

Investigação da relação entre o interesse dos alunos e a abrangência de materiais didáticos sobre Relatividade Geral em português usando técnicas de aprendizado de máquina

Matheus Borrone de Freitas (IC)^{1,2}, Alan Bendasoli Pavan (PQ)^{1,2}

¹Universidade Federal de Itajubá.

²Grupo de Estudo e Pesquisa em Machine Learning - GREMLING

Palavras-chave: Processamento de Linguagem Natural. Python. Web Scraping.

Introdução

A proposta principal deste projeto é a utilização de técnicas de aprendizado de máquina, com processamento de linguagem natural, e web scraping para a criação de um algoritmo em Python. O processamento de linguagem natural é o ramo da inteligência artificial que busca ensinar a máquina a interpretar a linguagem humana, enquanto o web scraping tem por finalidade fazer pesquisas automatizadas e colher dados na web.

O ramo de processamento de linguagem natural se mostra promissor para aplicações tecnológicas em educação, já que é capaz de gerar ferramentas poderosas para interpretação de textos e materiais educacionais em geral. Esse potencial é abordado em alguns trabalhos interessantes, como o artigo *Natural Language Processing for Enhancing Teaching and Learning*, de Diane Litman, da *University of Pittsburgh*, onde são retratadas algumas pesquisas na área, mostrando novos horizontes para o desenvolvimento e seus desafios.

Nesse trabalho, o enfoque é a criação de um programa que investigue a abrangência de materiais sobre um determinado tema disponíveis na rede, e como isso se relaciona com o interesse dos estudantes. A aplicação será no contexto de relatividade geral, interpretando os resultados obtidos e estudando a relação que há entre eles e o interesse dos alunos dos cursos de física da UNIFEI pelo assunto em questão.

Metodologia

A construção do projeto foi feita em duas etapas. A primeira delas foi mapear o interesse dos estudantes de física da Universidade Federal de Itajubá em relatividade geral. Para isso foi feita uma pesquisa por meio da ferramenta Google Formulários. Foi elaborado um questionário composto de 7 questões variando entre abertas e fechadas e disponibilizado aos estudantes.

A segunda etapa consistiu na elaboração de um programa computacional capaz de investigar a abrangência e diversidade de materiais sobre o assunto disponíveis na internet, especificamente em língua portuguesa. O conhecimento obtido através de alguns cursos online ajudaram na construção do programa, sendo eles os cursos *Python* e *Inteligência Artificial e Computacional*, oferecidos pela Faculdade de Informática e Administração Paulista, além do curso *Python para Processamento de Linguagem Natural*, oferecido pela Universidade de São Paulo.

O programa foi modelado para analisar livros como fontes de referência sobre o tema, a saber: *Introducing Einstein's Relativity*, de Ray D'Inverno; *General Relativity*, de Robert Wald e *An Introduction to General Relativity Spacetime and Geometry*, de Sean Carroll. Dessas fontes foram extraídos os termos mais comuns presente em todas elas, para pesquisas automatizadas no Google. Para a obtenção desses termos comuns e extração foram utilizadas ferramentas de processamento de linguagem natural, através de funções presentes nas bibliotecas *NLTK* e *Spacy*, referências nesse tipo de abordagem.

Com técnicas de web scraping, utilizando os pacotes *Selenium* e *Beautiful Soup*, foram realizadas diversas buscas em inglês e em português, a fim de comparar a quantidade de resultados gerados pelas buscas em cada idioma, para assim se ter uma ideia do quão vasta é a disponibilidade de materiais sobre o assunto no nosso idioma, quando comparada com a do idioma mais falado do mundo. O livro *Web Scraping com Python: Coletando Mais Dados na Web Moderna*, de Ryan Mitchell foi de grande valia nessa parte.

Por fim, buscou-se relacionar esses resultados ao interesse demonstrado pelos estudantes no tema relatividade geral e investigar se a linguagem é uma possível barreira no acesso a esse conhecimento.

Resultados e discussão

O questionário foi respondido por 31 estudantes, e

cerca de 93,5% desses demonstrou algum tipo de interesse na área. Baseado nas respostas, também se constatou que a maioria deles conseguem ler bem em inglês.

Quanto aos resultados obtidos pelo programa, ficou nítido que a abrangência de materiais sobre relatividade geral na internet é muito maior em inglês do que em português (em média mais de 100 vezes maior de acordo com as buscas realizadas). Isso sugere que realmente a questão do idioma pode ser uma barreira importante. O fato dos estudantes consultados serem bons leitores em inglês pode ter influência no alto interesse pelo assunto, uma vez que essa dificuldade não seria um empecilho, mas claro que esse ponto ainda carece de investigações mais profundas.

```
space / espaço = 90.95394736842105
hypersurface / hipersuperfície = 128.0
einstein / einstein = 48.84038199181446
singularity / singularidade = 57.95275590551181
derivatives / derivadas = 53.97727272727273
universe / universo = 23.695652173913043
observer / observador = 54.54545454545455
past / passado = 65.3225806451613
terms / termos = 108.03324099722991
tensors / tensores = 93.73040752351098
properties / propriedades = 70.12987012987013
invariant / invariante = 173.98373983739836
special / especial = 68.1710213776722
definition / definição = 63.07053941908714
radiation / radiação = 65.98360655737704
vacuum / vácuo = 82.71604938271605
flat / plano = 49.50819672131148
Em média, a quantidade de resultados em inglês é 101.41815688378037 vezes a de português
```

Figura 1 – Amostra do retrato da diferença entre resultados nas buscas em inglês e português

Conclusões

A ideia de investigar a abrangência e distribuição na rede de materiais envolvendo relatividade geral se justificou promissora, e revelou alguns resultados interessantes ao cruzar os dados obtidos com o interesse demonstrado pelos estudantes e suas justificativas. Considerar atrelar uma possível baixa tendência de interesse na área à questão de dificuldades com idioma parece fazer certo sentido, ainda que seja necessário uma investigação mais profunda.

O uso de técnicas de aprendizado de máquina para os fins do projeto se mostrou eficiente; e um aprimoramento do que foi construído, considerando até utilização de outras ferramentas computacionais, pode resultar em uma análise mais profunda buscando também explorar outros fatores influentes na questão abordada.

Ademais, o que foi construído serve de valia não só

para a temática aqui discutida, mas também se mostra um bom aliado em outras aplicações, pois o programa é aplicável a qualquer assunto que se deseje fazer um estudo do gênero.

Todo o conhecimento obtido durante o processo e a experiência também foram engrandecedores, e certamente servirão de bagagem para projetos futuros.

Agradecimento

Ao CNPq pelo apoio financeiro durante o projeto.

À nossa instituição Universidade Federal de Itajubá.

Referências

MITCHELL, Ryan. Web Scraping com Python: Coletando Mais Dados na Web Moderna. 2ª edição. São Paulo: Novatec Editora, 2019.

D'INVERNO, R.A. INTRODUCING EINSTEIN'S RELATIVITY. Nova York: Oxford University Press Inc, 1992.

WALD, R.M. GENERAL RELATIVITY. Chicago: The University of Chicago, 1984.

CARROLL, Sean. SPACETIME AND GEOMETRY: An Introduction to General Relativity. São Francisco: Pearson Education, INC. 2004

Fiap On. Programa Eu Capacito: Inteligência Artificial e Computacional. Faculdade de Informática e Administração Paulista. Disponível em: <https://on.fiap.com.br/local/nanocourses/index.php>. Acesso em: setembro de 2021.

Fiap On. Programa Eu Capacito: Python. Faculdade de Informática e Administração Paulista. Disponível em: <https://on.fiap.com.br/local/nanocourses/index.php>. Acesso em: setembro de 2021.

Cursos de Extensão USP. Python para Processamento de Linguagem Natural: Edição Comemorativa dos 50 anos do ICMC. Universidade de São Paulo. Disponível em: <https://cursosextensao.usp.br/course/view.php?id=2721>. Acesso em: setembro de 2021.