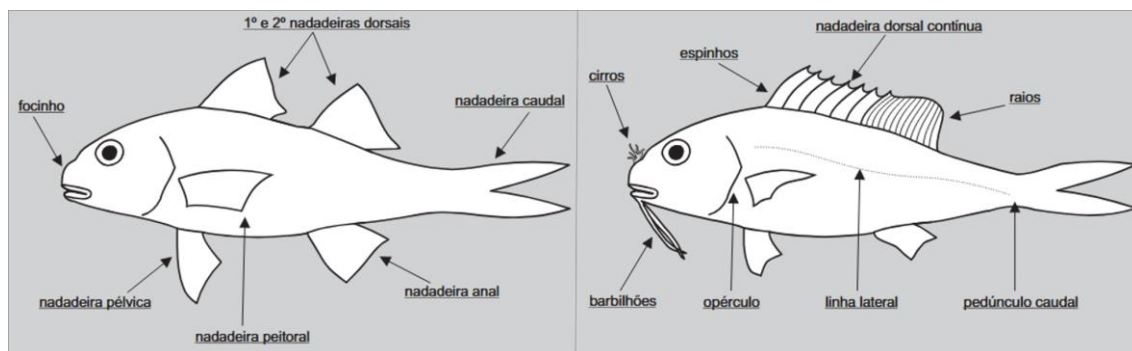


Figura 1. Ilustração das principais estruturas morfológicas externas dos peixes. Fonte: Sampaio e Nottingham (2008).

Adicionalmente, o guia reúne as principais legislações sobre crimes ambientais envolvendo a pesca, além de normas específicas para peixes ornamentais, pesca amadora, espécies ameaçadas de extinção e aqüicultura.

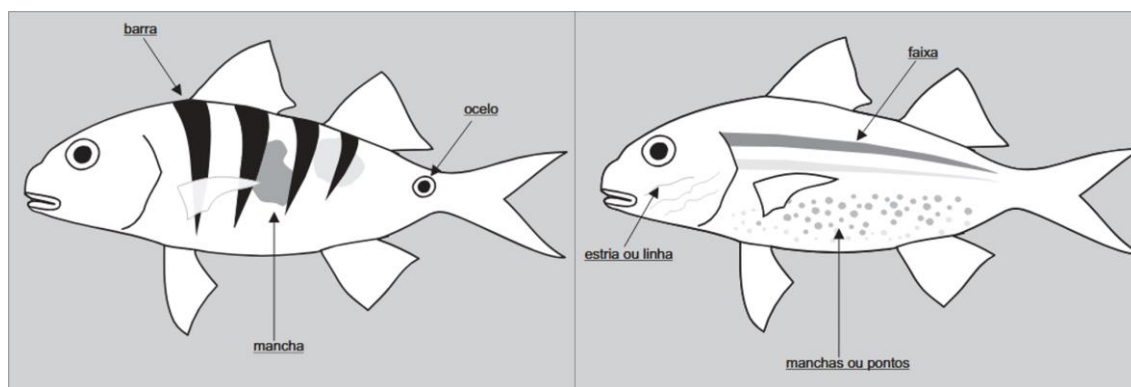


Figura 2. Exemplos de padrões de coloração em peixes ornamentais. Fonte: Sampaio e Nottingham (2008).

3. Aplicações Pedagógicas em Contextos Formais e Não Formais

A aplicação do léxico em contextos educativos, tanto formais quanto informais, abre possibilidades significativas de uso didático. Sua integração em ambientes como escolas, universidades, museus e oficinas pode enriquecer processos de ensino e aprendizagem, tornando o conteúdo científico mais acessível, interativo e relevante.

Nesse sentido, destaca-se o potencial dos léxicos computacionais quando integrados às plataformas educacionais. Essa combinação pode permitir que educadores monitorem o progresso dos alunos a partir da análise de suas interações com os termos do léxico, identifiquem dificuldades específicas e adaptem os conteúdos de forma personalizada. Além disso, os dados extraídos automaticamente podem ser utilizados para gerar relatórios pedagógicos, apoiar a tomada de decisão em sala de aula e enriquecer atividades de ensino contextualizadas com temas ambientais.

Uma abordagem semelhante é observada em Silva, Carvalho e Maciel (2021), que propuseram um *Learning Analytics Dashboard* (LAD) para organizar e visualizar dados educacionais obtidos no ambiente Moodle, com o objetivo de apoiar o processo de ensino-aprendizagem por meio da análise de padrões de comportamento dos estudantes. De modo análogo, o léxico proposto neste trabalho pode fornecer subsídios computacionais relevantes para análises pedagógicas automatizadas, promovendo experiências de aprendizagem mais personalizadas e eficazes.

De forma complementar, o estudo de Gomes *et al.* (2025) demonstra como dados estruturados podem ser utilizados para criar ferramentas de apoio à gestão educacional. Os autores desenvolveram uma aplicação voltada à análise de informações coletadas por meio de avaliações institucionais em uma universidade, permitindo a visualização de métricas e padrões que apoiam a tomada de decisões estratégicas em contextos acadêmicos. Embora com finalidades distintas, o léxico computacional apresentado neste trabalho, voltado à organização de informações sobre peixes ornamentais marinhos, compartilha o mesmo princípio orientador da estruturação e do uso inteligente de dados para fins educacionais. Enquanto a solução de Gomes *et al.* (2025) auxilia gestores no planejamento institucional, o léxico proposto visa apoiar práticas pedagógicas, oferecendo recursos interativos que facilitam o acesso a informações precisas, o acompanhamento da aprendizagem e o estímulo à educação ambiental.

Além do uso em ambientes formais de ensino, o léxico também pode ser aplicado em contextos de educação não formal, como museus de ciências, centros de visitação ambiental e aquários. Nesses cenários, o léxico pode ser adaptado para compor sistemas interativos de consulta, totens digitais ou guias informativos que ofereçam ao público

informações acessíveis e cientificamente fundamentadas sobre os organismos marinhos locais. Tais aplicações não apenas democratizam o acesso ao conhecimento científico, como também fortalecem a consciência ambiental e a valorização da biodiversidade.

4. Diferenças entre Léxico Computacional e Dicionário Tradicional

No campo do Processamento de Linguagem Natural (PLN), a distinção entre léxicos computacionais e dicionários tradicionais é importante para o desenvolvimento de sistemas eficientes de análise e compreensão textual. Embora ambos os recursos envolvam o armazenamento de informações linguísticas, suas finalidades, estrutura e formas de acesso são distintas.

O dicionário tradicional, mesmo quando digitalizado, tem como público-alvo os seres humanos. Seu conteúdo é voltado para consulta por falantes da língua, apresentando definições, pronúncia e uso em contexto. Quando armazenado digitalmente, seu objetivo principal é facilitar o acesso, mas sua estrutura permanece voltada para leitura humana e não é diretamente utilizável por sistemas automatizados de PLN (Smarsaro, 2004).

Por outro lado, o léxico eletrônico ou base de dados lexical é um componente técnico desenvolvido com o objetivo específico de ser lido e processado por máquinas. Esses léxicos são estruturados como bases de dados que codificam informações linguísticas de maneira formal e padronizada, incluindo dados morfológicos, sintáticos, semânticos e pragmáticos, com o intuito de auxiliar tarefas como análise sintática, tradução automática, indexação e recuperação de informações (Gonzales e Lima, 2003), (Smarsaro, 2004).

Essa distinção torna-se particularmente relevante em contextos especializados, como no projeto Léxico de Orquídeas, desenvolvido por Lisbôa e Barbosa (2013), cujo objetivo é a catalogação e disponibilização de terminologia técnica relacionada às orquídeas. Embora o léxico possua valor documental e científico voltado ao público humano, sua estrutura em banco de dados relacional o qualifica também para aplicações computacionais, como a extração e análise automática de unidades terminológicas.

Além do domínio linguístico, estruturas como os léxicos computacionais também ganham relevância no contexto educacional, como discutido anteriormente, ao possibilitar o acompanhamento do progresso dos alunos e a personalização do aprendizado. Essa aplicabilidade se amplia quando consideramos estudos como o de Araújo e Lima (2025), que exemplificam como a organização computacional de dados pode ser aplicada em ambientes virtuais de aprendizagem para promover análises pedagógicas mais precisas e adaptativas, evidenciando a versatilidade dessa abordagem.

O presente trabalho, que tem potencial de aplicação em contextos educacionais, científicos e de fiscalização, também se beneficia dessa perspectiva ao propor a criação de um léxico para peixes ornamentais marinhos.

Dessa forma, compreender a diferença entre léxico e dicionário não é apenas uma questão conceitual, mas uma etapa importante para fundamentar o uso de estruturas computacionais no desenvolvimento de aplicações baseadas em PLN. No contexto deste trabalho, essa distinção justifica a escolha por uma abordagem lexical estruturada, com potencial para integrar tanto aspectos técnicos quanto educativos e informativos.

5. Funções Hash

A utilização de funções hash neste trabalho visa otimizar a organização e o acesso aos dados lexicais de peixes ornamentais marinhos do Brasil. Essas funções oferecem uma solução eficiente para o armazenamento e recuperação de informações, particularmente em aplicações que exigem alta performance com grandes conjuntos de dados. Funções hash são algoritmos que recebem uma entrada, denominada chave, e a transforma em um valor de tamanho fixo, conhecido como *hash code*.

É fundamental que essa transformação seja determinística, ou seja, a mesma chave deve sempre gerar o mesmo *hash code*. Nesse cenário, as funções hash são essenciais na construção de estruturas como as *tabelas de dispersão (hash tables)*, sendo amplamente utilizadas devido à sua capacidade de realizar operações de inserção, busca e remoção em tempo constante, na maioria das vezes (Gregório e Goya, 2019). Tais características tornam essa abordagem aplicável em diversas áreas da computação, incluindo armazenamento de dados, entre outras aplicações. Para garantir um desempenho adequado, uma função hash ideal deve apresentar algumas características essenciais, tais como:

- distribuição uniforme dos valores gerados minimizando colisões, ou seja, situações em que duas entradas distintas geram o mesmo código;
- baixa complexidade computacional assegurando que o tempo para gerar o código seja reduzido;
- simplicidade de implementação, o que permite sua aplicação em sistemas com diferentes capacidades computacionais.

Essas propriedades são essenciais para o desempenho eficiente de algoritmos que utilizam funções hash. Meneses e Henriques (2023) mostram que otimizar a distribuição uniforme e a simplicidade da função hash acelera significativamente o processamento em algoritmos complexos, tornando-os adequados para consultas rápidas em léxicos computacionais, como para acessar o nome científico de um peixe. Além disso, Moreno (2017) destaca que, em PLN e áreas educacionais, tabelas hash são soluções eficazes para organizar e consultar grandes volumes de dados, favorecendo o desenvolvimento de ferramentas lexicais que aprimoram o aprendizado, a compreensão e a tradução, especialmente para a Língua Portuguesa. Dessa forma, funções hash bem definidas contribuem para a criação de sistemas ágeis e funcionais.

No que tange à construção de léxicos é necessário empregar metodologias que organizem os termos de forma eficiente, viabilizando seu uso em tarefas como indexação, pesquisa, classificação e mineração de dados. Nesse cenário, as funções hash são utilizadas como estruturas de apoio para o armazenamento organizado dos termos, o que possibilita, por exemplo, a verificação rápida da existência de um item, sua associação a informações morfológicas e taxonômicas, e a apresentação desses dados por meio de interfaces amigáveis.

Portanto, as 10 funções hash (AHO, CRC, Divisão Cormen, Divisão Drozdek, Dobra, Multiplicação, *One-at-a-Time*, PJW, Universal e Zobrist) utilizadas por Moreno (2017) e Moreno, Barbosa e Manfio (2021) e adotadas no presente trabalho não se justificam apenas pela eficiência algorítmica, mas também pela compatibilidade com os objetivos do trabalho que envolvem acessibilidade, precisão e escalabilidade. Ao assegurar respostas rápidas às consultas realizadas por diferentes usuários, como pesquisadores, estudantes ou agentes de fiscalização, essa solução demonstra-se eficaz.

6. Trabalhos Correlatos

Uma outra iniciativa similar a esta pesquisa se trata do Visual Tahs, desenvolvido por Moreno, Barbosa e Manfio (2021). Essa ferramenta propôs a criação de um Léxico de Ervas com o acesso aos dados por meio de uma interface gráfica. Similarmente ao léxico proposto nesta presente pesquisa, o Visual Tahs utilizou um livro da área de Biologia sobre estudo de ervas, como fonte principal pelo caráter rico e organizado que essas obras oferecem.

O Léxico das Orquídeas, elaborado por Lisbôa e Barbosa (2013), também empregou um livro com 130 espécies de orquídeas descritas para sua construção. Porém, esse utilizou apenas 30 dessas espécies com 11 atributos cada.

O trabalho de Moreno, Barbosa e Manfio (2021) conta com 104 espécies e 12 atributos. O Léxico das Ervas e o das Orquídeas compreendem características também abrangidas pelo Léxico de Peixes Ornamentais Marinhos, como: nome científico, características físicas, nomes populares e foto da espécie. O Léxico das Ervas conta com a utilização de funções hash e a possibilidade de geração, via interface, de relatórios e gráficos sobre seu uso, servindo como inspiração para a aplicação nesta pesquisa. Ressalta-se que, enquanto o Léxico das Orquídeas de Lisbôa e Barbosa (2013) utilizou o Microsoft Access 2007 para o armazenamento dos dados, o Léxico das Ervas recorreu ao XML por sua simplicidade na manipulação dos dados.

Outra contribuição relevante é o trabalho de Santos Junior e Isquendo (2022), que apresenta a construção de um banco de dados lexicográfico em XML a partir de dados dialetais do Projeto ALiB (Atlas Linguístico do Brasil), com vistas à realização de PLN. Diferentemente da proposta de Moreno, Barbosa e Manfio (2021), que também emprega XML, esse estudo foca na organização e extração automática de informações linguísticas dialetais, além da futura integração com aplicações web. O Banco de Dados foi alimentado com informações orais transcritas de 24 localidades da região Norte do Brasil e tem como objetivo subsidiar a criação de um vocabulário dialetal *online*, promovendo a intersecção entre os campos da Linguística e da Tecnologia da Informação. Embora não utilize funções hash, esse trabalho integra léxicos especializados a estruturas computacionais voltadas à interface e à manipulação eficiente dos dados, o que possibilita a automatização de processos linguísticos e, assim como no presente estudo, a criação de um sistema de consulta lexical mais acessível.

De forma complementar, Manfio, Moreno e Barbosa (2014) desenvolveram o sistema Professor Tical, fundamentado no uso de tabelas hash para otimizar o acesso a informações armazenadas em XML. A integração entre essas tecnologias possibilitou buscas rápidas em memória, viabilizando chatbots e sistemas interativos voltados à Linguística Geral, Dialetologia e Sociolinguística em português brasileiro. De maneira análoga, a proposta atual adota estratégias computacionais semelhantes para estruturar um léxico ambiental consultável de forma ágil, favorecendo tanto o ensino quanto as investigações científicas.

7. Metodologia

A primeira etapa consistiu na definição das categorias de informação que seriam contempladas no léxico. Essa foi baseada em uma análise de conteúdo do *Guia para Identificação de Peixes Ornamentais Brasileiros* (Sampaio e Nottingham, 2008) e do site

internacional *FishBase* (Froese e Pauly, 2025), amplamente reconhecido por sua base taxonômica e ecológica confiável.

Foram identificadas 14 categorias relevantes para a descrição das espécies: nome científico, nomes populares, família taxonômica, nome do cientista que catalogou a espécie, ano da catalogação científica, formato corporal, coloração, distribuição geográfica, tamanho médio e/ou máximo, hábitos alimentares, comportamento e exigências em cativeiro, permissão para o comércio ornamental e análise morfológica dos nomes populares. Para assegurar a representatividade dos dados e a viabilidade da coleta, foram incluídas todas as 150 espécies catalogadas no guia de Sampaio e Nottingham (2008), o que confere ao estudo um recorte consistente e abrangente.

As informações de peixes foram coletadas manualmente a partir dessas duas referidas fontes e organizadas em uma planilha eletrônica. Esse processo envolveu a padronização de nomenclatura e da estrutura de preenchimento de outros campos, correção de inconsistências e preenchimento de lacunas informacionais sempre que possível, respeitando a fidelidade às fontes originais. Ressalta-se que foi priorizada a obtenção das imagens por meio do trabalho de Sampaio e Nottingham (2008), porém algumas dessas imagens possuíam marcações ou sobreposições por outras imagens. Dessa forma o site *FishBase* (Froese e Pauly, 2025) foi a fonte nesses casos.

Posteriormente, a planilha no formato XLSX foi convertida para o formato XML, sendo esse o arquivo utilizado para a consulta de dados pelo sistema construído. Foi desenvolvida uma interface gráfica com o uso da biblioteca *Tkinter* do Python que permite a consulta pelo nome científico e exibe as informações da espécie em questão na janela principal. Adicionalmente, empregou-se a biblioteca *Matplotlib* para a exibição de gráficos referentes à aplicação das 10 funções hash implementadas, sendo elas AHO, CRC, Divisão Cormen, Divisão Drozdek, Dobra, Multiplicação, *One-at-a-Time*, PJW, Universal e Zobrist, selecionadas com base em sua recorrência e desempenho descritos em Moreno (2017) e Moreno, Barbosa e Manfio (2021). Entre essas, destacam-se a função hash Universal e a função *One-at-a-Time*, escolhidas por sua adequação ao problema de distribuição uniforme de chaves em estruturas de dicionário.

8. Interface Gráfica

A Figura 3 exibe a interface gráfica construída pelos autores do presente trabalho. Observa-se um campo de entrada destinado à pesquisa pelo nome científico do peixe que se deseja obter seus dados, sendo esses exibidos no campo em branco juntamente com uma imagem armazenada da espécie em questão. Menus para a seleção entre as 10 funções hash disponíveis e para a escolha entre o gráfico de “Colisões por Endereço” e de “Endereços Vazios x Ocupados” estão disponíveis na barra superior da interface. Por fim, torna-se possível a impressão da tabela hash, contendo a informação dos peixes presentes em cada índice e a quantidade de colisões por índice por meio do botão “Imprimir Hash”, estando esse também presente na barra superior.

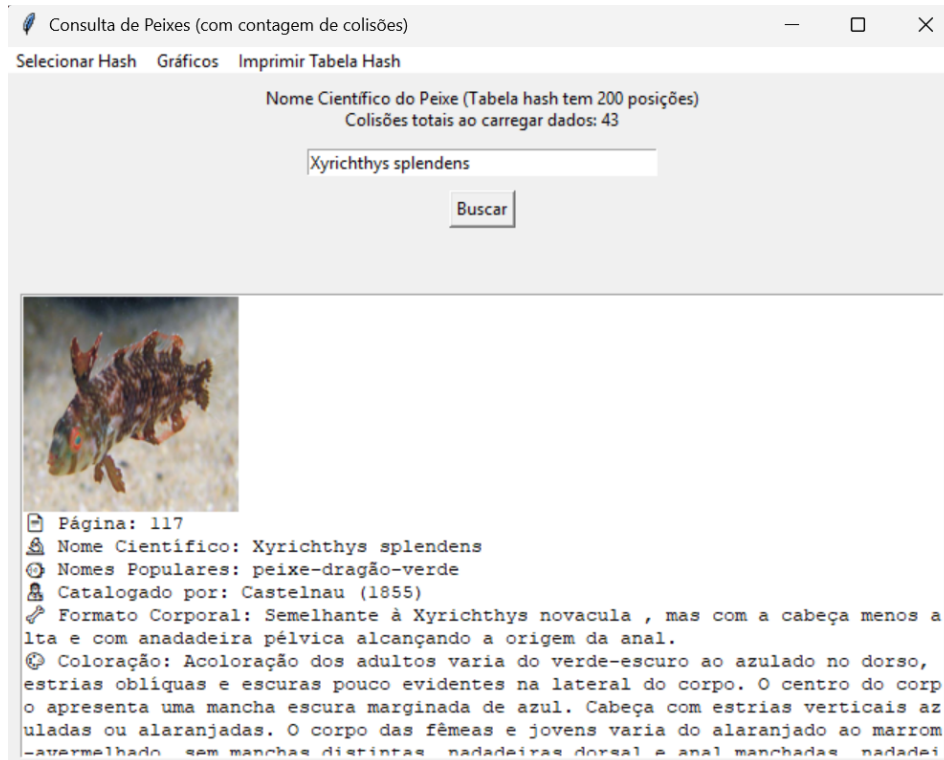


Figura 3. Tela da interface exibindo os resultados da busca no léxico. Fonte: Os autores.

9. Resultados

Assim como feito por Moreno (2017) e Moreno, Barbosa e Manfio (2021), foi realizada uma comparação entre diferentes funções hash para avaliar:

- dispersão dos dados em uma tabela hash com 200 posições;
- tempos médios do cálculo do índice e tempo de busca.

Os procedimentos de teste são os seguintes:

- 5 medições para cada função hash foram feitas;
- o ambiente foi controlado para evitar interferências, reiniciando o sistema e limpando cache antes de cada teste;
- o equipamento utilizado foi um notebook com Windows 10, Intel Core i7, 8 GB RAM, SSD 480 GB

A função hash *One at a Time* (Moreno, 2017) apresentou menor tempo de cálculo e de busca a partir do nome científico dos peixes, como pode ser visto na Figura 4. Embora a dispersão dos dados tenha sido melhor na função hash Universal (Moreno, 2017), (Figura 5) do que na *One at a time* (Figura 6), esse resultado pode ser atribuído à maior velocidade de cálculo da *One at a time*, que realiza somas e operações simples de bit a bit, como o operador XOR (ou exclusivo) e deslocamentos binários (*shifts*).

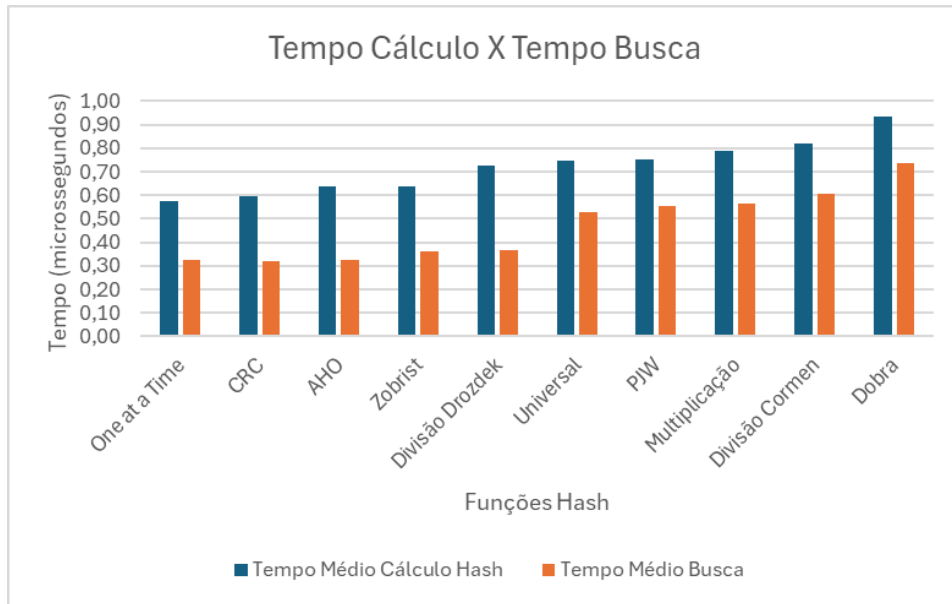


Figura 4. Gráfico de desempenho das funções de dispersão (hash) implementadas. Fonte: Os autores.

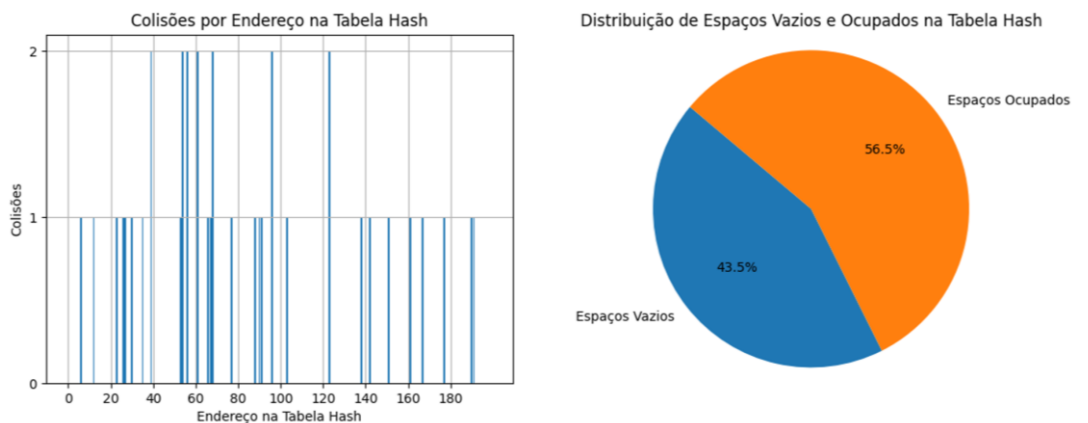


Figura 5. Visualização de colisões e distribuição de valores usando a função hash Universal. Fonte: Os autores.

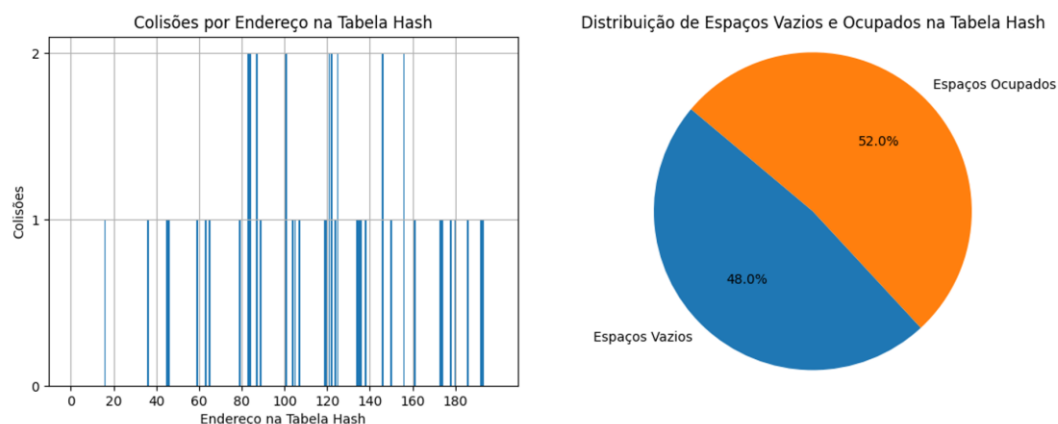


Figura 6. Visualização de colisões e distribuição de valores usando a função hash *One at a Time*. Fonte: Os autores.

Considerando o maior tempo médio de cálculo e busca, destaca-se a função Dobra. Nos experimentos de Moreno (2017) a partir do Léxico das Ervas, a função hash Dobra também apresentou o maior tempo de cálculo do índice. As funções hash Multiplicação e Divisão (Cormen *et al.*, 2012), por outro lado, apresentavam tempos de cálculo consideravelmente menores no Léxico das Ervas quando comparadas às demais

funções. O comportamento oposto foi observado com a função CRC (Moreno, 2017), que obteve melhor desempenho em relação ao tempo de cálculo no Léxico dos Peixes.

10. Considerações finais

Ressalta-se que a estruturação de um léxico organizado e de fácil compreensão e utilização pode favorecer a disseminação e o entendimento de termos técnicos específicos de determinadas áreas. Dessa forma, esses benefícios atingem estudantes, especialistas e indivíduos interessados pela temática. Ademais, o desenvolvimento de léxicos estabelece a sustentação para sistemas e aplicações mais avançados que podem vir a atender interesses educacionais, profissionais e legais, como na fiscalização do comércio de peixes ornamentais.

Em vista disso, o presente estudo concluiu seu objetivo de construir uma base de dados organizada e padronizada, cujo acesso é viabilizado por uma interface gráfica simples e intuitiva que propicia o acesso às informações cadastradas. Outrossim, o sistema projetado fornece ferramentas válidas para pessoas envolvidas com aplicações de funções hash, o que inclui estudantes da área de Ciência da Computação. Entre esses recursos estão a visualização de gráficos de colisões por endereço, a porcentagem de ocupação de endereços e os elementos presentes nas posições da tabela hash.

Acerca dos resultados obtidos, os gráficos construídos proporcionaram uma análise comparativa entre 10 funções hash diferentes no contexto do Léxico de Peixes Ornamentais Marinhos do Brasil. Observou-se diferenças de tempo médio de cálculo e busca entre as funções, por meio dos gráficos, bem como na ocupação de endereços entre as diferentes funções hash implementadas. Nesse sentido, ressalta-se a importância da escolha adequada, com o intuito de minimizar colisões, dispersar os dados de maneira eficaz e otimizar o sistema.

O presente estudo se limitou aos peixes expostos no guia de identificação de Sampaio e Nottingham (2008). À vista disso, em relação a implementações futuras, pode-se incluir novos atributos com origem de outras fontes, como status de conservação, peso, expectativa de vida, etc., detalhar mais características físicas e comportamentais. Adicionalmente, melhorias poderiam expandir a abrangência da base para outras espécies de Organismos Aquáticos Ornamentais marinhos e de água doce, como novos peixes, invertebrados e plantas.

Além disso, a integração com Bases de Dados abertas é uma possibilidade que facilitaria a expansão no número de espécies. Ademais, proporcionar a consulta não apenas pelo nome científico, mas também pelos nomes populares democratiza o acesso a essa Base de Dados, expandindo o público-alvo atingido. Em vista disso, a escalabilidade da proposta com certeza deve se valer das funções hash para garantir acesso rápido aos dados.

Adicionalmente, destaca-se a necessidade de validação da Base de Dados construída, uma vez que não foram descritos processos de verificação externa ou revisão por especialistas da área biológica. A inclusão de uma etapa de validação com ictiólogos, bem como a realização de testes com usuários finais, contribuiria para aumentar a confiabilidade e a utilidade prática do sistema. Do mesmo modo, recomenda-se a condução de testes de usabilidade e a aplicação de métricas empíricas que avaliem a interface gráfica implementada, considerando aspectos como acessibilidade, clareza e eficiência. Ainda, futuras versões do estudo podem aprofundar a discussão dos resultados

com interpretações mais densas e explorar suas implicações e benefícios em diferentes contextos de uso. Com esses aperfeiçoamentos, espera-se promover um sistema mais completo, robusto e conectado com demandas reais do meio acadêmico, ambiental e tecnológico.

Referências Bibliográficas

Araújo, M. D. S.; Lima, M. V. D. A. Aplicação de Ciência de Dados em um Ambiente Virtual de Aprendizagem: um estudo de caso na disciplina de Programação Orientada a Objetos. **Revista Novas Tecnologias na Educação**, Porto Alegre, v.22, n.3, p. 399–409, jun., 2025. Disponível em: <<https://seer.ufrgs.br/index.php/renote/article/view/145037>>.

Brasil. Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA). **Portaria Nº 102, De 20 De Setembro De 2022. Estabelece normas, critérios e padrões para exportação e importação de peixes de águas continentais, marinhas e estuarinas, com finalidade ornamental e de aquariofilia.** Disponível em: <<https://www.ibama.gov.br/component/legislacao/?view=legislacao&legislacao=139174#:~:text=Fica%20proibida%20a%20exporta%C3%A7%C3%A3o%20de,Par%C3%A1grafo%20%C3%BAnico>>.

Carvalho, T. L.; Costa, G. A.; Kojima, J. T.; Villamil-Rodríguez, J. F.; Freitas, T. M. D. Peixes ornamentais amazônicos: panorama de exportação. **Cadernos Técnicos de Veterinária e Zootecnia**, Belo Horizonte, n.105, 135-160, jan. 2023. Disponível em: <<https://vet.ufmg.br/wp-content/uploads/2022/12/cteletronico-105.pdf#page=135&zoom=100,0,0>>.

Cormen, R. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. **Algoritmos - Teoria e Prática**. Rio de Janeiro: Elsevier, 2012. p. 200-201.

Faria, C. R.; Barbosa, C. R. S. C. de. Ferramenta de Apoio aos Estudantes da Agricultura para Identificação de Invasores na Cultura da Soja. **Revista Novas Tecnologias na Educação**, Porto Alegre, v.18, n.2, p. 161-170, jan., 2020. Disponível em: <<https://seer.ufrgs.br/renote/article/view/110214>>.

Froese, R.; Pauly, D. (Ed.). **FishBase**. World Wide Web electronic publication, 2025. Disponível em: <<https://www.fishbase.se/search.php>>.

Gomes, Y. Y.; Mendes, F. R. M.; Mota, F. B. S.; Paula, M. M. V. de. Ferramenta de apoio à decisão para gestão educacional: uma abordagem com dados de avaliação institucional. **Revista Novas Tecnologias na Educação**, Porto Alegre, v. 22, n. 3, p. 410–420, jan., 2025. Disponível em: <https://seer.ufrgs.br/index.php/renote/article/view/145038>.

Gonzalez, M.; Lima, V. L. S. de. Recuperação de informação e processamento da linguagem natural. In: **Anais do XXIII Congresso da Sociedade Brasileira de Computação**. Campinas: SBC, 2003. p. 347-395.

Gregório, P.; Goya, D. Hash Criptográfico sobre Senhas e Aleatoriedade do Argon2*. In: **Anais Estendidos do XIX Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais**. São Paulo: SBC, 2019. p. 71-80.

Lisbôa, A. R.B.S.; Barbosa, C. R. S. C. de. Lexicon of Orchids. **Procedia - Social and Behavioral Sciences**, Amsterdã, v.95, p. 81–88, out., 2013. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S1877042813041451?via%3Dihub>>.

Manfio, E. R.; Moreno, F. C.; Barbosa, C. R. S. C. de. Professor Tical e Alib: Interação Humano Computador em Diferente Campo. In: **Anais da XIX Conferência**

Internacional sobre Informática na Educação. Fortaleza: Jaime Sánchez (Ed.), 2014. p. 782-787.

Meneses, R. D. D.; Henriques, M. A. A. Impacto da otimização de funções hash no desempenho do algoritmo de assinatura digital pós-quântica CRYSTALS-Dilithium. In: **Anais do XXXIII Workshop de Trabalhos de Iniciação Científica e de Graduação - Simpósio Brasileiro de Segurança da Informação e de Sistemas Computacionais.** Juiz de Fora: SBC, 2023. p. 267-272.

Moreno, F. C. **Visual Tahs: Ferramenta para analisar a eficácia de buscas das Funções Hash em um Léxico para Língua Natural.** Londrina: Programa de Pós-Graduação em Ciência da Computação da Universidade Estadual de Londrina, 2017. 187p. Dissertação (Mestrado em Ciência da Computação).

Moreno, F. C.; Barbosa, C. R. S. C. de., Manfio, E. R. Tabelas Hash para um Léxico Digital. **Revista de Informática Teórica e Aplicada**, Porto Alegre, v.28, n.2, p. 25–38, ago. 2021. Disponível em: <https://seer.ufrgs.br/index.php/rita/article/view/Vol28_Nr2_25/pdf_1>.

Rezende, F. P.; Fujimoto, R. Y. **Peixes Ornamentais no Brasil.** Brasília: Embrapa, 2021. 297p. Disponível em: <https://www.researchgate.net/profile/Patricia-Maciel-Honda/publication/354153701_Capitulo_4_Sanidade_In_Peixes_Ornamentais_no_Brasil_Volume_1_Mercado_legislacao_sistemas_de_producao_e_sanidade/links/613f39c599b5cc02743dcc58/Capitulo-4-Sanidade-In-Peixes-Ornamentais-no-Brasil-Volume-1-Mercado-legislacao-sistemas-de-producao-e-sanidade.pdf>

Sampaio, C. L. S.; Nottingham, M. C. **Guia para identificação de peixes ornamentais brasileiros: espécies marinhas.** Brasília: Ibama, 2008. 205p. Disponível em: <<https://www.ibama.gov.br/phocadownload/peixesornamentais/2008/guia-para-identificacao-de-peixes-ornamentais-marinhos-ibama.pdf>>.

Santos Junior, J. L. N. dos; Isquerdo, A. N. A construção de um banco de dados lexicográfico em XML a partir de dados dialetais: o Processamento Automático de Linguagem Natural (PLN). **Domínios de Lingu@gem**, Uberlândia, v.16, n.4, p 1544–1570, set., 2022.

Silva, G. L. B.; Carvalho, J. A. de; Maciel, A. M. A. Desenvolvimento de um Learning Analytics Dashboard a partir de Modelos de Mineração de Dados Educacionais. **Revista de Engenharia e Pesquisa Aplicada**, Recife, v.6, n.3, p 59-69, abr., 2021.

Smarsaro, A. D. D. **Descrição e formalização de palavras compostas do português do Brasil para elaboração de um dicionário eletrônico.** Rio de Janeiro: Programa de Pós-Graduação em Estudos da Linguagem da Pontifícia Universidade Católica do Rio de Janeiro. 2004. 154p. Tese (Doutorado em Letras).