

Criação de plugin para o *Wolfram Language & Mathematica*: Integração via *WEB API* para busca de equações no *SearchOnMath*

Gabriel P. Soares¹, Davi C. Bernardes¹, Flavio B. Gonzaga¹

¹Instituto de Ciências Exatas – Ciência da Computação
Universidade Federal de Alfenas (UNIFAL)
CEP 37.133-840 – Alfenas – MG – Brasil

{gabriel.soares,davi.bernardes}@sou.unifal-mg.edu.br

Resumo. Atualmente, o uso de plugins tem se tornado cada vez mais comum, para estender funcionalidades computacionais, combinando dois sistemas sofisticados como a *Wolfram Language* e o *SearchOnMath*. Neste trabalho, propõe-se o desenvolvimento de um plugin para integrar a *SearchOnMath*, uma ferramenta de busca de conteúdo matemático, com o *Wolfram Mathematica*. O *Wolfram Mathematica* é um software de álgebra computacional e programação. O objetivo é permitir a pesquisa por fórmulas matemáticas da *SearchOnMath* dentro do *Wolfram Mathematica* e visualizar os resultados formatados com notação matemática. O trabalho envolve a utilização da *WEB API* da *SearchOnMath*, e a criação de uma função em *Wolfram Language* para o *Mathematica*. Espera-se que esta integração amplie as funcionalidades destas ferramentas e facilite o acesso a conteúdo matemático.

1. Introdução

Com o avanço da tecnologia, o termo “*plugin*” tem se difundido, estando presente em diversas áreas da ciência da computação. Este termo refere-se a um programa de computador que fornece funcionalidades específicas a sistemas mais amplos, integrando sistemas diferentes e permitindo a criação de recursos avançados.

O presente trabalho observou que a *SearchOnMath*¹, uma ferramenta de busca específica para conteúdo matemático, ainda não possuía integração com o sistema *Wolfram Mathematica*². Assim, propôs-se como objetivo do projeto o desenvolvimento de um plugin de modo a ser possível a integração dessas ferramentas. No ambiente do *Mathematica*, a maneira primária de desenvolver plugins é por meio da escrita de funções, que ficam disponíveis no repositório da wolfram (*FunctionRepository*)³. Elas são blocos de código reutilizáveis que executam tarefas específicas. Essas funções podem ser agrupadas e organizadas de forma modular para formar um plugin completo. A seguir são apresentadas mais informações sobre as soluções mencionadas.

1.1. *Wolfram Mathematica*

O *Wolfram Mathematica* é uma ferramenta de software versátil criada pela empresa *Wolfram Research*, fundada por Stephen Wolfram. Desenvolvido inicialmente como um

¹<https://www.searchonmath.com>

²<https://www.wolfram.com/mathematica/>

³<https://resources.wolframcloud.com/FunctionRepository/>

sistema de álgebra computacional, o *Mathematica* expandiu suas funcionalidades para se tornar uma linguagem de programação completa, com uma ampla variedade de bibliotecas disponíveis em diversas áreas das ciências exatas, incluindo aprendizado de máquina, estatística, matemática simbólica, manipulação de dados, grafos, análise de série temporal, processamento de linguagem natural, otimização, plotagem de funções, entre outras. Além de, estar disponível para os ambientes Mac, Windows, Linux, Android, Iphone e em nuvem. Desta forma, o plugin desenvolvido também torna-se ao serviço nestas plataformas.

O software é composto por um núcleo interpretador de expressões e uma interface gráfica intuitiva. O núcleo interpreta os códigos da linguagem *Wolfram*, enquanto a interface gráfica, originalmente desenvolvida por Theodore Gray em 1988, oferece uma experiência de bloco de anotações (*Wolfram GUI notebook*) que permite a criação e edição de documentos contendo código, texto, imagens e gráficos. [Wolfram 2024]

Com mais de 6.000 funções e símbolos integrados na *Wolfram Language*, o *Mathematica* continua a crescer, com iniciativas como o *Wolfram Function Repository*, onde a comunidade pode contribuir com suas próprias funções. Além disso, o *Wolfram Data Repository* e o *Wolfram Neural Net Repository* oferecem conjuntos de dados e modelos de aprendizado de máquina para enriquecer ainda mais as capacidades da plataforma [Wolfram 2023b]. Detalhes específicos sobre o uso do *Wolfram Mathematica*, no contexto da função para a *SearchOnMath* serão abordados na Seção 3 (Metodologia).

1.2. SearchOnMath

A *SearchOnMath* é um mecanismo de busca com foco na pesquisa de conteúdo matemático em diversas áreas do conhecimento. Ao contrário das ferramentas de busca convencionais, que restringem suas pesquisas a texto, a *SearchOnMath* possibilita também a busca por fórmulas matemáticas. [SearchOnMath 2023]

Sendo o *Wolfram Mathematica* uma ferramenta amplamente utilizada pela comunidade científica, em especial na área de Ciências Exatas, os autores deste projeto entenderam que poderia ser de interesse da comunidade de usuários do *Mathematica*, pesquisadores, engenheiros, cientistas, startups, entre outros, o desenvolvimento de uma função que possibilitasse a sua integração com o serviço da *SearchOnMath*. Em 2022 foi apresentado então um protótipo da função da *SearchOnMath* na *Wolfram Technology Conference - WTC* [Gonzaga 2022], de modo que o presente trabalho fornece mais detalhes sobre a continuação do desenvolvimento da função (pós *WTC*). Atualmente, a função desenvolvida se encontra totalmente funcional e disponível aos usuários do *Mathematica*.

1.3. Objetivo geral

Este trabalho teve como objetivo principal a implementação de uma função da *SearchOnMath* dentro da plataforma *Wolfram Mathematica*, contribuindo com a Recuperação de Informação Matemática, colaborando com a comunidade acadêmica e científica.

1.4. Objetivos específicos

Para atingir o objetivo geral, foram estabelecidos objetivos específicos que norteiam o desenvolvimento e a implementação da função. Estes objetivos visam garantir que a

integração entre a *SearchOnMath* e o *Wolfram Mathematica* ocorra de maneira eficiente e funcional, proporcionando uma experiência aprimorada para os usuários. Abaixo, são detalhados os principais objetivos específicos deste trabalho:

- Criar uma função que permita que os usuários submetam pesquisas a serem realizadas na *SearchOnMath* diretamente a partir do *Wolfram Mathematica*.
- Publicar a função desenvolvida no repositório oficial de funções da *Wolfram*.

O restante do texto segue organizado da seguinte forma. Na Seção 2 (Trabalhos Relacionados) é apresentada uma revisão bibliográfica referente à área de Recuperação de Informação Matemática. A seção 3 (Metodologia) aborda detalhadamente o processo de integração com o *Wolfram Mathematica*, destacando as estratégias adotadas para alcançar a melhor compatibilidade entre as duas plataformas. Na seção 4 (Resultados), são apresentados os desdobramentos dessa integração, incluindo uma análise das funcionalidades e exemplos que demonstram de forma prática como ela é utilizada.

2. Trabalhos Relacionados

A seguir são apresentados alguns trabalhos recentes na área de Recuperação de Informação Matemática, com o objetivo de ao final destacar quais são as contribuições do trabalho aqui realizado considerando-se o estado da arte nesta área.

Em [Gao et al. 2017] são discutidos o uso de técnicas de representação neural para tarefas de Recuperação de Informação Matemática (*Math Information Retrieval - MIR*), bem como a investigação da possibilidade de se incorporar fórmulas matemáticas da mesma forma que as linguagens naturais. Os autores começam destacando a importância das fórmulas matemáticas na disseminação de informações científicas e o desafio da MIR em encontrar fórmulas específicas em documentos. Enquanto as técnicas de redes neurais têm obtido sucesso em linguagens naturais, seu uso em tarefas relacionadas à matemática tem sido limitado. Os principais questionamentos de pesquisa abordados são:

- Podemos aplicar abordagens de representação neural (como a incorporação de palavras) na “linguagem matemática”?
- Podemos usar técnicas de representação neural para melhorar o desempenho da Recuperação de Informação Matemática?

O artigo reconhece diferenças significativas entre a linguagem natural e a linguagem matemática. As fórmulas matemáticas são altamente estruturadas, têm símbolos ambíguos e são frequentemente apresentadas em formato de layout (por exemplo, $\text{\LaTeX}$ [Lamport 1986] ou *MathML* [MathML 2023]). Essas características tornam desafiadora a aplicação de abordagens de linguagem natural à linguagem matemática. Para investigar essas questões, os autores desenvolvem um método chamado “symbol2vec” para aprender representações vetoriais de símbolos matemáticos, como números, variáveis, operadores e funções. Além disso, propõem uma abordagem baseada em “formula2vec” para MIR e avaliam seu desempenho por meio de experimentos preliminares. Os resultados preliminares mostram que a incorporação de símbolos de fórmulas apresenta um desempenho promissor, abrindo possibilidades para sua aplicação na representação da linguagem matemática e em tarefas de MIR. A pesquisa também destaca a importância da informação textual em conjunto com a incorporação de fórmulas para obter o melhor desempenho na MIR. Os autores sugerem que futuros estudos possam explorar modelos de representação mais

sofisticados e técnicas avançadas de tokenização de fórmulas para melhorar ainda mais o desempenho da MIR.

Em [Amador et al. 2023] é discutido o crescente interesse da comunidade acadêmica na busca matemática em função das suas aplicações em diversos campos da ciência e tecnologia. É mencionado ainda que mecanismos de busca de texto padrão, como *Google*, *Bing* e *DuckDuckGo*, têm suporte limitado para fórmulas matemáticas. Este trabalho apresenta a demonstração de uma interface de busca matemática chamada *Math-Deck*⁴, projetada para facilitar a recuperação de texto e fórmulas, sendo amigável tanto para especialistas quanto para não especialistas. Para aprimorar a integração da reutilização e edição de fórmulas na formulação de consultas, o artigo propõe a implementação de uma nova barra de consulta. Nessa barra, as fórmulas podem ser inseridas tanto em formato $\text{\LaTeX}$ bruto quanto por meio de uma interface gráfica intuitiva. Isso possibilita a manipulação e anotação das fórmulas de forma mais eficiente e integrada. Além disso, a página de resultados de pesquisa realça as fórmulas correspondentes e os termos nas páginas de PDF, além de possibilitar que os usuários naveguem e exportem as fórmulas para uso em consultas e documentos.

No contexto de [Li et al. 2023], é apresentado um método inovador para recuperação de literatura científica, baseado em *Graph Convolutional Network* (GCN) e *Hesitant Fuzzy Set* (HFS). Esse método supera as limitações de abordagens anteriores, que focavam principalmente em expressões matemáticas, negligenciando os atributos da literatura e a relação entre os trabalhos, o que impactava a precisão dos resultados. Ao empregar o GCN para capturar as dependências entre os nós de literatura e integrar o HFS na avaliação de similaridade entre as expressões matemáticas de consulta e os resultados de recuperação, o método alcançou alta precisão nas buscas. Além disso, o artigo também realizou uma comparação com modelos existentes, como o *Tangent-cft* e o *SearchOnMath*, evidenciando as vantagens da abordagem proposta.

O estudo menciona que a *SearchOnMath* é uma ferramenta especializada na recuperação de fórmulas matemáticas em documentos científicos. No entanto, observa-se que seu desempenho é limitado pela abordagem de busca única por consulta, onde as condições de texto e fórmulas matemáticas são tratadas isoladamente. Por exemplo, se um usuário buscar pela expressão “ $\backslash (E=mc^2 \backslash)$ ” juntamente com o texto “relatividade”, a *SearchOnMath* não realiza uma busca conjunta que integre ambos os termos, mas sim duas buscas separadas – uma para a expressão matemática e outra para o texto – sem integrá-las em uma única consulta. Essa limitação afeta a precisão e relevância dos resultados fornecidos. Em contraste, a nova metodologia proposta, que utiliza GCN e HFS, supera essa limitação ao combinar condições de busca de forma integrada, demonstrando uma maior eficácia na recuperação de literatura científica relacionada a expressões matemáticas.

A respeito de [Mansouri et al. 2019] é abordado a importância das representações vetoriais para fórmulas matemáticas e seu uso em tarefas de recuperação de informações. Os autores destacam que, embora muitas tarefas de recuperação de informações e processamento de linguagem natural se beneficiem de representações distribuídas (vetores densos) de palavras, modelos de incorporação para produzir representações vetoriais para fórmulas matemáticas ainda não foram totalmente investigados. Salientam também que, houve

⁴<https://www.mathdeck.org/>

uma quantidade considerável de pesquisa sobre representações contínuas para palavras e parágrafos em recuperação de informações (IR) e processamento de linguagem natural (NLP), mas pouco trabalho foi feito em representações de fórmulas matemáticas.

O objetivo principal do projeto é investigar se as incorporações são adequadas para a recuperação de informações em tarefas que envolvem fórmulas matemáticas. Os autores apresentam o *Tangent-CFT*, um modelo de incorporação para fórmulas matemáticas usando representações *Symbol Layout Trees* (SLT) e *Operator Trees* (OPT), e usam o fastText para produzir as representações distribuídas com base na linearização e tokenização das fórmulas.

No artigo, os autores demonstram que o modelo apresentado combina incorporações de OPT, SLT e SLT-Type para obter uma relevância parcial maior do que os modelos de ponta para a tarefa de busca de fórmulas NTCIR-12. Além disso, foi evidenciado que combinar os resultados de um modelo de incorporação, como o *Tangent-CFT*, com os resultados de uma abordagem de correspondência de estrutura (por exemplo, Approach0⁵), pode resultar em pontuações de relevância parcial e total mais altas do que as abordagens anteriores.

Apesar das pesquisas apresentadas, a maior parte das ferramentas não está disponível publicamente online. Apenas *SearchOnMath*, *Approach0* e *MathDeck* estão disponíveis. No entanto, até a realização desse trabalho, nenhuma das ferramentas citadas possuía integração com o *Wolfram Mathematica*. A realização do trabalho, portanto, faz com que o universo de busca matemática passe a estar disponível dentro de tais ferramentas com base no serviço da *SearchOnMath*.

3. Metodologia

3.1. Função *SearchOnMath* para o *Wolfram Mathematica*

O primeiro passo para o desenvolvimento da função *SearchOnMath* consistiu no entendimento da sua *WEB API* [Gonzaga 2023], especialmente desenvolvida de modo a atender ao *Wolfram Mathematica*. A *WEB API*, implementada pela equipe da *SearchOnMath*, está livremente disponível⁶, e recebe apenas um único parâmetro denominado *query*. Assim, uma requisição completa para a *WEB API* buscando pela seguinte *query*: “Einstein’s formula \ (E=mc²) revolutionized physics” será algo semelhante a este exemplo⁷.

Uma vez submetida a consulta, a *SearchOnMath* retorna os resultados em formato JSON [Pezoa et al. 2016], contendo os seguintes campos:

- **currentPage**: Página atual de resultados. No caso desta *WEB API*, será retornada sempre a primeira página.
- **searchTime**: Tempo gasto pelo servidor (em segundos) para processar a requisição.
- **totalResults**: Quantidade total de resultados encontrados.

⁵<https://approach0.xyz/search/>

⁶<https://www.searchonmath.com/webservice/wolfram?query>

⁷[https://www.searchonmath.com/webservice/wolfram?query=Einstein's%20formula%2C%20%5C\(%20E%3Dmc%5E2%20%5C\)%2C%20revolutionized%20physics](https://www.searchonmath.com/webservice/wolfram?query=Einstein's%20formula%2C%20%5C(%20E%3Dmc%5E2%20%5C)%2C%20revolutionized%20physics)

- **totalPages**: Este campo informa o número de páginas encontradas, o que tornaria possível a paginação entre os resultados encontrados. Contudo, apenas os cinco primeiros resultados serão retornados nesta versão.
- **results**: Um vetor contendo os cinco primeiros resultados retornados pela *SearchOnMath*. Este vetor contém internamente os seguintes campos:
 - context:
 - formula:
 - source:
 - title:
 - url:

A Figura 1 exibe os resultados retornados pela *WEB API* da *SearchOnMath*. Na Figura 2 é apresentado o detalhamento de um dos resultados contidos no vetor *results*.

```
{
  "currentPage": 1,
  "results": [ ...
],
  "searchTime": 1.63,
  "totalPages": 76,
  "totalResults": 377
}
```

Figura 1. Resultados retornados pela *WEB API* da *SearchOnMath* em formato JSON.

```
"results": [
  {
    "context": "Forums <b>Physics</b> Special and General Relativity <b>Einsteins</b> <b>formula</b> Thread starter AfRoMaNn Start date Oct 17,2010 Oct 17, 2010 #1 AfRoMaNn 15 0 im pretty sure everyone's familiar with <b>einsteins</b> <b>formula</b> E...ut what is this formula for?  $E^2=m^2c^4$  Answers and Replies Oct 17, 2010 #2 bcrowell Staff Emeritus Science Advisor Insights Author Gold Member 6,723 426 It's the same equation but with both sides squared. It's like  $7E=7mc^2$  -- not an independent piece of information. Oct 17, 2010 #3 Janus Staff Emeritus Science Advisor Insights Author Gold Member 3,739 1,686 Who might be thinking of the formula 
$$E^2 = m^2 c^4 + p^2 c^2$$
 is fine for determining the energy equivalent for the rest mass of a particle. The formula above is more general as it also includes the momentum of the particle and also works for the photon, which has no rest mass. Share: Share Suggested for: Einsteins formula I Einstein aberration formula Last Post Mar 9, 2020 Replies 7 Views 694 A Quadrupole formula - polarization tensor Last Post Jun 29, 2021 Replies 4 Views 469 A Kerr metric time dilation formula Last Post Jun 19, 2021 Replies 19 Views 62 ...",
    "formula": "<math xmlns='\"http://www.w3.org/1998/Math/MathML\"' display='\"block\"'> <mi>E</mi> <mo>=</mo> <mi>m</mi> <sup> <mi>c</mi> <mn>2</mn> </sup> </math>",
    "source": "Physics Forums",
    "title": "Einsteins formula",
    "url": "https://www.physicsforums.com/threads/einsteins-formula.439065/"
  },
]
```

Figura 2. Campos internos do vetor *results* contendo os 5 primeiros resultados.

Função SearchOnMath em Wolfram Language

```

1 SearchOnMath[query_String, printCells_ : "" ] := Module[
2   {somMessage =
3     Import[
4       "https://www.searchonmath.com/web service/wolfram?query=" <>
5       URLEncode[query], "RawJSON"
6     ]
7   },
8   DsReturn =
9     AssociationThread[{"Title", "URL", "Source",
10      "Formula", "Context"}, {
11      somMessage["results"][[All, "title"]],
12      somMessage["results"][[All, "url"]],
13      somMessage["results"][[All, "source"]],
14      somMessage["results"][[All, "formula"]],
15      somMessage["results"][[All, "context"]]
16    }
17  ];
18  If[printCells === "Print",
19    CellPrint[TextCell[Column[
20      {
21        Hyperlink[
22          Style[Style[#[ "title"], "Section"], Underlined], #[ "url"]
23        ],
24        Row[{Text[Style["Source: ", "Subsection"]], #[ "source"]}],
25        Style["Formula found: ", "Subsection"],
26        DisplayForm[ImportString[#[ "formula"], "MathML"]],
27        Style["Formula context: ", "Subsection"],
28        StringReplace[
29          StringReplace[#[ "context"],
30            RegularExpression["(<b>)((.|\\s)+?)(<b>)"] :>
31            "\\!\\(\\*StyleBox[\\\"$2\\\",\\\"Subsubsection\\\"]\\)\"
32          ],
33          RegularExpression["(<math(.|\\s)+?</math>)"] :>
34          DisplayForm[ImportString["$1", "MathML"]]
35        ]
36      }
37    ], "Text", ShowStringCharacters -> False, FontSize -> 18,
38      FontColor -> Black, "Section"
39  ]
40  & /@ somMessage["results"]
41 ]
42 ];
43 Transpose[Dataset[DsReturn]]
44 ]

```

As linhas 2-7 da função têm o papel de realizar uma requisição a uma *URL* na web e importar os dados resultantes em formato *JSON* bruto (como mostrado nas figuras 1 e 2). A variável “*somMessage*” é utilizada para armazenar esses dados após a importação.

A *URL* acessada é composta por duas partes: a primeira parte é a *URL* base, que é “https://www.searchonmath.com/web service/wolfram?query=” e a segunda parte é o valor do parâmetro *query*. Conforme determina o padrão de passagem de parâmetros na *URL*, o valor da *query* é codificado pela função “URLEncode”.

O argumento adicional na função “*Import*”, especificado como “*RawJSON*”, define o formato de importação. Nesse caso, o formato “*RawJSON*” é utilizado, indicando que o conteúdo da *URL* acessada é importado diretamente em formato *JSON* bruto, sem qualquer análise ou transformação adicional.

O trecho de código das linhas **8-17** cria uma associação denominada “*DsReturn*”. No *Mathematica*, uma associação é uma estrutura de dados usada para armazenar informações em pares chave-valor. Neste caso, estamos criando uma associação com as chaves “Title”, “URL”, “Source”, “Formula”, e “Context”, e cada uma dessas chaves será associada a um valor específico (retornado pela *WEB API* da *SearchOnMath*).

Os valores para essas chaves são obtidos da variável *somMessage* da seguinte forma:

- `somMessage[“results”][[All, “title”]]` extrai a lista de títulos dos resultados encontrados em *somMessage*. A expressão `somMessage[“results”]` acessa a lista de resultados e `[[All, “title”]]` seleciona todos os títulos;
- Da mesma forma, `somMessage[“results”][[All, “url”]]` extrai a lista de URLs dos resultados.
- `somMessage[“results”][[All, “source”]]` extrai a lista de fontes dos resultados;
- `somMessage[“results”][[All, “formula”]]` extrai a lista de fórmulas dos resultados;
- `somMessage[“results”][[All, “context”]]` extrai a lista de contextos dos resultados.

Em resumo, esse bloco de código cria uma associação com as chaves “Title”, “URL”, “Source”, “Formula”, e “Context”, e associa cada chave a uma lista de valores obtidos da variável *somMessage*, que contém os resultados da consulta. Isso permite organizar e acessar facilmente as informações dos resultados encontrados.

A seção que abrange as linhas **18-42** da função é estruturada com uma condicional “If”. Essa instrução verifica se a variável *printCells* é igual a “Print”. Se essa condição for verdadeira, o código dentro do bloco “If” será executado. A seguir é apresentado um detalhamento deste trecho do código.

- `CellPrint[TextCell[Column[ ... ]]]`: Cria e imprime células de texto formatadas no *Mathematica* para cada resultado da pesquisa. Detalhando internamente:
 - `Hyperlink[Style[Style#[[“title”]], “Section”, Underlined], #[[“url”]]]`: Cria um hiperlink onde o título do resultado é exibido como um cabeçalho de seção em negrito e sublinhado, e o URL é o destino do link;
 - `Row[Text[Style[“Source: ”, “Subsection”]], #[[“source”]]]`: Exibe a fonte do resultado no formato “Source: [nome da fonte]”, onde o nome da fonte é exibido em um tamanho menor;
 - `DisplayForm[ImportString#[[“formula”]], “MathML”]`: Importa a fórmula do resultado como código *MathML* e a exibe no notebook;
 - `StringReplace[ ... ]`: O resultado retornado pela *SearchOnMath* no campo *context* contém as palavras relevantes destacadas utilizando-se da tag *HTML* `<b>` as fórmulas, por sua vez, são retornadas contidas em tags `<math>`. Nesta seção de código, substituições são feitas nestes elementos de modo a fazer a conversão apropriada para o formato do *Mathematica*.
- `Transpose[Dataset[DsReturn]]`: Após a impressão de todas as células de texto, os dados na associação *DsReturn* são transpostos e convertidos em um objeto *Dataset*,

uma estrutura de dados (*DataFrame* [Databricks 2024]) amplamente utilizado em ciência dos dados.

- `& /@ somMessage["results"]`: O operador `& /@` é usado para aplicar uma função a cada elemento de uma lista isto é chamado de mapeamento de função. Ele percorre a lista de resultados em `somMessage["results"]` e aplica a função no bloco de código dentro de `"If"` para cada resultado individual.

4. Resultados

Nesta seção, apresentaremos os resultados obtidos no desenvolvimento da função, destinados a integrar a *SearchOnMath* com a ferramenta *Wolfram Mathematica*. Como mencionado anteriormente, o *Wolfram Mathematica* é uma renomada plataforma de álgebra computacional e programação. A integração dessa ferramenta foi concebida com o propósito de aprimorar a acessibilidade ao conteúdo matemático, possibilitando pesquisas por fórmulas matemáticas na *SearchOnMath* a partir do *Wolfram Mathematica*, além de apresentar os resultados de maneira formatada e compreensível.

A seguir, será exibida uma visão detalhada dos resultados obtidos nesse processo de integração, destacando os aspectos técnicos, funcionais e os benefícios práticos que essa colaboração entre as plataformas trouxe para a comunidade acadêmica e científica.

4.1. Função *SearchOnMath* para o *Wolfram Mathematica*

Nas primeiras versões da *SearchOnMath*, as fórmulas matemáticas estavam disponíveis apenas em dois formatos: $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ou AsciiMath. Contudo, para se renderizar as fórmulas nos navegadores utilizando-se de quaisquer destes formatos, era necessário o uso de uma biblioteca JavaScript chamada MathJax.

Buscando então um menor tempo de carregamento da página de resultados, a equipe da *SearchOnMath* começou então a pré-renderizar as fórmulas (inicialmente em formato *Scalable Vector Graphic* (SVG) [Consortium 1999]). Desta forma, a tela de resultados passou a exibir as fórmulas como imagens, reduzindo consideravelmente a necessidade de se renderizar código $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ no lado cliente.

Apesar de então ter as fórmulas em $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$, AsciiMath e SVG, nenhum destes formatos se mostrou plenamente compatível com o *Wolfram Mathematica*, conforme explicado a seguir:

- AsciiMath: Não é suportado.
- $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$: Parcialmente suportado. Existe uma função no *Mathematica* capaz de converter $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ em *Wolfram Language* (linguagem oficial da *Wolfram*). Mas através de testes, a equipe da *SearchOnMath* constatou a ocorrência de muitos erros de renderização. Tais erros foram causados, principalmente, em virtude do uso de comandos $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ ainda não suportados pela função de conversão da *Wolfram*.
- SVG: Não suportado nativamente. Existe uma função no repositório de funções da *Wolfram*, chamada `SVGImport` [Hennigan 2022] que é capaz de carregar arquivos SVG. Contudo, o uso da função requer a configuração prévia do Python.

Assim, antes que a função da *SearchOnMath* para o *Wolfram Mathematica* pudesse ser implementada, foi necessário que a *SearchOnMath* oferecesse as fórmulas em um formato com melhor compatibilidade. Nesse sentido, a equipe da *SearchOnMath*

implementou a conversão das fórmulas em um novo formato: *MathML Presentation* [MathML 2023]. O *MathML* (*Mathematical Markup Language*), é uma linguagem de marcação desenvolvida para representar fórmulas matemáticas na web. Foi criado pelo *World Wide Web Consortium* (W3C) com o objetivo de permitir a exibição e a manipulação precisa de expressões matemáticas em documentos HTML.

Dessa forma, com as fórmulas da *SearchOnMath* disponíveis também em *MathML*, o desenvolvimento do trabalho proposto tornou-se viável.

4.1.1. Repositório de funções da Wolfram

O repositório de funções da *Wolfram*⁸ é uma plataforma online que oferece uma ampla variedade de recursos para uso em programação e análise de dados. Ao disponibilizar um conjunto diversificado de funções prontas para uso, este repositório facilita o desenvolvimento de aplicações e a exploração de dados complexos, agilizando o processo de programação e permitindo resultados de alta qualidade. Ele oferece aos usuários a possibilidade de publicar suas próprias funções e recursos para uso pela comunidade. Essa funcionalidade permite que desenvolvedores, pesquisadores e entusiastas compartilhem suas contribuições com outros usuários da *Wolfram*, promovendo a colaboração e o intercâmbio de conhecimento. [Wolfram 2023a]

Conforme então explicado no parágrafo anterior, a maneira adequada de se desenvolver plugins para o *Wolfram Mathematica* é por meio da escrita de funções, que podem então depois ser oficialmente incorporadas ao repositório de funções. A seção seguinte detalha o funcionamento da função *SearchOnMath*, desenvolvida como objetivo do projeto, e que foi oficialmente avaliada (*peer review*) e incorporada no repositório de funções da *Wolfram*.

4.1.2. Função *SearchOnMath*

A função *SearchOnMath* para a plataforma *Wolfram Mathematica* [Gonzaga 2023] retorna os 5 primeiros resultados da *SearchOnMath* com base na consulta submetida. O retorno padrão da função apresenta os dados formatados em um *dataset*. Contudo, é possível ainda passar um parâmetro para a função, requisitando que os resultados sejam não apenas retornados, mas também impressos formatados dentro do *Wolfram Mathematica*.

Para fazer uma busca, as fórmulas devem ser escritas no formato $\text{\LaTeX}$ e colocadas entre os delimitadores $\text{\textbackslash (. . . \textbackslash)}$, como em: *Einstein's formula* $\text{\textbackslash (E=mc^2 \textbackslash)}$ *revolutionized physics*. É importante ressaltar que os delimitadores não são considerados como parte do conteúdo a ser buscado. Eles apenas diferenciam qual parte da consulta deve ser tratada como texto, e qual deve ser tratada como fórmula matemática. Adicionalmente, elementos textuais obrigatórios podem ser definidos usando aspas, como no exemplo: *Einstein's formula* $\text{\textbackslash (E=mc^2 \textbackslash)}$ “*revolutionized physics*”.

O conjunto de dados retornados no *dataset* apresentam os seguintes campos:

⁸<https://resources.wolframcloud.com/FunctionRepository/resources/SearchOnMath/>

- *Title*: Contém o título da página;
- *URL*: URL da página;
- *Source*: Compõe-se da fonte (*Wikipedia* ⁹, *Stack Exchange Mathematics* ¹⁰, ...);
- *Formula*: Compreende a fórmula em formato de apresentação *MathML*;
- *Context*: Inclui o trecho da página que contém a fórmula encontrada (em formato *HTML*).

A seguir é exibido um exemplo, onde pode-se ver a consulta submetida ($\frac{\partial \phi}{\partial t} = k \nabla^2 \phi$) e parte do dataset retornado.

In[1]: SearchOnMath[" $\frac{\partial \phi}{\partial t} = k \nabla^2 \phi$ "]

Title	URL
Mixing in connected gases	https://physics.stackexchange.com/questions/35677
How is melting time affected by flow rate and temperature of surroundings?	https://physics.stackexchange.com/questions/57351
Regularity of fundamental solution to heat equation	https://math.stackexchange.com/questions/1887836
Critic of heat equation	https://www.physicsforums.com/threads/critic-of-h
Heat Conduction	https://www.physicsforums.com/threads/heat-condi

Out[1]:

Figura 3. Dataset retornado no notebook dentro do *Mathematica*

O parâmetro “*Print*” exibe o resultado da consulta já formatado na página do *Mathematica*, como a seguir, note que as funções matemáticas são renderizadas corretamente:

In[]: SearchOnMath[" $\frac{\partial \phi}{\partial t} = k \nabla^2 \phi$ ", "Print"]

Figura 4. Entrada dos dados com o parâmetro “*Print*”

Mixing in connected gases

Source: StackExchange Physics

Formula found:

$$\frac{\partial \phi}{\partial t} = D \nabla^2 \phi,$$

Formula context:

...other. Eventually equilibrium (same humidity) will be reached. Has any effect the height of the hole (relative to the base surface)? Unlikely, though that depends on what the printer is doing. For an empty box the influence of gravity on the process can probably be neglected. How long time will this process take? For diffusion driven mixing, my rough estimate attempt (see below) is a half life longer than a month(!). You should be able to obtain that from an expression like Fick's second law $\left(\frac{\partial \phi}{\partial t} = D \nabla^2 \phi\right)$ where ϕ is the concentration of water vapor and D the diffusion coefficient. I believe that you can take the size of the orifice into account by using an effective diffusion coefficient, perhaps $(D_{\text{eff}} = A_{\text{rel}} D)$, where A_{rel} ...

How is melting time affected by flow rate and temperature of surroundings?

Source: StackExchange Physics

Formula found:

$$\frac{\partial T}{\partial t} = k \nabla^2 T$$

Formula context:

... GridBox[{{w h e r e, {m, mass of the sphere}, {C_p, Specific heat of the solid}, {T_m, Mean temperature of the sphere}, {T_s, Surface temperature of the sphere}, {T_{amb}, Ambient temperature}, {h_{nat}, Heat transfer coeff. (natural convection)}}, ColumnSpacings → 1., RowSpacings → 0.694444] The above combined with internal transient conduction equation for the sphere with thermal conductivity (k) $\left(\frac{\partial T}{\partial t} = k \nabla^2 T\right)$

Figura 5. Dados formatados dentro do *Mathematica*

⁹<https://www.wikipedia.org>

¹⁰<https://math.stackexchange.com>

A rotina *SearchOnMath* tem a capacidade de lidar com consultas complexas envolvendo múltiplas fórmulas. O próximo exemplo demonstra uma consulta composta por múltiplas fórmulas, e também definindo “*magnetic field*” como termo obrigatório:

```
SearchOnMath [
  "Maxwell's equations in a vacuum describe a wave propagating at speed  $c$  as the
  interaction of electric field  $E$  and "magnetic field"  $B$ :  $\nabla \cdot E = 0$ ,  $\nabla \times E = -\frac{\partial B}{\partial t}$ ,  $\nabla \cdot B = 0$ ,  $\nabla \times B = \frac{1}{c^2} \frac{\partial E}{\partial t}$ ",
  "Print"]
```

Figura 6. Utilizando a função *SearchOnMath* com múltiplas fórmulas

[Do the electric and magnetic components of an electromagnetic wave really generate each other?](#)

Source: StackExchange Physics

Formula found:

$$\nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}$$

Formula context:

...Maxwell's equations in vacuum are: $\nabla \cdot \mathbf{E} = 0$, $\nabla \cdot \mathbf{B} = 0$, $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$, $\nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}$.
It's the last two of these that give rise to the interpretation that a changing magnetic field generates an electric field and vice versa. But you could always take the curl of both these equations, use the other two, and get simple, straightforward wave equations separately for $\mathbf{E}$ and $\mathbf{B}$: $\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}$ and $\nabla^2 \mathbf{B} = \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2}$.

[Deriving maxwell equations to obtain plane wave](#)

Source: StackExchange Physics

Formula found:

$$\nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}$$

Formula context:

...ng maxwell equations to obtain plane wave The Maxwell Equation is written the below form . **IN VACUUM** $\nabla \cdot \mathbf{E} = 0$, $\nabla \cdot \mathbf{B} = 0$, $\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}$, $\nabla \times \mathbf{B} = \frac{1}{c^2} \frac{\partial \mathbf{E}}{\partial t}$. By organizing the structure we can get , $\nabla^2 \mathbf{E} = \frac{1}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}$ and $\nabla^2 \mathbf{B} = \frac{1}{c^2} \frac{\partial^2 \mathbf{B}}{\partial t^2}$.
In order to obtain the relationship between the electric field and magnetic field also to prove the plane wave condition , we set the electric field of the $\mathbf{E} = E_0 \sin(ky - \omega t)$ and $\mathbf{B} = B_0 \sin(ky - \omega t)$.

Figura 7. Dados resultantes da busca com múltiplas fórmulas já formatados dentro do *Mathematica* utilizando o parâmetro “*Print*”

5. Continuidade do Projeto de Pesquisa

A *SearchOnMath* fornece a estudantes e pesquisadores uma plataforma dedicada para busca de conteúdo matemático, preenchendo uma lacuna que os mecanismos de busca textual tradicionais muitas vezes ignoram. Com a *SearchOnMath*, os alunos podem conectar-se facilmente com colegas que podem ajudá-los a resolver problemas ou fornecer informações adicionais sobre fórmulas específicas. Os pesquisadores podem aproveitar a plataforma para explorar o repositório *arXiv*¹¹, descobrindo vários usos de uma fórmula específica e determinando se ela já apareceu em áreas específicas de estudo.

Na versão inicial, a ferramenta recupera os cinco principais resultados de seu índice web. No entanto, como a *SearchOnMath* é um serviço pago e a função está sendo disponibilizada gratuitamente dentro do *Mathematica*, fornecer todos os resultados diretamente no *Mathematica* poderia levar à possibilidade de que os usuários que são assinantes tanto da *SearchOnMath* quanto do *Mathematica* deixassem de assinar a *SearchOnMath*. Isso ocorreria porque o serviço estaria disponível de forma completa e gratuita para os usuários do *Mathematica*.

¹¹<https://arxiv.org>

Para melhorar a funcionalidade da rotina num futuro próximo, foi planejada a incorporação de parâmetros adicionais, tais como:

- Permitir que os usuários definam um número diferente de resultados por solicitação, proporcionando mais flexibilidade na quantidade de informações exibidas;
- Implementar uma funcionalidade de paginação, permitindo aos usuários navegar pelos resultados da pesquisa com mais facilidade;
- Introduzir filtros que permitem aos usuários refinar sua pesquisa por domínios ou fontes específicas, garantindo resultados mais direcionados e relevantes;
- Permitir que os assinantes da SearchOnMath forneçam sua chave de acesso para obter resultados mais avançados.
- Habilitar a versão científica do *SearchOnMath*, que permitirá pesquisar todas as pré publicações acadêmicas do *arXiv*, ampliando consideravelmente a gama de conteúdos disponíveis para pesquisa.

Referências

- Amador, B., Langsenkamp, M., Dey, A., Shah, A. K., and Zanibbi, R. (2023). Searching the ACL Anthology with Math Formulas and Text. *Proc. ACM SIGIR*. p. to appear.
- Consortium, W. W. W. (1999). Svg - scalable vector graphics. <https://www.w3.org/Graphics/SVG/>. Acesso em 03 de julho de 2024.
- Databricks (2024). What are dataframes? *Databricks*.
- Gao, L., Jiang, Z., Yin, Y., Yuan, K., Yan, Z., and Tang, Z. (2017). Preliminary Exploration of Formula Embedding for Mathematical Information Retrieval: can mathematical formulae be embedded like a natural language?
- Gonzaga, F. (2023). SearchOnMath Function in the Wolfram Function Repository. <https://resources.wolframcloud.com/FunctionRepository/resources/SearchOnMath/>. Acesso em 04 de outubro de 2023.
- Gonzaga, F. B. (2022). Enabling Mathematical Search inside Wolfram Mathematica. Wolfram Technology Conference.
- Hennigan, R. (2022). SVGImport. <https://resources.wolframcloud.com/FunctionRepository/resources/SVGImport/>. Acesso em 25 de junho de 2023.
- Lamport, L. (1986). Latex: A document preparation system, adison. <https://www.latex-project.org>.
- Li, X., Tian, B., and Tian, X. (2023). Scientific Documents Retrieval Based on Graph Convolutional Network and Hesitant Fuzzy Set. *IEEE Access*, 11:27942–27954.
- Mansouri, B., Rohatgi, S., Oard, D. W., Wu, J., Giles, C. L., and Zanibbi, R. (2019). Tangent-CFT: An embedding model for mathematical formulas. In *Proceedings of the 2019 ACM SIGIR international conference on theory of information retrieval*, pages 11–18.
- MathML (2023). MathML Presentation. <https://www.w3.org/TR/2023/SPSD-MathML2-20230307/chapter3.html>. Acesso em 25 de junho de 2023.
- Pezoa, F., Reutter, J. L., Suarez, F., Ugarte, M., and Vrgoč, D. (2016). Foundations of json schema. In *Proceedings of the 25th International Conference on World Wide Web*, pages 263–273. International World Wide Web Conferences Steering Committee.
- SearchOnMath (2023). SearchOnMath - About page. <https://www.searchonmath.com/about>. Acesso em 23 de agosto de 2023.
- Wolfram, S. (2023a). Wolfram Function Repository. <https://resources.wolframcloud.com/FunctionRepository>. Acesso em 03 de outubro de 2023.
- Wolfram, S. (2023b). Wolfram Mathematica. https://en.wikipedia.org/wiki/Wolfram_Mathematica. Acesso em 20 de junho de 2023.
- Wolfram, S. (2024). The Story Continues: Announcing Version 14 of Wolfram Language and Mathematica. <https://writings.stephenwolfram.com/2024/01/the-story-continues-announcing-version-14-of-wolfram-language-and-mathematica/>. Acesso em 20 de abril de 2024.